ISBN 978-82-326-3402-6 (printed version)
ISBN 978-82-326-3403-3 (electronic version)
ISSN 1503-8181

Doctoral Thesis
Jean-Philippe Banon

Doctoral theses at NTNU, 2018:306

Jean-Philippe Banon

On the Simulation of Electromagnetic Wave Scattering by Periodic and Randomly Rough Layered Structures based on the Reduced Rayleigh Equations

Theory, numerical analysis and applications

NTNU
Norwegian University of
Science and Technology
Faculty of Natural Sciences
Department of Physics

NTNU
Norwegian University of
Science and Technology

Doctoral theses at NTNU, 2018:306

NTNU

NTNU
Norwegian University of
Science and Technology

Jean-Philippe Banon

On the Simulation of Electromagnetic Wave Scattering by Periodic and Randomly Rough Layered Structures based on the Reduced Rayleigh Equations

Theory, numerical analysis and applications

Thesis for the degree of Philosophiae Doctor

Trondheim, October 2018

Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Physics

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the degree of Philosophiae Doctor

Faculty of Natural Sciences
Department of Physics

ISBN 978-82-326-3402-6 (printed version)
ISBN 978-82-326-3403-3 (electronic version)
ISSN 1503-8181

Doctoral theses at NTNU, 2018:306

Printed by Skipnes Kommunikasjon as

À mes parents, Florence et Jean-Louis, à ma sœur Stéphanie et à la mémoire de tonton Marcou et mamie Gerda.

Abstract

This thesis presents research that was carried out on the topic of electromagnetic wave scattering by randomly rough surfaces and plasmonic surfaces. The manuscript is based both on published and unpublished results in the form of an extensive monograph intended to present a coherent and self sufficient overview of the theoretical framework of the reduced Rayleigh equations, the numerical techniques used to solve them and their applications to the study of light scattering from randomly rough surfaces and plasmonic systems. The published papers represent the core scientific contribution of this thesis, but the unpublished results, in particular those related to the numerical analysis of the presented methods, are also of non-negligible interest.

The research articles contained in this thesis can be categorized in two main topics. The four first papers are concerned with the scattering of light from randomly rough systems, while the two last papers deal with the characterization and optical metrology of photonic plasmonic systems.

On the topic of light scattering from randomly rough systems, the focus of the three first papers is mainly on the understanding of the physical mechanisms involved in the scattering by weakly rough surfaces and rough films. The first paper is a combined experimental and numerical study of the optical Yoneda ring phenomenon. The so-called Yoneda ring is a ring of enhanced intensity in the diffusely scattered light observed in the medium of highest refractive index. The phenomenon is well-known in the x-ray community although no clear satisfactory theory seems to have been suggested to explain it. This first paper is an experimental demonstration of the existence of the Yoneda phenomenon at optical frequencies, and the numerical results in agreement with the experimental data, give a first indication of the single scattering nature of the effect. The second paper develops a single scattering theory to explain the physical mechanisms behind the Yoneda phenomenon and the Brewster scattering phenomenon. The proposed theory explains these two phenomena respectively in terms of coupled progressive-evanescent modes and in terms of the dipole radiation from the microscopic scatterers in the materials. The theory of the Yoneda phenomenon can be viewed as a generalization of the theory of Rayleigh anomalies for periodic gratings. The Brewster scattering phenomenon can be viewed as a generalization of the Brewster reflection by a planar surface. In the latter, the fundamental role of Snell-conjugate waves is discussed and a simple and powerful geometrical interpretation is presented. Two new phenomena are also predicted by the theory for light scattering under total internal reflection: (i) a *s-black-out* phenomenon for which the diffusely scattered light is purely p-polarized independently of the scattering direction, and which occurs exactly at the critical incidence for total internal reflection and which is associated with the alignment of the elementary oscillating dipoles along the axis normal to the average surface plane; (ii) a linearly to circularly polarized Brewster scattering effect for incidences beyond the total internal reflection incidence which was shown to be caused by a regime in which dipoles in the materials are no longer oscillating but

rotating. The third paper revisits the physics of Selényi rings, which are interference rings of intensity in the diffusely scattered light from a rough dielectric film. A single scattering theory is developed and an effect of rings enhancement and attenuation induced by the cross-correlation between the two-rough surfaces of the dielectric film is demonstrated. The forth paper is rather of computational interest and presents a technique of reduction of variance for Monte Carlo simulations of light scattering from dielectric randomly rough surfaces.

Concerning the topic of optical characterization of plasmonic and photonic surfaces, the emphasis is put on applications to critical dimension metrology from optical measurements. The first paper presents experimental measurements of angle resolved spectrocopic Mueller matrix ellipsometry of a plasmonic photonic surface and some first results of data inversion using the commercial software COMSOL for the retrieval of the characteristic geometrical parameters of the sample. The second paper demonstrates a powerful technique for parameter retrieval based on the reduced Rayleigh equations. The method was applied successfully to the plasmonic photonic surface studied in the previous paper, with a demonstrated speed-up factor of a hundred compared with the use of COMSOL. The method based on the reduced Rayleigh equations has, however, a limited range of validity.

Finally, let us mention non-published results presented in the numerical analysis part of this thesis. The range of validity of the so-called Rayleigh hypothesis was explored numerically for both perfectly conducting and dielectric sinusoidal surfaces. An extensive parametric study was carried out and a limit of validity was found to be consistent with known analytical results for perfectly conducting sinusoidal surfaces. In addition, an alternative mathematical analysis, simpler than the conventional analysis based on complex analysis found in the literature, was developed and lead to the correct range of validity for sinusoidal surface. These results can be considered as premises to a more extensive research article on the topic.

Sammendrag

Denne avhandlingen består av seks vitenskapelige artikler innenfor fagområdet av elektromagnetisk bølgespredning fra rue overflater, periodiske overflater, og plasmoniske systemer. Manuskriptet inneholder også bakgrunnsteorien bak de reduserte Rayleigh ligningene og den numeriske analysen av metoder som brukes til beregningen og den fysiske forklaringen av de optiske fenomenene som vises og diskuteres i artiklene. Upubliserte resultater av numeriske analyse angående den så-kalte Rayleigh hypotesen og som begrenser metoden til overflater av lav ruhet er av spesiell interesse.

De tre første artiklene handler om utviklingen av den fysiske forståelsen av optiske fenomener som oppstår ved lysspredning fra rue dielektriske overflater eller rue dielektriske lag. I den første artikkelen bevises eksperimentelt og numerisk at Yoneda fenomenet eksisterer for optiske frekvenser. Yoneda fenomenet er karakterisert med en skarp ring av økende spredt lysintensitet. Yoneda effekten har vært vel kjent for Røntgenstråling spredning siden 60-tallet men den hadde ikke vært eksperimentelt bevist for optiske frekvenser før. Teorien som forklarer de fysiske mekanismer som står bak Yoneda og Brewster spredningseffektene utvikles i den andre artikkelen. Det vises at Yoneda effekten er en kontinuerlig generalisering av det som kalles Rayleigh anomalier for periodiske diffraksjonsgitter og at Brewster spredningseffekten oppstår på grunn av dipolsvingning stråling og er en generalisering av Brewster refleksjonen for en plan overflate. I tillegg forutser teorien to nye effekter som, til vår kunnskap, var ukjente. Ved totalrefleksjonsbetingelser vises det at lys som sprer seg når innfallsvinkelen er lik den grensevinkelen er rent p-polarisert uavhengig av spredningsretning. Denne effekten er forklart med å vise at dipoler i materialer svinger langs aksen som er normal den gjenomsnittlige overflate. For innfallsvinkler større enn grensevinkelen observeres en lineær til sirkulær polarisering Brewster effekt. Denne effekten forklares med å vise at dipoler i materialer ikke svinger lenger men snur i stedet og de fører til en typisk usymmetrisk sirkulær-polarisert stråling. Selényi interferensringer var studerte i den tredje artikkelen. Der vises det at korrelasjoner mellom ruhet til de to overflatene til et dielektrisk lag fører til selektiv økning eller demping av noen av ringene avhengig av korrelasjoner. Den fjerde artikkelen handler om en metode for å redusere det statistiske avviket til Monte Carlo simuleringer i tilfelle av lysspreding fra rue overflater.

Optisk karakterisering av plasmoniske overflater presenteres i de to siste artiklene. Der var optiske målinger av vinkelavhengig spektroskopisk Mueller matrise ellipsometri av en fotonik plasmonik system brukt sammen med simuleringer til å bestemme dimensjonene til de geometriske parameterne av systemet som, for eksempel, avstanden mellom partikler og partikkelstørrelsen. Først var inversjon av eksperimentelle data oppnådd med COMSOL og da var metoden forbedret ved å bruke koden basert på de reduserte Rayleigh ligningene. Den var bevist til å fungere hundre ganger raskere enn COMSOL. Metoden basert på de reduserte Rayleigh ligningene har derimot et begrenset gyldighetsområde som vises i den numersike analyse-delen av avhandlingen.

List of papers

Paper [1] Alma K. González-Alcalde, Jean-Philippe Banon, Øyvind S. Hetland, Alexei A. Maradudin, Eugenio R. Mendez, Tor Nordam, and Ingve Simonsen
Experimental and numerical studies of the scattering of light from a two-dimensional randomly rough interface in presence of total internal reflection: optical Yoneda peaks
Optics Express **24**, 25995-26005 (2016)

Paper [2] Jean-Philippe Banon, Øyvind S. Hetland, and Ingve Simonsen
On the physics of polarized light scattering from weakly rough dielectric surfaces: Yoneda and Brewster scattering phenomena
Submitted to Physical Review A

Paper [3] Jean-Philippe Banon, Øyvind S. Hetland, and Ingve Simonsen
Selective enhancement of Selényi rings induced by the cross-correlation between the interfaces of a two-dimensional randomly rough dielectric film
Annals of Physics **389**, 352-382 (2018)

Paper [4] Jean-Philippe Banon, Øyvind S. Hetland, and Ingve Simonsen
A reduction of variance technique applied to the simulation of light scattering from randomly rough surfaces
Manuscript under preparation

Paper [5] Morten Kildemo, Jean-Philippe Banon, Alexandre Baron, Brage B. Svendsen, Thomas Brakstad, and Ingve Simonsen
Optical response of gold hemispheroidal lattices on transparent substrates
Applied Surface Science **421**, 593-600 (2017)

Paper [6] Jean-Philippe Banon, Torstein Nesse, Zahra Ghadyani, Morten Kildemo, and Ingve Simonsen
Critical dimension metrology of a plasmonic photonic crystal based on Mueller matrix ellispsometry and the reduced Rayleigh equation
Optics Letters **42**, 2631-2634 (2017)

Other published papers not included in the thesis

Paper Torstein Nesse, Jean-Philippe Banon, Bodil Holst, and Ingve Simonsen
Optimal design of grid-based binary holograms for matter-wave lithography
Physical Review Applied **8**, 024011 (2017)

My contribution to the papers

Paper [1]

I derived the reduced Rayleigh equation for the film system. I implemented its numerical resolution together with Øyvind S. Hetland and contributed to the discussion on the manuscript text.

Paper [2]

I derived the theory and developed the analysis. I made the figures and wrote the manuscript together with Øyvind S. Hetland.

Paper [3]

I derived the theory for the reduced Rayleigh equations for a film system with two-rough interfaces. I implemented the RREs in the Monte Carlo solver together with Øyvind S. Hetland and I wrote a dedicated code for small amplitude perturbation theory (SAPT) to first order for the case of multilayer systems with rough interfaces. The Monte Carlo simulation were ran by Øyvind S. Hetland and the SAPT simulations were ran by both of us. I made some of the figures. I wrote the theory section, the appendices, and made the supplementary materials. The remaining sections were written both by Øyvind S. Hetland and myself.

Paper [4]

I derived the theory, implemented the numerical methods, ran the simulations and wrote most of the manuscript.

Paper [5]

I carried out a parametric numerical study with the reduced Rayleigh equations code to guide the parameters used for the COMSOL simulations. I made Figure 3 and contributed to the discussion on the manuscript text.

Paper [6]

I derived the theory for the reduced Rayleigh equations for plasmonic systems, and wrote the Rayleigh code. I wrote the optimization scheme together with Torstein Nesse. I ran the simulations and I made all the figures. I wrote most of the manuscript text.

Acknowledgments

I would like to express my gratitude to the people who have made my PhD studies in Trondheim both a fruitful and a pleasant experience.

First, I am naturally thankful to Prof. Ingve Simonsen for his supervision, his patience, his supportive spirit and for giving me the chance to meet collaborators around the world, and to Prof. Alex Hansen for his co-supervision. Among these collaborators, I am grateful to Prof. Morten Kildemo and his team for fruitful discussions and for providing experimental data of high quality. I am also grateful to Dr. Iryna Gozhyk and her team in Saint-Gobain Recherche who have made my short research stay in Aubervilliers a particularly pleasant one and for sharing their experimental expertise which has been a great source of inspiration. My greatest considerations go to Prof. Alexei A. Maradudin, leading expert in the field of wave scattering by random surfaces, and to Peggy, his wife, who made the honor of welcoming Øyvind S. Hetland, Torstein Nesse and myself at UC Irvine for a short research stay. I will always remember your kindness and the interesting discussions around a delightful diner.

I am grateful to Prof. Jean-Jacques Greffet and Prof. Christophe Geuzaine, for taking the time to evaluate my thesis and for joining the assessment committee for my defense along with Prof. Justin W. Wells.

The working days would not have been as enjoyable as they were without cheerful colleagues. Thank you Torstein for being such an enthusiastic office mate, a talented chef, a great SWIA night host, a tenacious tennis player, and for the many scientific discussions around our tiny white board. Thank you Øyvind for being this unalterably joyful person, always ready to help, a great conference mate, and for constantly challenging my understanding of physics with simple questions. You have reminded me that physics is not only about equations but also of making them intelligible with simple pictures. Thank you Morten Stornes for the barbecue evenings, and the thrilling AH nights, and Morten Vassvik for distracting me with geometrical problems, for delicious raspeballer and video games afternoon. Thank you Tor Nordam for nice discussions and for delicious 17th of Mai breakfasts. Takk Jacob Kryvi for å tvinge meg til å snakke norsk. Thank you more generally to all the people at Department of Physics with whom I had the chance to talk to.

I would like to take a few lines to express my gratitude to my former Master thesis team at the Alfvén Laboratory in Stockholm, Prof. Svetlana Ratynskaya, Dr. Panagiotis Tolias and Dr. Ladislas Vignitchouk for keeping in touch, for always welcoming me for my occasional visits and for keeping me up to date on the impressive progress on the project of dust dynamics in fusion plasmas. A special thanks to Ladislas for the yearly diner full of entertaining scientific and less scientific discussions.

Because sometimes one needs to be distracted from work, I would like to thank my friends for the great time spent together. I am glad for having practiced with all aikidokas from NTNUI Aikido Tekisuikan and from TAK. In particular, thank you to Izzie for the talking and the epic tennis matches, and to Seiniz for memorable sushi and samurai movie evenings. Thank you Viktória and Wolfgang for the many extra trainings, the occasional bruises, and for the many hours spent relaxing at your place. Og takk Stephanie for at du tilfeldig stoppet ved dojoen på Gløshaugen på en søndag Februar 12. Jeg gleder meg til å dele flere deilige cookies, å spile flere biljard og shuffleboard kamper, og å nyte flere bobler og andre søte øyeblikker med deg.

Let us not forget the non Trondheimers. A big thank to my old friends, who, despite the distance take the time to keep in touch. Thank you daddy Mathieu, Gilles (salut copain), mummy Elia, Laura, David, and Cyril for the "Héraultais". Thank you captain Nico and Quentin for the "Nantais". Thank you Harsha, Song Hong, Sarah, Sarah, Anna, Ana, and Anders for the "Stockholmers". Et un énorme merci à Alice, Thomas et Elsa pour les innoubliables randonnées. Sans vous, je n'aurais probablement pas autant explorer la Norvège.

Enfin, mes remerciements les plus chers vont à ma famille. Merci papa et maman pour votre soutiens constant, et votre patience à savoir votre grand si loin de la maison. Merci Stéphanie pour être la plus merveilleuse des petites sœurs, toi qui me donnes toujours chaud au cœur quand tu m'offres ton plus beau sourire. Merci à mes grands parents pour la joie et la fierté qu'ils me montrent et à la famille au sens large pour les repas toujours animés qui font que l'on se sent bien et que l'on n'oublie jamais ses racines.

Contents

Notations

Quantifiers

- Conventional symbols may be used for quantifiers. $\forall$ for all, $\in$ (resp. $\notin$) belongs to (resp. does not belong to), $|$ such that.

Common sets

- The empty set is denoted $\varnothing$.
- $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ denote respectively the sets of natural integers, relative integers, rational numbers, real numbers and complex numbers.
- For two sets A and B, we denote by $A \cup B$, $A \cap B$ and $A \setminus B$ respectively the union, intersection and difference of A and B.
- A numeric set to which the element 0 is removed, is denoted with a star. Examples: $\mathbb{N}^* \equiv \mathbb{N} \setminus \{0\}$, $\mathbb{R}^* \equiv \mathbb{R} \setminus \{0\}$.
- The part of an ordered numeric set composed of the positive (resp. negative) elements is denoted by a $_+$ (resp. $_-$). Examples: $\mathbb{R}_+ \equiv \{x \in \mathbb{R} \mid x \geq 0\}$, $\mathbb{R}_+^* \equiv \{x \in \mathbb{R} \mid x > 0\}$.

Brackets

- The following brackets are used:
  - to define a set $\{\cdots\}$;
  - to define an interval of $\mathbb{R}$, $[a, b]$, $[a, b[$, $]a, b]$, $]a, b[$ (where $a, b \in \mathbb{R} \mid a < b$);
  - as parenthesis in a formula to make factors explicit, $(\cdots)$, $[\cdots]$;
  - to define a set of integers between $n_1 < n_2$, $n_1, n_2 \in \mathbb{Z}$, $[\![n_1, n_2]\!] \equiv \{n \in \mathbb{Z} \mid n_1 \leq n \leq n_2\}$;
  - to denote an ensemble average over realizations of a stochastic process, or random variable $\langle \cdot \rangle$ (the variance and covariance are however denoted $\mathbb{V}\mathrm{ar}[\cdot]$ and $\mathbb{C}\mathrm{ov}[\cdot, \cdot]$);
  - to denote the modulus of a complex number or the Euclidean norm of a vector, or the length of a multi-index $|\cdot|$;
  - to denote a norm (except for the Euclidean norm) $\|\cdot\|$;
  - to denote the average of a function with respect to some variables over a domain, e.g. $\langle f \rangle_\phi \overset{\text{def}}{=} \frac{1}{2\pi} \int_0^{2\pi} f(\theta, \phi) \, \mathrm{d}\phi$;
  - to denote a functional $\mathcal{T}$ applied on a function f, $\langle \mathcal{T}, f \rangle$.

Common functions

- The exponential function may be denoted either $\exp(\cdot)$ or $e^{\cdot}$.
- The real and imaginary part of a complex number z are denoted respectively $\mathrm{Re}(z)$ and $\mathrm{Im}(z)$.
- Bessel function of first kind of order ν, J_ν.
- Kronecker delta: for $i, j \in \mathbb{Z}$, $\delta_{ij} = 1$ if $i = j$, 0 if $i \neq j$.
- Characteristic or indicator function of a subset A of E: for $x \in E$, $\mathbb{1}_A(x) = 1$ if $x \in A$, 0 if $x \notin A$.
- The following definition of the sinc function will be used: $\operatorname{sinc} x = \frac{\sin \pi x}{\pi x}$ for $x \neq 0$, and extended by continuity at 0 as $\operatorname{sinc} 0 = 1$.

Differential operators

- An orthonormal basis vector for a cartesian coordinate system will be denoted $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3)$.
- ∇ denotes the gradient operator. For a scalar field $\phi(\mathbf{x})$, where $\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3$, the gradient is a vector and can be expressed in cartesian coordinates as $\nabla\phi = (\frac{\partial\phi}{\partial x_1}, \frac{\partial\phi}{\partial x_2}, \frac{\partial\phi}{\partial x_3})$.
- $\nabla\cdot$ denotes the divergence operator. For a vector field $\mathbf{V}(\mathbf{x})$, where $\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3$, the divergence is a scalar and can be expressed in cartesian coordinates as $\nabla \cdot \mathbf{V} = \sum_{i=1}^{3} \frac{\partial V_i}{\partial x_i}$.
- $\nabla\times$ denotes the curl operator. For a vector field $\mathbf{V}(\mathbf{x})$, where $\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3$, the curl is a vector and can be expressed in cartesian coordinates as $\nabla \times \mathbf{V} = (\frac{\partial V_3}{\partial x_2} - \frac{\partial V_2}{\partial x_3}, \frac{\partial V_1}{\partial x_3} - \frac{\partial V_3}{\partial x_1}, \frac{\partial V_2}{\partial x_1} - \frac{\partial V_1}{\partial x_2})$.
- Δ denotes both the scalar Laplace operator and the vecorial Laplace operator. For a scalar field ϕ, the Laplace operator applied on ϕ may be expressed in a cartesian coordinate system as $\Delta\phi = \sum_{i=1}^{3} \frac{\partial^2\phi}{\partial x_i^2}$. For a vector field $\mathbf{V}$, the Laplace operator applied on $\mathbf{V}$ may be expressed in a cartesian coordinate system as $\Delta\,\mathbf{V} \cdot \hat{\mathbf{e}}_i = \Delta V_i = \sum_{j=1}^{3} \frac{\partial^2 V_i}{\partial x_j^2}$, for $i \in \{1, 2, 3\}$.
- Note that the symbol Δ may at time be used to denote a step size or a deviation of some quantity. We believe that no confusion with the Laplace operator should occur since the context will make the distinction clear.

Fourier transform

Let f be a function of a spatial variable $\mathbf{r} \in \mathbb{R}^3$ and a time variable $t \in \mathbb{R}$. f may be a scalar or a vector function. We define the Fourier transform of f with respect to the spatial variable as

$$\mathcal{F}_r[f](\mathbf{k}, t) = f(\mathbf{k}, t) = \int_{\mathbb{R}^3} f(\mathbf{r}, t)\, e^{-i\mathbf{k}\cdot\mathbf{r}}\, \mathrm{d}^3 r\,,$$

and the Fourier transform of f with respect to the time variable as

$$\mathcal{F}_t[f](\mathbf{r},\omega) = f(\mathbf{r},\omega) = \int_{\mathbb{R}} f(\mathbf{r},t)\, e^{i\omega t}\, \mathrm{d}t\,.$$

Note the different signs in the definitions of the spatial and time Fourier transforms. This choice is motivated by the physical interpretation of a plane wave of frequency ω propagating along a wave vector $\mathbf{k}$ for increasing time. Also note that the Fourier transform will often be denoted with the same symbol as the function in the direct space, when the context in non-ambiguous and only the argument is changed. With the adopted conventions, the inverse Fourier transforms are respectively given by

$$\mathcal{F}_r^{-1}[f](\mathbf{r},t) = \int_{\mathbb{R}^3} f(\mathbf{k},t)\, e^{i\mathbf{k}\cdot\mathbf{r}}\, \frac{\mathrm{d}^3k}{(2\pi)^3}\,,$$

and

$$\mathcal{F}_t^{-1}[f](\mathbf{r},t) = \int_{\mathbb{R}} f(\mathbf{r},\omega)\, e^{-i\omega t}\, \frac{\mathrm{d}\omega}{2\pi}\,.$$

Multi-index

A multi-index is a n-tuple of non-negative integers, i.e. an element $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_n) \in \mathbb{N}^n$. We will use the following definitions for calculating with multi-indices. The sum of two multi-indices $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_n)$ and $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_n)$ is

$$\boldsymbol{\alpha} + \boldsymbol{\beta} = (\alpha_1 + \beta_1, \cdots, \alpha_n + \beta_n)\,.$$

The *length* of a multi-index, denoted $|\boldsymbol{\alpha}|$ is the sum of its elements, i.e.

$$|\boldsymbol{\alpha}| = \sum_{i=1}^{n} \alpha_i\,.$$

The factorial of a multi-index is defined as

$$\boldsymbol{\alpha}! = \alpha_1!\alpha_2!\cdots\alpha_n!\,.$$

With this definition of the factorial for a multi-index, the multinomial coefficient reads

$$\binom{|\boldsymbol{\alpha}|}{\boldsymbol{\alpha}} = \frac{|\boldsymbol{\alpha}|!}{\boldsymbol{\alpha}!}\,,$$

and the binomial coefficient for two multi-indices α and β is defined as

$$\binom{\boldsymbol{\alpha}}{\boldsymbol{\beta}} = \prod_{i=1}^{n} \binom{\alpha_i}{\beta_i} = \frac{\boldsymbol{\alpha}!}{\boldsymbol{\beta}!(\boldsymbol{\alpha}-\boldsymbol{\beta})!}\,.$$

List of acronyms

p polarization or (TM) transverse magnetic polarization
s polarization or (TE) transverse electric polarization
TIR: Total internal reflection
RRE(s): Reduced Rayleigh equation(s)
(M)DRC, (M)DTC: (Mean) differential reflection/transmission coefficient also known as
BRDF, BTDF : bidirectional reflectance/transmittance distribution function
SOR: Successive over-relaxation
SAPT: Small amplitude perturbation theory
FICRE: Fixed point iterates of the coupled Rayleigh equations
FIRES: Fixed point iterates of the reduced Rayleigh equations of the second kind
MC: Monte Carlo
SEM: Scanning electron microscopy
AFM: Atomic force microscopy

Introduction

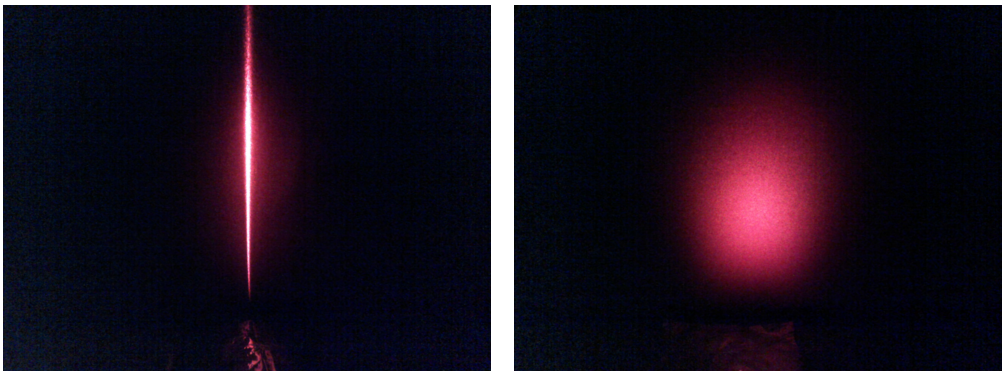
"Le couloir de la cuisine était clair, vitré des deux côtés, et un soleil brillait de chaque côté, car Colin aimait la lumière. [...] Les jeux des soleils sur les robinets produisaient des effets féeriques. Les souris de la cuisine aimaient danser au son des chocs des rayons de soleil sur les robinets, et couraient après les petites boules que formaient les rayons en achevant de se pulvériser sur le sol, comme des jets de mercure jaune."
L'Écume des jours, Boris Vian (1947).

A great number of natural phenomena that we observe, as well as numerous man-made objects we use in our everyday life, are made possible thanks to the interactions between light and matter. The magnificence of the multitude of colors of the sky and the clouds during a sunset, for our greatest pleasure, is due to the scattering of light by the molecules constituting the atmosphere, and by the droplets of water in the clouds. The photosynthesis which permits the conversion of light energy into chemical energy, and consequently plants to live, is an astonishing example of light-matter interaction and the mechanisms behind the efficiency of the energy transport is still an object of active scientific research. Your eyes, which are probably the organs you use most for sensing your environment, to admire romantic sunsets or to read these less romantic lines, are a marvelous piece of biological engineering shaped over millions of years of evolution. The back of your eyes is covered with cells dedicated to the conversion of light of different frequencies into electrical signals that your brain will interpret as colors. Your cornea is also a quite surprising part of your eyes. A mere transparent lens that focuses light on the back of the eye would you say? At first sight, yes, but by taking a closer look at the cornea, with a microscope, we can see that the cornea is far from being a uniform homogeneous transparent medium. It contains a multitude of randomly arranged collagen particles which should in principle scatter the light entering your eye and should make your cornea opaque! In fact, the cornea can become opaque due to some diseases or complications, leading to a so-called leukoma, a milky white patch on your cornea. The main difference, in terms of optics, between a transparent and an opaque cornea is simply the arrangement of the collagen particles while the density and size of the particles remain identical. This curious phenomenon is due to the interference of light scattered by each particle which results in a transparent cornea or an opaque one depending on the type of correlations between the position of the particles.

Your eyes, you certainly use them when watching TV, writing a message on your phone, or browsing the Internet. The information that you send to and receive from all around the world when using these devices is carried over long distances by light, but not the visible kind of light. Visible light, that is what you and I simply call light in our daily life, only represents a narrow spectrum of electromagnetic waves. Electromagnetic waves come in different frequencies or wavelengths. From radar detection to x-ray imaging in medicine, electromagnetic waves span a huge range of frequencies. Mastering the manipulation of these waves have been the object of

extensive research and development efforts leading to new technologies during the last century. New fields of intense active research emerged such as photonics, plasmonics, biomedical imaging, wave propagation in random media, to cite a few, promising new applications beneficial for society. As an example, plasmonics is the field of physics that studies the resonances, excited by light, of free electrons in metallic particles or on metallic surfaces, or in systems combining dielectric and metallic structures [7, 8]. The broadly called plasmon resonances have been used for centuries, or even millennium, to make colorful stained glass. The knowledge of plasmon resonances was not available at the time, but nowadays we know that the color of stained glass is due to plasmon resonances of metallic particles trapped inside the glass. Different sizes and shapes of the particles lead to different resonant frequencies and hence different colors of the reflected and transmitted light. Plasmonic surfaces are now used for optical sensing of specific molecules in a biological sample [9]. A metallic or composite surface is designed in such a way that it can adsorb a specific molecule, if present in your blood for example. Upon adsorption of the molecules, the frequency of the plasmon resonance shifts compared with that of the clean surface. This shift of resonant frequency can be measured optically very accurately to determine the presence or absence of a molecule in your blood, and helps assessing your medical diagnosis. Recently, the possibility of engineering thermal emission with plasmonic systems has been shown. Contrary to the long standing belief, one can design coherent, narrow band, anisotropic, polarized thermal emitters by shaping appropriately metallic particles, thus opening a road for new applications ranging from low consumption gas detectors, passing to engineering heat transfer at the nanoscale, thermal near-field imaging, to passive cooling [10, 11, 12, 13, 14].

We understand now that the study of the scattering of light of various frequencies by various objects can be both of fundamental scientific and technological interests. This thesis deals with the understanding of light scattering by surfaces. The term surface will encompass both the interface between two homogeneous media, like the interface between the air and a glass window, or a composite system made of a substrate on which some particles are deposited, like a plasmonic surface for example. We will be interested in understanding how light is reflected from and transmitted trough interfaces which are either structured or randomly rough. The stereotypical example of a structured surface is a periodic grating. Periodic gratings are key elements in spectrometers due to their property of scattering light of different frequencies with different well defined angles of scattering, hence separating the spectral components of a light signal. This is the behavior you can observe when looking at the rainbow pattern reflected from your favorite music CDs at home for example. Typical examples of randomly rough surfaces are the surfaces of a brushed steel, of paint, of biological tissues, or it could be the surface of the ocean and more generally most of the naturally occurring surfaces. The level of roughness of a surface must always be associated with the length scale at which we look at it, or to be more accurate when it comes to wave scattering, with the wavelength of the light. A light beam, such as that of a laser pointer, will be reflected mainly in the specular direction from a surface where the typical size of the profile amplitudes is small compared with the wavelength. This is what happens for example when you shine a laser pointer on a polished mirror. The surface hence looks smooth for this wavelength. However, if the wavelength of the light is reduced to be of the same order as the size of the corrugations, or equivalently if the corrugation of the surface is made larger, light will be scattered significantly in all directions, creating a diffuse speckle pattern (if coherent light is used). To observe this, you can shine your laser pointer on a sheet of aluminium foil you may find in your kitchen and try the two sides of it. You will observe a striking difference between the resulting reflected intensity patterns. The highly reflecting side will reflect the laser beam by broadening it only in a



(a) Highly reflecting side. (b) Poorly reflecting side.

Figure 1: Laser scattering experiment in the kitchen. Snapshot of the light of a red laser pointer reflected from the highly (a) or poorly (b) reflecting side of a piece of aluminium foil projected onto a white paper screen.

certain direction (see Fig. 1(a)), while the poorly reflecting side will broaden it more or less isotropically (see Fig. 1(b)). These two patterns are characteristic of the microscopic structure of the surface. The reflecting side of the aluminium foil has been polished or brushed in a certain direction, so that we find rather straight grooves on the surface. This is the reason why, although reflecting, you see your reflection rather deformed from the aluminium foil compared with a well polished mirror. On the other hand, the surface of the poorly reflecting side is made of random corrugations but not exhibiting grooves aligned in some specific direction.

Some of the questions we wish to answer then are the following. For a given surface and a given incident wave, what fraction of the incident power goes in the different scattering directions? How can we relate this intensity distribution to the shape of the surface profile and the property of the materials? Can we understand some basic physical mechanisms at the origin of some features seen in the intensity distribution of the scattered light? These are typical questions for a so-called *forward* scattering problem, for which the characteristics of the scatterer and the source are known, and we wish to predict and understand the scattering response. We can also ask the reverse questions, or so-called *inverse* scattering problem. If we are given a sample and we are allowed to make a set of scattering experiments on this sample to characterize it, how do we recover information about the profile of the surface by only knowing its scattering response? Another question, which is related to the previous one in practice, is that of the design of a surface. Given a desired scattering response, how should we design the surface profile leading to the target response, or a response as close as possible to the targeted one given some constraints?

In order to predict the scattering response of a surface, to understand the physics involved, to characterize a surface or to design it, we need some tools. We need a theory which describes the interaction between electromagnetic waves and matter. In most of the aforementioned applications, the light scattering phenomena can be described in terms of classical electrodynamics, i.e. Maxwell equations together with some constitutive models for the response of a bulk material at a given frequency. Solving Maxwell equations for the case of a wave interacting with some

arbitrarily shaped object is a rather complicated task which can only be done with the help of numerical simulations apart from some academic cases such as the scattering by a planar surface or by a sphere, which can be solved analytically. There exists a whole zoo of numerical tools for the study of light scattering which all have their advantages and drawbacks depending on the types of systems to be studied or the types of questions to be answered. One may, for example, discretize space and time and solve numerically Maxwell equations by propagating a light pulse in time step by step. This is known as the finite difference time domain method (FDTD). This may work well in practice for systems having a geometry that can be discretized easily on a rectangular grid, to observe the time evolution of a pulse in such a system or to study at once the response for several frequency components. However, if the structure becomes difficult to discretize on a rectangular grid so that the detailed structure of the surface requires a very fine grid, the computational burden will become quickly unpractical. Finite elements methods (FEM) or methods based on volume integral representations of the solution to the scattering problem may then be better suited since space can be discretized in small tetrahedral volumes which are well suited to discretizing complex shapes. Nevertheless, as the size of the system becomes large compared with the wavelength, and the fine details of the structure are smaller or on the order of the wavelength, the mesh size, which typically requires at least to have about ten nodes within a wavelength, may become rather dense, and the resulting linear system which must be solved may become unpractically large. In cases where the system can be considered as composed of a set of scatterers having homogeneous refractive indices, one can use some integral theorems in such a way that it is sufficient to only discretize the boundaries between these scatterers, which may yield to significant speed-up. These methods are known as surface integral methods or boundary elements methods (BEM), and are particularly well suited for scattering by a surface between two homogeneous media. The numerical method analyzed and used in the present thesis can be categorized as a surface integral method. The method consists in solving integral equations known as the reduced Rayleigh equations. The unknowns of these integral equations are the Fourier field amplitudes of the reflected and transmitted field, the kernels of the integral equations encodes the information about the scattering system and the right hand side encodes the source. The reduced Rayleigh equations are not particularly new, as they take their roots in the method developed by Lord Rayleigh in 1907 [15] for the scattering of waves by a perfectly conducting periodic grating. The reduced Rayleigh equations can be seen as a generalization of the Rayleigh method for penetrable media and for non-periodic surface.

The present thesis consists of three main parts named *theory*, *numerical analysis* and *applications*. Part I, starts with a warming up chapter to recall basic concepts of electromagnetic scattering, notations and vocabulary that we will need in the rest of this work (Chapter 1). Then the derivation of the fundamental transfer equations for the fields across a single interface from which one deduces the reduced Rayleigh equations is exposed in details in Chapter 2. This is probably the most critical chapter since everything else depends on it. It is thus important to spend some time on it, before going on. In Chapters 3 and 4, we present a systematic way to derive the reduced Rayleigh equations for diverse classes of systems within a unified framework. More specifically, Chapter 3 extends the reduced Rayleigh equations to multilayer structures where each interface is *a priori* rough and Chapter 4 treats the case of systems made of a single periodic interface first, which can be read independently of Chapter 3, before dealing with the case of systems made of independently periodic interfaces for which knowledge of Chapter 3 is assumed. It must be stressed that in Part I, we are only interested in formally deriving equations, which are *integral equations*, whose unknowns are

scattering amplitudes. We do not discuss how to solve these equations, apart from some remarks.

Because solving equations is often as important as to derive them, we devote Part II to numerical methods that can be used to solve the reduced Rayleigh equations. We present in Chapter 5 the most elementary method to solve the reduced Rayleigh equations. We have named it the *direct method* and it basically boils down to solving a linear system. In addition to explaining its implementation, we present a complexity analysis of the method and study its convergence as the size of the system grows and this for a few different physical systems. Other methods, grouped under the appellation of *iterative methods*, are presented in Chapter 6. The reader experienced with numerical linear algebra will not be surprised to find there standard iterative methods for solving linear systems, but may be interested in the derivation and implementation of *perturbative solutions* of the reduced Rayleigh equations, and a comparative study of the convergence rate of all methods for a few classes of physical systems. In Chapter 7, we address the issue of the consistency of the reduced Rayleigh equations with the original scattering problem. There we probe numerically the validity of the so-called *Rayleigh hypothesis* by carrying out an extensive parametric study. Finally, Chapters 8 and 9 deal respectively with approximation methods specifically tailored for randomly rough systems, where the average response is considered and further mathematical details on perturbation theory for multilayer systems and perturbed gratings are given.

The reader who is not interested in understanding the details of implementation or the numerical analysis can skip Part II, or most of it, in a first reading and rather take a look at Part III where applications illustrate what the methods can be used for. There, various systems such as *randomly rough systems* (Chapters 10 and 11), *plasmonic photonic surfaces* (Chapter 12), are studied and discussed based on more physical ground. The *inverse scattering problem* finally closes this work (Chapter 12) with a concrete case study based on experimental data.

Part I

Theory

Chapter 1

Scattering of electromagnetic waves

"Un mince rayon de lune passait par le trou du volet, et faisait briller le verre, sur ma table de nuit. Le trou était rond, le rayon était plat. Je me promis de demander à mon père l'explication de ce phénomène."
La Gloire de mon père, Marcel Pagnol (1957).

In this introductory chapter we present the problem of scattering of electromagnetic waves by an object and recall some basic notions of electromagnetism that we will use in the following chapters. This chapter is mostly meant to set the stage, to fix some vocabulary and notations as well as to present a few results on the resolution of the scattering problem in the academic case of a planar interface between two media. The latter will be useful for two reasons. The first one is to remind the different type of waves that can exist and familiarize ourselves with the physics involved before treating more complex systems. The second reason is that these results will enable us to verify that the approach used for the resolution of the scattering problem in the case of a structured interface or a randomly rough surface yields the expected result in the limit of the planar interface.

1.1 Electromagnetism in a nutshell

This first section gives elements of classical electromagnetism. We start by recalling Maxwell equations in material medias and define the different fields involved. The response of the materials will then be discussed in terms of electric and magnetic susceptibility or equivalently permittivity and permeability. The consequences of assumptions such as linearity, homogeneity and isotropy on the model will be presented and related to the structure of the possible electromagnetic waves. The discussion in this section is greatly inspired by that of Refs. [16, 17, 18].

1.1.1 Maxwell equations

Let $(\mathbf{r}, t) \in \mathbb{R}^3 \times \mathbb{R}$ represent a point in the Euclidean space endowed with a coordinate system $(O, \hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3)$ at a time t. The space is filled by a material whose state at $(\mathbf{r}, t)$ is characterized by distributions of free electric charge density $\rho(\mathbf{r}, t)$ and current density $\mathbf{J}(\mathbf{r}, t)$ as well as bounded charges and currents modeled as an equivalent polarization field $\mathbf{P}(\mathbf{r}, t)$ and magnetization field

$\mathbf{M}(\mathbf{r}, t)$. The evolution of the electromagnetic system is described by Maxwell equations:

$$\nabla \cdot \mathbf{D} = \rho \tag{1.1a}$$
$$\nabla \cdot \mathbf{B} = 0 \tag{1.1b}$$
$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{1.1c}$$
$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \,. \tag{1.1d}$$

Here $\mathbf{E}$ and $\mathbf{H}$ are respectively the electric and magnetic fields, and $\mathbf{D}$ and $\mathbf{B}$ are respectively the electric displacement and magnetic induction fields defined as

$$\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P} \tag{1.2a}$$
$$\mathbf{B} = \mu_0 (\mathbf{H} + \mathbf{M}) \,, \tag{1.2b}$$

where ε_0 and μ_0 are respectively the permittivity and permeability of vacuum and are physical constants. We recall that the speed of light in vacuum c is related to ε_0 and μ_0 by $\varepsilon_0 \mu_0 = 1/c^2$.

The time evolution of the system of fields, charges and currents could in principle be solved by solving the aforementioned system of partial differential equations (together with the equation of motion for the charges not presented here) and given models for the polarization and magnetization fields $\mathbf{P}$ and $\mathbf{M}$. In the following, we will focus on the case where the medium of interest is somewhat bounded in space and no free charges nor free currents are present in it nor in its vicinity, i.e. $\rho = 0$ and $\mathbf{J} = \mathbf{0}$ identically, but some external incident fields are present and we wish to study the response of the system. Of course, in order to have an external incident field, free charges and/or currents may be present far away from the system of interest, and the external incident field must be understood as the result of a modeling of the effect of these sources.

1.1.2 Response of a material medium

Different materials give different responses when an electric or a magnetic field is applied on them. The response of the material is modeled by the polarization and magnetization fields. Also, it is observed, quite naturally, that the polarization and magnetization depend on the applied electric and magnetic fields. An applied electric field will affect the polarization which in turn will change the local electric field and so on. It is then clear that the problem of determining the response of the material is coupled to the resolution of Maxwell equations. Predicting the response of a material would in principle require to solve simultaneously Maxwell equations and the motion of charges within the material, the latter requiring, in general, quantum mechanics.

In practice, this is a difficult task and although some simple models such as Drude or Lorentz models give reasonable physical explanation of the response of materials, a precise prediction remains costly. Thus models of responses are often of semi-classical form and fitted to experimental data. Here, we simply describe the consequence of a few assumptions on the response of a material medium. For simplicity, we will base the discussion on the polarization, but similar consideration can be said for the magnetization.

Locality in space and non-locality in time

The polarization $\mathbf{P}(\mathbf{r},t)$ at a point $\mathbf{r}$ at time t is induced by the electric field $\mathbf{E}$ [1] but *a priori* it can be induced by the electric field from another location $\mathbf{r}$ or from another time t' (say earlier time if causality should hold). This general consideration is known as non local response. A standard assumption is to assume *locality in space* but *non-locality in time*, meaning that the polarization at $\mathbf{r}$ is induced by the electric field at the same position but at a different time due to some latency of the response. We will make this assumption in our work.

On linearity, homogeneity, time invariance, causality and anisotropy

In general the polarization depends on the electric field on a non-linear fashion. This yields a rich variety of phenomena like frequency doubling for instance. In this work, the electric field will be assumed to be small enough in magnitude such that only the first order in a possible expansion of the polarization in powers of the electric field is necessary to describe the response of the system. Taking into account locality in space and non-locality in time, the most general form for the linear polarization reads

$$\mathbf{P}(\mathbf{r},t) = \varepsilon_0 \int_{\mathbb{R}} \boldsymbol{\chi}(t,t') \ \mathbf{E}(\mathbf{r},t') \, \mathrm{d}t' \, . \tag{1.3}$$

Here the response kernel $\boldsymbol{\chi}$ is an order 2 tensor, i.e. that the above equation can be written in terms of the vectors components as, for $i \in [\![1,3]\!]$,

$$P_i(\mathbf{r},t) = \varepsilon_0 \int_{\mathbb{R}} \chi_{ij}(t,t') \, E_j(\mathbf{r},t') \, \mathrm{d}t' \, , \tag{1.4}$$

where the Einstein summation convention is used for repeated indices. The factor ε_0 is there for later convenience. Note that another assumption has been made here. The fact that $\boldsymbol{\chi}$ does not depend on the position $\mathbf{r}$ shows the *homogeneity* of the material, i.e. that the response is the same at any point. However, the response can be different along different directions as the tensorial form shows, and is said to be *anisotropic*. By further assuming *time invariance*, meaning that $\boldsymbol{\chi}$ does not specifically depend on times t and t' but on their difference, we have

$$\mathbf{P}(\mathbf{r},t) = \varepsilon_0 \int_{\mathbb{R}} \boldsymbol{\chi}(t-t') \ \mathbf{E}(\mathbf{r},t') \, \mathrm{d}t' = \varepsilon_0 \int_{\mathbb{R}} \boldsymbol{\chi}(\tau) \ \mathbf{E}(\mathbf{r},t-\tau) \, \mathrm{d}\tau = \varepsilon_0 \, \boldsymbol{\chi} * \mathbf{E} \, (\mathbf{r},t) \, . \tag{1.5}$$

The polarization is then the convolution product (denoted by the $*$) in time of a response kernel $\boldsymbol{\chi}$ and the electric field $\mathbf{E}$. This is the most general expression for the polarization as a function of the electric field under the assumptions of *locality of the response in space*, *homogeneity*, *linearity* and *time invariance*. Note that the integration in the convolution product can be restricted to $\tau \in \mathbb{R}_+$, or equivalently $\boldsymbol{\chi}$ vanishes identically for negative arguments, if *causality* holds, i.e. that the response at time t only depends on the electric field at times earlier than t.

In the case of a so-called *isotropic* material, i.e. that the response is independent of the direction, the order 2 tensor $\boldsymbol{\chi}$ becomes a scalar χ.

Susceptibility and permittivity

Equation (1.5) can be expressed as a point-wise product in Fourier reciprocal space. Indeed, by taking the Fourier transform with respect to the time variable in Eq. (1.5), we obtain

$$P_i(\mathbf{r},\omega) = \varepsilon_0 \, \chi_{ij}(\omega) \, E_j \, (\mathbf{r},\omega) \, , \tag{1.6}$$

[1] and can *a priori* also be induced by the magnetic field but we drop this for simplicity.

where we define the electric susceptibility $\chi_{ij}(\omega)$ as the Fourier transform of the response kernel $\chi_{ij}(t)$, for $i, j \in [\![1, 3]\!]$,

$$\chi_{ij}(\omega) = \int_{\mathbb{R}} \chi_{ij}(t)\, e^{i\omega t}\, \mathrm{d}t\,. \tag{1.7}$$

Note that we chose not to change notations for a function and its Fourier transform but simply change the symbol in the argument. Knowing if one deals with a direct quantity or its Fourier transform will implicitly be clear from the context and the argument.

Under the aforementioned assumptions, the Fourier transform of Eq. (1.2a) with respect to the time variable yields

$$\mathbf{D}(\mathbf{r}, \omega) = \varepsilon_0\, \mathbf{E}(\mathbf{r}, \omega) + \mathbf{P}(\mathbf{r}, \omega) = \varepsilon_0\, \boldsymbol{\epsilon}(\omega)\ \mathbf{E}(\mathbf{r}, \omega)\,, \tag{1.8}$$

where we have used Eq. (1.6) and defined the *relative permittivity tensor*, also called *dielectric tensor*, $\boldsymbol{\epsilon}$ as

$$\boldsymbol{\epsilon}(\omega) = \mathbf{I}_3 + \boldsymbol{\chi}(\omega)\,, \tag{1.9}$$

or equivalently in index notations, for $i, j \in [\![1, 3]\!]$

$$\epsilon_{ij}(\omega) = \delta_{ij} + \chi_{ij}(\omega)\,, \tag{1.10}$$

where δ_{ij} is the Kronecker delta. Note that as the susceptibility, the permittivity takes complex values and is frequency dependent. In the case of an isotropic medium the permittivity tensor reduces to a scalar $\epsilon(\omega)$.

Drude and Lorentz models

We present now the simplest frequency dependent model for a material response to an external electric field, namely the Drude model [19]. We assume that the response is mainly due to a free electron gas subjected to the Lorentz force $-e\,\mathbf{E}$, where $-e$ is the elementary charge of the electron, and scattering from immobile ions modeled as a friction force $-m_e\,\gamma\,\mathbf{v}$, where m_e is the electron mass, γ is a characteristic collision frequency and $\mathbf{v}$ is the electron velocity. To the collision frequency γ is associated a characteristic time between consecutive collisions $\tau = 1/\gamma$. To simplify the discussion we restrict the modeling to one dimension. The equation of motion according the Newtonian mechanics reads

$$\frac{\mathrm{d}v}{\mathrm{d}t} = -\frac{e}{m_e}\, E - \gamma\, v\,. \tag{1.11}$$

By taking the Fourier transform with respect to time of the above equation we obtain

$$v(\omega) = -\frac{e}{m_e\,(\gamma - i\omega)}\, E(\omega)\,. \tag{1.12}$$

The associated current density is then given by

$$J = -e\, n_e v = \frac{\sigma_0}{1 - i\omega\tau}\, E \tag{1.13}$$

where n_e is the electron density and we define the dc (direct current) electric conductivity σ_0 as

$$\sigma_0 = \frac{n_e e^2}{m_e \gamma}\,. \tag{1.14}$$

Equation (1.13) has the form of Ohm's law where $\sigma(\omega) = \frac{\sigma_0}{1 - i\omega\tau}$ is a frequency dependent

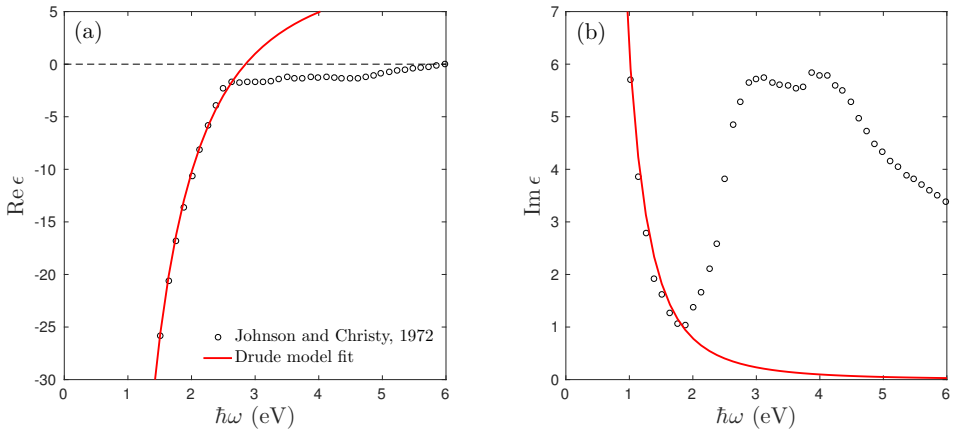


Figure 1.1: Real part (a) and imaginary part (b) of the dielectric function of gold as a function of photon energy $\hbar\omega$. The data points extracted from Ref. [20] are shown as open circles and a Drude model fitted to the data at low energy is shown as the solid red line. The parameter found for the fit were $\epsilon_\infty = 10.03$, $\hbar\omega_p = 9.02$ eV, and $\hbar\gamma = 0.08$ eV.

conductivity. Note that when $\omega \to 0$, i.e. the direct current limit, we obtain $\sigma(\omega) \to \sigma_0$ which justifies the denomination of dc conductivity for σ_0. Recalling that for general linear non-magnetic media the electric susceptibility tensor is related to the conductivity tensor by (see e.g. [16])

$$\chi_{ij}(\omega) = \frac{i\sigma_{ij}(\omega)}{\omega\varepsilon_0}\,, \tag{1.15}$$

(since we have $\mathbf{J}(\omega) = -i\omega\,\mathbf{P}(\omega)$ for non-magnetic media) the relative permittivity for the Drude model then reads

$$\epsilon(\omega) = 1 - \frac{\omega_p^2}{\omega^2 + i\gamma\omega} = 1 - \frac{\omega_p^2}{\omega^2+\gamma^2} + i\,\frac{\omega_p^2\gamma}{\omega\,(\omega^2+\gamma^2)}\,, \tag{1.16}$$

where we have defined the electron plasma frequency ω_p as $\omega_p^2 = \frac{n_e\,e^2}{m_e\,\varepsilon_0}$. The physical picture associated with the electron plasma frequency is that of a cloud of electrons initially displaced from its equilibrium around positive ions by an external constant electric field and then left to oscillate due to the electric field induced by the polarization of charges. The frequency of the oscillation is then defined as the electron plasma frequency. In the absence of dissipative processes, $\gamma = 0$, the relative permittivity is real and reads

$$\epsilon(\omega) = 1 - \frac{\omega_p^2}{\omega^2}\,. \tag{1.17}$$

We can remark that in the above equation, the plasma frequency ω_p acts as a threshold between positive and negative values of the permittivity. Indeed, $\epsilon > 0$ for $\omega > \omega_p$, $\epsilon < 0$ for $\omega < \omega_p$, and ϵ vanishes for $\omega = \omega_p$. We will discuss the physical interpretation of these three regimes with respect to electromagnetic waves in the next subsection.

A direct comparison between the dielectric function given by the Drude model fitted to experimental data and the experimental data itself shows that a satisfactory agreement may be found only at low photon energies $\hbar\omega$ (or equivalently large wavelengths). Figure 1.1 presents the real and imaginary parts of the dielectric function of gold as measured by Johnson and Christy [20] and the corresponding fit using a Drude model. A satisfactory agreement is found at low photon energies, but large discrepancies occur from around 2 eV and above, which is a signature of electron interband transitions. The effect of these interband transitions may be modeled by adding terms corresponding to elastically bound electrons with resonant frequency ω_j in Eq. (1.11), i.e. that it becomes

$$\frac{\mathrm{d}^2 x_j}{\mathrm{d}t^2} + \gamma_j \frac{\mathrm{d}x_j}{\mathrm{d}t} + \omega_j^2 x_j = -\frac{e}{m_e} E \tag{1.18}$$

where we now denote by x_j the displacement of such a bound electron from its equilibrium position. Such a model, when added up to the free electron model (Drude) yields the so-called Drude-Lorentz model

$$\epsilon(\omega) = \epsilon_\infty - \frac{\omega_p^2}{\omega^2 + i\gamma\omega} + \sum_j \frac{A_j}{\omega_j^2 - \omega^2 - i\gamma_j\omega}\,, \tag{1.19}$$

where the coefficients ϵ_∞, ω_p^2, γ, ω_j, γ_j, and A_j are to be fitted to experimental data. Values for these coefficients have been reported for different metals for example in Refs. [21, 22]. In the rest of this manuscript, the Drude model may be used to get qualitative physical insight due to its simplicity and physical interpretation, but when dealing with quantitative comparisons of simulation with real experiments, the dielectric functions will be taken directly from fits to experiments (see in particular Chapter 12).

1.1.3 Structure of electromagnetic waves in unbounded media

In the previous sub-section we have discussed some aspects of the modeling of the response of a material medium. In particular, we have introduced the notion of relative permittivity. We will now discuss different types of waves that can be present under certain conditions on the permittivity. We consider here the case of a non magnetic ($\mathbf{M} = \mathbf{0}$) material medium filling the whole Euclidian space. In addition, we assume the absence of free sources. Under these hypotheses, we can derive the wave equation for electromagnetic waves by taking the curl of Eq. (1.1c) and by using Eq. (1.1d). We obtain

$$\Delta\,\mathbf{E} - \nabla(\nabla\cdot\mathbf{E}) - \frac{1}{c^2}\,\frac{\partial^2\,\mathbf{E}}{\partial t^2} = \mu_0\,\frac{\partial^2\,\mathbf{P}}{\partial t^2}\,. \tag{1.20}$$

Here we have used the identity $\nabla\times\nabla\times\mathbf{E} = \nabla(\nabla\cdot\mathbf{E}) - \Delta\,\mathbf{E}$ where $\Delta\,\mathbf{E}$ denotes the vectorial Laplace operator, in the sense of a vector whose components are the Laplacians of the respective components of $\mathbf{E}$. We have also used the definitions of the magnetic induction and electric displacement. If we make the assumptions that the response of the medium obeys to the locality in space, homogeneity, and linearity, the above equation can be simplified after taking the Fourier transform with respect to time as

$$\Delta\,\mathbf{E} - \nabla(\nabla\cdot\mathbf{E}) + \frac{\omega^2}{c^2}\,\boldsymbol{\epsilon}\;\mathbf{E} = \mathbf{0}\,. \tag{1.21}$$

Where we have used Eq. (1.6) and the definition of the relative permittivity Eq. (1.9). Note that the permittivity is frequency dependent.

Isotropic media

In the case of an *isotropic* medium, we have $0 = \nabla\cdot\mathbf{D} = \varepsilon_0\epsilon\nabla\cdot\mathbf{E}$, hence for non zero permittivity, $\nabla\cdot\mathbf{E} = 0$ and we obtain the so-called Helmholtz equation

$$\Delta\,\mathbf{E} + \frac{\omega^2}{c^2}\,\epsilon\;\mathbf{E} = \mathbf{0}\,. \tag{1.22}$$

By taking the Fourier transform of Eq. (1.22) with respect to the spatial variables, we obtain

$$\left(\epsilon\,\frac{\omega^2}{c^2} - \mathbf{k}^2\right)\;\mathbf{E} = \mathbf{0}\,. \tag{1.23}$$

Thus the Helmholtz equation admits non-trivial solutions if and only if the dispersion relation

$$\epsilon\,\frac{\omega^2}{c^2} - \mathbf{k}^2 = 0 \tag{1.24}$$

is satisfied. In other words, in the Fourier expansion of the field only plane wave components with a wave vector lying on a sphere of radius $\sqrt{\epsilon}\,\frac{\omega}{c}$ can be present. In addition, the relation $\nabla\cdot\mathbf{E} = 0$ translates to $\mathbf{k}\cdot\mathbf{E} = 0$ in Fourier space, which means that the electric field is necessarily orthogonal to its wave vector. This type of wave is known as a transverse wave. However, this is only true for non-zero permittivity. Indeed, for zero permittivity there is no condition on $\nabla\cdot\mathbf{E}$ anymore and the electric field can possess a component along its wave vector. A wave whose electric field is along the wave vector is called a longitudinal wave. This type of wave is associated with the oscillation of charges within the material. This can be understood by considering the Drude model. We have seen that the zero of the permittivity occurs at the plasma frequency, whose physical interpretation is a collective oscillation of electrons.

Anisotropic media

Anisotropic media are characterized by a susceptibility tensor, or dielectric tensor, which does not reduce to a scalar. The first consequence of the anisotropy is that $\nabla\cdot\mathbf{E} = 0$ does not necessary hold anymore. In particular, $\mathbf{E}$ and $\mathbf{D}$ are not necessarily colinear, and the wave vector, which is orthogonal to $\mathbf{D}$ and the Poynting's vector which is orthogonal to $\mathbf{E}$ are no longer colinear. This means that the direction of propagation is not orthogonal to the planes of constant phase anymore. Such a wave is known as an *extraordinary wave* by opposition to an *ordinary wave* whose properties resemble that of a wave in an isotropic medium. It can be shown that the dielectric tensor is symmetric and is therefore diagonalizable. Assuming the tensor to be represented in a basis in which it is diagonal, the Fourier transform of Eq. (1.21) yields for $i \in [\![1, 3]\!]$

$$\sum_{j=1}^{3}\left[\left(\epsilon_j\,\frac{\omega^2}{c^2} - \mathbf{k}^2\right)\delta_{ij} + k_i\,k_j\right]E_j = 0\,, \tag{1.25}$$

where the ϵ_j are the dielectric constants along the principal axis of the medium. The homogeneous system of linear equations, Eq. (1.25), admits non-trivial solutions if and only if the determinant vanishes, which yields after some algebra to the dispersion relation [18]

$$\sum_{i=1}^{3}\frac{k_i^2}{\frac{\mathbf{k}^2}{\epsilon_i} - \frac{\omega^2}{c^2}} = \sum_{i=1}^{3}\frac{\hat{k}_i^2}{\frac{1}{\epsilon_i} - \frac{1}{\epsilon(\mathbf{k})}} = 0\,. \tag{1.26}$$

In the last equation we have defined $\hat{\mathbf{k}} = \mathbf{k}\,/|\,\mathbf{k}\,|$ and the effective dielectric constant $\epsilon(\mathbf{k})$ seen by the wave vector $\mathbf{k}$ defined as $|\,\mathbf{k}\,| \stackrel{\text{def}}{=} \sqrt{\epsilon(\mathbf{k})}\,\omega/c$. By multiplying Eq. (1.26) by $\prod_{i=1}^{3}(1/\epsilon_i - 1/\epsilon)$

we obtain a polynomial equation of degree two in $1/\epsilon$ which admits two solutions. The two solutions describe two surfaces in ϵ-space, or equivalently in the index space $n^2 = \epsilon$, as $\hat{\mathbf{k}}$ sweeps over the unit sphere. In the case of a uniaxial anisotropic material, $\epsilon_1 = \epsilon_2 \neq \epsilon_3$, it can be shown that one of the index surface is a sphere, and hence correspond to the aforementioned ordinary waves while the second surface is an ellipsoid, corresponding to the aforementioned extraordinary waves.

In this work, only systems made of optically isotropic material will be considered. This does not mean, however, that we will obtain isotropic behavior as the systems of interest may be composed of structured surfaces or of particles supported on a substrate which, as composite systems, are anisotropic. The anisotropy considered in this work is then a *structural anisotropy* rather than an anisotropy associated with the bulk response of the constituents of the system.

1.2 Scattering of electromagnetic waves

We formulate now the problem that we will deal with in the rest of this work, namely that of the scattering of electromagnetic waves by arbitrary surfaces or composite systems. Let us first precise that we will deal exclusively with time harmonic problems, i.e. that we consider a stationary regime where the time contribution to the electromagnetic field enters only via a phase factor $\exp(-i\omega t)$, where ω is the angular frequency of the excitation. Second, we will only consider isotropic media separated by arbitrary interfaces. The prototypical system will be that of two semi-infinite media separated by an arbitrary interface, but we will also consider stack of layers with distinct optical properties separated by arbitrary interfaces and special, or limiting cases resulting in particles supported by a substrate or embedded in a layer. We formulate in the following the problem for a system made of two semi-infinite media separated by an arbitrary interface, as sketched in Fig. 1.2(a), but the reader will extend it without difficulty to the remaining systems.

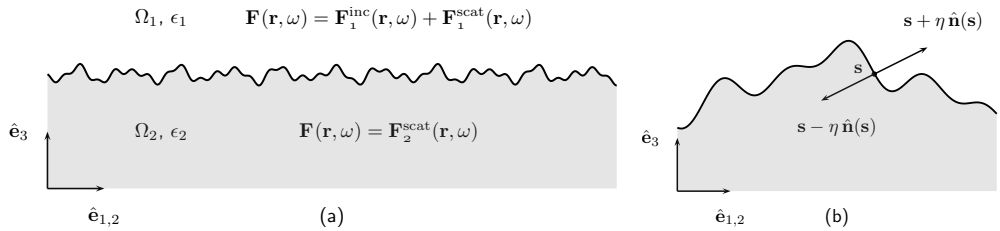


Figure 1.2: (a) A side view illustration of the scattering system, (b) a close up for the points involved in the boundary conditions.

1.2.1 Boundary and radiation conditions

We will denote by Ω_1 (resp. Ω_2) the open half-space situated above (resp. below) the interface $\partial\Omega$, as sketch in Fig. 1.2(a) and hosting a material of dielectric function ϵ_1 (resp. ϵ_2). Given an incident electromagnetic field $(\mathbf{E}^{\mathrm{inc}}, \mathbf{H}^{\mathrm{inc}})$, the scattering problem is a boundary value problem

and reads: find $(\mathbf{E}^{\text{scat}}, \mathbf{H}^{\text{scat}})$ such that

$$\begin{cases} \Delta\,\mathbf{F}(\mathbf{r},\omega) + \frac{\omega^2}{c^2}\,\epsilon_1(\omega)\;\mathbf{F}(\mathbf{r},\omega) = 0\,, & \forall\,\mathbf{r}\in\Omega_1\,,\forall\,\omega>0\,,\\ \Delta\,\mathbf{F}(\mathbf{r},\omega) + \frac{\omega^2}{c^2}\,\epsilon_2(\omega)\;\mathbf{F}(\mathbf{r},\omega) = 0\,, & \forall\,\mathbf{r}\in\Omega_2\,,\forall\,\omega>0\,,\\ \hat{\mathbf{n}}(\mathbf{s})\times\lim\limits_{\eta\to 0}\,[\mathbf{F}(\mathbf{s}+\eta\,\hat{\mathbf{n}}(\mathbf{s}),\omega) - \mathbf{F}(\mathbf{s}-\eta\,\hat{\mathbf{n}}(\mathbf{s}),\omega)] = 0\,, & \forall\,\mathbf{s}\in\partial\Omega\,,\forall\,\omega>0\,. \end{cases} \tag{1.27}$$

Here $\mathbf{F}$ is either the total electric field $\mathbf{E}$ or magnetic field $\mathbf{H}$, $\mathbf{s}$ is a point on the interface $\partial\Omega$, η is a positive real number and $\hat{\mathbf{n}}(\mathbf{s})$ is the surface normal oriented from medium 2 to medium 1 at the point $\mathbf{s}$. The last equation, is the condition of continuity of the tangential component of the electric and magnetic fields at the boundary. In practice, the incident field will be an incident wave incoming from infinity in the first medium. The total field hence reads

$$\mathbf{F}(\mathbf{r},\omega) = \begin{cases} \mathbf{F}_1^{\text{inc}}(\mathbf{r},\omega) + \mathbf{F}_1^{\text{scat}}(\mathbf{r},\omega)\,, & \forall\,\mathbf{r}\in\Omega_1\,,\forall\,\omega>0\,,\\ \mathbf{F}_2^{\text{scat}}(\mathbf{r},\omega)\,, & \forall\,\mathbf{r}\in\Omega_2\,,\forall\,\omega>0\,, \end{cases} \tag{1.28}$$

and in the case of an incident plane wave, the electric field will be of the form $\mathbf{E}_1^{\text{inc}}(\mathbf{r},\omega) = \mathbf{E}_0\,\exp(i\,\mathbf{k}\cdot\mathbf{r} - i\omega t)$, where $\mathbf{k}$ is a wave vector satisfying the dispersion relation in medium 1 and directed downward. In addition to the boundary condition at the interface between the two media, we need a radiation condition at infinity. For large $r = |\,\mathbf{r}\,|$, and for loss-less dielectric media, the scattered field must satisfy the Silver-Müller[2] radiation condition [23]

$$\left|\frac{\partial\,\mathbf{F}_j^{\text{scat}}}{\partial r} - ik_j\,\mathbf{F}^{\text{scat}}\right| \leq \frac{M}{r^2}\,, \tag{1.29}$$

where M is some positive constant and $k_j = \epsilon_j^{1/2}\omega/c$ (for $j\in\{1,2\}$) is the wave number in medium j. This radiation condition together with the boundary conditions ensures a unique solution to the scattering problem. In addition, we have the following behavior as $|\,\mathbf{r}\,|\to\infty$ [23]

$$|\,\mathbf{F}^{\text{scat}}\,| \leq \frac{M}{r}\,. \tag{1.30}$$

This can also be restated by saying that the radiated flux decays as $1/r^2$. For absorbing media, the field decays exponentially at infinity. An interpretation of the Silver-Müller radiation condition, Eq. (1.29), is that for large r, the field behaves as an outgoing polarized spherical wave. Another way of phrasing this is that r times the field in $1/r$ behaves as a plane wave. This claim will be made clearer when we will derive the far-field expansion from the expansion in plane waves of the scattered field in Section 1.2.3.

Remark 1.1. Note that an outgoing plane wave *does not* satisfy the Silver-Müller radiation condition. In the case of an incident plane wave on an infinite surface, we will see that the Silver-Müller radiation condition may not be appropriate. Indeed, it suffices to consider the case of reflection by a planar surface to foresee an issue. For reflection by a planar surface, it is well known that the incident plane wave will be reflected as a plane wave, which is outgoing with respect to the interface but does not satisfy the Silver-Müller radiation condition Eq. (1.29). The source of the problem is that the integrated flux received by the scatterer is infinite, or in other words, the plane wave is not a wave of finite energy (e.g. not normalizable in $L^2(\mathbb{R}^3)$). In such a case, the Silver-Müller radiation condition should be relaxed in some sense to allow for solutions containing outgoing plane waves. From a physical point of view, this means that we consider

[2]The name of Sommerfeld radiation condition is also used although it seems to be more associated to the problem of scattering of scalar waves.

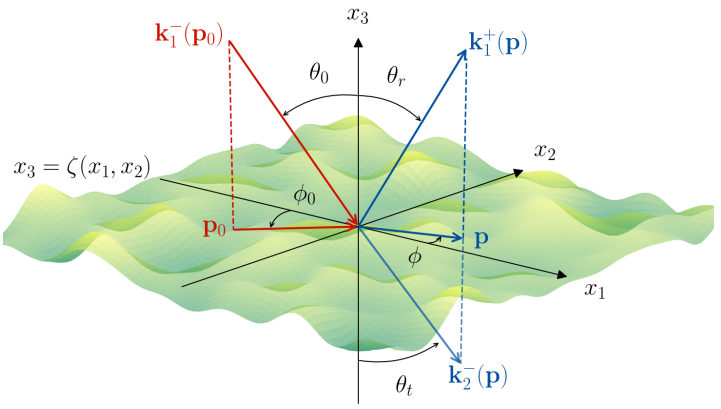


Figure 1.3: Sketch of the scattering system and the definitions of the angles of incidence (θ_0, ϕ_0) and scattering $(\theta_r, \theta_t, \phi)$, together with the relevant wave vectors. The vector $\mathbf{p}_0$ (resp. $\mathbf{p}$) is the projection of the wave vector of the incident wave $\mathbf{k}_1^-(\mathbf{p}_0)$ (reps. reflected or transmitted wave $\mathbf{k}_1^+(\mathbf{p})$ or $\mathbf{k}_2^-(\mathbf{p})$) in the vector plane $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2)$.

the scattering of an incident plane wave as a tool to further compute more physically relevant cases such as the scattering of a finite size beam (and normalizable) which can be constructed as a linear superposition of plane waves (Fourier representation). The response being linear, the scattered wave will hence be the superposition of the response for each individual plane wave leading to an overall scattered field which satisfies the Silver-Müller radiation condition.

1.2.2 Reflection and transmission amplitudes

It will be convenient, in our work, to expand in plane waves the fields in each medium. Indeed, plane waves satisfy by construction the Helmholtz equation and solving the scattering problem will hence reduce to finding the set of amplitudes of the field expansions via the use of the boundary conditions. The expansion in plane waves of the electric field in medium j reads

$$\mathbf{E}_j(\mathbf{r},\omega) = \sum_{a=\pm} \int_{\mathbb{R}^2} \left[\mathcal{E}^a_{j,p}(\mathbf{q},\omega)\,\hat{\mathbf{e}}^a_{p,j}(\mathbf{q}) + \mathcal{E}^a_{j,s}(\mathbf{q},\omega)\,\hat{\mathbf{e}}_s(\mathbf{q})\right] \exp\left(i\,\mathbf{k}^a_j(\mathbf{q})\cdot\mathbf{r} - i\omega t\right) \frac{\mathrm{d}^2 q}{(2\pi)^2}\,, \tag{1.31}$$

where we define

$$\mathbf{k}^\pm_j(\mathbf{q}) \overset{\text{def}}{=} \mathbf{q} \pm \alpha_j(\mathbf{q})\,\hat{\mathbf{e}}_3\,, \tag{1.32a}$$

$$\alpha_j(\mathbf{q}) \overset{\text{def}}{=} \sqrt{\epsilon_j\left(\frac{\omega}{c}\right)^2 - \mathbf{q}^2}, \qquad \mathrm{Re}\,(\alpha_j), \mathrm{Im}\,(\alpha_j) \geq 0\,, \tag{1.32b}$$

$$\hat{\mathbf{e}}_s(\mathbf{q}) \overset{\text{def}}{=} \frac{\hat{\mathbf{e}}_3 \times \mathbf{k}^\pm_j(\mathbf{q})}{|\hat{\mathbf{e}}_3 \times \mathbf{k}^\pm_j(\mathbf{q})|} = \hat{\mathbf{e}}_3 \times \hat{\mathbf{q}}\,, \tag{1.32c}$$

$$\hat{\mathbf{e}}^\pm_{p,j}(\mathbf{q}) \overset{\text{def}}{=} \frac{\hat{\mathbf{e}}_s \times \mathbf{k}^\pm_j(\mathbf{q})}{|\hat{\mathbf{e}}_s \times \mathbf{k}^\pm_j(\mathbf{q})|} = \frac{c}{\sqrt{\epsilon_j}\omega}\left(\pm\alpha_j(\mathbf{q})\,\hat{\mathbf{q}} - |\mathbf{q}|\,\hat{\mathbf{e}}_3\right). \tag{1.32d}$$

In other words, the wave vector $\mathbf{k}^\pm_j(\mathbf{q})$ of an elementary plane wave is decomposed into its projection $\mathbf{q}$ in the lateral vector plane $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2)$ and the component $\pm\alpha_j(\mathbf{q})$ along $\hat{\mathbf{e}}_3$ (see

Fig. 1.3). The sum for $a = \pm$ takes into account both upwards and downwards propagating and evanescent (and possibly growing) waves. The field amplitude is decomposed in the *local polarization basis* $(\hat{\mathbf{e}}^a_{p,j}(\mathbf{q}), \hat{\mathbf{e}}_s(\mathbf{q}))$, hence $\mathcal{E}^a_{j,\alpha}(\mathbf{q})$ denotes the component of the field amplitude in the polarization state α of the mode characterized by a and $\mathbf{q}$. In this basis, the directions given by $\hat{\mathbf{e}}^{\pm}_{p,j}(\mathbf{q})$, and $\hat{\mathbf{e}}_s(\mathbf{q})$ are respectively the directions of the p- and s-polarization of the electric field amplitude. From Eq. (1.32) it is clear that the vectors $(\hat{\mathbf{k}}^{\pm}_j(\mathbf{q}), \hat{\mathbf{e}}^{\pm}_{p,j}(\mathbf{q}), \hat{\mathbf{e}}_s(\mathbf{q}))$ (with $\hat{\mathbf{k}}^{\pm}_j(\mathbf{q}) = \mathbf{k}^{\pm}_j(\mathbf{q}) / |\,\mathbf{k}^{\pm}_j(\mathbf{q})\,|$) define a direct orthonormal basis of the three dimensional Euclidean space.

Assume now that the incident excitation consists solely of a monochromatic plane wave in medium 1 defined as

$$\mathbf{E}_0(\mathbf{r},\omega) = \exp\left(i\,\mathbf{k}^-_1(\mathbf{p}_0)\cdot\mathbf{r} - i\omega t\right)\ \left[\mathcal{E}_{0,p}\ \hat{\mathbf{e}}^-_{p,1}(\mathbf{p}_0) + \mathcal{E}_{0,s}\ \hat{\mathbf{e}}_s(\mathbf{p}_0)\right]\ , \tag{1.33}$$

where $\mathbf{p}_0$ is the projection of the incident wave's wave vector in the $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2)$ plane, with the property $|\,\mathbf{p}_0\,| \leq \sqrt{\epsilon_1}\,\omega/c$, i.e. that we consider an incident wave which propagates (also called progressive). The fact that this is the only incident wave considered together with an outgoing wave radiation condition gives that, apart from the incident field, the only elementary waves allowed in the scattered field are those with wave vectors of the form $\mathbf{k}^+_1(\mathbf{p})$ and $\mathbf{k}^-_2(\mathbf{p})$ respectively in media 1 and 2. This is summarized in the following equations for the field amplitudes

$$\boldsymbol{\mathcal{E}}^-_1(\mathbf{q}) = (2\pi)^2\,\delta(\mathbf{q}-\mathbf{p}_0)\ \boldsymbol{\mathcal{E}}_0\ , \tag{1.34a}$$
$$\boldsymbol{\mathcal{E}}^+_2(\mathbf{q}) = \mathbf{0}\ , \tag{1.34b}$$

where we have denoted $\boldsymbol{\mathcal{E}}_j = (\mathcal{E}_{j,p}, \mathcal{E}_{j,s})^{\mathrm{T}}$ the vector of p- and s- components. Next, we assume that the scattered field amplitudes are linearly related to the incident field amplitude $\boldsymbol{\mathcal{E}}_0$ via the reflection and transmission amplitudes, $\mathbf{R}(\mathbf{q} \mid \mathbf{p}_0)$ and $\mathbf{T}(\mathbf{q} \mid \mathbf{p}_0)$, defined as

$$\boldsymbol{\mathcal{E}}^+_1(\mathbf{q}) = \mathbf{R}(\mathbf{q}\,|\,\mathbf{p}_0)\ \boldsymbol{\mathcal{E}}_0\ , \tag{1.35a}$$
$$\boldsymbol{\mathcal{E}}^-_2(\mathbf{q}) = \mathbf{T}(\mathbf{q}\,|\,\mathbf{p}_0)\ \boldsymbol{\mathcal{E}}_0\ . \tag{1.35b}$$

The reflection and transmission amplitudes are therefore described by 2×2 matrices with complex entries of the form

$$\mathbf{X} = \begin{pmatrix} X_{pp} & X_{ps} \\ X_{sp} & X_{ss} \end{pmatrix}\ , \tag{1.36}$$

for $\mathbf{X} = \mathbf{R}$ or $\mathbf{T}$. From a physical point of view, the coefficient $R_{\alpha\beta}(\mathbf{q}\,|\,\mathbf{p}_0)$ (resp. $T_{\alpha\beta}(\mathbf{q}\,|\,\mathbf{p}_0)$) for $\alpha, \beta \in \{p, s\}$ is the field amplitude reflected (resp. transmitted) with lateral wave vector $\mathbf{q}$ in the polarization state α from a unit incident field with lateral wave vector $\mathbf{p}_0$ in the polarization state β. The reflection and transmission amplitudes are then the unknowns in our scattering problem. The aim of the next chapter will be to derive integral equations satisfied by the reflection and transmission amplitudes.

Remark 1.2. In subsequent chapters, when the discussion is made general for either the reflection or transmission amplitudes, we may simply refer to them as *scattering amplitudes*.

1.2.3 Far-field

In a scattering experiment, one can often only observe the intensity scattered far way from the scatterer. Before defining in details the different observables of interest for a scattering experiment, which we will do in the next section, we need to define what is meant by *far-field*. Intuitively, the name speaks for itself, it is the field far way from the scatterer, say many wavelengths away, so in a region where $\lambda = 2\pi c/\sqrt{\epsilon}\omega \ll r$, where r is the distance between the point of observation and the scatterer, and ϵ denotes the dielectric constant in the medium of observation. In such a region, we can derive an asymptotic expression for the field that proves to be very useful for defining far-field observables based on the reflection and transmission amplitudes.

Consider the scattered field $\mathbf{E}_1^+(\mathbf{r})$ which we assume to be square integrable and can be expanded in plane waves as

$$\mathbf{E}_1^+(\mathbf{r}) = \int_{\mathbb{R}^2} \mathbf{E}_1^+(\mathbf{q}) \exp\left(i\, \mathbf{k}_1^+(\mathbf{q}) \cdot \mathbf{r}\right) \frac{\mathrm{d}^2 q}{(2\pi)^2}\,, \tag{1.37}$$

where we drop the time dependence that only contributes by an overall phase factor, and $\mathbf{E}_1^+(\mathbf{q}) = \mathcal{E}_{1,p}^+(\mathbf{q},\omega)\,\hat{\mathbf{e}}_{p,1}^+(\mathbf{q}) + \mathcal{E}_{1,s}^+(\mathbf{q},\omega)\,\hat{\mathbf{e}}_s(\mathbf{q})$ as compared to the definition of the plane wave expansion in Eq. (1.31). Since we are interested in an observation point $\mathbf{r}$ far away from the scatterer, we can restrict the domain of integration to $\{\mathbf{q} \in \mathbb{R}^2 \mid |\,\mathbf{q}\,| < \sqrt{\epsilon_1}\omega/c\}$ since for in-plane wave vector outside this disk the wave components are evanescent, decay exponentially away from the surface and do not contribute to the far-field. Let $\mathbf{p}$ denote the in-plane wave vector constructed in such a way that $\mathbf{k}_1^+(\mathbf{p})$ is colinear with the observation point $\mathbf{r}$. In other words, $\mathbf{k}_1^+(\mathbf{p}) = \sqrt{\epsilon_1}\omega/c\,\hat{\mathbf{e}}_r$, with $\hat{\mathbf{e}}_r = \mathbf{r}\,/|\,\mathbf{r}\,|$ which we may write as

$$\hat{\mathbf{e}}_r = \sin\theta_r \cos\phi_r\, \hat{\mathbf{e}}_1 + \sin\theta_r \sin\phi_r\, \hat{\mathbf{e}}_2 + \cos\theta_r\, \hat{\mathbf{e}}_3\,. \tag{1.38}$$

The polar angle θ_r and the azimuthal angle ϕ_r then parametrize the direction of observation. Equation (1.37) can be written as

$$\mathbf{E}_1^+(\mathbf{r}) = \frac{k^2}{(2\pi)^2} \int_{|\,\tilde{\mathbf{q}}\,|<1} \mathbf{E}_1^+(\mathbf{q}) \exp\left(ikr\, \hat{\mathbf{k}}_1^+(\mathbf{q}) \cdot \hat{\mathbf{e}}_r\right) \mathrm{d}^2\tilde{q}\,, \tag{1.39}$$

where we have introduced the wave number $k = \sqrt{\epsilon_1}\omega/c$, $r = |\,\mathbf{r}\,|$ and $\hat{\mathbf{k}}_1^+ = \mathbf{k}_1^+/k$. In addition, we have made a change of variable, $\mathbf{q} = k\,\tilde{\mathbf{q}}$. The scalar product in the exponential reads in terms of the angles of observation, also called *angles of reflection*, as

$$\hat{\mathbf{k}}_1^+(\mathbf{q}) \cdot \hat{\mathbf{e}}_r = \tilde{q}_1 \sin\theta_r \cos\phi_r + \tilde{q}_2 \sin\theta_r \sin\phi_r + (1 - \tilde{q}_1^2 - \tilde{q}_2^2)^{1/2} \cos\theta_r\,. \tag{1.40}$$

The form of the right-hand side in the above equation invites us to make a first change of variable to cylindrical coordinates $\tilde{\mathbf{q}} = \tilde{q}(\cos\phi\,\hat{\mathbf{e}}_1 + \sin\phi\,\hat{\mathbf{e}}_2)$ so that

$$\begin{aligned} \hat{\mathbf{k}}_1^+(\mathbf{q}) \cdot \hat{\mathbf{e}}_r &= \tilde{q}\sin\theta_r\,(\cos\phi\cos\phi_r + \sin\phi\sin\phi_r) + (1-\tilde{q}^2)^{1/2}\cos\theta_r\,, \\ &= \tilde{q}\sin\theta_r\cos(\phi-\phi_r) + (1-\tilde{q}^2)^{1/2}\cos\theta_r\,, \end{aligned} \tag{1.41}$$

and then a second change of variable $\tilde{q} = \sin\theta$ to finally obtain

$$\hat{\mathbf{k}}_1^+(\mathbf{q}) \cdot \hat{\mathbf{e}}_r = \sin\theta\sin\theta_r\cos(\phi-\phi_r) + \cos\theta\cos\theta_r\,. \tag{1.42}$$

Note that for the angle $\phi = \phi_r$, $\cos(\phi - \phi_r)$ becomes unity and the scalar product hence reads $\cos(\theta - \theta_r)$, which is unity for $\theta = \theta_r$. In other words, the scalar product is unity when $\mathbf{q} = \mathbf{p}$. This is a clear indication that the observation angles ϕ_r and θ_r play roles of angles of reference,

and that the wave component $\mathbf{E}_1^+(\mathbf{q}=\mathbf{p})$ is likely to contribute the most to the integral. After the aforementioned changes of variables, the field expansion reads

$$\mathbf{E}_1^+(\mathbf{r}) = \frac{k^2}{(2\pi)^2}\int_0^{\pi/2} e^{ikr\,\cos\theta\cos\theta_r}\int_0^{2\pi}\mathbf{E}_1^+(\mathbf{q}(\theta,\phi))\,e^{ikr\,\sin\theta\sin\theta_r\cos(\phi-\phi_r)}\,\mathrm{d}\phi\,\sin\theta\,\cos\theta\,\mathrm{d}\theta\,. \tag{1.43}$$

It is now time to use the fact that the point of observation is far away from the surface, so that $1 \ll kr$, and use the method of *stationary phase* to approximate the integrals [24]. For an integral of the type

$$I(x) = \int_a^b f(t)\exp\left(ix\psi(t)\right)\mathrm{d}t\,, \tag{1.44}$$

where f and ψ are real valued, continuously differentiable functions on $[a,b]$, and if in addition ψ' vanishes at only one point $\tau \in]a,b[$ with $\psi''(\tau) < 0$, then as $x \to \infty$, we have

$$I(x) = \left(\frac{2\pi}{-x\psi''(\tau)}\right)^{1/2} f(\tau)\,\exp\left(ix\psi(\tau) - i\frac{\pi}{4}\right) + \mathcal{O}(1/x)\,. \tag{1.45}$$

Applying this result first to the integral over ϕ in Eq. (1.43) with $x = kr$, $f(\phi) = \mathbf{E}_1^+(\mathbf{q}(\theta,\phi))$ and $\psi(\phi) = \sin\theta\sin\theta_r\cos(\phi-\phi_r)$, we find that: $\psi'(\phi) = -\sin\theta\sin\theta_r\sin(\phi-\phi_r)$ vanishes at $\phi = \phi_r$ and $\psi''(\phi_r) = -\sin\theta\sin\theta_r < 0$, hence

$$\begin{aligned}&\int_0^{2\pi}\mathbf{E}_1^+(\mathbf{q}(\theta,\phi))\,e^{ikr\,\sin\theta\sin\theta_r\cos(\phi-\phi_r)}\,\mathrm{d}\phi\\ &\quad= \left(\frac{2\pi}{kr\sin\theta\sin\theta_r}\right)^{1/2}\mathbf{E}_1^+(\mathbf{q}(\theta,\phi_r))\,\exp\left(ikr\sin\theta\sin\theta_r - i\frac{\pi}{4}\right) + \mathcal{O}(1/kr)\,.\end{aligned} \tag{1.46}$$

Substituting the above result into Eq. (1.43) yields after some simplifications

$$\mathbf{E}_1^+(\mathbf{r}) = \frac{k^{3/2}}{(2\pi)^{3/2}r^{1/2}}\int_0^{\pi/2}\mathbf{E}_1^+(\mathbf{q}(\theta,\phi_r))\,e^{ikr\,\cos(\theta-\theta_r)-i\pi/4}\,\frac{\sin^{1/2}\theta\,\cos\theta}{\sin^{1/2}\theta_r}\,\mathrm{d}\theta + \mathcal{O}\big((kr)^{-3/2}\big)\,. \tag{1.47}$$

Repeating the stationary phase method for the integral over θ for $f(\theta) = \mathbf{E}_1^+(\mathbf{q}(\theta,\phi_r))e^{-i\pi/4}\sin^{1/2}\theta\,\cos\theta/\sin^{1/2}\theta_r$ and $\psi(\theta) = \cos(\theta-\theta_r)$, we find that: $\psi'(\theta) = -\sin(\theta-\theta_r)$ vanishes at $\theta = \theta_r$ and $\psi''(\theta_r) = -1 < 0$, hence after simplifications

$$\mathbf{E}_1^+(\mathbf{r}) = -i\frac{k}{2\pi r}\,\exp\left(ikr\right)\cos\theta_r\;\mathbf{E}_1^+(\mathbf{p}) + \mathcal{O}\big((kr)^{-3/2}\big)\,, \tag{1.48}$$

where we remind the reader that $k = \sqrt{\epsilon_1}\omega/c$. Equation (1.48) is a well known asymptotic expansion of the far-field that can be found in the literature (see e.g. [25, 26]).

Remark 1.3. The reader will easily verify that the far-field given in Eq. (1.48) satisfies the Silver-Müller radiation condition Eq. (1.29), and as claimed in Section 1.2.1, that the far-field times r behaves as a plane wave, or in other words the far field behaves locally as a spherical wave.

We are now ready to define far-field observables for scattering experiments.

1.2.4 Observables

The reflection and transmission amplitudes are complex valued and do not correspond to physical quantities that can be measured experimentally[3]. An important role that one expects from a theory is to make prediction that can be tested experimentally. Therefore, we need to define now what we will call *observables*, or physical quantities that can be experimentally measured. These will be defined theoretically through the scattering amplitudes but they can be obtained experimentally by measuring the intensity of the scattered light within some solid angles around some specific scattering directions. By placing a polarizer between the source and the system, and by placing an analyzer in front of the detector, the different quantities for co- or cross-polarized states can be obtained.

Differential scattering coefficients

Assuming we have obtained the reflection amplitudes $R_{\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0)$, we can now proceed to express the *differential reflection coefficient*[4] (DRC) defined as the time-averaged flux radiated around a given scattering direction (θ_r, ϕ_r) per unit solid angle and per unit incident flux and denoted $\partial R/\partial\Omega_r(\mathbf{p} \mid \mathbf{p}_0)$. Let a virtual hemisphere of radius $r \gg \lambda_1$ ($\lambda_1 = 2\pi c/(\epsilon_1^{1/2}\omega)$) lie on the plane $x_3 = 0$ on top of the scattering system. The support of this hemisphere is a disk of area $S = \pi r^2$. We consider the scattering from a *truncated* version of the scattering system in which the surface profile is set to be flat outside the disk support. Consequently, the field amplitudes we will manipulate are not strictly speaking those of the full system of interest but will converge to them as $r \to \infty$. For this reason we denote the reflection amplitudes for the truncated system $R^{(S)}_{\alpha\beta}$, where the super-script indicates that only the disk of area S supports the non-planar part of the surface. The time-averaged flux incident on this disk is given by

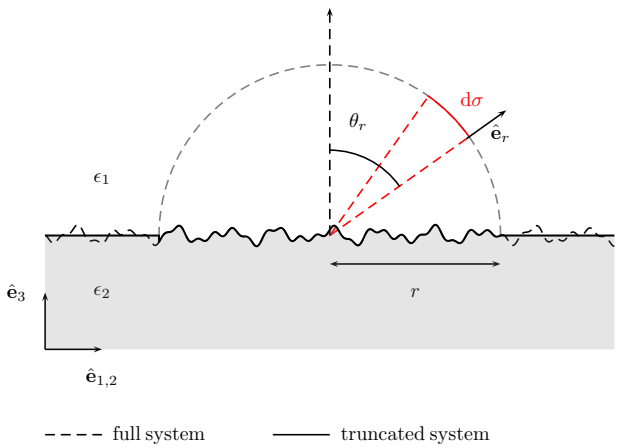


Figure 1.4: A side view illustration of the geometry considered for the determination of the differential reflection coefficient. The truncated profile is represented as a solid line, while the full profile is sketched as a dashed line; they coincide within the disk support of radius r.

[3] We exclude the use of interferometry.

[4] The differential reflection coefficient is also known as *bidirectional reflectance distribution function* (BRDF).

$$
\begin{aligned}
P_{\mathrm{inc}/S} &= -\frac{\varepsilon_0 c^2}{2\omega}\,\mathrm{Re}\int_S \left[\mathbf{E}_0^*(\mathbf{p}_0)\times\left(\mathbf{k}_1^-(\mathbf{p}_0)\times\mathbf{E}_0(\mathbf{p}_0)\right)\right]\cdot\hat{\mathbf{e}}_3\,\exp\left[-i(\mathbf{k}_1^{-*}(\mathbf{p}_0)-\mathbf{k}_1^-(\mathbf{p}_0))\cdot\mathbf{x}\right]\mathrm{d}^2x \\
&= -\frac{\varepsilon_0 c^2}{2\omega}\,\mathrm{Re}\int_S \left[|\,\mathbf{E}_0(\mathbf{p}_0)|^2\,\mathbf{k}_1^-(\mathbf{p}_0)-\left(\mathbf{E}_0^*(\mathbf{p}_0)\cdot\mathbf{k}_1^-(\mathbf{p}_0)\right)\mathbf{E}_0(\mathbf{p}_0)\right]\cdot\hat{\mathbf{e}}_3\,\mathrm{d}^2x \\
&= S\,\frac{\varepsilon_0 c^2}{2\omega}\,\alpha_1(\mathbf{p}_0)\,|\,\mathbf{E}_0(\mathbf{p}_0)|^2 = S\,\frac{\varepsilon_0 c}{2}\,\cos\theta_0\left(|\,\mathcal{E}_{0,p}\,|^2+|\,\mathcal{E}_{0,s}\,|^2\right). \qquad (1.49)
\end{aligned}
$$

Here, the * denotes the complex conjugate, the incident field amplitude is given by $\mathbf{E}_0(\mathbf{p}_0) = \mathcal{E}_{0,p}\,\hat{\mathbf{e}}_p^-(\mathbf{p}_0)+\mathcal{E}_{0,s}\,\hat{\mathbf{e}}_s(\mathbf{p}_0)$ as defined in Eq. (1.33). Moreover, the vector identity $\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = (\mathbf{a}\cdot\mathbf{c})\mathbf{b}-(\mathbf{a}\cdot\mathbf{b})\mathbf{c}$ and the orthogonality between the field and the wave vector $\mathbf{E}_0^*(\mathbf{p}_0)\cdot\mathbf{k}_1^-(\mathbf{p}_0)=0$ have been used. Note that the flux incident on the disk is proportional to the disk area. Let us now consider the outgoing flux crossing an elementary surface $\mathrm{d}\sigma = r^2\sin\theta_r\mathrm{d}\theta_r\mathrm{d}\phi_r = r^2\mathrm{d}\Omega_r$ around a point $\mathbf{r} = r\,(\sin\theta_r\cos\phi_r\,\hat{\mathbf{e}}_1+\sin\theta_r\sin\phi_r\,\hat{\mathbf{e}}_2+\cos\theta_r\,\hat{\mathbf{e}}_3) = r\,\hat{\mathbf{e}}_r$. The flux crossing this elementary surface is given by

$$
P_{\mathrm{d}\sigma} = \frac{1}{2\mu_0}\,\mathrm{Re}\left[\mathbf{E}_1^{+*}(\mathbf{r})\times\mathbf{B}_1^+(\mathbf{r})\right]\cdot\hat{\mathbf{e}}_r\,\mathrm{d}\sigma\,. \qquad (1.50)
$$

We then use the asymptotic expansion of the field in the far-field region given by Eq. (1.48) (see also [25, 26])

$$
\mathbf{E}_1^+(\mathbf{r}) \sim -i\,\epsilon_1^{1/2}\,\frac{\omega}{2\pi\,c}\cos\theta_r\,\frac{\exp(i\epsilon_1^{1/2}\frac{\omega}{c}r)}{r}\;\mathbf{E}_1^+(\mathbf{p})\,, \qquad (1.51a)
$$

$$
\mathbf{B}_1^+(\mathbf{r}) \sim -i\,\epsilon_1\,\frac{\omega}{2\pi\,c^2}\cos\theta_r\,\frac{\exp(i\epsilon_1^{1/2}\frac{\omega}{c}r)}{r}\,\hat{\mathbf{e}}_r\times\mathbf{E}_1^+(\mathbf{p})\,, \qquad (1.51b)
$$

where $\mathbf{p} = \sqrt{\epsilon_1}\frac{\omega}{c}(\sin\theta_r\cos\phi_r\,\hat{\mathbf{e}}_1+\sin\theta_r\sin\phi_r\,\hat{\mathbf{e}}_2)$. This asymptotic approximation will become more and more accurate as we let $r\to\infty$. Plugging Eq. (1.51) into Eq. (1.50) we obtain

$$
\begin{aligned}
P_{\mathrm{d}\sigma} &= \frac{\varepsilon_0 c}{2}\,\epsilon_1^{3/2}\left(\frac{\omega}{2\pi\,c}\right)^2\cos^2\theta_r\,|\,\mathbf{E}_1^+(\mathbf{p})|^2\,\mathrm{d}\Omega_r \\
&= \frac{\varepsilon_0 c}{2}\,\epsilon_1^{3/2}\left(\frac{\omega}{2\pi\,c}\right)^2\cos^2\theta_r\,\left(|\,\mathcal{E}_{1,p}^+(\mathbf{p})|^2+|\,\mathcal{E}_{1,s}^+(\mathbf{p})|^2\right)\mathrm{d}\Omega_r\,. \qquad (1.52)
\end{aligned}
$$

(Note: $1/(\mu_0 c) = \varepsilon_0 c$). The total differential reflection coefficient is then given by

$$
\frac{\partial R}{\partial\Omega_r}(\mathbf{p}\mid\mathbf{p}_0) = \lim_{r\to\infty}\frac{P_{\mathrm{d}\sigma}}{P_{\mathrm{inc}/S}\,\mathrm{d}\Omega_r} = \lim_{r\to\infty}\frac{\epsilon_1\left(\frac{\omega}{2\pi\,c}\right)^2\cos^2\theta_r}{S\,\cos\theta_0}\,\frac{|\,\mathcal{E}_{1,p}^+(\mathbf{p})|^2+|\,\mathcal{E}_{1,s}^+(\mathbf{p})|^2}{|\,\mathcal{E}_{0,p}\,|^2+|\,\mathcal{E}_{0,s}\,|^2}\,. \qquad (1.53)
$$

From the total differential reflection coefficient given by Eq. (1.53), we deduce the differential reflection coefficient when an incident plane wave of polarization β, with in-plane wave vector $\mathbf{p}_0$ is reflected into a plane wave of polarization α with in-plane wave vector $\mathbf{p}$ given as

$$
\frac{\partial R_{\alpha\beta}}{\partial\Omega_r}(\mathbf{p}\mid\mathbf{p}_0) = \lim_{r\to\infty}\frac{\epsilon_1\left(\frac{\omega}{2\pi\,c}\right)^2\cos^2\theta_r}{S\,\cos\theta_0}\,|R_{\alpha\beta}^{(S)}(\mathbf{p}\mid\mathbf{p}_0)|^2 = \lim_{r\to\infty}\frac{\partial R_{\alpha\beta}^{(S)}}{\partial\Omega_r}(\mathbf{p}\mid\mathbf{p}_0)\,. \qquad (1.54)
$$

Here we have defined

$$
\frac{\partial R_{\alpha\beta}^{(S)}}{\partial\Omega_r}(\mathbf{p}\mid\mathbf{p}_0) = \frac{\epsilon_1\left(\frac{\omega}{2\pi\,c}\right)^2\cos^2\theta_r}{S\,\cos\theta_0}\,|R_{\alpha\beta}^{(S)}(\mathbf{p}\mid\mathbf{p}_0)|^2\,. \qquad (1.55)
$$

A similar analysis for the *differential transmission coefficients*[5] (DTC) yields

$$\frac{\partial T_{\alpha\beta}}{\partial \Omega_t}(\mathbf{p} \mid \mathbf{p}_0) = \lim_{r\to\infty} \frac{\partial T^{(S)}_{\alpha\beta}}{\partial \Omega_t}(\mathbf{p} \mid \mathbf{p}_0)\,. \tag{1.56}$$

Here we have defined

$$\frac{\partial T^{(S)}_{\alpha\beta}}{\partial \Omega_t}(\mathbf{p} \mid \mathbf{p}_0) = \frac{\epsilon_2^{3/2}\left(\frac{\omega}{2\pi\, c}\right)^2 \cos^2\theta_t}{S\,\epsilon_1^{1/2}\,\cos\theta_0}\,|T^{(S)}_{\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0)|^2\,, \tag{1.57}$$

where now $\mathbf{p} = \sqrt{\epsilon_2}\frac{\omega}{c}(\sin\theta_t\cos\phi_t\,\hat{\mathbf{e}}_1 + \sin\theta_t\sin\phi_t\,\hat{\mathbf{e}}_2)$ and θ_t is defined as the angle between $\hat{\mathbf{e}}_r$ and $-\,\hat{\mathbf{e}}_3$ (i.e. $\theta_t = \pi - \theta_r$).

Diffraction efficiencies

Can we still define differential scattering coefficients in the case of the scattering of a plane wave by a planar interface or a grating? We will see that in such cases, the field amplitude, say $\mathbf{E}_1^+(\mathbf{p})$, concentrates all its mass on discrete wave vectors and is not a function as was implicitly assumed in the previous section. Take the case of a planar interface for example. The reflected field amplitude would take the form $\mathbf{E}_1^+(\mathbf{p}) = \mathbf{E}_r(\mathbf{p}_0)\,\delta(\mathbf{p}-\mathbf{p}_0)$. If no care is taken, we may try to plug this expression into the far-field expansion and get

$$\mathbf{E}_1^+(\mathbf{r}) \sim -i\,\epsilon_1^{1/2}\frac{\omega}{2\pi\, c}\cos\theta_r\,\frac{\exp(i\epsilon_1^{1/2}\frac{\omega}{c}r)}{r}\;\mathbf{E}_r(\mathbf{p}_0)\,\delta(\mathbf{p}-\mathbf{p}_0)\,, \tag{1.58}$$

which we may interpret as saying that the far-field in zero for every direction but that given by $\mathbf{k}_1^+(\mathbf{p}_0)$, which would be a wrong interpretation. Indeed, the reflected plane wave fills the whole space above the surface and therefore takes non-zero values everywhere (up to the nodes due to time oscillation of course). This simply means that the far-field expansion is not valid. In fact, we do not need to make any approximation to obtain the far-field in this case since it is given by the expression of the plane wave itself. Assume that the reflected field is a discrete sum of plane waves, typically produced by the scattering of a plane wave by a periodic grating, and let us analyze what one would obtain as the differential reflection if we consider one wave at a time. One such wave may be written as

$$\mathbf{E}_n^+(\mathbf{r}) = \mathbf{E}_n(\mathbf{p}_0)\,\exp(i\,\mathbf{k}_1^+(\mathbf{p}_n)\cdot\mathbf{r})\,, \tag{1.59}$$

where the index n simply denotes that we picked one such wave whose in-plane wave vector is some $\mathbf{p}_n$. We do not need to worry for now what $\mathbf{p}_n$ may be nor how to compute the amplitude $\mathbf{E}_n(\mathbf{p}_0)$ of such a wave (this will be the aim of the rest of this thesis). Let us use the electric field expressed in Eq. (1.59) and the corresponding magnetic field in Eq, (1.50) to get the flux crossing the elementary surface $\mathrm{d}\sigma$. We obtain in a straightforward manner

$$P_{\mathrm{d}\sigma} = \frac{\varepsilon_0 c}{2}\,\epsilon_1^{1/2}\,|\,\mathbf{E}_n(\mathbf{p}_0)|^2\;\hat{\mathbf{k}}_1^+(\mathbf{p}_n)\cdot\hat{\mathbf{e}}_r\,r^2\mathrm{d}\Omega_r\,, \tag{1.60}$$

under the constraint that the field is not evanescent. It is clear from Eq. (1.60) that the flux is proportional to $\hat{\mathbf{k}}_1^+(\mathbf{p}_n)\cdot\hat{\mathbf{e}}_r$, which means that the flux is maximal in the direction of the wave vector. However, note that this flux does not vanish at other directions (unless $\hat{\mathbf{k}}_1^+(\mathbf{p}_n)\cdot\hat{\mathbf{e}}_r = 0$). The total differential reflection coefficient is then given by

$$\frac{\partial R}{\partial \Omega_r}(\mathbf{p} \mid \mathbf{p}_0) = \lim_{r\to\infty}\frac{P_{\mathrm{d}\sigma}}{P_{\mathrm{inc}/S}\,\mathrm{d}\Omega_r} = \frac{|\,\mathbf{E}_n(\mathbf{p}_0)|^2}{|\,\mathbf{E}_0\,|^2}\,\frac{\hat{\mathbf{k}}_1^+(\mathbf{p}_n)\cdot\hat{\mathbf{e}}_r}{\pi\cos\theta_0}\,. \tag{1.61}$$

[5] Also known as *bidirectional transmittance distribution function* (BTDF).

Note that the ratio is actually independent of S. In such cases, one is rather more interested in the *total flux fraction* taken away by *each* reflected wave, and this independently of the direction. We are thus interested in the quantity

$$e^{(R,n)} \stackrel{\text{def}}{=} \int_0^{2\pi} \int_0^{\pi/2} \frac{\partial R}{\partial \Omega_r}(\mathbf{p} \mid \mathbf{p}_0)\, \mathrm{d}\Omega_r \,, \tag{1.62}$$

known as *diffraction efficiency* of the reflected wave n. Let us now express the dot product in Eq. (1.61) with the help of polar and azimuthal angles,

$$\hat{\mathbf{k}}_1^+(\mathbf{p}_n) \cdot \hat{\mathbf{e}}_r = \sin\theta_n \sin\theta_r \cos(\phi_n - \phi_r) + \cos\theta_n \cos\theta_r \,, \tag{1.63}$$

where the angles θ_r and ϕ_r have been defined previously and θ_n and ϕ_n, associated with the wave vector $\hat{\mathbf{k}}_1^+(\mathbf{p}_n)$, are defined in an obvious way, similar to what we already did for deriving the far-field in Section 1.2.3. The diffraction efficiency hence reads

$$e^{(R,n)} = \frac{|\,\mathbf{E}_n(\mathbf{p}_0)|^2}{|\,\mathbf{E}_0\,|^2} \int_0^{2\pi} \int_0^{\pi/2} \frac{\sin\theta_n \sin\theta_r \cos(\phi_n - \phi_r) + \cos\theta_n \cos\theta_r}{\pi \cos\theta_0} \sin\theta_r \mathrm{d}\theta_r \mathrm{d}\phi_r \,. \tag{1.64}$$

The integration does not require much efforts after some simple trigonometric simplifications and yields

$$e^{(R,n)} = \frac{\cos\theta_n |\,\mathbf{E}_n(\mathbf{p}_0)|^2}{\cos\theta_0 |\,\mathbf{E}_0\,|^2} \,. \tag{1.65}$$

If we had treated the case of a transmitted wave then the final expression for the efficiency would read

$$e^{(T,n)} = \frac{\epsilon_2^{1/2} \cos\theta_n |\,\mathbf{E}_n(\mathbf{p}_0)|^2}{\epsilon_1^{1/2} \cos\theta_0 |\,\mathbf{E}_0\,|^2} \,, \tag{1.66}$$

hence in general we can simply write

$$e^{(X,n)} = \frac{\alpha_{j_X}(\mathbf{p}_n) |\,\mathbf{E}_n(\mathbf{p}_0)|^2}{\alpha_1(\mathbf{p}_0) |\,\mathbf{E}_0\,|^2} \,, \tag{1.67}$$

where the superscript is $X = R$ or T to denote reflection or transmission, and the corresponding medium is indexed by j_X is 1 or 2 respectively. In a similar fashion as what was done for the differential reflection and transmission coefficients, we can separate the contribution from the different polarization and define

$$e^{(X,n)}_{\alpha\beta}(\mathbf{p}_0) = \frac{\alpha_{j_X}(\mathbf{p}_n)}{\alpha_1(\mathbf{p}_0)} |X^{(n)}_{\alpha\beta}(\mathbf{p}_0)|^2, \tag{1.68}$$

where $(X^{(n)}_{\alpha\beta}(\mathbf{p}_0))_{\alpha\beta\in\{p,s\}}\, \delta(\mathbf{p}_n - \mathbf{p}_0)$ are the reflection or transmission amplitudes of the wave n (see more precision in Section 1.5) and the efficiency hence corresponds to the *fraction of the incident flux carried away by the plane wave indexed by n, polarized in the state α given that the incident plane wave had a unit field, and an in-plane wave vector $\mathbf{p}_0$ and was polarized in the state β.*

Stokes vectors and Mueller matrices

The polarization state of a transverse monochromatic plane wave can be characterized by its Stokes vector $\mathbf{S} \in \mathbb{R}^4$. If an orthonormal basis $(\hat{\mathbf{e}}_x, \hat{\mathbf{e}}_y, \hat{\mathbf{k}})$ is attached to the wave vector $\mathbf{k}$ of the

wave, the Stokes vector is defined as

$$\mathbf{S} = \begin{pmatrix} S_1 \\ S_2 \\ S_3 \\ S_4 \end{pmatrix} \stackrel{\text{def}}{=} \frac{\varepsilon_0 c}{2} \epsilon_j^{1/2} \cos\theta(\mathbf{k}) \begin{pmatrix} |E_x|^2 + |E_y|^2 \\ |E_x|^2 - |E_y|^2 \\ |E_{x'}|^2 - |E_{y'}|^2 \\ |E_{\sigma^+}|^2 - |E_{\sigma^-}|^2 \end{pmatrix}, \tag{1.69}$$

where E_x, E_y are the components of the complex amplitude of the electric field expressed in the basis $(\hat{\mathbf{e}}_x, \hat{\mathbf{e}}_y)$, $E_{x'}$, $E_{y'}$ are the components of the complex amplitude of the electric field expressed in the basis $(\hat{\mathbf{e}}_{x'}, \hat{\mathbf{e}}_{y'})$ which is the image of the basis $(\hat{\mathbf{e}}_x, \hat{\mathbf{e}}_y)$ rotated by $\pi/4$ radians around $\hat{\mathbf{k}}$ and E_{σ^+}, E_{σ^-} are the components of the complex amplitude of the electric field expressed in the basis $(\hat{\mathbf{e}}_{\sigma^+}, \hat{\mathbf{e}}_{\sigma^-})$, i.e. that E_{σ^+}, E_{σ^-} correspond to the left and right circularly polarized components of the electric field respectively. The equations for change of basis are given by

$$\begin{pmatrix} E_{x'} \\ E_{y'} \end{pmatrix} = \frac{\sqrt{2}}{2} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} E_x \\ E_y \end{pmatrix} \tag{1.70a}$$

$$\begin{pmatrix} E_{\sigma^+} \\ E_{\sigma^-} \end{pmatrix} = \frac{\sqrt{2}}{2} \begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix} \begin{pmatrix} E_x \\ E_y \end{pmatrix}. \tag{1.70b}$$

The factor $\varepsilon_0 c \epsilon_j^{1/2} \cos\theta(\mathbf{k})/2$, where $\theta(\mathbf{k})$ is the angle between the wave vector and a fixed x_3-axis, is a matter of definition which we adopt here as it is convenient for the study of scattering by surfaces and is reminiscent of the flux per unit area when we derived the efficiencies of the diffracted modes. The interpretation of the elements of the Stokes vector is then that of the total flux per unit area for the first element and differences of fluxes per unit area for the component along two ortonormal polarization states for the remaining elements. With the above definitions, the Stokes vector can be expressed solely in terms of E_x and E_y as

$$\mathbf{S} = \frac{\varepsilon_0 c}{2} \epsilon_j^{1/2} \cos\theta(\mathbf{k}) \begin{pmatrix} |E_x|^2 + |E_y|^2 \\ |E_x|^2 - |E_y|^2 \\ 2\mathrm{Re}(E_x E_y^*) \\ 2\mathrm{Im}(E_x E_y^*) \end{pmatrix}. \tag{1.71}$$

Note, however, that in an experimental measurement, one cannot measure directly the complex amplitudes E_x and E_y but only intensities. It is therefore the definition, Eq. (1.69), which is used in practice, since it requires the measurement of the total (i.e. all polarizations) intensity and of differences of intensity when the wave is analyzed after passing through different analyzers (which are polarizers). From a computational point of view, it is rather convenient to work directly with the complex amplitudes.

Now that we have explained how to characterize the polarization state of a plane wave, we can characterize a sample by the way it influences the polarization state of an incident wave reflected from or transmitted through it. Let an incident plane wave, whose in-plane wave vector in the $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2)$-plane is $\mathbf{p}_0$ be scattered by a periodic surface whose average plane is $x_3 = 0$. The Stokes vector characterizing the incident plane wave is given by $\mathbf{S}^{(\mathrm{inc})}$. Let $\mathbf{S}^{(\mathrm{X},n)}(\mathbf{p}_0)$ be the Stokes vector characterizing the polarization state of the plane wave scattered by the periodic

surface in reflection ($X = R$) or transmission ($X = T$) whose in-plane wave vector is $\mathbf{p}_n$. Using the notation from the previous subsection, the scattering amplitudes $X^{(n)}_{\alpha\beta}(\mathbf{p}_0)$ yields that for an arbitrary polarized incident field decomposed into its p- and s-polarization component $\mathbf{E}_0(\mathbf{p}_0) = \mathcal{E}_{0,s}\,\hat{\mathbf{e}}_s(\mathbf{p}_0) + \mathcal{E}_{0,p}\,\hat{\mathbf{e}}^-_{p,1}(\mathbf{p}_0)$, the field amplitude components of the nth diffractive mode are given by

$$\mathcal{E}^{(n)}_{j,p}(\mathbf{p}_0) = X^{(n)}_{pp}(\mathbf{p}_0)\;\mathcal{E}_{0,p}(\mathbf{p}_0) + X^{(n)}_{ps}(\mathbf{p}_0)\;\mathcal{E}_{0,s}(\mathbf{p}_0) \tag{1.72a}$$

$$\mathcal{E}^{(n)}_{j,s}(\mathbf{p}_0) = X^{(n)}_{sp}(\mathbf{p}_0)\;\mathcal{E}_{0,p}(\mathbf{p}_0) + X^{(n)}_{ss}(\mathbf{p}_0)\;\mathcal{E}_{0,s}(\mathbf{p}_0)\,, \tag{1.72b}$$

where $j = 1$ or 2 and $X = R$ or T correspondingly. According to our definition of the local p- and s-polarization vectors, they correspond to the local $(\hat{\mathbf{e}}_x, \hat{\mathbf{e}}_y)$ basis used in the definition of the Stokes vector. Using Eq. (1.71) for the Stokes vector of the diffractive mode and plugging Eq. (1.72) we obtain for example

$$\begin{aligned}
\frac{2\omega}{\varepsilon_0 c^2\,\alpha_{j_X}(\mathbf{p}_n)}\,S^{(\mathrm{X},n)}_1 &= |\,\mathcal{E}^{(n)}_{j,p}\,|^2 + |\,\mathcal{E}^{(n)}_{j,s}\,|^2 \\
&= |X^{(n)}_{pp}\;\mathcal{E}_{0,p} + X^{(n)}_{ps}\;\mathcal{E}_{0,s}\,|^2 + |X^{(n)}_{sp}\;\mathcal{E}_{0,p} + X^{(n)}_{ss}\;\mathcal{E}_{0,s}\,|^2 \\
&= \left(|X^{(n)}_{pp}|^2 + |X^{(n)}_{sp}|^2\right)|\,\mathcal{E}_{0,p}\,|^2 + \left(|X^{(n)}_{ps}|^2 + |X^{(n)}_{ss}|^2\right)|\,\mathcal{E}_{0,s}\,|^2 \\
&\quad + \mathrm{Re}\left[\left(X^{(n)}_{pp}X^{(n)*}_{ps} + X^{(n)}_{sp}X^{(n)*}_{ss}\right)\,\mathcal{E}_{0,p}\,\mathcal{E}^*_{0,s}\right] \\
&= \frac{1}{2}\left[|X^{(n)}_{pp}|^2 + |X^{(n)}_{sp}|^2 + |X^{(n)}_{ps}|^2 + |X^{(n)}_{ss}|^2\right]\left(|\,\mathcal{E}_{0,p}\,|^2 + |\,\mathcal{E}_{0,s}\,|^2\right) \\
&\quad + \frac{1}{2}\left[|X^{(n)}_{pp}|^2 + |X^{(n)}_{sp}|^2 - |X^{(n)}_{ps}|^2 - |X^{(n)}_{ss}|^2\right]\left(|\,\mathcal{E}_{0,p}\,|^2 - |\,\mathcal{E}_{0,s}\,|^2\right) \\
&\quad + 2\mathrm{Re}\left[X^{(n)}_{pp}X^{(n)*}_{ps} + X^{(n)}_{sp}X^{(n)*}_{ss}\right]\,\mathrm{Re}\left[\mathcal{E}_{0,p}\,\mathcal{E}^*_{0,s}\right] \\
&\quad + 2\mathrm{Im}\left[X^{(n)}_{pp}X^{(n)*}_{ps} + X^{(n)}_{sp}X^{(n)*}_{ss}\right]\,\mathrm{Im}\left[\mathcal{E}_{0,p}\,\mathcal{E}^*_{0,s}\right]
\end{aligned}$$

hence we can write the Stokes element $S^{(\mathrm{X},n)}_1$ as a linear combination of the Stokes elements of the incident wave as

$$S^{(\mathrm{X},n)}_1 = M^{(\mathrm{X},n)}_{11}\,S^{(\mathrm{inc})}_1 + M^{(\mathrm{X},n)}_{12}\,S^{(\mathrm{inc})}_2 + M^{(\mathrm{X},n)}_{13}\,S^{(\mathrm{inc})}_3 + M^{(\mathrm{X},n)}_{14}\,S^{(\mathrm{inc})}_3\,.$$

In the above derivation, we have used that $\mathrm{Re}(zz') = \mathrm{Re}(z)\mathrm{Re}(z') - \mathrm{Im}(z)\mathrm{Im}(z')$ and the useful identity $ax + by = (a+b)(x+y)/2 + (a-b)(x-y)/2$. We see in the above equation that the first element of the Stokes vector for the diffractive mode can be written as a linear combination of the components of the Stokes vector of the incident wave. Similarly to what we have done above, we can express the Stokes vector of the diffracted wave as the product of a 4 by 4 real matrix, known as the *Mueller matrix*, and the Stokes vector of the incident wave:

$$\mathbf{S}^{(\mathrm{X},n)}(\mathbf{p}_0) = \mathbf{M}^{(\mathrm{X},n)}(\mathbf{p}_0)\;\mathbf{S}^{(\mathrm{inc})}(\mathbf{p}_0)\,, \tag{1.73}$$

where the elements of the Mueller matrix are given by

$$\frac{\alpha_1(\mathbf{p}_0)}{\alpha_{jX}(\mathbf{p}_n)}\, M_{11}^{(\mathrm{X},n)} = \frac{1}{2}\left[|X_{pp}^{(n)}|^2 + |X_{sp}^{(n)}|^2 + |X_{ps}^{(n)}|^2 + |X_{ss}^{(n)}|^2\right] \tag{1.74a}$$

$$\frac{\alpha_1(\mathbf{p}_0)}{\alpha_{jX}(\mathbf{p}_n)}\, M_{12}^{(\mathrm{X},n)} = \frac{1}{2}\left[|X_{pp}^{(n)}|^2 + |X_{sp}^{(n)}|^2 - |X_{ps}^{(n)}|^2 + |X_{ss}^{(n)}|^2\right] \tag{1.74b}$$

$$\frac{\alpha_1(\mathbf{p}_0)}{\alpha_{jX}(\mathbf{p}_n)}\, M_{13}^{(\mathrm{X},n)} = \mathrm{Re}\left[X_{pp}^{(n)} X_{ps}^{(n)*} + X_{sp}^{(n)} X_{ss}^{(n)*}\right] \tag{1.74c}$$

$$\frac{\alpha_1(\mathbf{p}_0)}{\alpha_{jX}(\mathbf{p}_n)}\, M_{14}^{(\mathrm{X},n)} = \mathrm{Im}\left[X_{pp}^{(n)} X_{ps}^{(n)*} + X_{sp}^{(n)} X_{ss}^{(n)*}\right] \tag{1.74d}$$

$$\frac{\alpha_1(\mathbf{p}_0)}{\alpha_{jX}(\mathbf{p}_n)}\, M_{21}^{(\mathrm{X},n)} = \frac{1}{2}\left[|X_{pp}^{(n)}|^2 - |X_{sp}^{(n)}|^2 + |X_{ps}^{(n)}|^2 - |X_{ss}^{(n)}|^2\right] \tag{1.74e}$$

$$\frac{\alpha_1(\mathbf{p}_0)}{\alpha_{jX}(\mathbf{p}_n)}\, M_{22}^{(\mathrm{X},n)} = \frac{1}{2}\left[|X_{pp}^{(n)}|^2 - |X_{sp}^{(n)}|^2 - |X_{ps}^{(n)}|^2 + |X_{ss}^{(n)}|^2\right] \tag{1.74f}$$

$$\frac{\alpha_1(\mathbf{p}_0)}{\alpha_{jX}(\mathbf{p}_n)}\, M_{23}^{(\mathrm{X},n)} = \mathrm{Re}\left[X_{pp}^{(n)} X_{ps}^{(n)*} - X_{sp}^{(n)} X_{ss}^{(n)*}\right] \tag{1.74g}$$

$$\frac{\alpha_1(\mathbf{p}_0)}{\alpha_{jX}(\mathbf{p}_n)}\, M_{24}^{(\mathrm{X},n)} = \mathrm{Im}\left[X_{pp}^{(n)} X_{ps}^{(n)*} - X_{sp}^{(n)} X_{ss}^{(n)*}\right] \tag{1.74h}$$

$$\frac{\alpha_1(\mathbf{p}_0)}{\alpha_{jX}(\mathbf{p}_n)}\, M_{31}^{(\mathrm{X},n)} = \mathrm{Re}\left[X_{pp}^{(n)} X_{sp}^{(n)*} + X_{ps}^{(n)} X_{ss}\right] \tag{1.74i}$$

$$\frac{\alpha_1(\mathbf{p}_0)}{\alpha_{jX}(\mathbf{p}_n)}\, M_{32}^{(\mathrm{X},n)} = \mathrm{Re}\left[X_{pp}^{(n)} X_{sp}^{(n)*} - X_{ps}^{(n)} X_{ss}^{(n)*}\right] \tag{1.74j}$$

$$\frac{\alpha_1(\mathbf{p}_0)}{\alpha_{jX}(\mathbf{p}_n)}\, M_{33}^{(\mathrm{X},n)} = \mathrm{Re}\left[X_{pp}^{(n)} X_{ss}^{(n)*} + X_{ps}^{(n)} X_{sp}^{(n)*}\right] \tag{1.74k}$$

$$\frac{\alpha_1(\mathbf{p}_0)}{\alpha_{jX}(\mathbf{p}_n)}\, M_{34}^{(\mathrm{X},n)} = \mathrm{Im}\left[X_{pp}^{(n)} X_{ss}^{(n)*} - X_{ps}^{(n)} X_{sp}^{(n)*}\right] \tag{1.74l}$$

$$\frac{\alpha_1(\mathbf{p}_0)}{\alpha_{jX}(\mathbf{p}_n)}\, M_{41}^{(\mathrm{X},n)} = -\mathrm{Im}\left[X_{pp}^{(n)} X_{sp}^{(n)*} + X_{ps}^{(n)} X_{ss}^{(n)*}\right] \tag{1.74m}$$

$$\frac{\alpha_1(\mathbf{p}_0)}{\alpha_{jX}(\mathbf{p}_n)}\, M_{42}^{(\mathrm{X},n)} = -\mathrm{Im}\left[X_{pp}^{(n)} X_{sp}^{(n)*} - X_{ps}^{(n)} X_{ss}^{(n)*}\right] \tag{1.74n}$$

$$\frac{\alpha_1(\mathbf{p}_0)}{\alpha_{jX}(\mathbf{p}_n)}\, M_{43}^{(\mathrm{X},n)} = -\mathrm{Im}\left[X_{pp}^{(n)} X_{ss}^{(n)*} + X_{ps}^{(n)} X_{sp}^{(n)*}\right] \tag{1.74o}$$

$$\frac{\alpha_1(\mathbf{p}_0)}{\alpha_{jX}(\mathbf{p}_n)}\, M_{44}^{(\mathrm{X},n)} = \mathrm{Re}\left[X_{pp}^{(n)} X_{ss}^{(n)*} - X_{ps}^{(n)} X_{sp}^{(n)*}\right] . \tag{1.74p}$$

We can thus define a Mueller matrix for each reflected and transmitted diffractive order. The entries of the Mueller matrix depend of the sample and on the way the sample is oriented with respect to the incident wave. Indeed, the reflection and transmission amplitudes are dependent on the angles of incidence and so are the Mueller matrices. More generally, we can define two Mueller matrices which are functions of the in-plane wave vector of the outgoing elementary wave reflected or transmitted by a sample. The definition is then analogous to the one presented here and we have

$$\mathbf{S}^{(\mathrm{X})}(\mathbf{p}\,|\,\mathbf{p}_0) = \mathbf{M}^{(\mathrm{X})}(\mathbf{p}\,|\,\mathbf{p}_0)\; \mathbf{S}^{(\mathrm{inc})}(\mathbf{p}_0)\,, \tag{1.75}$$

and the pre-factor is adapted to $\alpha_{jX}(\mathbf{p})/\alpha_1(\mathbf{p}_0)$ instead of $\alpha_{jX}(\mathbf{p}_n)/\alpha_1(\mathbf{p}_0)$.

1.3 Radiation of oscillating and rotating dipoles

The present section is devoted to revisiting some useful results about the polarization properties of dipole radiation in free space. Why do we need to know about dipole radiation when the aim of this work is to study scattering from surfaces? The answer is: understanding dipole radiation is crucial for understanding the physics of the scattering from a surface. The knowledgeable reader may have in mind the dipolar resonance that occurs in metallic particle for example, hence justifying the present section, as we will study in Chapter 12 the optical response of metallic particles deposited on a substrate for example. However, we will see in Chapter 10, that dipole radiation is crucial for understanding the scattering from a simple rough dielectric surface. The reason is quite simple. A dielectric medium can microscopically be described by a dense array of scatterers (the atoms) which respond to an electromagnetic excitation in a dipolar way. This elementary dipole feature of the scatterers will leave a clear signature in the light scattered from a dielectric system.

1.3.1 Polarization of an oscillating dipole radiation in free space with respect to the local $(\hat{\mathbf{e}}_p, \hat{\mathbf{e}}_s)$ basis

We consider here the radiation emitted by a single oscillating dipole in free space. The dipole of dipole moment $\mathbf{D}(\vartheta) = d(\sin\vartheta\ \hat{\mathbf{e}}_1 + \cos\vartheta\ \hat{\mathbf{e}}_3) = d\,\hat{\mathbf{e}}_\vartheta$ is chosen to lie in the $(O, \hat{\mathbf{e}}_1, \hat{\mathbf{e}}_3)$-plane and is allowed to be tilted from the x_3-axis by an angle $\vartheta \in [0, \pi/2]$ radians, as is illustrated in Fig. 1.5(a). The dipole is placed in free space at the origin of the coordinate system, oscillates with angular frequency ω and radiates in the far-filed the electric field [17]

$$\mathbf{E}_{\rm dip}(\mathbf{r}, t) = -\frac{\omega^2}{4\pi\varepsilon_0 c^2}\ \frac{\hat{\mathbf{e}}_r \times [\hat{\mathbf{e}}_r \times \mathbf{D}(\vartheta)]}{r}\ e^{-i\omega(t-r/c)}\,, \tag{1.76}$$

where $\mathbf{r} = r\ \hat{\mathbf{e}}_r = r(\sin\theta\cos\phi\,\hat{\mathbf{e}}_1 + \sin\theta\sin\phi\,\hat{\mathbf{e}}_2 + \cos\theta\,\hat{\mathbf{e}}_3)$ is the point of observation, and $r = |\,\mathbf{r}\,|$. A well known result is that no power is radiated along the axis of oscillation of the dipole as can be seen from Eq. (1.76) since $\hat{\mathbf{e}}_r \times \mathbf{D}(\vartheta)$ vanishes in when $\hat{\mathbf{e}}_r \parallel \hat{\mathbf{e}}_\vartheta$. In addition, the radiation is polarized as can be seen from the cross products in Eq. (1.76). Indeed, the electric field is polarized along the vector $\hat{\mathbf{e}}_\theta{}'$ which is the basis vector tangent to a meridian in a spherical coordinate system (r, θ', ϕ') attached to the direction of the dipole moment. Here we are rather interested in analyzing the polarization of the dipole radiation with respect to a local polarization basis $(\hat{\mathbf{e}}_p, \hat{\mathbf{e}}_s)$ which mimics that defined in Eq. (1.32), and which is defined with respect to the propagation direction of the radiation *and* the plane $x_3 = 0$. These polarization vectors are defined as $\hat{\mathbf{e}}_s = \frac{\hat{\mathbf{e}}_3 \times \hat{\mathbf{e}}_r}{|\,\hat{\mathbf{e}}_3 \times \hat{\mathbf{e}}_r\,|}$ and $\hat{\mathbf{e}}_p = \frac{\hat{\mathbf{e}}_s \times \hat{\mathbf{e}}_r}{|\,\hat{\mathbf{e}}_s \times \hat{\mathbf{e}}_r\,|}$ are defined with respect to $\hat{\mathbf{e}}_r$ in order to mimic the local s- and p- polarization vectors attached to a scattering direction along $\hat{\mathbf{e}}_r$.

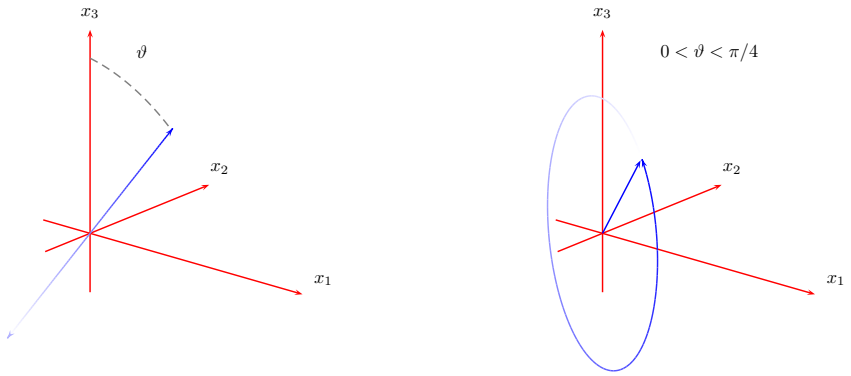
Remark 1.4. Note that we have

$$\hat{\mathbf{e}}_p = \hat{\mathbf{e}}_\theta = \frac{\mathrm{d}\,\hat{\mathbf{e}}_r}{\mathrm{d}\theta} = \cos\theta\cos\phi\ \hat{\mathbf{e}}_1 + \cos\theta\sin\phi\ \hat{\mathbf{e}}_2 - \sin\theta\ \hat{\mathbf{e}}_3 \tag{1.77a}$$

$$\hat{\mathbf{e}}_s = \hat{\mathbf{e}}_\phi = \frac{1}{\sin\theta}\,\frac{\mathrm{d}\,\hat{\mathbf{e}}_r}{\mathrm{d}\phi} = -\sin\phi\ \hat{\mathbf{e}}_1 + \cos\phi\ \hat{\mathbf{e}}_2\,, \tag{1.77b}$$

which are nothing but the conventional basis vector in spherical coordinates.

The angular dependence of the polarization components of the radiation emitted by the oscillating dipole are then entirely controlled by the following dot products: $\hat{\mathbf{e}}_r \times [\hat{\mathbf{e}}_r \times \hat{\mathbf{e}}_\vartheta] \cdot \hat{\mathbf{e}}_s$ and $\hat{\mathbf{e}}_r \times [\hat{\mathbf{e}}_r \times \hat{\mathbf{e}}_\vartheta] \cdot \hat{\mathbf{e}}_p$. First, it is worth noting that $\hat{\mathbf{e}}_r \times [\hat{\mathbf{e}}_r \times \hat{\mathbf{e}}_\vartheta] \cdot \hat{\mathbf{e}}_s$ and $\hat{\mathbf{e}}_r \times [\hat{\mathbf{e}}_r \times \hat{\mathbf{e}}_\vartheta] \cdot \hat{\mathbf{e}}_p$ are invariant



(a) Oscillating dipole (b) Rotating dipole

Figure 1.5: Illustrations of (a) an oscillating and (b) a rotating dipole in the $(O, \hat{\mathbf{e}}_1, \hat{\mathbf{e}}_3)$-plane. The dipole moment of the oscillating dipole is aligned with a direction making an angle ϑ with the x_3-axis. The dipole moment of the rotating dipole describes an ellipse in the $(O, \hat{\mathbf{e}}_1, \hat{\mathbf{e}}_3)$-plane with its long axis being aligned with x_3-axis for $0 < \vartheta < \pi/4$ and along the x_1-axis for $\pi/4 < \vartheta < \pi/2$ according to the definition Eq. (1.83). The special case $\vartheta = \pi/4$ corresponds to a circular motion. In all cases, the ellipse is swept clockwise when viewed from $x_2 > 0$ (i.e. $\hat{\mathbf{e}}_2$ pointing towards the observer).

under the transformation $\hat{\mathbf{e}}_r \mapsto -\hat{\mathbf{e}}_r$, hence the p- and s-polarization dependence of the dipole radiation are symmetric with respect to the origin as $\hat{\mathbf{e}}_r$ runs over the unit sphere. Second, for $\vartheta \in]0, \pi/2]$ radians, the vector identity $\mathbf{a} \times [\mathbf{b} \times \mathbf{c}] = (\mathbf{a} \cdot \mathbf{c})\, \mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\, \mathbf{c}$ yields

$$\hat{\mathbf{e}}_r \times [\hat{\mathbf{e}}_r \times \hat{\mathbf{e}}_\vartheta] = (\hat{\mathbf{e}}_r \cdot \hat{\mathbf{e}}_\vartheta)\, \hat{\mathbf{e}}_r - \hat{\mathbf{e}}_\vartheta \,, \tag{1.78}$$

hence the s-polarization dependence of the dipole radiation reads

$$\hat{\mathbf{e}}_r \times [\hat{\mathbf{e}}_r \times \hat{\mathbf{e}}_\vartheta] \cdot \hat{\mathbf{e}}_s = -\hat{\mathbf{e}}_\vartheta \cdot \hat{\mathbf{e}}_\phi = -\sin\vartheta \sin\phi \,. \tag{1.79}$$

Direct consequences of Eq. (1.79) is that $\hat{\mathbf{e}}_r \times [\hat{\mathbf{e}}_r \times \hat{\mathbf{e}}_\vartheta] \cdot \hat{\mathbf{e}}_s$ does not depend on the polar angle θ and that it vanishes for all $\hat{\mathbf{e}}_r$ in the $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_3)$-plane [see Fig. 1.6(d)]. The p-polarization component reads

$$\hat{\mathbf{e}}_r \times [\hat{\mathbf{e}}_r \times \hat{\mathbf{e}}_\vartheta] \cdot \hat{\mathbf{e}}_p = -\hat{\mathbf{e}}_\vartheta \cdot \hat{\mathbf{e}}_\theta \,, \tag{1.80}$$

which is a quantity that depends on ϑ, θ, and ϕ. In the particular case where $\hat{\mathbf{e}}_\theta$ belongs to the $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_3)$-plane, there are two directions, $\hat{\mathbf{e}}_r = \pm\, \hat{\mathbf{e}}_\vartheta$, which make the dot product in Eq.(1.80) vanish. They correspond to the two intersection of the dipole moment direction with the unit sphere [see Figs. 1.6(a-c)]. We simply recover the known fact that no power is emitted along the direction of oscillation of the dipole, and this independently of the polarization. More interesting are cases for which $\hat{\mathbf{e}}_\theta$, and hence $\hat{\mathbf{e}}_r$, do not belong to the $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_3)$-plane. By expanding the dot product in Eq. (1.80) in terms of the angles ϑ, θ, and ϕ, we obtain the following implicit equation for the set of point on the unit sphere for which the p-polarization component of the dipole radiation vanishes

$$\sin\vartheta \cos\theta \cos\phi - \cos\vartheta \sin\theta = 0 \,, \tag{1.81}$$

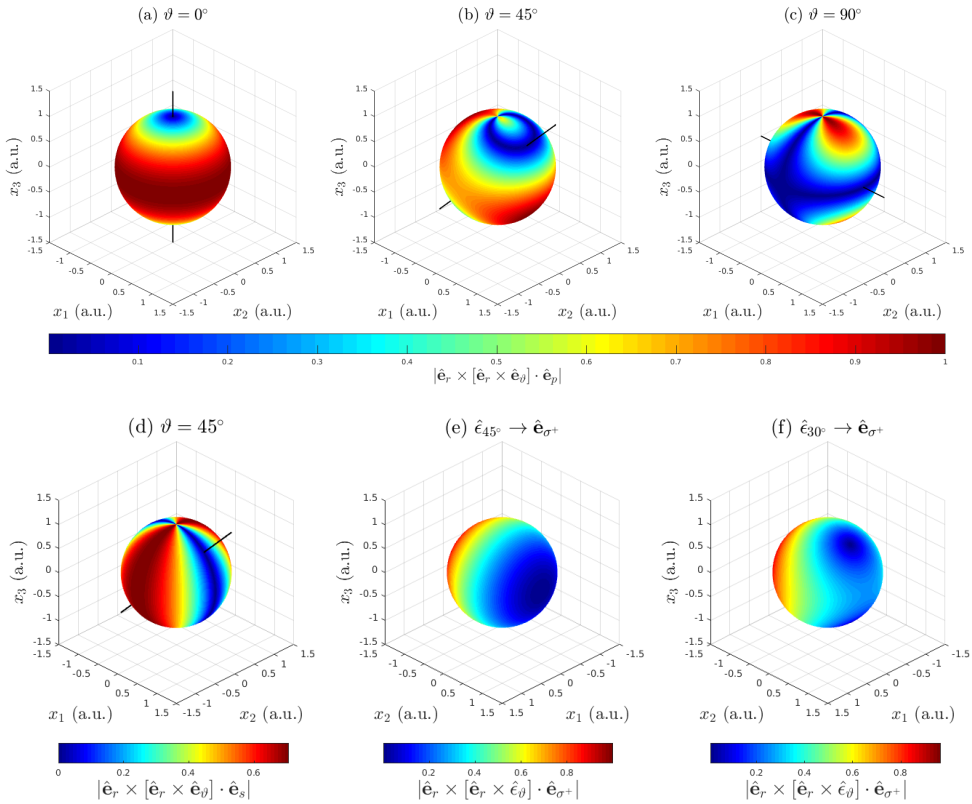


Figure 1.6: (a-c) P-polarization dependence of the radiation of a tilted dipole in free space, $|\,\hat{\mathbf{e}}_r \times [\hat{\mathbf{e}}_r \times \hat{\mathbf{e}}_\vartheta] \cdot \hat{\mathbf{e}}_p\,|$, as $\hat{\mathbf{e}}_r$ runs over the unit sphere for different tilting angles $\vartheta \in \{0°, 45°, 90°\}$. (d) S-polarization dependence of the radiation of a tilted oscillating dipole for $\vartheta = 45°$. The black line in panels (a-d) indicates the direction of the dipole moment. (e-f) σ^+-polarization dependence of the radiation of a rotating dipole parametrized by $\vartheta = 45°$ and $\vartheta = 30°$ respectively (note the orientation of the coordinate system). Figure taken from **Paper** [2].

or equivalently for non-pathologic cases

$$\frac{\tan\vartheta}{\tan\theta} = \frac{1}{\cos\phi}\,. \tag{1.82}$$

We verify that for the cases $\phi = 0$ and $\phi = \pi$ radians, we recover that $\theta = \vartheta$ and $\theta = \pi - \vartheta$, i.e. the points of intersection of the dipole moment direction and the unit sphere. For $\phi \in]-\pi/2, \pi/2[$, we have $0 < \cos\phi < 1$. The first inequality implies that $\tan\theta > 0$ (recall that $0 < \vartheta < \pi/2$ hence $\tan\vartheta > 0$), and the second implies that $\tan\vartheta > \tan\theta$. By the monotony of the tangent function in the interval $[0, \pi[$, and the continuity of Eq. (1.82) with respect to the variables θ and ϕ, we deduce that when ϕ varies in $]-\pi/2, \pi/2[$ the set of the points of zero traces a curve on the unit sphere and this curve is latitude-bounded by $\theta < \vartheta$. By using the aforementioned symmetry of the polarization dependence of the dipole radiation with respect to the origin (invariance under the mapping $\hat{\mathbf{e}}_r \mapsto -\hat{\mathbf{e}}_r$), we immediately deduce that when ϕ varies in $]\pi/2, 3\pi/2[$ the set of the points of zero traces a curve on the unit sphere which is latitude-bounded by $\theta > \pi - \vartheta$. This

is well illustrated in Fig. 1.6(b) where $|\,\hat{\mathbf{e}}_r \times[\hat{\mathbf{e}}_r \times \hat{\mathbf{e}}_\vartheta]\cdot\hat{\mathbf{e}}_p\,|$ is shown as the color map on the unit sphere. For $\vartheta = 45°$, one can appreciate the curve of zero p-polarized emission which passes through the north pole of the unit sphere and the intersection point of the dipole moment direction with the north hemisphere. The degenerate cases $\vartheta = 0°$ and $\vartheta = 90°$ are also illustrated in Figs. 1.6(a) and (c). For $\vartheta = 0°$ the curves of zero polarization emission reduces to two points (the poles). For $\vartheta = 90°$ the curves merge to become union of the equator ($\theta = \pi/2$) and meridians $\phi = \pm\pi/2$. This can be understood from Eq. (1.82) since θ must go to zero when $\vartheta \to 0$ as $\tan\vartheta$ vanishes, and, *either* ϕ must go towards $\pm\pi/2$ *or* θ must go towards $\pi/2$ when $\vartheta \to \pi/2$ as $\tan\vartheta$ diverges.

1.3.2 Polarization of a rotating dipole radiation in free space with respect to the local $(\hat{\mathbf{e}}_{\sigma^+}, \hat{\mathbf{e}}_{\sigma^-})$ basis

We consider now the radiation of a rotating dipole lying in the $(O, \hat{\mathbf{e}}_1, \hat{\mathbf{e}}_3)$-plane. We assume that Eq. (1.76) still holds but we need to modify the dipole moment. The dipole moment can be expressed in complex form as

$$\mathbf{D}(\vartheta) = d\,(\sin\vartheta\;\hat{\mathbf{e}}_1 + i\cos\vartheta\;\hat{\mathbf{e}}_3) = d\,\hat{\boldsymbol{\epsilon}}_\vartheta\,. \tag{1.83}$$

The real vector $\mathrm{Re}[\hat{\boldsymbol{\epsilon}}_\vartheta \exp(-i\omega t)]$ hence describes an ellipse in the $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_3)$-plane whose excentricity is parametrized by ϑ. In the limit cases $\vartheta = 0$ and $\vartheta = \pi/2$ radians we obtain an oscillating dipole along $\hat{\mathbf{e}}_3$ and $\hat{\mathbf{e}}_1$ respectively. For $\vartheta = \pi/4$, we obtain a circularly rotating dipole. For $0 < \vartheta < \pi/4$ (resp. $\pi/4 < \vartheta < \pi/2$) the long axis of the ellipse is aligned with the x_3-axis (resp. the x_1-axis). We wish to study the polarization of the radiation from such an elliptically rotating dipole with respect the local left and right circularly polarized basis $\hat{\mathbf{e}}_{\sigma^+}$ and $\hat{\mathbf{e}}_{\sigma^-}$ defined as

$$\hat{\mathbf{e}}_{\sigma^\pm} = \frac{1}{\sqrt{2}}\,(\hat{\mathbf{e}}_p \pm i\,\hat{\mathbf{e}}_s)\,. \tag{1.84}$$

The σ^+-polarization component of the rotating dipole radiation is then measured by

$$\hat{\mathbf{e}}_r \times[\hat{\mathbf{e}}_r \times\hat{\boldsymbol{\epsilon}}_\vartheta]\cdot\hat{\mathbf{e}}_{\sigma^+} = -\hat{\boldsymbol{\epsilon}}_\vartheta\cdot\hat{\mathbf{e}}_{\sigma^+}\,, \tag{1.85}$$

which can be expressed in terms of the angles as[6]

$$\hat{\mathbf{e}}_r \times[\hat{\mathbf{e}}_r \times\hat{\boldsymbol{\epsilon}}_\vartheta]\cdot\hat{\mathbf{e}}_{\sigma^+} = -\frac{1}{\sqrt{2}}\,\sin\vartheta\cos\theta\cos\phi - \frac{i}{\sqrt{2}}\,(\cos\vartheta\sin\theta - \sin\vartheta\sin\phi)\,. \tag{1.86}$$

The modulus square of Eq. (1.86) hence reads

$$|\,\hat{\mathbf{e}}_r \times[\hat{\mathbf{e}}_r \times\hat{\boldsymbol{\epsilon}}_\vartheta]\cdot\hat{\mathbf{e}}_{\sigma^+}\,|^2 = \frac{1}{2}\,\sin^2\vartheta\cos^2\theta\cos^2\phi + \frac{1}{2}\,(\cos\vartheta\sin\theta - \sin\vartheta\sin\phi)^2\,. \tag{1.87}$$

The directions of zero σ^+-polarized light emission are then obtained if and only if both terms in the right-hand side of Eq. (1.87) are zero. By setting the first term in Eq. (1.87) to zero, there are three cases to be analyzed. We need to distinguish whether $\sin\vartheta$, $\cos\theta$ or $\sin\phi$ vanishes. First if $\vartheta = 0$, then the second term is zero if and only if the condition $\sin\theta = 0$ is satisfied. Such a case corresponds to a dipole oscillating along the x_3-axis and its zeros of emission are the poles of the unit sphere. The cases for which $\vartheta \neq 0$ are more interesting. If $\vartheta \neq 0$, then either $\theta = \pi/2$ (recall $\theta \in [0, \pi]$) or $\phi = \pm\pi/2$. Let us first assume that $\phi = \pm\pi/2$. The second term in Eq. (1.87) vanishes if and only if

$$\sin\theta = \pm\tan\vartheta\,. \tag{1.88}$$

[6]The dot product here must be taken as the Hermitian inner product for complex vectors $\mathbf{a}\cdot\mathbf{b} = \sum_j a_j^* b_j$.

Since $\sin\theta < 1$, this last condition imposes a constraint on ϑ which must then lie between 0 and $\pi/4$ in order for $\tan\vartheta$ (and hence $\sin\theta$) to be less than unity. Since $\theta \in [0,\pi]$, $\sin\theta > 0$, and only the case $\phi = \pi/2$ yields two solutions, θ_1 and θ_2 symmetric with respect to $\theta = \pi/2$. This is well illustrated in Fig. 1.6(f). Assuming now that $\theta = \pi/2$, the second term in Eq. (1.87) vanishes if and only if

$$\sin\phi = \operatorname{cotan}\vartheta\,. \tag{1.89}$$

The above condition can only be satisfied if $\vartheta \in]\pi/4, \pi/2[$ for $\sin\phi$ to be less than unity. Then there are two solutions lying between 0 and π (since $\sin\phi > 0$ for $\vartheta \in]\pi/4, \pi/2[$), and symmetric with respect to $\pi/2$. In fact, it can be shown that the polarization of the radiation of the rotating dipole for $\vartheta \in]\pi/4, \pi/2[$ corresponds to that of a rotating dipole for which $\vartheta' = \pi/2 - \vartheta$ (e.g. that from Fig. 1.6(f)) but rotated by 90° with respect to the x_2-axis. Indeed, it is clear from Eq. (1.83) that the mapping $\vartheta \mapsto \pi/2 - \vartheta$ rotates the ellipse by 90° around the x_2-axis. The above analysis allows for a very intuitive geometrical interpretation. For $\vartheta < \pi/4$, the rotating dipole describes an ellipse whose long axis is along the x_3-axis. The two directions of zero σ^+-polarized emission correspond to the two directions from which the ellipse is seen as a circle and with the orientation of the dipole rotation seen as opposite to that of the measured polarization. In fact, these two directions correspond to pure σ^--polarized emission. This explains why the zeros of emission are found on the meridian $\phi = \pi/2$ in that case. For $\vartheta > \pi/4$ the ellipse's long axis is along the x_1-direction, which explains the fact that zeros are found on the equator. By symmetry, the zero scattering direction of the σ^--polarized emission are symmetric to those of the σ^+-polarized emission with respect to the $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_3)$-plane. In the degenerate case $\vartheta = \pi/4$, the two solutions merge into one, $(\theta, \phi) = (\pi/2, \pi/2)$ i.e. that $\hat{\mathbf{e}}_r = \hat{\mathbf{e}}_2$, as illustrated in Fig. 1.6(e).

1.4 Reflection and refraction at a planar interface

1.4.1 Reflection and transmission amplitudes

In the case of a planar interface the scattering problem for an incident monochromatic plane wave reads: find $(\mathbf{E}_{\text{scat}}, \mathbf{H}_{\text{scat}})$ such that

$$\begin{cases} \Delta\,\mathbf{F}(\mathbf{r}) + \frac{\omega^2}{c^2}\,\epsilon_1\;\mathbf{F}(\mathbf{r}) = 0 & \forall\,\mathbf{r} \in \mathbb{R}^2 \times \mathbb{R}_+^*\,, \\ \Delta\,\mathbf{F}(\mathbf{r}) + \frac{\omega^2}{c^2}\,\epsilon_2\;\mathbf{F}(\mathbf{r}) = 0 & \forall\,\mathbf{r} \in \mathbb{R}^2 \times \mathbb{R}_-^*\,, \\ \hat{\mathbf{e}}_3 \times \lim\limits_{\eta \to 0}\,[\mathbf{F}(\mathbf{x} + \eta\,\hat{\mathbf{e}}_3) - \mathbf{F}(\mathbf{x} - \eta\,\hat{\mathbf{e}}_3)] = 0 & \forall\,\mathbf{x} \in \mathbb{R}^2 \times \{0\}\,. \end{cases} \tag{1.90a}$$

Here $\mathbf{F}$ is either the total electric field $\mathbf{E}$ or magnetic field $\mathbf{H}$, the notation $\mathbf{x} \in \mathbb{R}^2 \times \{0\}$ must be understood as $\mathbf{x}$ a point in the plane $x_3 = 0$ and η is a positive real number. The last equation, is the condition of continuity of the tangential component of the electric and magnetic fields at the boundary. Note that we have dropped the frequency dependence as we study the scattering of a monochromatic plane wave in the linear regime. To be more accurate, as explained in the previous subsection, the solution of Eq. (1.90) is not unique. Indeed, we have not yet explicitly stated the incident field. The incident electric field is described by a monochromatic plane wave incident from medium 1 ($x_3 > 0$) and given by Eq. (1.33), which we recall here omitting the time dependence

$$\mathbf{E}_0(\mathbf{r}) = \exp\left(i\,\mathbf{k}_1^-(\mathbf{p}_0)\cdot\mathbf{r}\right)\;\left[\mathcal{E}_{0,p}\;\hat{\mathbf{e}}_{p,1}^-(\mathbf{p}_0) + \mathcal{E}_{0,s}\;\hat{\mathbf{e}}_s(\mathbf{p}_0)\right]\,.$$

We introduce $\mathbf{F}_{1,2}$ the restriction of $\mathbf{F}$ respectively in the domain $\mathbb{R}^2 \times \mathbb{R}^*_+$ and $\mathbb{R}^2 \times \mathbb{R}^*_-$. It can be shown that the reflected field and the refracted field consist each of a plane wave and read

$$\mathbf{E}_{\text{scat},1}(\mathbf{r}) = \exp\left(i\,\mathbf{k}_1^+(\mathbf{p}_0)\cdot\mathbf{r}\right)\ \left[\rho_p(\mathbf{p}_0)\,\mathcal{E}_{0,p}\ \hat{\mathbf{e}}^+_{p,1}(\mathbf{p}_0) + \rho_s(\mathbf{p}_0)\,\mathcal{E}_{0,s}\ \hat{\mathbf{e}}_s(\mathbf{p}_0)\right] \tag{1.91a}$$

$$\mathbf{E}_{\text{scat},2}(\mathbf{r}) = \exp\left(i\,\mathbf{k}_2^-(\mathbf{p}_0)\cdot\mathbf{r}\right)\ \left[\tau_p(\mathbf{p}_0)\,\mathcal{E}_{0,p}\ \hat{\mathbf{e}}^-_{p,2}(\mathbf{p}_0) + \tau_s(\mathbf{p}_0)\,\mathcal{E}_{0,s}\ \hat{\mathbf{e}}_s(\mathbf{p}_0)\right]\ . \tag{1.91b}$$

Remark 1.5. Equation (1.91) contains a few results worth mentioning. First, in view of the wave vectors of the reflected and refracted wave, $\mathbf{k}_1^+(\mathbf{p}_0)$ and $\mathbf{k}_2^-(\mathbf{p}_0)$ respectively, we note that their projection onto the $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2)$-plane are identical to the projection of the wave vector of the incident wave, namely $\mathbf{p}_0$. In other words, the projection of the wave vector onto the plane of the interface is conserved under reflection and refraction. This is the well-known, *Snell-Descartes* law. Second, the amplitudes for the p- and s-polarization components are proportional to the amplitudes of the p- and s-polarization components of the incident wave respectively. In other words, the p and s components of the scattering are decoupled (i.e. $R_{sp} = R_{ps} = T_{sp} = T_{ps} = 0$, compare Eq. (1.91) with Eq. (1.35)). These properties can be summarized in terms of the reflection and transmission amplitudes (i.e. in Fourier space, see Eq. (1.35)) as

$$\mathbf{R}(\mathbf{p}\,|\,\mathbf{p}_0) = (2\pi)^2\,\delta(\mathbf{p}-\mathbf{p}_0)\begin{pmatrix}\rho_p(\mathbf{p}_0) & 0\\ 0 & \rho_s(\mathbf{p}_0)\end{pmatrix}\ , \tag{1.92a}$$

$$\mathbf{T}(\mathbf{p}\,|\,\mathbf{p}_0) = (2\pi)^2\,\delta(\mathbf{p}-\mathbf{p}_0)\begin{pmatrix}\tau_p(\mathbf{p}_0) & 0\\ 0 & \tau_s(\mathbf{p}_0)\end{pmatrix}\ . \tag{1.92b}$$

The complex amplitudes ρ_α and τ_α are known as the *Fresnel amplitudes* and are given by [17]

$$\rho_s(\mathbf{p}_0) = \frac{\alpha_1(\mathbf{p}_0) - \alpha_2(\mathbf{p}_0)}{\alpha_1(\mathbf{p}_0) + \alpha_2(\mathbf{p}_0)} \tag{1.93a}$$

$$\rho_p(\mathbf{p}_0) = \frac{\epsilon_2\alpha_1(\mathbf{p}_0) - \epsilon_1\alpha_2(\mathbf{p}_0)}{\epsilon_2\alpha_1(\mathbf{p}_0) + \epsilon_1\alpha_2(\mathbf{p}_0)} \tag{1.93b}$$

$$\tau_s(\mathbf{p}_0) = \frac{2\,\alpha_1(\mathbf{p}_0)}{\alpha_1(\mathbf{p}_0) + \alpha_2(\mathbf{p}_0)} \tag{1.93c}$$

$$\tau_p(\mathbf{p}_0) = \frac{2\,\sqrt{\epsilon_1\epsilon_2}\,\alpha_1(\mathbf{p}_0)}{\epsilon_2\alpha_1(\mathbf{p}_0) + \epsilon_1\alpha_2(\mathbf{p}_0)}\ . \tag{1.93d}$$

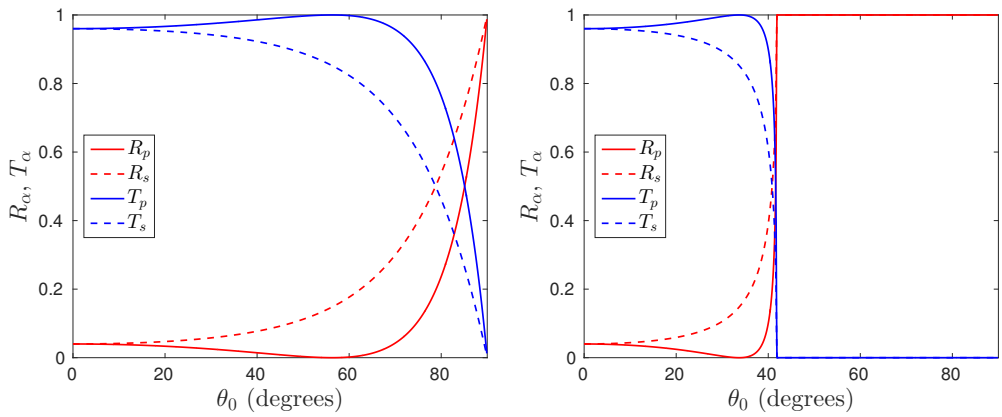
For loss-less media ($\epsilon \in \mathbb{R}$), we can verify that the sum of the efficiencies of the reflected and refracted wave is unity, i.e. that the electromagnetic energy is conserved, that is

$$R_\alpha + T_\alpha = 1\,, \tag{1.94}$$

where here we have denoted by R_α and T_α the efficiencies of the reflected and transmitted wave (see Eq. (1.68) for the definition of the efficiency), which are also known as reflectivity and transmittivity in this context, and read

$$R_\alpha(\mathbf{p}_0) = |\rho_\alpha(\mathbf{p}_0)|^2 \tag{1.95a}$$

$$T_\alpha(\mathbf{p}_0) = \frac{\alpha_2(\mathbf{p}_0)}{\alpha_1(\mathbf{p}_0)}\,|\tau_\alpha(\mathbf{p}_0)|^2\,. \tag{1.95b}$$



(a) $\epsilon_1 = 1$, $\epsilon_2 = 2.25$. (b) $\epsilon_1 = 2.25$, $\epsilon_2 = 1$.

Figure 1.7: Reflectivity and transmissivity for the scattering of a p- and s-polarized plane wave by a planar interface between two dielectric media as a function of the angle of incidence θ_0.

For dissipative media ($\mathrm{Im}\,\epsilon > 0$) the sum of efficiencies is smaller than unity, as electromagnetic energy is absorbed in the dissipative medium and eventually converted into heat. We illustrate in Fig. 1.7 the variation of the reflectivity and transmittivity for p- and s-polarized waves as functions of the angle of incidence θ_0 for dielectric media, both in the case where the excitation is incident in the optically less dense medium ($\epsilon_1 < \epsilon_2$, Fig. 1.7(a)) and in the case where the excitation is incident in the optically denser medium ($\epsilon_1 > \epsilon_2$, Fig. 1.7(b)). Two well-known phenomena are illustrated in Fig. 1.7. First, the phenomenon of *total internal reflection* can be observed in Fig. 1.7(b) in the case of reflection in the optically denser medium. There exists, in this case, a critical angle of incidence, $\theta_0 = \theta_c = \arcsin(n_2/n_1)$, above which the power incident on the surface is entirely reflected. This does not mean, however, that no wave is present in medium 2. The wave in medium 2 is evanescent as the norm of its in-plane wave vector is larger than $n_2\,\omega/c$, which translates the fact that the phase velocity of the wave is larger than that allowed in medium 2. The wave is hence bounded to the surface. The presence of the evanescent wave can be revealed for example with the phenomenon of frustrated total internal reflection in which a third medium, say with $\epsilon_3 = \epsilon_1$, is placed in the vicinity of the interface between media 1 and 2. The power of the incident wave is hence redistributed between the reflected wave and the one transmitted in medium 3, as photons are allowed to tunnel through the thin gap between media 1 and 2 with appreciable probability; the thinner the gap, the larger the probability (we take here a quantum mechanical picture for simplicity). The second phenomenon of interest which is present in both configurations considered in Fig. 1.7 is the so-called Brewster or polarizing angle. For p-polarized light, there exists an angle of incidence $\theta_0 = \theta_\mathrm{B} = \arctan(n_2/n_1)$, known as Brewster's angle, for which the reflected wave disappears. Contrarily to the phenomenon of total internal reflection, it is really the reflected wave which disappears in the sense that the reflection amplitude vanishes, and not even an evanescent wave is present. At the Brewster angle of incidence, unpolarized light is then reflected with a purely s-polarized state, hence the denomination of polarizing angle. This phenomenon can be interpreted at different level of depths of understanding. One way of understanding the Brewster effect is to first acknowledge the zero of the reflection amplitude for p-polarized light

and to ask ourselves how were the Fresnel coefficients derived and what is the fundamental difference between p- and s-polarized light. The Fresnel coefficients are usually derived by assuming the presence of three plane waves of same linear polarization (p or s): the incident, the reflected, and refracted waves. What determines the precise expression of the Fresnel coefficient for s- or p- polarized light is directly linked to the boundary conditions. In both cases, the problem of vector wave scattering reduces to a scalar problem but the boundary conditions at the interface are different. The zero of reflection amplitude for p-polarized light at the Brewster angle can then be interpreted as a situation of impedance matching, i.e. a configuration in which the refracted wave only is necessary to satisfy the boundary conditions together with the incident wave. This picture can be taken even further by claiming that the incident wave and the refracted wave, seen as one unified mode, is in some sense a not bounded eigenmode of the system, similar to a surface plasmon polariton which is a eigenmode bounded to the surface on both side of the interface (see next section). Such an interesting picture is motivated and well illustrated in [7]. Although these considerations give a somewhat clarified mathematical intuition of the phenomenon, the underlying physical mechanism remains rather obscure. It is thus time for the next questions: where do the different boundary conditions come from, and what do they physically mean? The boundary conditions across an interface between two media are encoded in the Maxwell equations, Eq. (1.1), taken in the sense of distribution, equations which we have deliberately taken for granted. In fact, these equations are the result of a modeling of the response of a medium, composed of atoms, from a microscopic picture to a macroscopic picture averaged over many atoms. The details of such a modeling would take us too far beyond the scope of this work, and we only need here to make one simple but important remark. An atom is composed a positively charged nucleus and negatively charged electrons. Hence, from the point of view of its interaction with light, an atom can be modeled at the simplest level of sophistication as a dipole. It is thus expected that appropriate macroscopic description of continuous media should retain the fundamental dipole character of the underlying discrete constituents of the medium. This is well hidden in Maxwell equations, but it is there. We can hence hope for a microscopic picture of the Brewster phenomenon in terms of dipole radiation. The commonly accepted picture for explaining the Brewster phenomenon in the case of a plane wave incident on a vacuum/dielectric interface is that the dipoles in the dielectric medium oscillate along the electric field of the refracted wave. Since an oscillating dipole does not radiate any power along the direction of oscillation, a vanishing reflected wave is observed when the direction of propagation of the reflected wave (which is produced by the response of the dielectric medium, i.e. the sum of all dipoles radiation) is colinear to the direction of oscillation of the dipoles in the dielectric medium. In other words, the Brewster phenomenon is observed when the wave vectors of the reflected and refracted waves are orthogonal. Although the above picture is somewhat satisfactory, it must be complemented with some remarks and a more subtle analysis. In the case of internal reflection, the Brewster phenomenon also occurs even though no dipole is present in the vacuum. This seems to indicate that the reflected wave, which is still produced by the response of the dielectrics but as the medium of incidence this time, vanishes although the dipoles supposedly oscillating now along the direction of the electric field of the incident wave are not oriented along the wave vector of the reflected wave. This issue has been clarified, to the best of our knowledge, by Doyle in 1985 [27] based on the concept of the Ewald wave-triad introduced by Ewald back in 1916 [28] when deriving the Ewald-Oseen extinction theorem based on principles of microscopic optics.

We would like to stress here that we have chosen to take the above discussion at a very conceptual level without motivating our claims with too much mathematical details. The reason for this

choice is that we only want to prepare the reader to a more in-depth analysis and generalization of this discussion in the case of light scattering by a weakly rough interface in Chapter 10. In particular, the Fresnel coefficients will be revisited in details using a scalar-polarization factorization which follows very naturally from the reduced Rayleigh equations and which were discussed by Doyle in [27]. Such a factorization of the Fresnel coefficients is fruitful for understanding the physics from a microscopic picture and will be analyzed in details in Chapter 10.

1.4.2 Surface plasmon polariton

So far we have been paying attention to waves reflected from and refracted through a planar interface assuming that the incident wave was propagating, as coming from infinity. We have seen that in the case of total internal reflection the refracted wave is evanescent in the optically less dense medium while the reflected wave propagates away from the surface. We are now asking ourselves whether evanescent waves, bound to the surface from the two sides can exist. The answer is yes, but under some conditions. In particular, we are looking for non-trivial solutions of the homogeneous problem in which no sources are present. In other words, we are looking for surface waves that are propagating freely along the surface without coupling out to modes propagating away from the surface nor from incoming propagating waves. Assuming loss-less non-magnetic isotropic media, it can be shown that such a mode can exist if (i) the dielectric constants of the two media are of opposite sign, and (ii) that the wave is p-polarized. In addition, a specific condition is required for the wave vector of the wave as we will soon see. Assuming $\epsilon_1 > 0$ and $\epsilon_2 < 0$, all these conditions can be derived by assuming a field taking the form of a plane wave evanescent on both side of the surface as

$$\mathbf{E}_1(\mathbf{r}) = \exp\left(i\,\mathbf{k}_1^+(\mathbf{p})\cdot\mathbf{r}\right)\mathbf{A}_1 \tag{1.96a}$$

$$\mathbf{E}_2(\mathbf{r}) = \exp\left(i\,\mathbf{k}_2^-(\mathbf{p})\cdot\mathbf{r}\right)\mathbf{A}_2 \tag{1.96b}$$

$$|\,\mathbf{p}\,| > \sqrt{\epsilon_1}\,\frac{\omega}{c}\,, \tag{1.96c}$$

where $\mathbf{A}_1$ and $\mathbf{A}_2$ are constant amplitudes. We then can infer the required conditions for satisfying the boundary conditions (see e.g. [8]). In order to avoid a lengthy derivation which can be found in textbooks on plasmonics [8], we simply take a short cut which makes use of the Fresnel coefficients and which we motivate by an analogy with the response of a harmonic oscillator. For a harmonic oscillator, it is known that when excited at an eigen-frequency with an external force, the response of the system is resonant and the oscillation becomes unbounded. The idea is then to look for conditions which would make the Fresnel coefficients diverge. It is clear from Eq. (1.93) that the denominator for s-polarized light, $\alpha_1+\alpha_2$, never vanishes however the dielectric constant and incident wave vector is chosen, which hints at the fact that there does not exist any s-polarized eigen-mode. However, the denominator for p-polarized light, $\epsilon_2\alpha_1+\epsilon_1\alpha_2$, may vanish if ϵ_1 and ϵ_2 have opposite signs. In addition, the corresponding in-plane wave vector, $\mathbf{p}_{\mathrm{SPP}}$, is thus such that

$$\epsilon_2\alpha_1(\mathbf{p}_{\mathrm{SPP}}) + \epsilon_1\alpha_2(\mathbf{p}_{\mathrm{SPP}}) = 0\,. \tag{1.97}$$

Together with the dispersion relations in both media, $|\,\mathbf{k}_j^{\pm}(\mathbf{p}_{\mathrm{SPP}})|^2 = \epsilon_j\omega^2/c^2$, we can solve for the in-plane wave vector as a function of the dielectric constants to obtain

$$|\,\mathbf{p}_{\mathrm{SPP}}\,| = \left(\frac{\epsilon_1\epsilon_2}{\epsilon_1+\epsilon_2}\right)^{1/2}\frac{\omega}{c}\,. \tag{1.98}$$

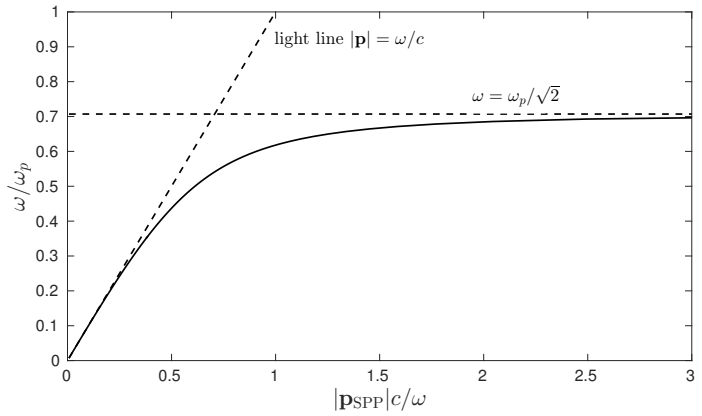


Figure 1.8: Dispersion curve for a surface plasmon polariton propagating along a planar surface between vacuum $\epsilon_1 = 1$ and a loss-less Drude metal $\epsilon_2 = 1-\omega_p^2/\omega^2$. The dashed lines indicate the light line, i.e. the dispersion curve for a grazing wave propagating in vacuum and the asymptotic line of the SPP dispersion curve.

Equation (1.98) is the dispersion relation for a surface plasmon polariton propagating along a planar surface between a dielectric and a metal. Note that due to the rotational invariance of the system, no privileged direction can be inferred which is the reason for only knowing the norm of the in-plane vector. However, when excited, the excitation may break the symmetry and lead to a specific direction for $\mathbf{p}_{\rm SPP}$. It is instructive to consider the dispersion curve of a surface plasmon polariton on a planar interface between vacuum and a metal whose frequency dependent dielectric function is modeled by a Drude model (see Eq. (1.17)). The dispersion curve is illustrated in Fig. 1.8. From the dispersion curve in Fig. 1.8, it is clear that the norm of the in-plane wave vector of a SPP is larger than that of a grazing wave propagating in vacuum, as the dispersion curve lies to the right of the so-called light line. This would also occur if we had a dielectric with $\epsilon_1 > 1$ as medium 1 and the dispersion curve would lie to the right of the line of equation $|\,\mathbf{p}\,| = \sqrt{\epsilon_1}\omega/c$. Consequently, a SPP on a planar interface cannot be excited by an incident progressive (propagating) wave. There exists essentially two ways to excite a SPP with electromagnetic waves. The first idea is to excite a SPP by the use of an evanescent field decaying away from another object which when taken close to the surface excites a SPP as the evanescent field may couple to a SPP. Examples of such a set up are the Otto and Kretschmann configurations [8]. The Otto configuration consists in approaching a prism close to the metallic surface and in shining light in such a way that the light reflected inside the prism would undergo total internal reflection in the absence of the metal. The evanescent field on the side of the prism facing the metallic surface leaks the metal and may excite a SPP at the vacuum/metal interface. Alternatively, a prism can be placed directly on top of a thin metallic film and a SPP may be excited on the metal/vacuum interface; this is the Kretschmann configuration. A second idea to excite a SPP is to consider a corrugated interface instead of a planar one. As will be discussed in the next section, a periodic grating diffracts an incident plane wave into a set of plane waves which propagate with specific wave vectors given by the grating formula, as well as a set of evanescent waves. In the case where such an evanescent mode has a in-plane wave

vector close to that of a SPP, a resonance may occur. More generally, SPP may be excited by corrugations or defects on a metallic surface.

Remark 1.6. We would like here to precise a terminology that may at time lead to confusion. We talk of an evanescent wave for a wave whose intensity decays exponentially in the direction orthogonal to the average plane of the interface. The term of superficial wave or bounded wave may also be found in the literature. These are *not* necessarily resonant modes such as the ones described above like a surface plasmon polariton which is special type of evanescent wave being solution of the homogeneous problem. The term of surface wave may be found in the literature to designate such specific resonant waves.

1.5 Scattering by a periodic surface

1.5.1 Diffractive orders and the grating formula

In the case of the scattering of a monochromatic plane wave by a periodic surface, or grating, it can be shown that only a discrete set of wave components are allowed in the scattered field. The in-plane wave vectors of the scattered waves are given by the grating formula, stating that the difference between the in-plane wave vector of a scattered wave and the in-plane wave vector of the incident wave must lie on the reciprocal lattice of the surface (see Fig. 1.9). Mathematically, for a surface whose periodicity is characterized by the two primitive lattice vectors $\mathbf{a}_1$ and $\mathbf{a}_2$, the scattered field may be written as

$$\mathbf{E}_l^{(\mathrm{scat})}(\mathbf{r}) = \sum_{\boldsymbol{\ell}\in\mathbb{Z}^2} \mathbf{E}_l^{(\boldsymbol{\ell})}(\mathbf{p}_0) e^{i\,\mathbf{k}_l^{\pm}(\mathbf{p}_{\boldsymbol{\ell}})\cdot\mathbf{r}} = \int_{\mathbb{R}^2} (2\pi)^2 \left(\sum_{\boldsymbol{\ell}\in\mathbb{Z}^2} \mathbf{E}_l^{(\boldsymbol{\ell})}(\mathbf{p}_0)\, \delta(\mathbf{p}-\mathbf{p}_{\boldsymbol{\ell}}) \right) e^{i\,\mathbf{k}_l^{\pm}(\mathbf{p})\cdot\mathbf{r}}\ \frac{\mathrm{d}^2 p}{(2\pi)^2}\,, \tag{1.99}$$

where the in-plane wave vectors $(\mathbf{p}_{\boldsymbol{\ell}})_{\boldsymbol{\ell}\in\mathbb{Z}^2}$ are given by

$$\mathbf{p}_{\boldsymbol{\ell}} = \mathbf{p}_0 + \mathbf{G}^{(\boldsymbol{\ell})}\,, \tag{1.100}$$

and where $\mathbf{p}_0$ denotes the incident in-plane wave vector and $\mathbf{G}^{(\boldsymbol{\ell})}$ is a reciprocal lattice vector defined by

$$\mathbf{G}^{(\boldsymbol{\ell})} = \ell_1\ \mathbf{b}_1 + \ell_2\ \mathbf{b}_2\ . \tag{1.101}$$

Here $\mathbf{b}_1$ and $\mathbf{b}_2$ are the primitive reciprocal lattice vector defined by the relation

$$\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\,\delta_{ij} \qquad \text{for } i,j \in \{1,2\}\,. \tag{1.102}$$

Based on Eq. (1.99), the reflection and transmission amplitudes are then of the form

$$\mathbf{X}\,(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}}) = (2\pi)^2 \sum_{\boldsymbol{\ell}\in\mathbb{Z}^2} \delta(\mathbf{p}-\mathbf{p}_{\boldsymbol{\ell}})\ \mathbf{X}^{(\boldsymbol{\ell})}(\mathbf{p}_0)\,, \tag{1.103}$$

where $\mathbf{X} = \mathbf{R}$ or $\mathbf{T}$. In addition, by using the dispersion relation $|\,\mathbf{k}_l^{\pm}(\mathbf{p}_{\boldsymbol{\ell}})|^2 = \epsilon_l\,\omega^2/c^2$, the only waves that have an in-plane wave vector smaller than $\sqrt{\epsilon_l}\,\omega/c$ can propagate in medium l while the remaining are evanescent waves. For the propagating waves, the polar and azimuthal

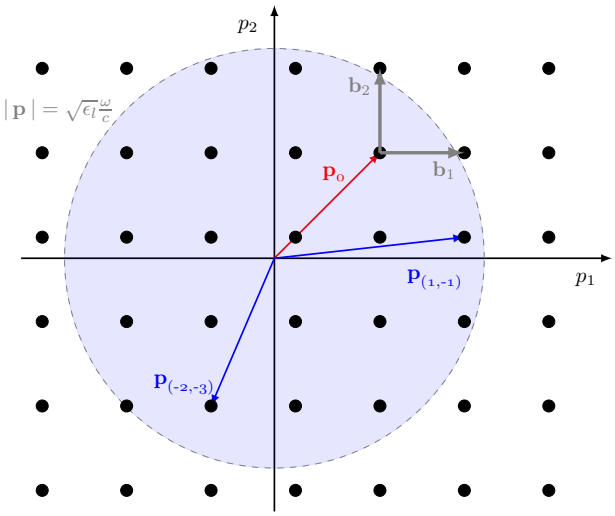


Figure 1.9: Illustration of the grating formula for a square grating. The in-plane wave vectors allowed in the scattered field expansion differs from the incident in-plane wave vector $\mathbf{p}_0$ by a reciprocal lattice vector. The wave components, or modes, whose in-plane wave vector lies inside the blue disk of radius $\sqrt{\epsilon_l}\,\omega/c$ are the only modes allowed to propagate in medium l, while the modes with in-plane wave vector outside the disk are evanescent in medium l.

angles of scattering, $\theta_l^{(\boldsymbol{\ell})}$ and $\phi_l^{(\boldsymbol{\ell})}$, between the normal to the average plane of the surface and the diffracted wave vector $\mathbf{k}_l^{\pm}(\mathbf{p}_{\boldsymbol{\ell}})$ are then given by

$$\sqrt{\epsilon_l}\,\frac{\omega}{c}\,\sin\theta_l^{(\boldsymbol{\ell})} = |\,\mathbf{p}_{\boldsymbol{\ell}}\,| \tag{1.104a}$$

$$\cos\phi_l^{(\boldsymbol{\ell})} = \frac{\mathbf{p}_{\boldsymbol{\ell}}\cdot\hat{\mathbf{e}}_1}{|\,\mathbf{p}_{\boldsymbol{\ell}}\,|} \tag{1.104b}$$

$$\sin\phi_l^{(\boldsymbol{\ell})} = \frac{\mathbf{p}_{\boldsymbol{\ell}}\cdot\hat{\mathbf{e}}_2}{|\,\mathbf{p}_{\boldsymbol{\ell}}\,|}\,. \tag{1.104c}$$

For a one dimensional grating illuminated under non-conical incidence (i.e. with the plane of incidence perpendicular to the principal direction of the corrugations) the grating formula for the scattered angles takes the well-known expression

$$n_l\,\sin\theta_l^{(\boldsymbol{\ell})} = n_1\,\sin\theta_0 + \ell_1\,\frac{\lambda}{a}\,. \tag{1.105}$$

1.5.2 Rayleigh and Wood anomalies

In a reflectivity measurement of a periodic grating, when the light intensity reflected specularly from the grating is recorded as one varies the angle of incidence or the wavelength of the source, the recorded intensity may experience sudden variations when crossing specific angles of incidence or specific wavelengths. Depending on the incident polarization and whether the material composing the grating is a metal or a dielectric, one may observe two types of such anomalies. Some occur as sharp changes of intensity, making edges in the intensity spectrum,

and can be observed regardless of polarization and the material constituting the grating. Other occur as broader variations over a spectral or angular range only for metallic grating illuminated under p-polarization. These two types of anomalies are known as Rayleigh (or Rayleigh-Wood) and Wood anomalies respectively. The next paragraph gives a brief historical development for the understanding of these two types of anomaly based on a more complete and excellent review written by Maystre in chapters 1 and 2 of Ref. [7] and the references therein. Then the physical interpretation of these phenomena will be illustrated in some more details.

The so-called *Wood anomalies* were observed by Wood in 1902 while studying the reflection of a spectrally continuous source of light by a metallic grating which he named "singular anomalies" and he made the crucial remark that they could only be observed for p-polarization [29]. Lord Rayleigh attempted to give an explanation of these anomalies using arguments based on the apparition or disappearance of propagating diffracted modes at grazing emergence and could predict within a few percent accuracy the spectral positions of the anomalies based on the grating formula [30, 15]. The small deviation between Lord Rayleigh's calculation and the positions of the anomalies observed by Wood were believed to be due to poor knowledge of the precise value of the grating period. Only thirty years later, did Strong [31] complement Rayleigh's conjecture by distinguishing two types of anomalies as he studied experimentally the light reflected by gratings with identical period but made of different materials and different profiles. Strong concluded from his study that on the one hand "the dark band [anomaly] has a sharp edge which falls at the wave-length predicted by Rayleigh's relation at a position which is independent of the nature of the film which develops the band" and on the other hand "the bright band is displaced from the dark band an amount which depends on the nature of the metal film" [31]. In other words, Rayleigh's conjecture only explains the sharp anomaly which only depends on the period of the grating and angle of incidence while the anomalies observed by Wood seemed to be rather of the second type described by Strong, and cannot be explained purely on geometrical consideration, such as the grating formula, since the positions of the anomalies depend on the material. Considering these new insights, Fano gave the first modern theoretical explanation of the Wood anomalies based on the excitation of surface plasmon polariton, although the term of surface plasmon polariton and its condensed matter interpretation appeared later. The terms used by Fano was "forced resonance" of "superficial stationary waves" and "quasi-stationary waves" ("superficial wave" must be read as a synonym of evanescent wave and "quasi-stationary" refers to damped SPP), but apart from the terminology, Fano's interpretation and calculations correspond to that of the excitation of a SPP [32]. As pointed out by Strong, Fano made the distinction between the Rayleigh anomaly, edge in the intensity spectrum, from the broad or "diffuse" Wood anomaly which starts from the Rayleigh edge and consists of a dark band followed by bright band red-shifted from the Rayleigh edge. Fano derived, based on a perturbative solution of the Rayleigh method to second order in the surface amplitude, that (i) the diffraction amplitudes of *all* modes should exhibit a step at the passing off of a grazing diffracted order (and it was also noted that the jump may or may not be significant depending on the parameters of the system), and (ii) that the Wood anomaly occurs as an evanescent diffracted mode has an in-plane wave vector approaching that of the SPP which was predicted to be always slightly larger than the wave number in vacuum hence explaining the red-shift with respect to the Rayleigh edge, and (iii) the broad nature of the Wood anomaly is analogous to that of the resonant amplitude of a damped oscillator. The theoretical and phenomenological understanding of the Wood anomalies were then subsequently developed through experimental and theoretical studies (see e.g. [33]).

Rayleigh anomalies

A Rayleigh anomaly may be observed as an edge in the efficiency of diffracted modes in either the medium of incidence (reflected mode) or in the medium of transmission (transmitted mode) as one varies the angle of incidence or the wavelength of the incident light. As an illustrative example, let us consider the situation depicted in Fig. 1.10. A one-dimensional grating is illuminated under non-conical incidence by a plane wave with a wavelength larger than the lattice constant and such that for not too large angles of incidence only the zero order reflected and transmitted modes are allowed to propagate (Fig. 1.10(a)). For such a small angle of incidence, the discrete modes with in-plane wave vectors $\mathbf{p}_{\boldsymbol{\ell}}$ for $\boldsymbol{\ell} \neq (0,0)$ lie outside the propagation disk both in the medium of incidence and the medium of transmission, and hence are evanescent modes in both media, as represented by the red fins in Fig. 1.10(a). Now, by increasing the angle of incidence, their will be a certain angle at which the mode $\mathbf{p}_{(-1,0)}$ will lie exactly on the boundary of the propagation disk in the transmitted medium (assuming here dielectric media with $\epsilon_1 < \epsilon_2$). In other words, for an angle of incidence just below this critical angle the mode given by $\mathbf{p}_{(-1,0)}$ is evanescent in medium 2, while for an angle of incidence just above this critical angle this mode propagates in medium 2 with grazing angle of scattering. This transition from an evanescent to a propagating mode is associated with the opening of an additional channel in which the radiated power can be transported. Consequently, we may observe a re-distribution of the scattered power shared between zero order reflected and refracted modes and the transmitted mode characterized by $\boldsymbol{\ell} = (-1,0)$, hence featuring a sudden variation in the reflectivity and transmissivity measurements (for which only the power radiated by the zero orders is recorded). Note that during this transition, the *reflected* mode $\boldsymbol{\ell} = (-1,0)$ is still evanescent. By further increasing the angle of incidence, we may reach other critical angles for which consecutive transitions from evanescent to propagating modes in either media can occur. Note that if we had started with a wavelength smaller than the lattice constant, some propagating diffractive order would already be present at normal incidence. By increasing the angle of incidence, some transitions from evanescent to propagating modes may occur, as previously illustrated, but also transitions from propagating to evanescent modes can occur as well. The reader may consider for instance the mode $\boldsymbol{\ell} = (1,0)$.

We have considered in the previous discussion the evanescent-propagating transitions when varying the angle of incidence and fixed wavelength. We may be interested by fixing an angle of incidence, and then varying the wavelength instead. In this case, the disk of propagation will have varying radius. Therefore, we way observe evanescent-propagating transitions at discrete wavelengths (or equivalently photon energies) when an in-plane wave vector $\mathbf{p}_{\boldsymbol{\ell}}$ is crossed by the varying boundary of the propagating disk.

Now extending the discussion to the more general case of a two-dimensional lattice, we may ask the following question. What is the set of points in the parameter space $(\theta_0, \phi_0, \lambda)$, or equivalently $(\theta_0, \phi_0, \hbar\omega)$, that are associated with an evanescent-propagating transition of a given mode $\boldsymbol{\ell}$ in a given medium? Such a set of points may in general draw a surface in the $(\theta_0, \phi_0, \lambda)$ space, and we will refer to such surfaces as *Rayleigh surfaces*. If one fixes one of the parameters, say λ, we may refer as *Rayleigh lines* the subset of a Rayleigh surface intersected by the plane characterizing the fixed parameter (e.g. $\lambda =$ constant). To obtain the equation of a Rayleigh surface, it suffices to take the square modulus of the grating formula and set $|\,\mathbf{p}_{\boldsymbol{\ell}}\,|^2 = \epsilon_l\,\omega^2/c^2$ to translate the fact that $\mathbf{p}_{\boldsymbol{\ell}}$ must lie on the boundary of the propagation disk for the medium l at the transition. We then obtain the equation of the Rayleigh surface for the mode $\boldsymbol{\ell}$ in medium

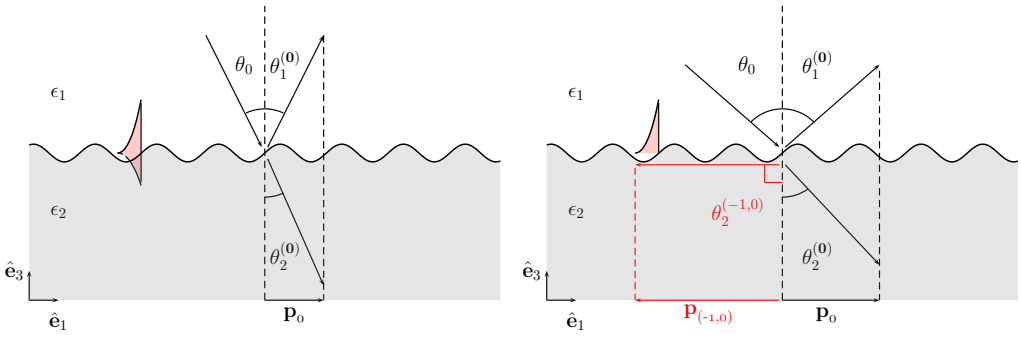


Figure 1.10: Illustration of a Rayleigh anomaly. Dielectric grating with period $a < \lambda$, illuminated at an angle θ_0 such that: (a) only the zero reflected and transmitted diffractive orders are allowed to propagate by the dispersion relation, (b) the transmitted diffractive order $\boldsymbol{\ell} = (-1, 0)$ is allowed to propagate with grazing direction. By increasing the angle of incidence from (a) to (b), the evanescent wave characterized by the order $\boldsymbol{\ell} = (-1, 0)$ in the substrate becomes a propagating wave.

l as

$$|\,\mathbf{p}_0 + \mathbf{G}_{\boldsymbol{\ell}}\,|^2 = \epsilon_l \,\frac{\omega^2}{c^2} \tag{1.106}$$

$$|\sqrt{\epsilon_1}\,\frac{\omega}{c}\,(\sin\theta_0\cos\phi_0\,\hat{\mathbf{e}}_1 + \sin\theta_0\sin\phi_0\,\hat{\mathbf{e}}_2) + \mathbf{G}_{\boldsymbol{\ell}}\,|^2 = \epsilon_l\,\frac{\omega^2}{c^2}\,. \tag{1.107}$$

It can also be written as

$$\left(\sqrt{\epsilon_1}\,\sin\theta_0\cos\phi_0 + \mathbf{G}_{\boldsymbol{\ell}}\cdot\hat{\mathbf{e}}_1\,\frac{\lambda}{2\pi}\right)^2 + \left(\sqrt{\epsilon_1}\,\sin\theta_0\sin\phi_0 + \mathbf{G}_{\boldsymbol{\ell}}\cdot\hat{\mathbf{e}}_2\,\frac{\lambda}{2\pi}\right)^2 = \epsilon_l\,. \tag{1.108}$$

In the special case of rectangular lattices, of lattice constants a_1 and a_2 along the x_1- and x_2-axis respectively, we have simply

$$\left(\sqrt{\epsilon_1}\,\sin\theta_0\cos\phi_0 + \ell_1\,\frac{\lambda}{a_1}\right)^2 + \left(\sqrt{\epsilon_1}\,\sin\theta_0\sin\phi_0 + \ell_2\,\frac{\lambda}{a_2}\right)^2 = \epsilon_l\,. \tag{1.109}$$

Note that in general, for a skewed lattice ℓ_1, ℓ_2, the norms a_1, a_2 of the primitive lattice vectors and the angle between them, ϑ, may contribute in both terms of Eq. (1.108). For a skewed lattice defined by the primitive lattice vectors

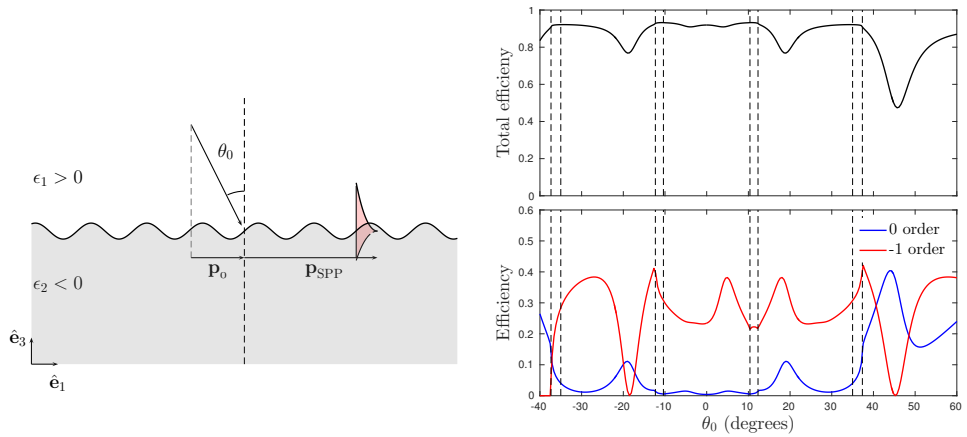
$$\mathbf{a}_1 = a_1\,\hat{\mathbf{e}}_1 \tag{1.110a}$$

$$\mathbf{a}_2 = a_2\,(\cos\vartheta\,\hat{\mathbf{e}}_1 + \sin\vartheta\,\hat{\mathbf{e}}_2)\,, \tag{1.110b}$$

and the corresponding primitive reciprocal lattice vectors read

$$\mathbf{b}_1 = \frac{2\pi}{a_1}\left(\hat{\mathbf{e}}_1 - \frac{1}{\tan\vartheta}\,\hat{\mathbf{e}}_2\right) \tag{1.111a}$$

$$\mathbf{b}_2 = \frac{2\pi}{a_2\sin\vartheta}\,\hat{\mathbf{e}}_2\,. \tag{1.111b}$$



(a) (b)

Figure 1.11: (a) Illustration of a Wood anomaly. Metallic grating illuminated at an angle θ_0 such that one of the evanescent diffracted mode excites a surface plasmon polariton as $\mathbf{p}_{\boldsymbol{\ell}} \approx \mathbf{p}_{\mathrm{SPP}}$. (b) Total efficiency and efficiencies of the zero and -1 orders reflected from a silver one-dimensional sinusoidal grating as a function of the angle of incidence for p-polarized light. The wavelength is equal to $\lambda = 476$ nm, the period is equal to $a = 1210$ nm and the amplitude of the sinusoid is $H = 95$ nm (i.e. that the peak to dip height is $2H = 190$ nm. The vertical dashed lines shows the expected positions of the Rayleigh anomalies according to Eq. (1.109). The results were obtained from computer simulation based on the reduced Rayleigh equations, a method which will be presented in the remaining of this work. The presented system was considered experimentally and numerically respectively by Hutley and Bird [34], and McPhedran and Maystre [35].

Hence the equation for the Rayleigh surface reads

$$\left(\sqrt{\epsilon_1}\,\sin\theta_0\cos\phi_0 + \ell_1\,\frac{\lambda}{a_1}\right)^2 + \left(\sqrt{\epsilon_1}\,\sin\theta_0\sin\phi_0 + \ell_2\,\frac{\lambda}{a_2\sin\vartheta} - \ell_1\,\frac{\lambda}{a_1\tan\vartheta}\right)^2 = \epsilon_l\,. \tag{1.112}$$

Remark 1.7. There is a Rayleigh surface associated with each choice of mode $\boldsymbol{\ell} = (\ell_1, \ell_2)$ and medium.

Remark 1.8. The media may be dispersive, hence $\epsilon_1 = \epsilon_1(\lambda)$ and $\epsilon_2 = \epsilon_2(\lambda)$, which indicates that finding the Rayleigh surfaces, or Rayleigh lines, may require solving Eq. (1.108) numerically. If the media are non-dispersive, solving Eq. (1.108) reduces to finding roots of a polynomial of degree two.

Wood anomalies

Let us consider a metallic periodic grating as depicted in Fig. 1.11(a). Consider first that one starts by setting an angle of incidence such that the $\boldsymbol{\ell} = (1, 0)$ mode propagates in medium 1, and then we increase the angle of incidence. For a certain angle of incidence, given by Eq. (1.112) for a fixed wavelength and azimuthal angle, the order $\boldsymbol{\ell} = (1, 0)$ turns evanescent and we obtain a Rayleigh anomaly as described in the previous paragraph and one observes a sharp edge in the efficiency as a function of the angle of incidence. At the Rayleigh anomaly

the norm of the in-plane wave vector for the mode $\boldsymbol{\ell} = (1, 0)$ is by definition equal to $\sqrt{\epsilon_1}\omega/c$. By increasing further the angle of incidence, the norm of the in-plane wave vector increases to be larger than $\sqrt{\epsilon_1}\omega/c$ and approaches the resonant in-plane wave vector for a surface plasmon polariton, $\mathbf{p}_{\mathrm{SPP}}$. Recall from Section 1.4.2 that we have derived that the norm of the in-plane wave vector for a SPP is larger than $\sqrt{\epsilon_1}\omega/c$ for a planar interface. This fact remains valid for a slightly corrugated grating where the corresponding resonant wave vector can be expected to be only slightly perturbed from that of a planar surface, or in other words, the dispersion curves for the SPP in each case would be rather close. In an analogous way to a harmonic oscillator, a divergence may be obtain in the scattering amplitudes at $\mathbf{p}_{\boldsymbol{\ell}} = \mathbf{p}_{\mathrm{SPP}}$ for a loss-less metal. In practice, metal are dissipative and it is rather the behavior of a damped oscillator which is observed, characterized by a broad variation of the scattering amplitudes, or of efficiencies for the modes which propagates. Indeed, for dissipative metals, the resonant in-plane wave vector has complex valued components and cannot be reached by $\mathbf{p}_{\boldsymbol{\ell}}$ which has real valued components, hence the broad and bounded nature of the response as $\mathbf{p}_{\boldsymbol{\ell}} \approx \mathbf{p}_{\mathrm{SPP}}$.

Figure 1.11(b) illustrates the above discussion in the case of a one-dimensional sinusoidal grating made of silver. The total efficiency (sum of efficiencies of all propagating modes) as well as the efficiency of the zero and -1 orders are shown as a function of the angle of incidence θ_0 for non-conical incidence. The sharp Rayleigh anomalies are visible in all curves at the passing on and off of various diffracted mode whose positions predicted from Eq. (1.112) are indicated by the vertical dashed lines. Furthermore, broader variations in the efficiencies are observed as the various evanescently diffracted modes excite a SPP. The results shown in Fig. 1.11(b) are computer simulations obtained based on the reduced Rayleigh equations, which is the method which will be discussed at length in the remaining of this work. This example can be considered as a first non-trivial benchmark test for the method since the obtained results are in excellent agreement with those obtained numerically by McPhedran and Maystre in Ref. [35] and the experimental measurements made by Hutley and Bird in Ref. [34] for the grating described in the figure caption. The reader is invited to compare the results provided in Fig. 1.11(b) with those obtained in the aforementioned references.

1.6 Scattering by a randomly rough surface

When an optically coherent source of light, such as a laser beam, is scattered from a randomly rough surface, we have all experienced that the intensity of the light scattered in all directions can be described, roughly speaking, as the sum of two components. One that corresponds to a strong specularly scattered peak, and one that is a *diffusely* scattered background of speckles. We need to precise what is meant by *diffuse* and to introduce some related jargon that is often encountered in the literature and that, unfortunately, may be source of confusion when describing some physical phenomena. In the following sections, we will first explain how random surfaces following the same statistical laws are described mathematically so that we have a better understanding of the type of system considered. Then some vocabulary will be introduced related to the scattering of light from randomly rough surfaces as well as the definition of the mean differential scattering coefficients, which will be the observables of interest when studying the scattering from randomly rough surfaces in this thesis.

1.6.1 Mathematical description of a random surface

In this thesis, devoted to the theory of the reduced Rayleigh equations, all the considered surfaces may be described by an equation $x_3 = \zeta(x_1, x_2) = \zeta(\mathbf{x})$, where $\mathbf{x} = x_1\,\hat{\mathbf{e}}_1 + x_2\,\hat{\mathbf{e}}_2$ represents a point in the plane $(O, \hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2)$. The surface is then characterized by a function ζ of the two in-plane variables (x_1, x_2) (see illustration in Fig. 1.3). To define random surfaces, we need a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where Ω is a sample space, $\mathcal{F}$ is a σ-algebra on Ω and $\mathbb{P}$ a probability measure on $(\Omega, \mathcal{F})$. Let the plane $\mathbb{R}^2$ be an index space, a stochastic process, $\{Z_{\mathbf{x}}\}_{\mathbf{x}\in\mathbb{R}^2}$, is a family of random variables $Z_{\mathbf{x}}$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and indexed by $\mathbf{x} \in \mathbb{R}^2$. A realization of the stochastic process is then the set of values, indexed by $\mathbb{R}^2$, taken by the random variables $Z_{\mathbf{x}}(\omega)$ for an outcome ω, i.e. that a *realization* of a stochastic process can be viewed as a function defined by $\zeta_\omega : \mathbf{x} \in \mathbb{R}^2 \mapsto \zeta_\omega(\mathbf{x}) = Z_{\mathbf{x}}(\omega)$. It is the latter notation which we will adopt, by rather having ω as an index and $\mathbf{x}$ as an argument of the realization function since we find it closer to the intuition that "for each outcome ω, one has a realization of the function ζ". It is merely a reshuffling of notations suiting better the physical intuition rather than the probabilistic one. What is important to remind is that for each outcome of the sampling space $\omega \in \Omega$, one obtains a function ζ_ω. The values taken by ζ_ω at a given point $\mathbf{x} \in \mathbb{R}^2$ describes a random variable when viewed as a function of ω, i.e. precisely $Z_{\mathbf{x}}$. Hence we will denote the corresponding random variable simply by $\zeta(\mathbf{x})$ instead of $Z_{\mathbf{x}}$, and it has to be understood as $\zeta(\mathbf{x}) : \omega \in \Omega \mapsto \zeta_\omega(\mathbf{x}) = Z_{\mathbf{x}}(\omega)$.

We will now make some assumptions on the properties we want a stochastic process to have. First, we want the realizations of the stochastic process to be continuous with respect to $\mathbf{x}$, i.e. that the realizations ζ_ω are continuous. This is indeed a rather natural property one may expect from a physical surface. Moreover, one may wish the realization functions to be differentiable. Second, we will assume that all the random variables (remember indexed by $\mathbf{x}$) $\zeta(\mathbf{x})$ follow the same law of probability, i.e. that the law of probability is independent on the point $\mathbf{x}$. For example , the random variable $\zeta(\mathbf{x})$ could follow a Gaussian law of probability, i.e. that the cumulative probability for $\zeta(\mathbf{x})$ would of the form

$$\mathbb{P}(\zeta(\mathbf{x}) < z) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{z} \exp\left(-\frac{(z'-\mu)^2}{2\sigma^2}\right) \mathrm{d}z' \,. \tag{1.113}$$

In particular, this implies that all the random variables have same average and variance. By a arbitrary choice of the origin of the coordinate system the average can be chosen to be zero and we have the following properties

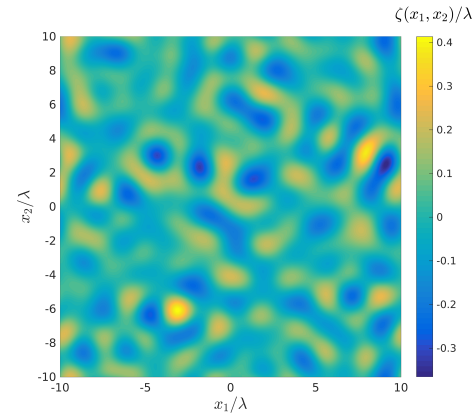
$$\langle \zeta(\mathbf{x}) \rangle = 0 \,, \tag{1.114}$$

$$\langle \zeta(\mathbf{x})^2 \rangle = \sigma^2 \,, \tag{1.115}$$

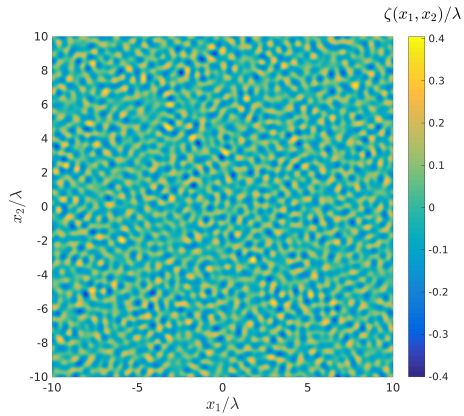
where the angle bracket denotes an ensemble average over realizations of the stochastic process. The constant $\sigma > 0$ hence denote the standard deviation of surface height with respect to the average plane $x_3 = 0$, and is often called *rms roughness* in the literature.

Remark 1.9. Note that in principle one may define a stochastic process such that the average $\langle \zeta(\mathbf{x}) \rangle$ may depend on $\mathbf{x}$ as well as the higher moments of the random variable. We discard this possibility for simplicity.

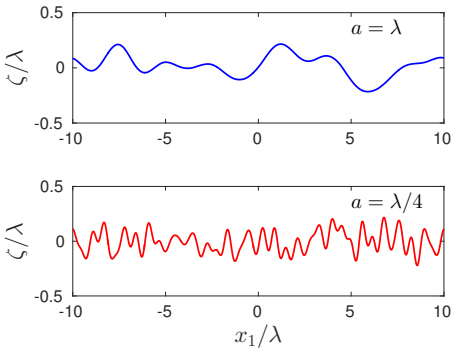
For a Gaussian height distribution, it is enough to give the two first moments to define completely the law of probability. We have now described how the probability measure of surface height



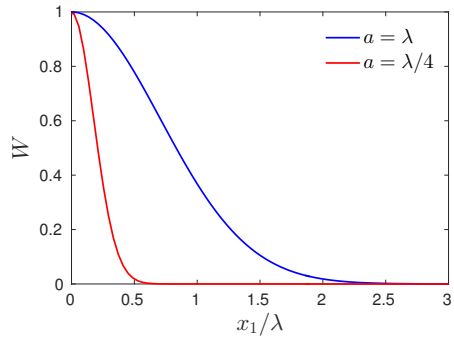
(a) Surface profile, $a = \lambda$



(b) Surface profile, $a = \lambda/4$



(c) Cuts along $x_2 =$ constant.



(d) Auto-correlation functions.

Figure 1.12: (a, b) Two dimensional portion of a computer generated realization of a random surface with Gaussian height probability density with zero mean and rms roughness $\sigma = \lambda/10$ with Gaussian auto-correlation function characterized by a correlation length equal to (a) $a = \lambda$ and (b) $a = \lambda/4$. (c) One dimensional cuts of the surfaces in (a) and (b) along a plan $x_2 =$ constant. (d) Graphs of the auto-correlation functions.

at a given point $\mathbf{x}$ may be defined. However, it remains an additional degree of freedom, which is the lateral correlation between two points on the surface. In terms of random variables, we must define the covariance between two random variables $\zeta(\mathbf{x})$ and $\zeta(\mathbf{x}')$ indexed by two points $\mathbf{x}$ and $\mathbf{x}'$. This defines the so-called auto-correlation function $W : (\mathbf{x}, \mathbf{x}') \in \mathbb{R}^2 \times \mathbb{R}^2 \mapsto \mathbb{R}$ as

$$\langle \zeta(\mathbf{x})\zeta(\mathbf{x}')\rangle = \sigma^2\, W(\mathbf{x}, \mathbf{x}')\,. \tag{1.116}$$

Note that with this definition, we have $W(\mathbf{x}, \mathbf{x}) = 1$. In practice, the auto-correlation may only depend on the difference of positions, i.e. that the auto-corralation can be of the form $W(\mathbf{x} - \mathbf{x}')$ which means that the absolute position on the surface does not matter, the surface looks homogeneous in terms of correlation, but it may still be anisotropic. If in addition the auto-correlation is anisotropic, the auto-correlation function only depends on the distance between

two positions, i.e. that the auto-correlation function can be of the form $W(|\mathbf{x}-\mathbf{x}'|)$. Typically, the auto-correlation will be a decaying function of the distance between two points. Physically, the heights at two points on a surface are expected to be more correlated if they are close than if they are distant. The characteristic decay length of the auto-correlation is called the *correlation length*. Examples of an isotropic Gaussian and an exponential auto-correlation functions are given in Eq. (1.117). The correlation length is denoted $a > 0$.

$$W_{\text{Gauss}}(\mathbf{x},\mathbf{x}') = \exp\left(-\frac{|\mathbf{x}-\mathbf{x}'|^2}{a^2}\right) , \tag{1.117a}$$

$$W_{\text{exp}}(\mathbf{x},\mathbf{x}') = \exp\left(-\frac{|\mathbf{x}-\mathbf{x}'|}{a}\right) . \tag{1.117b}$$

Moreover, a quantity of interest when the auto-correlation function only depends on $\mathbf{x}-\mathbf{x}'$ is the so-called *power spectrum* g, defined as the Fourier transform of the auto-correlation function

$$g(\mathbf{p}) = \int W(\mathbf{x})\ \exp\left(-i\,\mathbf{p}\cdot\mathbf{x}\right) \mathrm{d}^2x \tag{1.118}$$

For example the power spectrum for the Gaussian auto-correlation function defined in Eq. (1.117a) is given by

$$g(\mathbf{p}) = \pi a^2\ \exp\left(-\frac{|\mathbf{p}|^2a^2}{4}\right) . \tag{1.119}$$

The power spectrum is particularly useful numerically to generate realizations of a stochastic process with a specific auto-correlation function. A method known as the Fourier filtering method, and which we have used in the work presented in this thesis for generating randomly rough surfaces, is detailed for example in Ref. [36].

To sum up, the statistical properties of random surfaces can be determined essentially by two types of data: (i) the probability distribution of heights and (ii) the auto-correlation function. We illustrate in Fig. 1.12 examples of computer generated realizations of random surfaces for two stochastic processes characterized by a Gaussian height probability density with zero mean and rms roughness $\sigma = \lambda/10$ and different isotropic Gaussian auto-correlation functions characterized by a correlation length $a = \lambda$ and $a = \lambda/4$ whose graphs are plotted in Fig. 1.12(d). Note that we express all lengths normalized by the light wavelength, which is a rather natural way to set the physical scale of the problem. Although the two surfaces share the same height probability density and have heights amplitudes which would roughly varies by a few σ around the plane $x_3 = 0$, the surface for which the correlation length is shortest oscillates more rapidly than that for which the correlation length is largest, as can be seen in Figs. 1.12(a-c). It is then expected that the characteristic slopes of profiles with short correlation length will be larger than those of profiles with large correlation length. In terms of light scattering it seems natural to expect that surface profiles with short correlation length will scatter light more broadly than surface profiles with long correlation length. Such an intuitive behavior will be made more precise in later chapters.

1.6.2 Some vocabulary

Focusing on the reflection of light by a surface[7], a commonly accepted definition for *diffuse reflection* is the reflection of light from a surface such that an incident coherent ray is reflected

[7] to fix the idea but transmission can be treated similarly

at many angles rather than at just one angle as in the case of specular reflection. This definition sounds intuitively simple and legit. Note that this suggests though that the reflection from a grating (for which the lattice constant is larger than the wavelength) is a case of diffuse reflection. Do we really want to consider such a case as *"diffuse"*? Maybe, maybe not, that is a matter of opinion. Nevertheless, we feel that we might want to have different terms when talking about the scattering from a grating and that from a randomly rough surface. Indeed, in the first case light is scattered on a well defined discrete set of angles while, in the latter, light is scattered (on average) smoothly over *all* angles. So let us keep our simple (and maybe naive) definition of diffuse scattering and try to make sub-categories. What is the difference between a grating and a randomly rough surface from an optical point of view? In the first case, the diffraction peaks are the result of an interference phenomenon, only path scattered on specific directions will interfere constructively and the intensity diffracted on the remaining directions will vanish (for an infinite grating). So we have the feeling that optical coherence is a key point here. On the other side, on average, for a randomly rough surface, the phase of each path scattered in a given direction will be random and there will be no overall strong constructive or destructive interference. So all direction are somewhat equivalent and we observe this background of scattered light. It seems that this observation, interpreted as a loss of coherence, lead some authors to call it incoherent scattering as opposition to the coherent scattering of a grating. We will see that this appellation still is ambiguous. Indeed, what about the strong specular reflection we observe for weakly rough surface? What if we had a perturbed grating? We would expect still some peaks but with a background added to it. So both may happen for a given surface. The trick is to decompose the overall scattering in two pieces, a coherent and an incoherent component. If we look at the scattering from one single rough interface with a laser, we actually do not see a smooth intensity pattern, we see speckles. By changing surfaces, the speckle pattern will change, some bright spot may become dark and vice versa. However, the specular peak still remains bright. So this seems to indicate that for different surfaces, the field off-specular fluctuates between positive and negative values, so the average field may very well be zero, but the intensity being the absolute square of the field will oscillate between 0 and some positive value, so in average, the intensity will be strictly positive. For the specular direction, the same happens with the difference that the field itself rather oscillates around a non zero average, since we always see it bright whatever the surface realization. This gives our decomposition: *the coherent component of the average intensity is the intensity of the average field, and the incohorent component of the average intensity is the average intensity minus the intensity of the average field.*

1.6.3 Mean differential scattering coefficients

Based on the above discussion, and the definition of the differential scattering coefficients given in Section 1.2.4, we can decompose the mean differential scattering coefficients into a sum of a coherent and an incoherent component. Let us consider the following ensemble average for the differential reflection coefficient of the truncated system

$$\left\langle \frac{\partial R^{(S)}_{\alpha\beta}}{\partial \Omega_r}(\mathbf{p} \mid \mathbf{p}_0) \right\rangle = \frac{\epsilon_1 \left(\frac{\omega}{2\pi c}\right)^2 \cos^2\theta_r}{S \cos\theta_0} \left\langle |R^{(S)}_{\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0)|^2 \right\rangle . \tag{1.120}$$

By decomposing the reflection amplitudes as the sum of the mean and fluctuation (deviation from the mean)

$$R^{(S)}_{\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0) = \left\langle R^{(S)}_{\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0) \right\rangle + \left[R^{(S)}_{\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0) - \left\langle R^{(S)}_{\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0) \right\rangle \right] , \tag{1.121}$$

we can decompose the mean differential reflection coefficient as the sum of a *coherent* component and an *incoherent* component as

$$\left\langle \frac{\partial R_{\alpha\beta}^{(S)}}{\partial \Omega_r}(\mathbf{p} \mid \mathbf{p}_0) \right\rangle = \left\langle \frac{\partial R_{\alpha\beta}^{(S)}}{\partial \Omega_r}(\mathbf{p} \mid \mathbf{p}_0) \right\rangle_{\text{coh}} + \left\langle \frac{\partial R_{\alpha\beta}^{(S)}}{\partial \Omega_r}(\mathbf{p} \mid \mathbf{p}_0) \right\rangle_{\text{incoh}}, \tag{1.122}$$

where

$$\left\langle \frac{\partial R_{\alpha\beta}^{(S)}}{\partial \Omega_r}(\mathbf{p} \mid \mathbf{p}_0) \right\rangle_{\text{coh}} = \frac{\epsilon_1 \left(\frac{\omega}{2\pi c}\right)^2 \cos^2\theta_r}{S \cos\theta_0} \left| \left\langle R_{\alpha\beta}^{(S)}(\mathbf{p} \mid \mathbf{p}_0) \right\rangle \right|^2 \tag{1.123a}$$

$$\left\langle \frac{\partial R_{\alpha\beta}^{(S)}}{\partial \Omega_r}(\mathbf{p} \mid \mathbf{p}_0) \right\rangle_{\text{incoh}} = \frac{\epsilon_1 \left(\frac{\omega}{2\pi c}\right)^2 \cos^2\theta_r}{S \cos\theta_0} \left[\left\langle |R_{\alpha\beta}^{(S)}(\mathbf{p} \mid \mathbf{p}_0)|^2 \right\rangle - \left| \left\langle R_{\alpha\beta}^{(S)}(\mathbf{p} \mid \mathbf{p}_0) \right\rangle \right|^2 \right]. \tag{1.123b}$$

Similar expression are easily found for the differential transmission coefficients. The mean differential scattering coefficients will be our primary observables for the study of light scattering by randomly rough systems in Chapter 10. Note that in principle one should still take the limit of the above expressions as the surface S covers the entire plane. Numerically, however, it is clear that only finite surfaces can be studied. Equations (1.123a) and (1.123b) are then used as such for a finite S and it is understood that the surface area considered numerically is large enough so that Eqs. (1.123a) and (1.123b) yield a good approximation of the limit. When closed form approximation of the reflection and transmission amplitudes are known, and the differential scattering coefficient may be averaged analytically, the limit can also be taken analytically. These technical issues are detailed in Chapter 8 devoted to the simulation and approximations of the average optical response of randomly rough surfaces.

Remark 1.10. Are the words "*coherent*" and "*incoherent*" well chosen? If we come back to our speckle patterns that we used to construct the definition of the incoherent component of the intensity, we may ask the following question: Why are there speckles? Well, precisely because for one surface, the scattering is optically coherent and yields interference effects. So "incoherent scattering" is "coherent scattering" that washes out when you average over surfaces ... do you see the confusion coming? Even worse, for the scattering from a film, we will see in Chapter 10 that interference fringes, which is an *optically coherent* effect, can be observed in the *incoherent* component of the intensity. Our conclusion, is that the terms coherent and incoherent components of the intensity are sources of confusion, because they refer here to statistical properties and not optical properties. Nevertheless, we will use these terms to be consistent with the terminology found in the literature. Maybe, *statistically* coherent and incoherent components of the intensity may be better choice. At least, the reader is now warned.

1.7 Summary

Let us summarize the useful pieces of information discussed in the present chapter. First, we have recalled some basics of the electromagnetic theory of homogeneous media, and posed the problem of electromagnetic wave scattering by an interface between two media. We have introduced the Fourier representation of the fields and the notion of reflection and transmission amplitudes. Then typical scattering observables in the cases of periodic and randomly rough surfaces, which will be quantities of interest in later chapters, have been defined based on the scattering amplitudes. Some results, such as the scattering of a plane wave by a planar interface and the radiation of oscillating and rotating dipoles have been revisited and phenomenological

descriptions of Rayleigh and Wood anomalies have been presented as they will serve as basis bricks in further understanding of the phenomena observed in the scattering of more complex systems.

Now that the stage is set, we are ready to embark in the theory of the reduced Rayleigh equations, which aims at determining the unknown reflection and transmission amplitudes for the scattering problem.

Chapter 2

The reduced Rayleigh equations for a single rough surface

In this chapter, we present the derivation of what we will call the *transfer equations* and the reduced Rayleigh equations for harmonic electromagnetic waves satisfying the Helmholtz equation on both side of an arbitrary interface[1] between to homogeneous media. The transfer equations are integral relations between the amplitudes of the plane wave expansion of the fields in both media. The transfer equations are derived without specifying an incident field, and they hold for the total field. The reduced Rayleigh equations are decoupled inhomogeneous integral equations for the reflection and transmission amplitudes given an incident wave. The reflection and transmission amplitudes are related to the amplitudes of the plane wave expansion of the scattered field reflected from and transmitted through the interface.

The method based on the reduced Rayleigh equations can be considered as a generalization to penetrable media separated by an arbitrary interface of the Rayleigh method for periodic perfectly conducting grating suggested by Rayleigh in 1907 [15]. The reduced Rayleigh equations were first derived by Toigo *et al.* [37] and Brown *et al.* [38, 39, 40] and then used (among other) extensively by Maradudin and co-workers for the study of a light scattering from one- and two-dimensional perfectly conducting surfaces [41, 42], metallic [41, 43, 44] and dielectric surfaces [41, 43, 45, 46, 1], rough films supported by a metallic substrate [47] as well as for the study of the dispersion of surface plasmon polariton in polaritonic crystals [48] and the scattering of surface plasmon polaritons from surface defects [49, 50, 51]. The derivation of the reduced Rayleigh equations given in [39] is based on the extinction theorem while the derivation found in subsequent works is often closer to Rayleigh's original idea of expanding the field in plane wave and finding the unknown amplitudes by requiring the satisfaction of the boundary conditions. Such an approach was taken by Maradudin and co-worker. Here we present a derivation given by Soubret *et al.* in [52, 53] for the transfer equations. The derivation is similar to that of Maradudin and co-workers with the only difference that no specific form of the incident field is required. Hence the transfer equations allow one to derive the reduced Rayleigh equations in one line for any incident field. Moreover, the transfer equations are well suited to the generalization of the method to multilayer systems with arbitrary interfaces as we will see in Chapter 3.

The notations adopted here are somewhat inspired from that of Soubret and Maradudin and

[1] By "arbitrary" here we mean an interface which is planar in average whose profile can be described by a bounded regular function of the in-plane coordinate.

co-workers. The notation framework constructed during the derivation, especially the set of sub- and superscripts, will prove to be powerful in Chapter 3 to generalize the reduced Rayleigh equations to multi-layer systems in a few lines and will give us a composition rule for the computation of the appropriate multi-interface transfer kernels from single-interface kernels. It is therefore beneficial to spend some time to get used to these notations.

Even though the derivation is written in a very formal way (what one may call the physicist way), we will at time make a number of remarks that aim at emphasizing some limitations of the method, and pointing out some potential mathematical issues or difficulties. These are meant to warn the reader that the theory of the reduced Rayleigh equations have some open problems which, to the author's knowledge, are still unsolved. The reader need not worry at a first reading since some of these issues will be discussed and illustrated numerically in more details in Part II.

As a complement, the transfer equations and the reduced Rayleigh equations for scalar waves will also be presented briefly. The derivation of the transfer equations for scalar waves is simpler than that for electromagnetic waves but relies on the same steps. The derivation will then be kept short and the details left to the reader. The equations for scalar waves are of interest by themselves to treat problems of scattering of acoustic waves or of quantum matter waves. Including the results for scalar waves in this thesis is not merely due to our taste of completeness. We will see in Chapter 10 that it is beneficial to compare the equations for scalar waves and electromagnetic waves to get a deeper insight about the mechanisms at play for explaining phenomena observed for electromagnetic waves. Some of the phenomena will be identified as due to mechanisms that only require a scalar picture of the waves while other phenomena will require to take into account the polarization nature of electromagnetic waves.

2.1 Derivation of the transfer equations and the reduced Rayleigh equations

The system under consideration is composed of two media, of respective dielectric constant ϵ_1 and ϵ_2, separated by a rough interface. The Cartesian system of coordinates, defined by the orthonormal basis $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3)$, is chosen in such a way that the $(O, \hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2)$-plane is parallel to the average plane of the rough interface. The origin, O, is arbitrarily chosen. It is assumed that the surface profile, $\partial\Omega$, can be described by a bounded real valued differentiable function in that coordinated system, the equation of the surface being defined by $x_3 = \zeta(x_1, x_2) = \zeta(\mathbf{x})$, where in the following $\mathbf{x} = x_1\,\hat{\mathbf{e}}_1 + x_2\,\hat{\mathbf{e}}_2$. The media denoted 1 (resp. 2), occupies the upper (resp. lower) region defined by $\Omega_1 = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_3 > \zeta(x_1, x_2)\}$ (resp. $\Omega_2 = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_3 < \zeta(x_1, x_2)\}$).

Remark 2.1. The assumption that the surface can be represented by a function ζ with the aforementioned properties already gives a set of restrictions for the range of application of the reduced Rayleigh equations. The surface must have an average plane, and cannot have overhangs. Concerning the regularity of ζ, it will be assumed in the following that ζ is differentiable and bounded. Nevertheless, we will in the applications make use of the reduced Rayleigh equations in cases where the surface is only differentiable almost everywhere. Our belief, based on comparison between simulations and experimental data, is that such a practice can yield a good approximation of the solution of the scattering problem if the amplitude of the surface is not too large. This issue will be explored in more details in Part II.

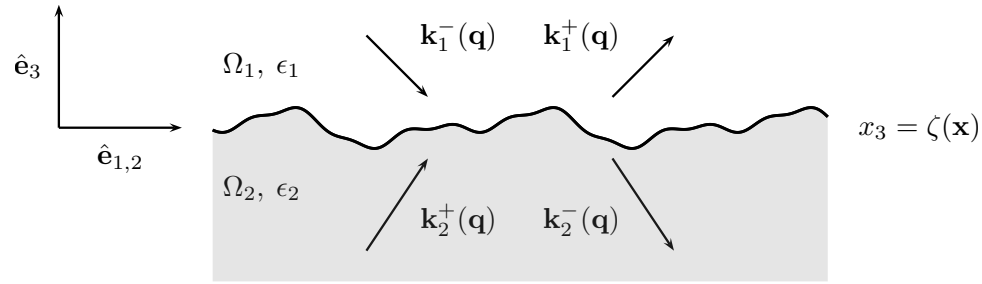


Figure 2.1: Scattering system composed of two half space separated by an arbitrary interface. The arrows are examples of wave vectors sharing the same in-plane projection $\mathbf{q}$.

We consider the presence of an electromagnetic field $(\mathbf{E}, \mathbf{H})$ in the whole space. The restriction of the fields in the region Ω_j, $j \in \{1,2\}$, will be denoted by a subscript j. For instance, the electric field evaluated at a point $\mathbf{r} \in \Omega_1$ at time t is denoted $\mathbf{E}_1(\mathbf{r}, t)$. The Maxwell equations, together with linear constitutive relations (in the frequency domain), result in the fact that the electric field satisfies the Helmholtz equation in each region. Namely for $j \in \{1,2\}$, we have

$$\Delta \mathbf{E}_j + \epsilon_j \, \frac{\omega^2}{c^2} \, \mathbf{E}_j = \mathbf{0} \, . \tag{2.1}$$

In the following, we drop the time, or frequency dependence, to lighten notation. In order to make the present chapter self-sufficient, we repeat here the definitions introduced in Chapter 1 concerning the expansion in plane waves of the fields. We assume that a solution of the Helmholtz equation can be written as a linear combination of plane waves, Eq. (1.31), i.e. for $j \in \{1,2\}$ and for all $\mathbf{r} \in \Omega_j$,

$$\mathbf{E}_j(\mathbf{r}) = \sum_{a=\pm} \int_{\mathbb{R}^2} \mathbf{E}_j^a(\mathbf{q}) \, \exp\left(i \, \mathbf{k}_j^a(\mathbf{q}) \cdot \mathbf{r}\right) \, \frac{\mathrm{d}^2 q}{(2\pi)^2} \, , \tag{2.2}$$

where we define

$$\mathbf{E}_j^a(\mathbf{q}) \stackrel{\text{def}}{=} \mathcal{E}_{j,p}^a(\mathbf{q}, \omega) \, \hat{\mathbf{e}}_{p,j}^a(\mathbf{q}) + \mathcal{E}_{j,s}^a(\mathbf{q}, \omega) \, \hat{\mathbf{e}}_s(\mathbf{q}) \, , \tag{2.3}$$

and we recall that we have the following definitions for the wave vectors and polarization vectors from Eq. (1.32)

$$\mathbf{k}_j^\pm(\mathbf{q}) \stackrel{\text{def}}{=} \mathbf{q} \pm \alpha_j(\mathbf{q}) \, \hat{\mathbf{e}}_3 \, , \tag{2.4a}$$

$$\alpha_j(\mathbf{q}) \stackrel{\text{def}}{=} \sqrt{\epsilon_j \left(\frac{\omega}{c}\right)^2 - \mathbf{q}^2}, \qquad \mathrm{Re}\,(\alpha_j), \mathrm{Im}\,(\alpha_j) \geq 0 \, , \tag{2.4b}$$

$$\hat{\mathbf{e}}_s(\mathbf{q}) \stackrel{\text{def}}{=} \frac{\hat{\mathbf{e}}_3 \times \mathbf{k}_j^\pm(\mathbf{q})}{|\hat{\mathbf{e}}_3 \times \mathbf{k}_j^\pm(\mathbf{q})|} = \hat{\mathbf{e}}_3 \times \hat{\mathbf{q}} \, , \tag{2.4c}$$

$$\hat{\mathbf{e}}_{p,j}^\pm(\mathbf{q}) \stackrel{\text{def}}{=} \frac{\hat{\mathbf{e}}_s \times \mathbf{k}_j^\pm(\mathbf{q})}{|\hat{\mathbf{e}}_s \times \mathbf{k}_j^\pm(\mathbf{q})|} = \frac{c}{\sqrt{\epsilon_j}\omega} \left(\pm \alpha_j(\mathbf{q}) \, \hat{\mathbf{q}} - |\mathbf{q}| \, \hat{\mathbf{e}}_3\right) \, . \tag{2.4d}$$

Then the wave vector $\mathbf{k}_j^\pm(\mathbf{q})$ of an elementary plane wave, satisfying the dispersion relation in medium j, is decomposed in the component $\mathbf{q}$ in the vector plane $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2)$ and the component $\pm\alpha_j(\mathbf{q})$ along $\hat{\mathbf{e}}_3$. In the following we will lighten notation by writing $\int \leftrightarrow \int_{\mathbb{R}^2}$ if not precised otherwise.

Remark 2.2. The plane wave expansion, Eq. (2.2), contains both progressive, evanescent and possibly exponentially growing waves. Later, the field will be split into an incident field which typically will be a plane wave propagating from infinity towards the interface and a scattered field which will be imposed to contain only outgoing elementary plane waves (i.e. progressive or evanescent away from the interface). In other words, the *scattered* field in medium 1 will contain only modes with wave vector of the form $\mathbf{k}_1^+(\mathbf{q})$ and the *scattered* field in medium 2 will contain only modes with wave vector of the form $\mathbf{k}_2^-(\mathbf{q})$, i.e. that $\mathbf{E}_1^-$ and $\mathbf{E}_2^+$ will be set to zero identically for the scattered field, which means that only the incident field will contribute to these modes. Therefore, the scattered field may be written

$$\mathbf{E}_1^{(\mathrm{scat})}(\mathbf{r}) = \int \mathbf{E}_1^+(\mathbf{q}) \exp\left(i\, \mathbf{k}_1^+(\mathbf{q}) \cdot \mathbf{r}\right) \frac{\mathrm{d}^2 q}{(2\pi)^2} \tag{2.5}$$

$$\mathbf{E}_2^{(\mathrm{scat})}(\mathbf{r}) = \int \mathbf{E}_2^-(\mathbf{q}) \exp\left(i\, \mathbf{k}_2^-(\mathbf{q}) \cdot \mathbf{r}\right) \frac{\mathrm{d}^2 q}{(2\pi)^2} \,. \tag{2.6}$$

It is commonly admitted that the plane wave expansion is valid away from the interface region, say for $x_3 > \max\zeta$ and $x_3 < \min\zeta$. Whether the expansion is valid also within the grooves depends in general on the surface profile. Indeed, considering a point above the surface but below the plane $x_3 = 0$, evanescent modes of the form $\mathbf{E}_1^+(\mathbf{q}) \exp\left(i\, \mathbf{k}_1^+(\mathbf{q}) \cdot \mathbf{r}\right) = \mathbf{E}_1^+(\mathbf{q}) \exp\left(i\, \mathbf{q} \cdot \mathbf{x}\right) \exp\left(-\beta_1(\mathbf{q}) x_3\right)$, with $\alpha_1(\mathbf{q}) = i\beta_1(\mathbf{q})$ with $\beta_1 > 0$. Since for such a point $x_3 < 0$, the x_3-dependent exponential factor grows with increasing modulus of the in-plane wave vector, $|\,\mathbf{q}\,|$. Hence only fields for which the corresponding amplitude $|\,\mathbf{E}_1^+(\mathbf{q})|$ decays fast enough with $|\,\mathbf{q}\,| \to \infty$ may be represented in the whole space above the surface (and similarly below the interface). As a rule of thumb we may expect that

$$\mathbf{E}_1^+(\mathbf{q}) = \mathcal{O}\Big[\exp\left(-\beta_1(\mathbf{q}) \min\zeta\right)\Big] \tag{2.7a}$$

$$\mathbf{E}_2^-(\mathbf{q}) = \mathcal{O}\Big[\exp\left(\beta_2(\mathbf{q}) \max\zeta\right)\Big] \tag{2.7b}$$

for the field to be representable by Eq. (2.2). Naturally, the field amplitudes $\mathbf{E}_1^+(\mathbf{q})$ and $\mathbf{E}_2^-(\mathbf{q})$ depend on the surface. In the following, we will make the assumption that such a plane wave expansion is valid also within the grooves: this is known as the *Rayleigh hypothesis* or Rayleigh assumption. The validity of the Rayleigh hypothesis has been the topic of numerous papers in the literature and still is a somewhat debated topic. A particularly relevant proof for the range of validity of the Rayleigh hypothesis in the case of diffraction by a one-dimensional perfectly conducting sinusoidal surface was given by Millar in [54, 55]. This issue will be discussed in details in Chapter 7.

The boundary conditions for Maxwell's equation across the interface read

$$\mathbf{n}(\mathbf{x}) \times \Big[\mathbf{E}_1(\mathbf{s}(\mathbf{x})) - \mathbf{E}_2(\mathbf{s}(\mathbf{x}))\Big] = \mathbf{0} \tag{2.8a}$$

$$\mathbf{n}(\mathbf{x}) \cdot \Big[\epsilon_1\, \mathbf{E}_1(\mathbf{s}(\mathbf{x})) - \epsilon_2\, \mathbf{E}_2(\mathbf{s}(\mathbf{x}))\Big] = 0 \tag{2.8b}$$

$$\mathbf{n}(\mathbf{x}) \times \Big[\mathbf{H}_1(\mathbf{s}(\mathbf{x})) - \mathbf{H}_2(\mathbf{s}(\mathbf{x}))\Big] = \mathbf{0} \tag{2.8c}$$

where $\mathbf{n}(\mathbf{x})$ is a vector normal to the surface at point $\mathbf{s}(\mathbf{x}) = \mathbf{x} + \zeta(\mathbf{x})\, \hat{\mathbf{e}}_3$ given by

$$\mathbf{n}(\mathbf{x}) = \hat{\mathbf{e}}_3 - \partial_1\zeta(\mathbf{x})\, \hat{\mathbf{e}}_1 - \partial_2\zeta(\mathbf{x})\, \hat{\mathbf{e}}_2 \,. \tag{2.9}$$

Here, ∂_k denotes the partial derivative with respect to the direction $\hat{\mathbf{e}}_k$. By inserting the plane wave expansion, Eq. (2.2), into the boundary conditions, one gets

$$\sum_{a=\pm}\int \mathbf{n}(\mathbf{x})\times\mathbf{E}_1^a(\mathbf{q})\, e^{i\,\mathbf{k}_1^a(\mathbf{q})\cdot\mathbf{s}(\mathbf{x})}\,\frac{\mathrm{d}^2q}{(2\pi)^2} = \sum_{a=\pm}\int \mathbf{n}(\mathbf{x})\times\mathbf{E}_2^a(\mathbf{q})\, e^{i\,\mathbf{k}_2^a(\mathbf{q})\cdot\mathbf{s}(\mathbf{x})}\,\frac{\mathrm{d}^2q}{(2\pi)^2} \tag{2.10a}$$

$$\frac{\epsilon_1}{\epsilon_2}\sum_{a=\pm}\int \mathbf{n}(\mathbf{x})\cdot\mathbf{E}_1^a(\mathbf{q})\, e^{i\,\mathbf{k}_1^a(\mathbf{q})\cdot\mathbf{s}(\mathbf{x})}\,\frac{\mathrm{d}^2q}{(2\pi)^2} = \sum_{a=\pm}\int \mathbf{n}(\mathbf{x})\cdot\mathbf{E}_2^a(\mathbf{q})\, e^{i\,\mathbf{k}_2^a(\mathbf{q})\cdot\mathbf{s}(\mathbf{x})}\,\frac{\mathrm{d}^2q}{(2\pi)^2} \tag{2.10b}$$

$$\sum_{a=\pm}\int \mathbf{n}(\mathbf{x})\times[\mathbf{k}_1^a(\mathbf{q})\times\mathbf{E}_1^a(\mathbf{q})]\, e^{i\,\mathbf{k}_1^a(\mathbf{q})\cdot\mathbf{s}(\mathbf{x})}\,\frac{\mathrm{d}^2q}{(2\pi)^2} = \sum_{a=\pm}\int \mathbf{n}(\mathbf{x})\times[\mathbf{k}_2^a(\mathbf{q})\times\mathbf{E}_2^a(\mathbf{q})]\, e^{i\,\mathbf{k}_2^a(\mathbf{q})\cdot\mathbf{s}(\mathbf{x})}\,\frac{\mathrm{d}^2q}{(2\pi)^2} \tag{2.10c}$$

$$\sum_{a=\pm}\int \mathbf{k}_1^a(\mathbf{q})\cdot\mathbf{E}_1^a(\mathbf{q})\, e^{i\,\mathbf{k}_1^a(\mathbf{q})\cdot\mathbf{s}(\mathbf{x})}\,\frac{\mathrm{d}^2q}{(2\pi)^2} = \sum_{a=\pm}\int \mathbf{k}_2^a(\mathbf{q})\cdot\mathbf{E}_2^a(\mathbf{q})\, e^{i\,\mathbf{k}_2^a(\mathbf{q})\cdot\mathbf{s}(\mathbf{x})}\,\frac{\mathrm{d}^2q}{(2\pi)^2} \tag{2.10d}$$

Notice that the last equation does not come from the boundary conditions but is the trivial equation $\nabla\cdot\mathbf{E}_1 = \nabla\cdot\mathbf{E}_2 = 0$ that proves to be useful to rewrite some equations in a more compact way as we will see later. In order to eliminate $\mathbf{E}_2^{\pm}$, consider the following linear combination:

$$\int e^{-i\,\mathbf{k}_2^b(\mathbf{p})\cdot\mathbf{s}(\mathbf{x})}\left(\mathbf{k}_2^b(\mathbf{p})\times[\text{Eq. (2.10a)}] + [\text{Eq. (2.10c)}] - [\text{Eq. (2.10b)}]\,\mathbf{k}_2^b(\mathbf{p}) - [\text{Eq. (2.10d)}]\,\mathbf{n}(\mathbf{x})\right)\mathrm{d}^2x \tag{2.11}$$

where $\mathbf{p} = p_1\,\hat{\mathbf{e}}_1 + p_2\,\hat{\mathbf{e}}_2$ is an arbitrary vector of the vector plane $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2)$, and $b = \pm 1$ may be arbitrarily chose. We recall that by definition we have $\mathbf{k}_2^b(\mathbf{p}) = \mathbf{p} + b\,\alpha_2(\mathbf{p})\,\hat{\mathbf{e}}_3$. The integrand on the right hand side of the considered linear combination, Eq. (2.11), is proportional to

$$\begin{aligned}
&\underbrace{\mathbf{k}_2^b(\mathbf{p})\times[\mathbf{n}(\mathbf{x})\times\mathbf{E}_2^a(\mathbf{q})]}_{\mathbf{k}_2^b(\mathbf{p})\cdot\mathbf{E}_2^a(\mathbf{q})\;\mathbf{n}(\mathbf{x}) - \mathbf{k}_2^b(\mathbf{p})\cdot\mathbf{n}(\mathbf{x})\;\mathbf{E}_2^a(\mathbf{q})} + \underbrace{\mathbf{n}(\mathbf{x})\times[\mathbf{k}_2^a(\mathbf{q})\times\mathbf{E}_2^a(\mathbf{q})]}_{\mathbf{n}(\mathbf{x})\cdot\mathbf{E}_2^a(\mathbf{q})\;\mathbf{k}_2^a(\mathbf{q}) - \mathbf{n}(\mathbf{x})\cdot\mathbf{k}_2^a(\mathbf{q})\;\mathbf{E}_2^a(\mathbf{q})}\\
&-\mathbf{n}(\mathbf{x})\cdot\mathbf{E}_2^a(\mathbf{q})\;\mathbf{k}_2^b(\mathbf{p}) - \mathbf{k}_2^a(\mathbf{q})\cdot\mathbf{E}_2^a(\mathbf{q})\;\mathbf{n}(\mathbf{x})\\
&= -(\mathbf{k}_2^b(\mathbf{p}) + \mathbf{k}_2^a(\mathbf{q}))\cdot\mathbf{n}(\mathbf{x})\,\mathbf{E}_2^a(\mathbf{q}) + \underbrace{(\mathbf{k}_2^b(\mathbf{p}) - \mathbf{k}_2^a(\mathbf{q}))\cdot\mathbf{E}_2^a(\mathbf{q})\,\mathbf{n}(\mathbf{x}) - \mathbf{n}(\mathbf{x})\cdot\mathbf{E}_2^a(\mathbf{q})(\mathbf{k}_2^b(\mathbf{p}) - \mathbf{k}_2^a(\mathbf{q}))}_{\mathbf{E}_2^a(\mathbf{q})\times(\mathbf{n}(\mathbf{x})\times(\mathbf{k}_2^b(\mathbf{p}) - \mathbf{k}_2^a(\mathbf{q})))}\\
&= -(\mathbf{k}_2^b(\mathbf{p}) + \mathbf{k}_2^a(\mathbf{q}))\cdot\mathbf{n}(\mathbf{x})\,\mathbf{E}_2^a(\mathbf{q}) + \mathbf{E}_2^a(\mathbf{q})\times\left(\mathbf{n}(\mathbf{x})\times(\mathbf{k}_2^b(\mathbf{p}) - \mathbf{k}_2^a(\mathbf{q}))\right)
\end{aligned}$$

Where we have used the vector identity $\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = (\mathbf{a}\cdot\mathbf{c})\mathbf{b} - (\mathbf{a}\cdot\mathbf{b})\mathbf{c}$. This is equivalent to Eq. (26) in Ref. [53] but in a slightly more compact form though. The right hand side of the linear combination Eq. (2.11) then reads

$$\begin{aligned}
\mathrm{RHS} = \sum_{a=\pm}\iint &\left[-(\mathbf{k}_2^b(\mathbf{p}) + \mathbf{k}_2^a(\mathbf{q}))\cdot\mathbf{n}(\mathbf{x})\,\mathbf{E}_2^a(\mathbf{q}) + \mathbf{E}_2^a(\mathbf{q})\times\left(\mathbf{n}(\mathbf{x})\times(\mathbf{k}_2^b(\mathbf{p}) - \mathbf{k}_2^a(\mathbf{q}))\right)\right]\\
&\times\exp\left[-i\left(\mathbf{k}_2^b(\mathbf{p}) - \mathbf{k}_2^a(\mathbf{q})\right)\cdot\mathbf{s}(\mathbf{x})\right]\,\frac{\mathrm{d}^2x\,\mathrm{d}^2q}{(2\pi)^2}
\end{aligned} \tag{2.12}$$

We will have to evaluate two terms, by writing $\mathrm{RHS} = \mathrm{RHS}_1 + \mathrm{RHS}_2$ with

$$\mathrm{RHS}_1 = -\sum_{a=\pm}\iint e^{-i(\mathbf{k}_2^b(\mathbf{p}) - \mathbf{k}_2^a(\mathbf{q}))\cdot\mathbf{s}(\mathbf{x})}(\mathbf{k}_2^b(\mathbf{p}) + \mathbf{k}_2^a(\mathbf{q}))\cdot\mathbf{n}(\mathbf{x})\,\mathbf{E}_2^a(\mathbf{q})\,\mathrm{d}^2x\,\frac{\mathrm{d}^2q}{(2\pi)^2}\,, \tag{2.13a}$$

$$\mathrm{RHS}_2 = \sum_{a=\pm}\iint e^{-i(\mathbf{k}_2^b(\mathbf{p}) - \mathbf{k}_2^a(\mathbf{q}))\cdot\mathbf{s}(\mathbf{x})}\;\mathbf{E}_2^a(\mathbf{q})\times\left[\mathbf{n}(\mathbf{x})\times\left(\mathbf{k}_2^b(\mathbf{p}) - \mathbf{k}_2^a(\mathbf{q})\right)\right]\mathrm{d}^2x\,\frac{\mathrm{d}^2q}{(2\pi)^2}\,. \tag{2.13b}$$

Similarly, the left hand side of Eq. (2.11) can be expressed as

$$\mathrm{LHS} = \sum_{a=\pm} \iint \Big[-(\mathbf{k}_2^b(\mathbf{p}) + \mathbf{k}_1^a(\mathbf{q})) \cdot \mathbf{n}(\mathbf{x})\, \mathbf{E}_1^a(\mathbf{q}) + (\mathbf{k}_2^b(\mathbf{p}) - \mathbf{k}_1^a(\mathbf{q})) \cdot \mathbf{E}_1^a(\mathbf{q})\, \mathbf{n}(\mathbf{x}) - \mathbf{n}(\mathbf{x}) \cdot \mathbf{E}_1^a(\mathbf{q}) \left(\frac{\epsilon_1}{\epsilon_2}\, \mathbf{k}_2^b(\mathbf{p}) - \mathbf{k}_1^a(\mathbf{q}) \right) \Big] e^{-i(\mathbf{k}_2^b(\mathbf{p}) - \mathbf{k}_1^a(\mathbf{q})) \cdot \mathbf{s}(\mathbf{x})}\, \mathrm{d}^2x\, \frac{\mathrm{d}^2q}{(2\pi)^2} \,. \tag{2.14}$$

In Eq. (2.14), we cannot directly use the identity $\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c}$ to group the two last terms of the integrand because of the $\frac{\epsilon_1}{\epsilon_2}$ factor. We can group part of it though, by writing $\frac{\epsilon_1}{\epsilon_2} = \frac{\epsilon_1 - \epsilon_2}{\epsilon_2} + 1$. Hence

$$\mathrm{LHS} = \sum_{a=\pm} \iint \Big[-\big(\mathbf{k}_2^b(\mathbf{p}) + \mathbf{k}_1^a(\mathbf{q})\big) \cdot \mathbf{n}(\mathbf{x})\, \mathbf{E}_1^a(\mathbf{q}) + \mathbf{E}_1^a(\mathbf{q}) \times \big(\mathbf{n}(\mathbf{x}) \times (\mathbf{k}_2^b(\mathbf{p}) - \mathbf{k}_1^a(\mathbf{q}))\big) - \frac{\epsilon_1 - \epsilon_2}{\epsilon_2}\, \mathbf{n}(\mathbf{x}) \cdot \mathbf{E}_1^a(\mathbf{q})\, \mathbf{k}_2^b(\mathbf{p}) \Big] e^{-i(\mathbf{k}_2^b(\mathbf{p}) - \mathbf{k}_1^a(\mathbf{q})) \cdot \mathbf{s}(\mathbf{x})}\, \mathrm{d}^2x\, \frac{\mathrm{d}^2q}{(2\pi)^2} \,. \tag{2.15}$$

We will have to evaluate three terms, by writing $\mathrm{LHS} = \mathrm{LHS}_1 + \mathrm{LHS}_2 + \mathrm{LHS}_3$ with

$$\mathrm{LHS}_1 = -\sum_{a=\pm} \iint e^{-i(\mathbf{k}_2^b(\mathbf{p}) - \mathbf{k}_1^a(\mathbf{q})) \cdot \mathbf{s}(\mathbf{x})} \big(\mathbf{k}_2^b(\mathbf{p}) + \mathbf{k}_1^a(\mathbf{q})\big) \cdot \mathbf{n}(\mathbf{x})\, \mathbf{E}_1^a(\mathbf{q})\, \mathrm{d}^2x\, \frac{\mathrm{d}^2q}{(2\pi)^2} \,, \tag{2.16a}$$

$$\mathrm{LHS}_2 = \sum_{a=\pm} \iint e^{-i(\mathbf{k}_2^b(\mathbf{p}) - \mathbf{k}_1^a(\mathbf{q})) \cdot \mathbf{s}(\mathbf{x})}\, \mathbf{E}_1^a(\mathbf{q}) \times \big(\mathbf{n}(\mathbf{x}) \times \big(\mathbf{k}_2^b(\mathbf{p}) - \mathbf{k}_1^a(\mathbf{q})\big)\big)\, \mathrm{d}^2x\, \frac{\mathrm{d}^2q}{(2\pi)^2} \,, \tag{2.16b}$$

$$\mathrm{LHS}_3 = -\frac{\epsilon_1 - \epsilon_2}{\epsilon_2} \sum_{a=\pm} \iint e^{-i(\mathbf{k}_2^b(\mathbf{p}) - \mathbf{k}_1^a(\mathbf{q})) \cdot \mathbf{s}(\mathbf{x})}\, \mathbf{n}(\mathbf{x}) \cdot \mathbf{E}_1^a(\mathbf{q})\, \mathbf{k}_2^b(\mathbf{p})\, \mathrm{d}^2x\, \frac{\mathrm{d}^2q}{(2\pi)^2} \,. \tag{2.16c}$$

We carry out first the integration with respect to $\mathbf{x}$ to recast Eqs. (2.13) and (2.16) as

$$\mathrm{RHS}_1 = -\sum_{a=\pm} \int (\mathbf{k}_2^b(\mathbf{p}) + \mathbf{k}_2^a(\mathbf{q})) \cdot\, \mathbf{w}_{2,2}^{b,a}(\mathbf{p} \,|\, \mathbf{q})\, \mathbf{E}_2^a(\mathbf{q})\, \frac{\mathrm{d}^2q}{(2\pi)^2} \,, \tag{2.17a}$$

$$\mathrm{RHS}_2 = \sum_{a=\pm} \int \mathbf{E}_2^a(\mathbf{q}) \times \Big(\mathbf{w}_{2,2}^{b,a}(\mathbf{p} \,|\, \mathbf{q}) \times (\mathbf{k}_2^b(\mathbf{p}) - \mathbf{k}_2^a(\mathbf{q}))\Big)\, \frac{\mathrm{d}^2q}{(2\pi)^2} \,, \tag{2.17b}$$

$$\mathrm{LHS}_1 = -\sum_{a=\pm} \int (\mathbf{k}_2^b(\mathbf{p}) + \mathbf{k}_1^a(\mathbf{q})) \cdot \mathbf{w}_{2,1}^{b,a}(\mathbf{p} \,|\, \mathbf{q})\, \mathbf{E}_1^a(\mathbf{q})\, \frac{\mathrm{d}^2q}{(2\pi)^2} \,, \tag{2.17c}$$

$$\mathrm{LHS}_2 = \sum_{a=\pm} \int \mathbf{E}_1^a(\mathbf{q}) \times \Big(\mathbf{w}_{2,1}^{b,a}(\mathbf{p} \,|\, \mathbf{q}) \times (\mathbf{k}_2^b(\mathbf{p}) - \mathbf{k}_1^a(\mathbf{q}))\Big)\, \frac{\mathrm{d}^2q}{(2\pi)^2} \,, \tag{2.17d}$$

$$\mathrm{LHS}_3 = \frac{\epsilon_2 - \epsilon_1}{\epsilon_2} \sum_{a=\pm} \int \mathbf{w}_{2,1}^{b,a}(\mathbf{p} \,|\, \mathbf{q}) \cdot \mathbf{E}_1^a(\mathbf{q})\, \mathbf{k}_2^b(\mathbf{p})\, \frac{\mathrm{d}^2q}{(2\pi)^2} \,, \tag{2.17e}$$

where we have defined, for $l, m \in \{1, 2\}$, $b, a \in \{\pm 1\}$ and $\mathbf{p}, \mathbf{q}$ in the vector plane $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2)$

$$\mathbf{w}_{l,m}^{b,a}(\mathbf{p} \,|\, \mathbf{q}) = \int \exp\big[-i(\mathbf{k}_l^b(\mathbf{p}) - \mathbf{k}_m^a(\mathbf{q})) \cdot \mathbf{s}(\mathbf{x})\big]\, \mathbf{n}(\mathbf{x})\, \mathrm{d}^2x \,. \tag{2.18}$$

Remark 2.3. The notation $(\cdot|\cdot)$ for the arguments of functions that depend on two wave-vectors is used in the physics literature on the reduced Rayleigh equations while *a priori* the notation $(\cdot, \cdot)$ may be better suited. It must not be confused with a scalar product. One may like to think of "$\mathbf{p}$ *given* $\mathbf{q}$" although a probabilistic interpretation would be premature at this stage. We just use here the notation without giving to much importance to its interpretation. We just mean a function of $\mathbf{p}$ and $\mathbf{q}$.

Remark 2.4. Under the assumption that the Rayleigh hypothesis is valid for the considered surface profile, it is not *a priori* obvious to know whether the point-wise boundary conditions Eq. (2.10) and Eq. (2.11) are equivalent. The point-wise boundary conditions implies the integrated boundary conditions Eq. (2.11): if the boundary conditions hold for every point on the surface, then Eq. (2.11) holds for *every plane wave test function* of the form $\phi(\mathbf{x}) = e^{-i\,\mathbf{k}_2^b(\mathbf{p})\cdot\mathbf{s}(\mathbf{x})}$. The reciprocal statement is *a priori* not necessarily true. Hence solving the scattering problem with Eq. (2.11) as a boundary condition may not be equivalent to the initial problem. We are not sure whether the reciprocal statement is true and how to prove it.

2.1.1 Analysis of $\mathbf{w}_{l,m}^{b,a}$

Now we need to evaluate $\mathbf{w}_{l,m}^{b,a}(\mathbf{p}\,|\,\mathbf{q})$, or to be more accurate, we need to know how to integrate against it. We have $\mathbf{n}(\mathbf{x}) = \hat{\mathbf{e}}_3 - \nabla\zeta(\mathbf{x})$, hence

$$\begin{aligned}
\mathbf{w}_{l,m}^{b,a}(\mathbf{p}\,|\,\mathbf{q}) &= \int e^{-i(\mathbf{k}_l^b(\mathbf{p})-\mathbf{k}_m^a(\mathbf{q}))\cdot\mathbf{s}(\mathbf{x})}\left(\hat{\mathbf{e}}_3 - \nabla\zeta(\mathbf{x})\right)\mathrm{d}^2x \\
&= \int e^{-i(\mathbf{k}_l^b(\mathbf{p})-\mathbf{k}_m^a(\mathbf{q}))\cdot\mathbf{s}(\mathbf{x})}\,\hat{\mathbf{e}}_3\,\mathrm{d}^2x - \int e^{-i(\mathbf{k}_l^b(\mathbf{p})-\mathbf{k}_m^a(\mathbf{q}))\cdot\mathbf{s}(\mathbf{x})}\,\nabla\zeta(\mathbf{x})\,\mathrm{d}^2x \\
&= \int e^{-i(\mathbf{k}_l^b(\mathbf{p})-\mathbf{k}_m^a(\mathbf{q}))\cdot\mathbf{s}(\mathbf{x})}\,\hat{\mathbf{e}}_3\,\mathrm{d}^2x \\
&\quad - \int \nabla\zeta(\mathbf{x})\; e^{-i[b\alpha_l(\mathbf{p})-a\alpha_m(\mathbf{q})]\cdot\zeta(\mathbf{x})}\,e^{-i(\mathbf{p}-\mathbf{q})\cdot\mathbf{x}}\,\mathrm{d}^2x\ . \qquad (2.19)
\end{aligned}$$

We can make an integration by parts under the condition $b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q}) \neq 0$. In that case, the second term Eq. (2.19) becomes[2]

- If $b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q}) \neq 0$ then

$$\begin{aligned}
\mathbf{w}_{l,m}^{b,a}(\mathbf{p}\,|\,\mathbf{q}) &= \int e^{-i(\mathbf{k}_l^b(\mathbf{p})-\mathbf{k}_m^a(\mathbf{q}))\cdot\mathbf{s}(\mathbf{x})}\,\hat{\mathbf{e}}_3\,\mathrm{d}^2x + \int \frac{\mathbf{p}-\mathbf{q}}{b\alpha_l(\mathbf{p})-a\alpha_m(\mathbf{q})}\,e^{-i(\mathbf{k}_l^b(\mathbf{p})-\mathbf{k}_m^a(\mathbf{q}))\cdot\mathbf{s}(\mathbf{x})}\,\mathrm{d}^2x \\
&= \int e^{-i(\mathbf{k}_l^b(\mathbf{p})-\mathbf{k}_m^a(\mathbf{q}))\cdot\mathbf{s}(\mathbf{x})}\,\frac{\mathbf{k}_l^b(\mathbf{p})-\mathbf{k}_m^a(\mathbf{q})}{b\alpha_l(\mathbf{p})-a\alpha_m(\mathbf{q})}\,\mathrm{d}^2x \\
&= \mathcal{J}_{l,m}^{b,a}(\mathbf{p}\,|\,\mathbf{q})\;\left(\mathbf{k}_l^b(\mathbf{p})-\mathbf{k}_m^a(\mathbf{q})\right)\ , \qquad (2.20)
\end{aligned}$$

  where we define

$$\mathcal{J}_{l,m}^{b,a}(\mathbf{p}\,|\,\mathbf{q}) \overset{\text{def}}{=} (b\alpha_l(\mathbf{p})-a\alpha_m(\mathbf{q}))^{-1}\int \exp\left[-i(\mathbf{k}_l^b(\mathbf{p})-\mathbf{k}_m^a(\mathbf{q}))\cdot\mathbf{s}(\mathbf{x})\right]\mathrm{d}^2x\ . \qquad (2.21)$$

- If $b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q}) = 0$ then $\mathbf{w}_{l,m}^{b,a}(\mathbf{p}\,|\,\mathbf{q})$ coincides with the Fourier transform of $\mathbf{n}$ and Eq. (2.19) reads

$$\begin{aligned}
\mathbf{w}_{l,m}^{b,a}(\mathbf{p}\,|\,\mathbf{q}) &= \int e^{-i(\mathbf{p}-\mathbf{q})\cdot\mathbf{x}}\,\hat{\mathbf{e}}_3\,\mathrm{d}^2x - \int \nabla\zeta(\mathbf{x})\;e^{-i(\mathbf{p}-\mathbf{q})\cdot\mathbf{x}}\,\mathrm{d}^2x \\
&= (2\pi)^2\delta(\mathbf{p}-\mathbf{q})\,\hat{\mathbf{e}}_3 - i\hat{\zeta}(\mathbf{p}-\mathbf{q})\,(\mathbf{p}-\mathbf{q})\,. \qquad (2.22)
\end{aligned}$$

[2] We use formally the integration by part in the sense of distribution although we are not entirely sure what mathematically rigorous sense to give the "integral" $\mathbf{w}_{l,m}^{b,a}(\mathbf{p}\,|\,\mathbf{q})$.

Here, we see that the Dirac δ appears naturally as the Fourier transform of 1 and $\hat{\zeta}$ denotes the Fourier transform of the surface profile.

Equation (2.22) can also be obtained from Eq. (2.20) by a Taylor expansion of $\exp[-i(b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q}))\zeta]$ around $b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q}) = 0$. Indeed, we have

$$\begin{aligned}
\mathcal{J}_{l,m}^{b,a}(\mathbf{p}\,|\,\mathbf{q}) &= (b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q}))^{-1} \int \exp\left[-i(\mathbf{k}_l^b(\mathbf{p}) - \mathbf{k}_m^a(\mathbf{q}))\cdot\mathbf{s}(\mathbf{x})\right] \mathrm{d}^2x \\
&= (b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q}))^{-1} \int \exp\left[-i(\mathbf{p}-\mathbf{q})\cdot\mathbf{x}\right] \exp\left[-i(b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q}))\zeta(\mathbf{x})\right] \mathrm{d}^2x \\
&= \sum_{k=0}^{\infty} \frac{(-i)^k}{k!} (b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q}))^{k-1}\, \hat{\zeta}^{(k)}(\mathbf{p}-\mathbf{q}) \\
&= \frac{(2\pi)^2\,\delta(\mathbf{p}-\mathbf{q})}{b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q})} - i\hat{\zeta}(\mathbf{p}-\mathbf{q}) + \sum_{k=2}^{\infty} \frac{(-i)^k}{k!} (b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q}))^{k-1}\, \hat{\zeta}^{(k)}(\mathbf{p}-\mathbf{q})\,. \qquad (2.23)
\end{aligned}$$

Here we have defined the k^{th}-*Fourier moment of the surface profile function*, $\hat{\zeta}^{(k)}$, as

$$\hat{\zeta}^{(k)}(\mathbf{q}) \stackrel{\text{def}}{=} \int \exp\left[-i(\mathbf{p}-\mathbf{q})\cdot\mathbf{x}\right]\, \zeta^k(\mathbf{x})\, \mathrm{d}^2x\ . \qquad (2.24)$$

Thus

$$\begin{aligned}
\mathcal{J}_{l,m}^{b,a}(\mathbf{p}\,|\,\mathbf{q})\ \left(\mathbf{k}_l^b(\mathbf{p}) - \mathbf{k}_m^a(\mathbf{q})\right) &= \Big[\mathbf{p} - \mathbf{q} + (b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q}))\,\hat{\mathbf{e}}_3\Big] \\
&\times \left[\frac{(2\pi)^2\,\delta(\mathbf{p}-\mathbf{q})}{b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q})} - i\hat{\zeta}(\mathbf{p}-\mathbf{q}) + \sum_{k=2}^{\infty} \frac{(-i)^k}{k!} (b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q}))^{k-1}\, \hat{\zeta}^{(k)}(\mathbf{p}-\mathbf{q})\right] \qquad (2.25) \\
&= (2\pi)^2\,\delta(\mathbf{p}-\mathbf{q})\ \hat{\mathbf{e}}_3 + \Big[\mathbf{p} - \mathbf{q} + (b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q}))\,\hat{\mathbf{e}}_3\Big] \\
&\times \left[-i\hat{\zeta}(\mathbf{p}-\mathbf{q}) + \sum_{k=2}^{\infty} \frac{(-i)^k}{k!} (b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q}))^{k-1}\, \hat{\zeta}^{(k)}(\mathbf{p}-\mathbf{q})\right] \\
&\stackrel{\text{def}}{=} (2\pi)^2\,\delta(\mathbf{p}-\mathbf{q})\ \hat{\mathbf{e}}_3 + \Big[\mathbf{k}_l^b(\mathbf{p}) - \mathbf{k}_m^a(\mathbf{q})\Big]\, \mathcal{L}_{l,m}^{b,a}(\mathbf{p}\,|\,\mathbf{q}) \qquad (2.26) \\
&\sim (2\pi)^2\,\delta(\mathbf{p}-\mathbf{q})\ \hat{\mathbf{e}}_3 - i\hat{\zeta}(\mathbf{p}-\mathbf{q})\,(\mathbf{p}-\mathbf{q})\,, \qquad (2.27)
\end{aligned}$$

where the asymptotic expression, Eq. (2.27), is to be understood as $b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q}) \to 0$. Equation (2.20) will then be considered as the "general" expression and in particular Eq. (2.26) will be particularly useful for carefully carrying out integrations against $\mathbf{w}_{l,m}^{b,a}$. Note that Eq. (2.26) is a defining equation for $\mathcal{L}_{l,m}^{b,a}$ which, under sufficient regularity of ζ, is a function.

2.1.2 The right hand side

By plugging Eq. (2.20) in the form given by Eq. (2.26) into Eqs. (2.17a) we obtain

$$\begin{aligned}
\mathrm{RHS}_1 &= -\sum_{a=\pm} \int \left(\mathbf{k}_2^b(\mathbf{p}) + \mathbf{k}_2^a(\mathbf{q})\right) \cdot \mathcal{J}_{22}^{b,a}(\mathbf{p}\,|\,\mathbf{q})\ \left(\mathbf{k}_2^b(\mathbf{p}) - \mathbf{k}_2^a(\mathbf{q})\right)\ \mathbf{E}_2^a(\mathbf{q})\, \frac{\mathrm{d}^2q}{(2\pi)^2} \\
&= -\sum_{a=\pm} \int \left(\mathbf{k}_2^b(\mathbf{p}) + \mathbf{k}_2^a(\mathbf{q})\right) \cdot \Big[(2\pi)^2\,\delta(\mathbf{p}-\mathbf{q})\ \hat{\mathbf{e}}_3 + \Big[\mathbf{k}_2^b(\mathbf{p}) - \mathbf{k}_2^a(\mathbf{q})\Big]\, \mathcal{L}_{22}^{b,a}(\mathbf{p}\,|\,\mathbf{q})\Big]\ \mathbf{E}_2^a(\mathbf{q})\, \frac{\mathrm{d}^2q}{(2\pi)^2} \\
&= -\sum_{a=\pm} \left(b\alpha_2(\mathbf{p}) + a\alpha_2(\mathbf{p})\right) \mathbf{E}_2^a(\mathbf{p}) - \sum_{a=\pm} \int \mathcal{L}_{22}^{b,a}(\mathbf{p}\,|\,\mathbf{q}) \Big[\underbrace{[\mathbf{k}_2^b(\mathbf{p})]^2 - [\mathbf{k}_2^a(\mathbf{q})]^2}_{0}\Big]\ \mathbf{E}_2^a(\mathbf{q})\, \frac{\mathrm{d}^2q}{(2\pi)^2} \\
&= -2\,b\,\alpha_2(\mathbf{p})\ \mathbf{E}_2^b(\mathbf{p})\ . \qquad (2.28)
\end{aligned}$$

Here we have used, in the last step, that both wave vectors $\mathbf{k}_2^b(\mathbf{p})$ and $\mathbf{k}_2^a(\mathbf{q})$ have same norm since they both satisfy the dispersion relation in the same medium (medium 2). Similarly, we have for RHS_2,

$$\begin{aligned}
\mathrm{RHS}_2 &= \sum_{a=\pm} \int \mathbf{E}_2^a(\mathbf{q}) \times \Big[\Big((2\pi)^2 \, \delta(\mathbf{p}-\mathbf{q}) \; \hat{\mathbf{e}}_3 + \big(\mathbf{k}_2^b(\mathbf{p}) - \mathbf{k}_2^a(\mathbf{q}) \big) \, \mathcal{L}_{22}^{b,a}(\mathbf{p} \,|\, \mathbf{q}) \Big) \times \big(\mathbf{k}_2^b(\mathbf{p}) - \mathbf{k}_2^a(\mathbf{q}) \big) \Big] \, \frac{\mathrm{d}^2 q}{(2\pi)^2} \\
&= \sum_{a=\pm} \int \mathbf{E}_2^a(\mathbf{q}) \times \Big[(2\pi)^2 \, \delta(\mathbf{p}-\mathbf{q}) \; \hat{\mathbf{e}}_3 \times \big(\mathbf{p} - \mathbf{q} \big) \\
&\quad + \underbrace{\big(\mathbf{k}_2^b(\mathbf{p}) - \mathbf{k}_2^a(\mathbf{q}) \big) \times \big(\mathbf{k}_2^b(\mathbf{p}) - \mathbf{k}_2^a(\mathbf{q}) \big)}_{0} \, \mathcal{L}_{22}^{b,a}(\mathbf{p} \,|\, \mathbf{q}) \Big] \, \frac{\mathrm{d}^2 q}{(2\pi)^2} \\
&= \mathbf{0} \; .
\end{aligned} \tag{2.29}$$

Finally, summing Eqs. (2.28) and (2.29), we obtain the full right hand side

$$\mathrm{RHS} = -2 \, b \, \alpha_2(\mathbf{p}) \; \mathbf{E}_2^b(\mathbf{p}) \; . \tag{2.30}$$

Remark 2.5. In the second step in the above evaluation of RHS_1 and RHS_2 (Eqs. (2.28) and (2.29)), the term coming from the integration against the Dirac mass implicitly assumes that the field amplitude $\mathbf{E}_2^a(\mathbf{q})$ is continuous at $\mathbf{q} = \mathbf{p}$. We will for now make this assumption. This may not always be true, and one should analyze this term more carefully in that case.

2.1.3 The left hand side

The left-hand-side can be treated in a similar fashion as the right-hand-side. In particular, within the same hypothesis for the regularity of $\mathbf{E}_1^a$ as that used for $\mathbf{E}_2^a$ (see Remark 2.5), we show easily that LHS_2 vanishes. Indeed, by plugging Eq. (2.26) into Eq. (2.17d) we obtain

$$\begin{aligned}
\mathrm{LHS}_2 &= \sum_{a=\pm} \int \mathbf{E}_1^a(\mathbf{q}) \times \Big[\Big((2\pi)^2 \, \delta(\mathbf{p}-\mathbf{q}) \; \hat{\mathbf{e}}_3 + \big(\mathbf{k}_2^b(\mathbf{p}) - \mathbf{k}_1^a(\mathbf{q}) \big) \, \mathcal{L}_{21}^{b,a}(\mathbf{p} \,|\, \mathbf{q}) \Big) \times \big(\mathbf{k}_2^b(\mathbf{p}) - \mathbf{k}_1^a(\mathbf{q}) \big) \Big] \, \frac{\mathrm{d}^2 q}{(2\pi)^2} \\
&= \sum_{a=\pm} \int \mathbf{E}_1^a(\mathbf{q}) \times \Big[(2\pi)^2 \, \delta(\mathbf{p}-\mathbf{q}) \; \hat{\mathbf{e}}_3 \times \big(\mathbf{p} - \mathbf{q} \big) \\
&\quad + \underbrace{\big(\mathbf{k}_2^b(\mathbf{p}) - \mathbf{k}_1^a(\mathbf{q}) \big) \times \big(\mathbf{k}_2^b(\mathbf{p}) - \mathbf{k}_1^a(\mathbf{q}) \big)}_{0} \, \mathcal{L}_{21}^{b,a}(\mathbf{p} \,|\, \mathbf{q}) \Big] \, \frac{\mathrm{d}^2 q}{(2\pi)^2} \\
&= \mathbf{0} \; .
\end{aligned} \tag{2.31}$$

There is no particularly interesting simplification that occurs in LHS_1 and LHS_3 by plugging Eq. (2.26) into Eqs. (2.17c) and (2.17e). Using Eq. (2.26) into Eq. (2.17c) we get

$$\begin{aligned}
\mathrm{LHS}_1 &= - \sum_{a=\pm} \int \big(\mathbf{k}_2^b(\mathbf{p}) + \mathbf{k}_1^a(\mathbf{q}) \big) \cdot \Big(\mathcal{J}_{21}^{b,a}(\mathbf{p} \,|\, \mathbf{q}) \; \big(\mathbf{k}_2^b(\mathbf{p}) - \mathbf{k}_1^a(\mathbf{q}) \big) \Big) \, \mathbf{E}_1^a(\mathbf{q}) \, \frac{\mathrm{d}^2 q}{(2\pi)^2} \\
&= - \sum_{a=\pm} \int \mathcal{J}_{21}^{b,a}(\mathbf{p} \,|\, \mathbf{q}) \; \big(|\, \mathbf{k}_2^b(\mathbf{p}) \,|^2 - |\, \mathbf{k}_1^a(\mathbf{q}) \,|^2 \big) \, \mathbf{E}_1^a(\mathbf{q}) \, \frac{\mathrm{d}^2 q}{(2\pi)^2} \\
&= - (\epsilon_2 - \epsilon_1) \, \frac{\omega^2}{c^2} \sum_{a=\pm} \int \mathcal{J}_{21}^{b,a}(\mathbf{p} \,|\, \mathbf{q}) \; \mathbf{E}_1^a(\mathbf{q}) \, \frac{\mathrm{d}^2 q}{(2\pi)^2} \; .
\end{aligned} \tag{2.32}$$

Here we have used the dispersion relation satisfied by the wave vector in media 1 and 2, $|\,\mathbf{k}_j^{\pm}\,|^2 = \epsilon_j\,\omega^2/c^2$. Using Eq. (2.26) into Eq. (2.17e) we simply get

$$\mathrm{LHS}_3 = \frac{\epsilon_2-\epsilon_1}{\epsilon_2}\sum_{a=\pm}\int \mathcal{J}_{21}^{b,a}\left(\mathbf{p}\,|\,\mathbf{q}\right)\,\left(\,\mathbf{k}_2^b\left(\mathbf{p}\right)-\mathbf{k}_1^a\left(\mathbf{q}\right)\,\right)\cdot\mathbf{E}_1^a\left(\mathbf{q}\right)\,\mathbf{k}_2^b\left(\mathbf{p}\right)\frac{\mathrm{d}^2q}{(2\pi)^2}\ . \tag{2.33}$$

After summing Eqs. (2.32), (2.31) and (2.33), the full left hand side reads

$$\mathrm{LHS} = -(\epsilon_2-\epsilon_1)\sum_{a=\pm}\int \mathcal{J}_{21}^{b,a}\left(\mathbf{p}\,|\,\mathbf{q}\right)\left(\frac{\omega^2}{c^2}\,\mathbf{E}_1^a\left(\mathbf{q}\right)-\frac{\left(\mathbf{k}_2^b\left(\mathbf{p}\right)-\mathbf{k}_1^a\left(\mathbf{q}\right)\right)\cdot\mathbf{E}_1^a\left(\mathbf{q}\right)}{\epsilon_2}\,\mathbf{k}_2^b\left(\mathbf{p}\right)\right)\frac{\mathrm{d}^2q}{(2\pi)^2}\ . \tag{2.34}$$

2.1.4 Transfer equations

Finally, following Eqs. (2.30) and (2.34), the resulting linear combination Eq. (2.11) is found to be

$$\sum_{a=\pm}\int \mathcal{J}_{2,1}^{b,a}\left(\mathbf{p}\,|\,\mathbf{q}\right)\,\left(\frac{\omega^2}{c^2}\,\mathbf{E}_1^a\left(\mathbf{q}\right)-\frac{\left(\mathbf{k}_2^b\left(\mathbf{p}\right)-\mathbf{k}_1^a\left(\mathbf{q}\right)\right)\cdot\mathbf{E}_1^a\left(\mathbf{q}\right)}{\epsilon_2}\,\mathbf{k}_2^b\left(\mathbf{p}\right)\right)\frac{\mathrm{d}^2q}{(2\pi)^2} = \frac{2\,b\,\alpha_2\left(\mathbf{p}\right)}{\epsilon_2-\epsilon_1}\,\mathbf{E}_2^b\left(\mathbf{p}\right)\ . \tag{2.35}$$

Using the fact that $\mathbf{k}_1^a\left(\mathbf{q}\right)\cdot\mathbf{E}_1^a\left(\mathbf{q}\right)=0$ and multiplying Eq. (2.35) by ϵ_2, we recognize the identity $\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = (\mathbf{a}\cdot\mathbf{c})\mathbf{b}-(\mathbf{a}\cdot\mathbf{b})\mathbf{c}$ in the integrand on the left hand side. Hence

$$\sum_{a=\pm}\int \mathcal{J}_{2,1}^{b,a}\left(\mathbf{p}\,|\,\mathbf{q}\right)\,\mathbf{k}_2^b\left(\mathbf{p}\right)\times\left[\mathbf{k}_2^b\left(\mathbf{p}\right)\times\mathbf{E}_1^a\left(\mathbf{q}\right)\right]\frac{\mathrm{d}^2q}{(2\pi)^2} = -\frac{2\,b\,\epsilon_2\,\alpha_2\left(\mathbf{p}\right)}{\epsilon_2-\epsilon_1}\,\mathbf{E}_2^b\left(\mathbf{p}\right) \tag{2.36}$$

We introduce the endomorphism $\mathbf{K}_{2,\mathbf{p}}^b$, equivalent to the operation $\mathbf{k}_2^b\left(\mathbf{p}\right)\times.$[3], defined in the basis $(\hat{\mathbf{e}}_1,\hat{\mathbf{e}}_2,\hat{\mathbf{e}}_3)$ by the matrix

$$\mathbf{K}_{2,\mathbf{p}}^b = \begin{pmatrix} 0 & -\kappa_3 & \kappa_2 \\ \kappa_3 & 0 & -\kappa_1 \\ -\kappa_2 & \kappa_1 & 0 \end{pmatrix}_{(\hat{\mathbf{e}}_1,\hat{\mathbf{e}}_2,\hat{\mathbf{e}}_3)}\ , \tag{2.37}$$

where we have used a shorthand notation $\kappa_j = \mathbf{k}_2^b\left(\mathbf{p}\right)\cdot\hat{\mathbf{e}}_j$, for $j\in[\![1,3]\!]$. Equation (2.36) becomes

$$\sum_{a=\pm}\int \mathcal{J}_{2,1}^{b,a}\left(\mathbf{p}\,|\,\mathbf{q}\right)\,\left[\mathbf{K}_{2,\mathbf{p}}^b\right]^2\,\mathbf{E}_1^a\left(\mathbf{q}\right)\frac{\mathrm{d}^2q}{(2\pi)^2} = -\frac{2\,b\,\epsilon_2\,\alpha_2\left(\mathbf{p}\right)}{\epsilon_2-\epsilon_1}\,\mathbf{E}_2^b\left(\mathbf{p}\right)\ , \tag{2.38}$$

with

$$\left[\mathbf{K}_{2,\mathbf{p}}^b\right]^2 = \begin{pmatrix} -\kappa_2^2-\kappa_3^2 & \kappa_1\kappa_2 & \kappa_1\kappa_3 \\ \kappa_2\kappa_1 & -\kappa_1^2-\kappa_3^2 & \kappa_2\kappa_3 \\ \kappa_3\kappa_1 & \kappa_3\kappa_2 & -\kappa_1^2-\kappa_2^2 \end{pmatrix}_{(\hat{\mathbf{e}}_1,\hat{\mathbf{e}}_2,\hat{\mathbf{e}}_3)}\ . \tag{2.39}$$

Let us now use the orthonormal polarization basis associated with the wave vector $\mathbf{k}_j^{\pm}\left(\mathbf{q}\right)$, namely $(\hat{\mathbf{k}}_j^{\pm}\left(\mathbf{q}\right),\hat{\mathbf{e}}_{p,j}^{\pm}\left(\mathbf{q}\right),\hat{\mathbf{e}}_s\left(\mathbf{q}\right))$, and defined in Eq. (2.4). In this basis, the directions given by $\hat{\mathbf{e}}_{p,j}^{\pm}\left(\mathbf{q}\right)$, $\hat{\mathbf{e}}_s\left(\mathbf{q}\right)$ are respectively the directions of the p- and s-polarization of the electric field

[3] i.e. $\forall\mathbf{X}\in\mathbb{R}^3$, $\mathbf{K}_{2,\mathbf{p}}^b\mathbf{X} = \mathbf{k}_2^b\left(\mathbf{p}\right)\times\mathbf{X}$

$\mathbf{E}_j^{\pm}(\mathbf{q})$. It is convenient to express Eq. (2.38) in the basis $(\hat{\mathbf{k}}_1^b(\mathbf{q}), \hat{\mathbf{e}}_{p,1}^b(\mathbf{q}), \hat{\mathbf{e}}_s(\mathbf{q}))$, since in this basis the $\left[\mathbf{K}_{\mathbf{p}}^{1b}\right]^2$ operator reduces to

$$\left[\mathbf{K}_{2,\mathbf{p}}^b\right]^2 = -|\,\mathbf{k}_2^b(\mathbf{p})\,|^2 \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}_{(\hat{\mathbf{k}}_2^b(\mathbf{q}), \hat{\mathbf{e}}_{p,2}^b(\mathbf{q}), \hat{\mathbf{e}}_s(\mathbf{q}))} . \tag{2.40}$$

Moreover $\mathbf{E}_2^b(\mathbf{p}) \cdot \mathbf{k}_2^b(\mathbf{p}) = 0$, so in this basis, the first line of the integral system Eq. (2.38) is trivial since we have

$$\mathbf{E}_2^b(\mathbf{p}) = \left(0, \mathcal{E}_{2,p}^b(\mathbf{q}), \mathcal{E}_{2,s}^b(\mathbf{q})\right)^{\mathrm{T}}_{(\hat{\mathbf{k}}_2^b(\mathbf{q}), \hat{\mathbf{e}}_{p,2}^b(\mathbf{q}), \hat{\mathbf{e}}_s(\mathbf{q}))} . \tag{2.41}$$

Therefore we can reduce the system to the plane $(\hat{\mathbf{e}}_{p,1}^b(\mathbf{p}), \hat{\mathbf{e}}_s(\mathbf{p}))$. Similarly, it is convenient to write $\mathbf{E}_1^a(\mathbf{q})$ in the basis $(\hat{\mathbf{e}}_{p,1}^a(\mathbf{q}), \hat{\mathbf{e}}_s(\mathbf{q}))$, by using the matrix

$$\mathbf{P}_{l,m}^{b,a}(\mathbf{p}\,|\,\mathbf{q}) = \begin{pmatrix} \hat{\mathbf{e}}_{p,l}^b(\mathbf{p}) \cdot \hat{\mathbf{e}}_{p,m}^a(\mathbf{q}) & \hat{\mathbf{e}}_{p,l}^b(\mathbf{p}) \cdot \hat{\mathbf{e}}_s(\mathbf{q}) \\ \hat{\mathbf{e}}_s(\mathbf{p}) \cdot \hat{\mathbf{e}}_{p,m}^a(\mathbf{q}) & \hat{\mathbf{e}}_s(\mathbf{p}) \cdot \hat{\mathbf{e}}_s(\mathbf{q}) \end{pmatrix} \tag{2.42}$$

to change from $(\hat{\mathbf{e}}_{p,l}^b(\mathbf{p}), \hat{\mathbf{e}}_s(\mathbf{p}))$ to $(\hat{\mathbf{e}}_{p,m}^a(\mathbf{q}), \hat{\mathbf{e}}_s(\mathbf{q}))$. By expanding Eq. (2.42) by using Eq. (1.32), and plugging it into Eq. (2.38), one gets the *forward transfer equation*

$$\sum_{a=\pm} \int \left(\epsilon_2 - \epsilon_1\right) \frac{\omega^2}{c^2}\, \mathcal{J}_{2,1}^{b,a}(\mathbf{p}\,|\,\mathbf{q})\; \mathbf{P}_{2,1}^{b,a}(\mathbf{p}\,|\,\mathbf{q})\; \boldsymbol{\mathcal{E}}_1^a(\mathbf{q})\, \frac{\mathrm{d}^2 q}{(2\pi)^2} = 2\, b\, \alpha_2(\mathbf{p})\; \boldsymbol{\mathcal{E}}_2^b(\mathbf{p}) \;, \tag{2.43}$$

with $\boldsymbol{\mathcal{E}}_j^{\pm}(\mathbf{q}) = (\mathcal{E}_{j,p}^{\pm}(\mathbf{q}), \mathcal{E}_{j,s}^{\pm}(\mathbf{q}))^{\mathrm{T}}$. By multiplying both sides of Eq. (2.43) by $\sqrt{\epsilon_1 \epsilon_2}$, the transfer equation can also be written as

$$\sum_{a=\pm} \int \mathcal{J}_{2,1}^{b,a}(\mathbf{p}\,|\,\mathbf{q})\; \mathbf{M}_{2,1}^{b,a}(\mathbf{p}\,|\,\mathbf{q})\; \boldsymbol{\mathcal{E}}_1^a(\mathbf{q})\, \frac{\mathrm{d}^2 q}{(2\pi)^2} = \frac{2\, b\, \sqrt{\epsilon_1 \epsilon_2}\, \alpha_2(\mathbf{p})}{\epsilon_2 - \epsilon_1}\; \boldsymbol{\mathcal{E}}_2^b(\mathbf{p}) \;, \tag{2.44}$$

where

$$\mathbf{M}_{l,m}^{b,a}(\mathbf{p}\,|\,\mathbf{q}) \overset{\text{def}}{=} \sqrt{\epsilon_1 \epsilon_2}\, \frac{\omega^2}{c^2}\; \mathbf{P}_{l,m}^{b,a}(\mathbf{p}\,|\,\mathbf{q}) \tag{2.45}$$

$$= \begin{pmatrix} |\mathbf{p}||\mathbf{q}| + ab\, \alpha_l(\mathbf{p})\, \alpha_m(\mathbf{q})\, \hat{\mathbf{p}} \cdot \hat{\mathbf{q}} & -b\sqrt{\epsilon_m}\, \frac{\omega}{c}\, \alpha_l(\mathbf{p}) [\hat{\mathbf{p}} \times \hat{\mathbf{q}}] \cdot \hat{\mathbf{e}}_3 \\ a\sqrt{\epsilon_l}\, \frac{\omega}{c}\, \alpha_m(\mathbf{q}) [\hat{\mathbf{p}} \times \hat{\mathbf{q}}] \cdot \hat{\mathbf{e}}_3 & \sqrt{\epsilon_l \epsilon_m}\, \frac{\omega^2}{c^2}\, \hat{\mathbf{p}} \cdot \hat{\mathbf{q}} \end{pmatrix} . \tag{2.46}$$

Equation (2.44) is the form which is maybe encountered most in the literature. However, we found the form of the transfer equation given by Eq. (2.43) somewhat clearer in the sense that the matrix of change of polarization basis is emphasized. In addition, Eq. (2.43) will be simpler to compare with the corresponding transfer equation for scalar waves (see Section 2.6). Nevertheless, since Eq. (2.44) is the form that is used in the literature, we will use the form given in Eq. (2.44) in the remaining of this work, and Eq. (2.43) will be used for occasional remarks. Notice that due to the symmetry of the boundary conditions, one may also derive in the same way the *backward transfer equation*

$$\sum_{a=\pm} \int \mathcal{J}_{1,2}^{b,a}(\mathbf{p}\,|\,\mathbf{q})\; \mathbf{M}_{1,2}^{b,a}(\mathbf{p}\,|\,\mathbf{q})\; \boldsymbol{\mathcal{E}}_2^a(\mathbf{q})\, \frac{\mathrm{d}^2 q}{(2\pi)^2} = \frac{2\, b\, \sqrt{\epsilon_1 \epsilon_2}\, \alpha_1(\mathbf{p})}{\epsilon_1 - \epsilon_2}\; \boldsymbol{\mathcal{E}}_1^b(\mathbf{p}) \;, \tag{2.47}$$

which ends up as interchanging the indices $1 \leftrightarrow 2$ in Eq. (2.44). Equations (2.44) and (2.47) are the so-called transfer equations. Typically, Eq. (2.44) is appropriate to solve the problem of reflection whereas Eq. (2.47) is appropriate to solve the problem transmission, as we will see in the next subsection.

2.2 Reduced Rayleigh equations

Consider now that we are interested in evaluating the angular distribution of light reflected (resp. transmitted) by (resp. through) a single interface separating two semi-infinite media, knowing that a plane wave is incident on it. The incident wave can be expressed as

$$\mathbf{E}_1^- (\mathbf{q}) = (2\pi)^2 \delta(\mathbf{q} - \mathbf{p}_0) \left[\mathcal{E}_{0,p} \, \hat{\mathbf{e}}_{p,1}^-(\mathbf{p}_0) + \mathcal{E}_{0,s} \, \hat{\mathbf{e}}_s(\mathbf{p}_0) \right] . \tag{2.48}$$

We assume that this is the only contribution of the electric field propagating towards the interface. We introduce the reflection (resp. transmission) amplitudes $\mathbf{R}$ (resp. $\mathbf{T}$) defined as

$$\boldsymbol{\mathcal{E}}_1^+ (\mathbf{q}) = \mathbf{R} (\mathbf{q} \,|\, \mathbf{p}_\mathrm{o}) \, \boldsymbol{\mathcal{E}}_0 \ , \tag{2.49}$$

$$\boldsymbol{\mathcal{E}}_2^- (\mathbf{q}) = \mathbf{T} (\mathbf{q} \,|\, \mathbf{p}_\mathrm{o}) \, \boldsymbol{\mathcal{E}}_0 \ , \tag{2.50}$$

where we have defined $\boldsymbol{\mathcal{E}}_0 = (\mathcal{E}_{0,p}, \mathcal{E}_{0,s})^\mathrm{T}$. The 2×2 complex valued matrix $\mathbf{R} (\mathbf{q} \,|\, \mathbf{p}_\mathrm{o})$ (resp. $\mathbf{T} (\mathbf{q} \,|\, \mathbf{p}_\mathrm{o})$) is then the operator that describes how the upwards (resp. downwards) plane waves components of the reflected (resp. transmitted) field are distributed knowing the incident field. It is closely related to the angular distribution of the reflected (resp. transmitted) power as we will see later on, and is the naturally the unknown of the problem. By plugging Eqs. (2.48) and (2.49) (resp. Eqs. (2.48) and (2.50)) in Eq. (2.44) (resp. Eq. (2.47)) and choosing $b = +1$ (resp. $b = -1$), we obtain an integral equation satisfied by the reflection (resp. transmission) amplitudes

$$\int \mathcal{J}_{2,1}^{+,+} (\mathbf{p} \,|\, \mathbf{q}) \ \mathbf{M}_{2,1}^{+,+} (\mathbf{p} \,|\, \mathbf{q}) \ \mathbf{R} (\mathbf{q} \,|\, \mathbf{p}_\mathrm{o}) \, \frac{\mathrm{d}^2 q}{(2\pi)^2} = - \, \mathcal{J}_{2,1}^{+,-} (\mathbf{p} \,|\, \mathbf{p}_\mathrm{o}) \ \mathbf{M}_{2,1}^{+,-} (\mathbf{p} \,|\, \mathbf{p}_\mathrm{o}) \ , \tag{2.51}$$

$$\int \mathcal{J}_{1,2}^{-,-} (\mathbf{p} \,|\, \mathbf{q}) \ \mathbf{M}_{1,2}^{-,-} (\mathbf{p} \,|\, \mathbf{q}) \ \mathbf{T} (\mathbf{q} \,|\, \mathbf{p}_\mathrm{o}) \, \frac{\mathrm{d}^2 q}{(2\pi)^2} = \frac{2\sqrt{\epsilon_1 \epsilon_2} \, \alpha_1(\mathbf{p}_0)}{\epsilon_2 - \epsilon_1} \, (2\pi)^2 \, \delta(\mathbf{p} - \mathbf{p}_0) \, \mathbf{I}_2 \, . \tag{2.52}$$

Here, the fundamental property of the Dirac delta has been used and $\mathbf{E}_0(\mathbf{p}_0)$ has been simplified on both side of the equation, and $\mathbf{I}_2$ is the identity matrix.

Finite size beam

Equations (2.51) and (2.52) are the reduced Rayleigh equations obtained by assuming that the incident excitation is a monochromatic plane wave. The optical response obtained for an incident plane wave can in certain cases model within reasonable accuracy real life experiments. In other cases, modeling the incident excitation as a finite size beam may be more realistic and appropriate. For a monochromatic incident finite size beam two options may be considered for modeling the scattering response. One may solve the scattering response from a sample of independent incident plane waves impinging with different wave vectors and then sum the corresponding responses weighted according to the weights of the Fourier components corresponding to the elementary plane waves contained in the finite size beam. This method is valid thanks to the linearity of the problem. An advantage of this method is that once all the elementary scattering problems have been solved (those for each incident plane waves) and stored, one deduces

immediately the solution for any finite size beam by linear superposition. Alternatively, if one is interested in the scattering response of a very specific finite size beam, one can simply solve a single scattering problem directly for the considered finite size beam. Indeed, an incident finite size beam may be written as a superposition of downward propagating plane wave as

$$\mathbf{E}_1^{(\mathrm{inc})}(\mathbf{r}) = \int \mathbf{E}_1^-(\mathbf{q}) \exp(i\,\mathbf{k}_1^-(\mathbf{q})\cdot\mathbf{r})\,\frac{\mathrm{d}^2q}{(2\pi)^2}\,, \tag{2.53}$$

where $\mathbf{E}_1^-(\mathbf{q}) = \mathcal{E}_{1,p}^-(\mathbf{q})\,\hat{\mathbf{e}}_{p,1}^-(\mathbf{q}) + \mathcal{E}_{1,s}^-(\mathbf{q})\,\hat{\mathbf{e}}_s(\mathbf{q})$ now represents an amplitude defining the finite size beam. The forward transfer equation, Eq. (2.44), applied for $b = +$ (with $\boldsymbol{\mathcal{E}}_2^b = 0$) then gives the following integral equation for the scattered field amplitude in medium 1, $\boldsymbol{\mathcal{E}}_1^+$,

$$\int \mathcal{J}_{2,1}^{+,+}(\mathbf{p}\,|\,\mathbf{q})\;\mathbf{M}_{2,1}^{+,+}(\mathbf{p}\,|\,\mathbf{q})\;\boldsymbol{\mathcal{E}}_1^+(\mathbf{q})\;\frac{\mathrm{d}^2q}{(2\pi)^2} = -\int \mathcal{J}_{2,1}^{+,-}(\mathbf{p}\,|\,\mathbf{q})\;\mathbf{M}_{2,1}^{+,-}(\mathbf{p}\,|\,\mathbf{q})\;\boldsymbol{\mathcal{E}}_1^-(\mathbf{q})\;\frac{\mathrm{d}^2q}{(2\pi)^2}\,. \tag{2.54}$$

Similarly, the backward transfer equation applied for $b = -$ yields the following integral equation for the scattered field amplitude in medium 2, $\boldsymbol{\mathcal{E}}_2^-$,

$$\int \mathcal{J}_{1,2}^{-,-}(\mathbf{p}\,|\,\mathbf{q})\;\mathbf{M}_{1,2}^{+,+}(\mathbf{p}\,|\,\mathbf{q})\;\boldsymbol{\mathcal{E}}_2^-(\mathbf{q})\;\frac{\mathrm{d}^2q}{(2\pi)^2} = -\frac{2\sqrt{\epsilon_1\epsilon_2}\,\alpha_1(\mathbf{p})}{\epsilon_1-\epsilon_2}\;\boldsymbol{\mathcal{E}}_1^-(\mathbf{p})\,. \tag{2.55}$$

2.3 Fresnel coefficients for a planar interface

An instructive case to consider is that of a flat interface for which the reflection and transmission amplitudes are well known, namely the Fresnel coefficients. We will verify that the integral equations (2.51) and (2.52) can be solved exactly in that case and yields the Fresnel coefficients. For a flat interface, the surface profile is simply a constant, i.e. $x_3 = \zeta(\mathbf{x}) = H$. For simplicity we will choose the origin of coordinate on the surface, which gives $H = 0$ (this choice is arbitrary and one could choose to do have a non-zero height which will simply result in a phase factor). Thus we have

$$\mathcal{J}_{l,m}^{b,a}(\mathbf{p}\,|\,\mathbf{q}) = \frac{\int e^{-i(\mathbf{p}-\mathbf{q})\cdot\mathbf{x}}\,\mathrm{d}^2x}{b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q})} = \frac{(2\pi)^2}{b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{p})}\,\delta(\mathbf{q}-\mathbf{p})\,. \tag{2.56}$$

By plugging Eq. (2.56) in Eqs. (2.51,2.52), the integration becomes straightforward thanks to to the Dirac delta and we can invert for the reflection and transmission amplitudes:

$$\mathbf{R}(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}) = \frac{\alpha_1(\mathbf{p}_0)-\alpha_2(\mathbf{p}_0)}{\alpha_1(\mathbf{p}_0)+\alpha_2(\mathbf{p}_0)}\left[\mathbf{M}_{2,1}^{+,+}(\mathbf{p}_\mathrm{o}\,|\,\mathbf{p}_\mathrm{o})\right]^{-1}\;\mathbf{M}_{2,1}^{+,-}(\mathbf{p}_\mathrm{o}\,|\,\mathbf{p}_\mathrm{o})\,(2\pi)^2\,\delta(\mathbf{p}-\mathbf{p}_0)\,, \tag{2.57}$$

$$\mathbf{T}(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}) = \frac{2\sqrt{\epsilon_1\epsilon_2}\,(\alpha_2(\mathbf{p}_0)-\alpha_1(\mathbf{p}_0))\,\alpha_1(\mathbf{p}_0)}{\epsilon_2-\epsilon_1}\;\left[\mathbf{M}_{1,2}^{-,-}(\mathbf{p}_\mathrm{o}\,|\,\mathbf{p}_\mathrm{o})\right]^{-1}\;(2\pi)^2\,\delta(\mathbf{p}-\mathbf{p}_0)\,. \tag{2.58}$$

Using the definition of the matrices $\mathbf{M}_{l,m}^{b,a}$, Eq. (2.46), we realize that only diagonal terms remain and we finally get

$$\mathbf{R}(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}) = \begin{pmatrix} \frac{\alpha_1(\mathbf{p}_0)-\alpha_2(\mathbf{p}_0)}{\alpha_1(\mathbf{p}_0)+\alpha_2(\mathbf{p}_0)} & \frac{|\,\mathbf{p}_0\,|^2-\alpha_2(\mathbf{p}_0)\alpha_1(\mathbf{p}_0)}{|\,\mathbf{p}_0\,|^2+\alpha_2(\mathbf{p}_0)\alpha_1(\mathbf{p}_0)} \\ 0 & \frac{\alpha_1(\mathbf{p}_0)-\alpha_2(\mathbf{p}_0)}{\alpha_1(\mathbf{p}_0)+\alpha_2(\mathbf{p}_0)} \end{pmatrix}(2\pi)^2\,\delta(\mathbf{p}-\mathbf{p}_0)\,, \tag{2.59}$$

$$\mathbf{T}(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}) = \begin{pmatrix} \frac{2\sqrt{\epsilon_1\epsilon_2}\,\alpha_1(\mathbf{p}_0)}{\epsilon_2-\epsilon_1} & \frac{(\alpha_2(\mathbf{p}_0)-\alpha_1(\mathbf{p}_0))}{|\,\mathbf{p}_0\,|^2+\alpha_2(\mathbf{p}_0)\alpha_1(\mathbf{p}_0)} \\ 0 & \frac{2\,(\alpha_2(\mathbf{p}_0)-\alpha_1(\mathbf{p}_0))\,\alpha_1(\mathbf{p}_0)}{\frac{\omega^2}{c^2}(\epsilon_2-\epsilon_1)} \end{pmatrix}(2\pi)^2\,\delta(\mathbf{p}-\mathbf{p}_0)\,. \tag{2.60}$$

We can simplify the expression using in the reflection amplitude that

$$\begin{aligned}
&[\alpha_1(\mathbf{p}_0) \pm \alpha_2(\mathbf{p}_0)] \left[|\,\mathbf{p}_0\,|^2 \pm \alpha_2(\mathbf{p}_0)\alpha_1(\mathbf{p}_0) \right] \\
&\qquad = [\alpha_1(\mathbf{p}_0) \pm \alpha_2(\mathbf{p}_0)]\,|\,\mathbf{p}_0\,|^2 \pm \alpha_2(\mathbf{p}_0)\alpha_1^2(\mathbf{p}_0) + \alpha_1(\mathbf{p}_0)\alpha_2^2(\mathbf{p}_0) \\
&\qquad = [\alpha_1(\mathbf{p}_0) \pm \alpha_2(\mathbf{p}_0)]\,|\,\mathbf{p}_0\,|^2 \pm \alpha_2(\mathbf{p}_0) \left[\epsilon_1 \frac{\omega^2}{c^2} - |\,\mathbf{p}_0\,|^2 \right] + \alpha_1(\mathbf{p}_0) \left[\epsilon_2 \frac{\omega^2}{c^2} - |\,\mathbf{p}_0\,|^2 \right] \\
&\qquad = [\epsilon_2\alpha_1(\mathbf{p}_0) \pm \epsilon_1\alpha_2(\mathbf{p}_0)] \frac{\omega^2}{c^2}\,,
\end{aligned}$$

and by making appear a factor $\frac{\alpha_1(\mathbf{p}_0)+\alpha_2(\mathbf{p}_0)}{\alpha_1(\mathbf{p}_0)+\alpha_2(\mathbf{p}_0)}$ in the transmission amplitude we can write

$$\mathbf{R}\,(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}) = \begin{pmatrix} \frac{\epsilon_2\alpha_1(\mathbf{p}_0)-\epsilon_1\alpha_2(\mathbf{p}_0)}{\epsilon_2\alpha_1(\mathbf{p}_0)+\epsilon_1\alpha_2(\mathbf{p}_0)} & 0 \\ 0 & \frac{\alpha_1(\mathbf{p}_0)-\alpha_2(\mathbf{p}_0)}{\alpha_1(\mathbf{p}_0)+\alpha_2(\mathbf{p}_0)} \end{pmatrix} (2\pi)^2\,\delta(\mathbf{p}-\mathbf{p}_0)\,, \tag{2.61}$$

$$\mathbf{T}\,(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}) = \begin{pmatrix} \frac{2\sqrt{\epsilon_1\epsilon_2}\,\alpha_1(\mathbf{p}_0)}{\epsilon_2\alpha_1(\mathbf{p}_0)+\epsilon_1\alpha_2(\mathbf{p}_0)} & 0 \\ 0 & \frac{2\,\alpha_1(\mathbf{p}_0)}{\alpha_1(\mathbf{p}_0)+\alpha_2(\mathbf{p}_0)} \end{pmatrix} (2\pi)^2\,\delta(\mathbf{p}-\mathbf{p}_0)\,. \tag{2.62}$$

We notice that we recover all the well known results for a single flat interface scattering of an incident plane wave. The Dirac delta indicates that scattering occurs only in the specular direction, and gives Snell-Descartes' law, namely that the component of the wave vector parallel to the surface must be conserved. The off-diagonal terms are zero indicates that cross-polarization scattering is forbidden and last but not least we recognize that the diagonal terms are the well known Fresnel reflection and transmission amplitudes for each $p-$ and $s-$ polarized light.

2.4 Coupled Rayleigh equations

It seems that in the literature the denomination of *reduced* Rayleigh equations for Eqs. (2.51) and (2.52) refers to the fact that the two equations are uncoupled. In other words, one can choose to solve the scattering problem only for the reflection or transmission amplitudes separately. It is worth noticing that one could choose to keep the system of equations for the reflection and transmission amplitudes coupled. Indeed, instead of taking $b = +1$ (resp. $b = -1$) in Eq. (2.44) (resp. Eq. (2.47)), one could take the opposite sign and end up with the following set of coupled integral equations for the reflection and transmission amplitudes

$$\begin{aligned}
\int \mathcal{J}_{2,1}^{-;+}\,(\mathbf{p}\,|\,\mathbf{q})\;\mathbf{M}_{2,1}^{-;+}\,(\mathbf{p}\,|\,\mathbf{q})\;\mathbf{R}\,(\mathbf{q}\,|\,\mathbf{p}_\mathrm{o}) + \mathcal{J}_{2,1}^{-;-}\,(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o})\;\mathbf{M}_{2,1}^{-;-}\,(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o})\,\frac{\mathrm{d}^2q}{(2\pi)^2} &= -\frac{2\sqrt{\epsilon_1\epsilon_2}\,\alpha_2(\mathbf{p})}{\epsilon_2-\epsilon_1}\,\mathbf{T}\,(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o})\,, \\
\int \mathcal{J}_{1,2}^{+;-}\,(\mathbf{p}\,|\,\mathbf{q})\;\mathbf{M}_{1,2}^{+;-}\,(\mathbf{p}\,|\,\mathbf{q})\;\mathbf{T}\,(\mathbf{q}\,|\,\mathbf{p}_\mathrm{o})\,\frac{\mathrm{d}^2q}{(2\pi)^2} &= \frac{2\sqrt{\epsilon_1\epsilon_2}\,\alpha_1(\mathbf{p})}{\epsilon_1-\epsilon_2}\,\mathbf{R}\,(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o})\,,
\end{aligned}$$

which can be written as

$$\mathbf{T}\,(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}) = \frac{\epsilon_1-\epsilon_2}{2\sqrt{\epsilon_1\epsilon_2}\,\alpha_2(\mathbf{p})} \left(\int \mathcal{J}_{2,1}^{-;+}\,(\mathbf{p}\,|\,\mathbf{q})\;\mathbf{M}_{2,1}^{-;+}\,(\mathbf{p}\,|\,\mathbf{q})\;\mathbf{R}\,(\mathbf{q}\,|\,\mathbf{p}_\mathrm{o})\,\frac{\mathrm{d}^2q}{(2\pi)^2} + \mathcal{J}_{2,1}^{-;-}\,(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o})\;\mathbf{M}_{2,1}^{-;-}\,(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}) \right)\,, \tag{2.63a}$$

$$\mathbf{R}\,(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}) = \frac{\epsilon_1-\epsilon_2}{2\sqrt{\epsilon_1\epsilon_2}\,\alpha_1(\mathbf{p})} \int \mathcal{J}_{1,2}^{+;-}\,(\mathbf{p}\,|\,\mathbf{q})\;\mathbf{M}_{1,2}^{+;-}\,(\mathbf{p}\,|\,\mathbf{q})\;\mathbf{T}\,(\mathbf{q}\,|\,\mathbf{p}_\mathrm{o})\,\frac{\mathrm{d}^2q}{(2\pi)^2}\,. \tag{2.63b}$$

Why would one choose to work with a coupled set of equations when a decoupled one is available? At first sight, it seems indeed like a not so clever idea to solve the coupled system instead of the decoupled one. We will see in the chapter devoted to numerical methods for solving the Rayleigh equations that a direct method consists in discretizing the integral equations and solving the resulting linear system. Using the coupled equations would result in a system size twice as big as each of the system for the separate decoupled equations. Since solving a linear with direct methods typically scales as the cube of the system size, using the coupled set of equations would be 8 times more costly to solve than one of the decoupled equations, so a factor 4 in the end, if both reflection and transmission amplitudes are needed. However there are two *a priori* interesting points to make about the coupled Rayleigh equations:

- The unknown in the right-hand side integral, is only either $\mathbf{R}$ or $\mathbf{T}$. In other words, $\mathbf{R}$ is given as a function of $\mathbf{T}$ only and vice versa.
- Equations (2.63a) and (2.63b) can be interpreted as a *fixed point* problem of the form $\mathbf{X} = f(\mathbf{X})$, where $\mathbf{X} = (\mathbf{R}, \mathbf{T})$.

A practical application of the first point is that if one knows $\mathbf{R}$ or $\mathbf{T}$, for example after solving *one* of the reduced Rayleigh equations, then it suffices to plug the known amplitudes into the appropriate equation of Eqs. (2.63a) and (2.63b) to get the remaining amplitudes. Numerically, this translates into a matrix-vector multiplication whose number of operations is of the order of the square of the system size, which is considerably cheaper than solving the remaining reduced Rayleigh equation.

The interpretation of the coupled Rayleigh equations as a fixed point problem invites us to consider an iterative way of solving the system. One starts with a guess for $\mathbf{R}$ and $\mathbf{T}$, for example Fresnel coefficients for a flat surface, and then plug these amplitudes in the right-hand side of Eqs. (2.63a) and (2.63b) to obtain a new guess and so on. One must be careful though, since this method is not guaranteed to converge if f (which we have not explicitly written) is not contractant for example. Another problem, purely numerical this time, is the amplification of round-off errors by successive iterations (matrix-vector operation numerically). We will see later that this is a serious issue which is related to the ill-conditioning of the iteration matrices.

2.5 Reduced Rayleigh equations of the second kind

2.5.1 Derivation of the RRE of the second kind

The reduced Rayleigh equations (2.51) and (2.52) take the form of Fredholm integral equations of the first kind. By splitting the scalar kernel factor $\mathcal{J}_{l,m}^{b,a}$ as a sum of a term proportional to a Dirac mass and a remainder, we can recast the reduced Rayleigh equations as Fredholm integral equations of the second kind instead. Such a splitting was suggested for example by Maradudin in Ref. [56] for the study of surface plasmon polariton on a one-dimensional grating as an more appropriate alternative to small amplitude perturbation theory (see Chapter 6). We can write

$$\mathcal{J}_{l,m}^{b,a}(\mathbf{p} \,|\, \mathbf{q}) = \left(b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q})\right)^{-1} (2\pi)^2 \delta(\mathbf{p} - \mathbf{q}) + \mathcal{K}_{l,m}^{b,a}(\mathbf{p} \,|\, \mathbf{q}) \,, \tag{2.64}$$

which defines

$$\mathcal{K}_{l,m}^{b,a}(\mathbf{p}\,|\,\mathbf{q}) \stackrel{\text{def}}{=} \int \frac{\exp[-i(b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q}))\zeta(\mathbf{x})] - 1}{b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q})} \exp[-i(\mathbf{p}-\mathbf{q})\cdot\mathbf{x}]\ \mathrm{d}^2x \tag{2.65}$$

$$= \sum_{n=1}^{\infty} \frac{(-i)^n}{n!} (b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q}))^{n-1}\, \hat{\zeta}^{(n)}(\mathbf{p}-\mathbf{q})\,. \tag{2.66}$$

In the last equation, we have expanded the exponential term at the numerator in a power series and obtained an expansion of $\mathcal{K}_{l,m}^{b,a}(\mathbf{p}\,|\,\mathbf{q})$ in Fourier transform of the powers of the surface profile function, or in short *Fourier moments*, defined as

$$\hat{\zeta}^{(n)}(\mathbf{p}) = \int \zeta^n(\mathbf{x}) \exp[-i\,\mathbf{p}\cdot\mathbf{x}]\ \mathrm{d}^2x\ . \tag{2.67}$$

Note that if the surface profile function vanishes identically (i.e. the surface is planar) then $\mathcal{K}_{l,m}^{b,a}(\mathbf{p}\,|\,\mathbf{q}) = 0$ and the term proportional to the Dirac mass in Eq. (2.64) corresponds to the scalar kernel factor $\mathcal{J}_{l,m}^{b,a}$ associated with a scattering system whose interface is planar. The term $\mathcal{K}_{l,m}^{b,a}$ thus represents deviation of the scalar kernel factor $\mathcal{J}_{l,m}^{b,a}$ from that of a planar interface. Let us now insert Eq. (2.64) into the reduced Rayleigh equation Eq. (2.51) for the reflection amplitudes. We obtain

$$(\alpha_2(\mathbf{p}) - \alpha_1(\mathbf{p}))^{-1}\ \mathbf{M}_{21}^{++}(\mathbf{p}\mid\mathbf{p})\ \mathbf{R}(\mathbf{p}\mid\mathbf{p}_0) + \int \mathcal{K}_{21}^{++}(\mathbf{p}\mid\mathbf{q})\ \mathbf{M}_{21}^{++}(\mathbf{p}\mid\mathbf{q})\ \mathbf{R}(\mathbf{q}\mid\mathbf{p}_0)\ \frac{\mathrm{d}^2q}{(2\pi)^2}$$
$$= -(\alpha_2(\mathbf{p}_0) + \alpha_1(\mathbf{p}_0))^{-1}\ \mathbf{M}_{21}^{+-}(\mathbf{p}_0\mid\mathbf{p}_0)\,(2\pi)^2\delta(\mathbf{p}-\mathbf{p}_0) - \mathcal{K}_{21}^{+-}(\mathbf{p}\mid\mathbf{p}_0)\ \mathbf{M}_{21}^{+-}(\mathbf{p}\mid\mathbf{p}_0)\,. \tag{2.68}$$

Equation (2.68) can be interpreted as an integral equation of the second kind. This invites us to compute successive approximations of $\mathbf{R}(\mathbf{p}\mid\mathbf{p}_0)$ by iterations. First, instead of working with the unknown $\mathbf{R}(\mathbf{p}\mid\mathbf{p}_0)$, we make the following change of variable

$$\Delta\,\mathbf{R}(\mathbf{p}\mid\mathbf{p}_0) = \mathbf{R}(\mathbf{p}\mid\mathbf{p}_0) - \frac{\alpha_1(\mathbf{p}_0) - \alpha_2(\mathbf{p}_0)}{\alpha_2(\mathbf{p}_0) + \alpha_1(\mathbf{p}_0)} \left[\mathbf{M}_{2,1}^{+,+}(\mathbf{p}_o\,|\,\mathbf{p}_o)\right]^{-1} \mathbf{M}_{2,1}^{+,-}(\mathbf{p}_o\,|\,\mathbf{p}_o)\,(2\pi)^2\,\delta(\mathbf{p}-\mathbf{p}_0)$$
$$= \mathbf{R}(\mathbf{p}\mid\mathbf{p}_0) - (2\pi)^2\,\delta(\mathbf{p}-\mathbf{p}_0)\,\boldsymbol{\rho}^{(0)}(\mathbf{p}_0) \tag{2.69}$$

which expresses the fact that we are looking for the deviation of the reflection amplitude from the corresponding amplitude for a similar system but whose interface is planar. Plugging the above equation in Eq. (2.68) we obtain an equation for $\Delta\,\mathbf{R}(\mathbf{p}\mid\mathbf{p}_0)$

$$(\alpha_2(\mathbf{p}) - \alpha_1(\mathbf{p}))^{-1}\ \mathbf{M}_{21}^{++}(\mathbf{p}\mid\mathbf{p})\,\Delta\,\mathbf{R}(\mathbf{p}\mid\mathbf{p}_0) + \int \mathcal{K}_{21}^{++}(\mathbf{p}\mid\mathbf{q})\ \mathbf{M}_{21}^{++}(\mathbf{p}\mid\mathbf{q})\,\Delta\,\mathbf{R}(\mathbf{q}\mid\mathbf{p}_0)\ \frac{\mathrm{d}^2q}{(2\pi)^2}$$
$$= -\,\mathcal{K}_{21}^{++}(\mathbf{p}\mid\mathbf{p}_0)\ \mathbf{M}_{21}^{++}(\mathbf{p}\mid\mathbf{p}_0)\,\boldsymbol{\rho}^{(0)}(\mathbf{p}_0) - \mathcal{K}_{21}^{+-}(\mathbf{p}\mid\mathbf{p}_0)\ \mathbf{M}_{21}^{+-}(\mathbf{p}\mid\mathbf{p}_0)\,. \tag{2.70}$$

Note that the Dirac mass disappears as it cancels with the reflection amplitude of the flat system, which was in fact what motivated the change of variable. The above equation can be interpreted as a fixed point problem and solved by successive iterations in the following way.

Initialize $\Delta\,\mathbf{R}(\mathbf{p}\mid\mathbf{p}_0) = 0$.

Loop until a desired precision, or a maximum number of iterations is reached

$$\Delta\,\mathbf{R}^{(k+1)}(\mathbf{p}\mid\mathbf{p}_0) = (\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p}))\ \left[\mathbf{M}_{21}^{++}(\mathbf{p}\mid\mathbf{p})\right]^{-1}$$
$$\Big(\int \mathcal{K}_{21}^{++}(\mathbf{p}\mid\mathbf{q})\ \mathbf{M}_{21}^{++}(\mathbf{p}\mid\mathbf{q})\,\Delta\,\mathbf{R}^{(k)}(\mathbf{q}\mid\mathbf{p}_0)\ \frac{\mathrm{d}^2q}{(2\pi)^2}$$
$$+\,\mathcal{K}_{21}^{++}(\mathbf{p}\mid\mathbf{p}_0)\ \mathbf{M}_{21}^{++}(\mathbf{p}\mid\mathbf{p}_0)\,\boldsymbol{\rho}^{(0)}(\mathbf{p}_0) + \mathcal{K}_{21}^{+-}(\mathbf{p}\mid\mathbf{p}_0)\ \mathbf{M}_{21}^{+-}(\mathbf{p}\mid\mathbf{p}_0)\Big)$$

End of the loop.

The corresponding integral equation of the second kind for the transmission amplitudes reads

$$(-\alpha_1(\mathbf{p})+\alpha_2(\mathbf{p}))^{-1}\ \mathbf{M}^{-,-}_{1,2}(\mathbf{p}\mid\mathbf{p})\ \mathbf{T}(\mathbf{p}\mid\mathbf{p}_0)+\int \mathcal{K}^{-,-}_{12}(\mathbf{p}\mid\mathbf{q})\ \mathbf{M}^{-,-}_{1,2}(\mathbf{p}\mid\mathbf{q})\ \mathbf{T}(\mathbf{q}\mid\mathbf{p}_0)\ \frac{\mathrm{d}^2q}{(2\pi)^2}$$
$$=\frac{2\sqrt{\epsilon_1\epsilon_2}\alpha_1(\mathbf{p}_0)}{\epsilon_2-\epsilon_1}\,(2\pi)^2\delta(\mathbf{p}-\mathbf{p}_0)\,, \tag{2.71}$$

and by making the change of variable

$$\begin{aligned}\Delta\,\mathbf{T}(\mathbf{p}\mid\mathbf{p}_0)&=\mathbf{T}(\mathbf{p}\mid\mathbf{p}_0)-\frac{2\sqrt{\epsilon_1\epsilon_2}\alpha_1(\mathbf{p}_0)}{\epsilon_2-\epsilon_1}(\alpha_2(\mathbf{p}_0)-\alpha_1(\mathbf{p}_0))\left[\mathbf{M}^{-,-}_{1,2}\,(\mathbf{p}_\mathrm{o}\,|\,\mathbf{p}_\mathrm{o})\right]^{-1}\,(2\pi)^2\,\delta(\mathbf{p}-\mathbf{p}_0)\\&=\mathbf{T}(\mathbf{p}\mid\mathbf{p}_0)-(2\pi)^2\,\delta(\mathbf{p}-\mathbf{p}_0)\,\boldsymbol{\tau}^{(0)}(\mathbf{p}_0)\,,\end{aligned} \tag{2.72}$$

we obtain the equation for $\Delta\,\mathbf{T}(\mathbf{p}\mid\mathbf{p}_0)$

$$(\alpha_2(\mathbf{p})-\alpha_1(\mathbf{p}))^{-1}\ \mathbf{M}^{-,-}_{12}(\mathbf{p}\mid\mathbf{p})\,\Delta\,\mathbf{T}(\mathbf{p}\mid\mathbf{p}_0)+\int \mathcal{K}^{-,-}_{12}(\mathbf{p}\mid\mathbf{q})\ \mathbf{M}^{-,-}_{12}(\mathbf{p}\mid\mathbf{q})\,\Delta\,\mathbf{T}(\mathbf{q}\mid\mathbf{p}_0)\ \frac{\mathrm{d}^2q}{(2\pi)^2}$$
$$=-\,\mathcal{K}^{-,-}_{12}(\mathbf{p}\mid\mathbf{p}_0)\ \mathbf{M}^{-,-}_{12}(\mathbf{p}\mid\mathbf{p}_0)\,\boldsymbol{\tau}^{(0)}(\mathbf{p}_0)\,. \tag{2.73}$$

Depending on the context we may refer to Eqs. (2.68) and (2.71) or equivalently Eqs. (2.70) and (2.73) to the *reduced Rayleigh equations of the second kind*. The associated fixed point algorithm will prove to be a powerful numerical tool (see Chapters 6 and 8).

2.5.2 Solution and physical interpretation

We now briefly comment on the physical interpretation of the solution one would obtain by iterating infinitely many times the aforementioned algorithm. For the reflection amplitude, the first iterate reads

$$\begin{aligned}\Delta\,\mathbf{R}^{(1)}(\mathbf{p}\mid\mathbf{p}_0)=&(\alpha_1(\mathbf{p})-\alpha_2(\mathbf{p}))\ \left[\mathbf{M}^{++}_{21}(\mathbf{p}\mid\mathbf{p})\right]^{-1}\\&\left(\mathcal{K}^{++}_{21}(\mathbf{p}\mid\mathbf{p}_0)\ \mathbf{M}^{++}_{21}(\mathbf{p}\mid\mathbf{p}_0)\,\boldsymbol{\rho}^{(0)}(\mathbf{p}_0)+\mathcal{K}^{+-}_{21}(\mathbf{p}\mid\mathbf{p}_0)\ \mathbf{M}^{+-}_{21}(\mathbf{p}\mid\mathbf{p}_0)\right),\end{aligned} \tag{2.74}$$

and the following iterates can be recast as

$$\Delta\,\mathbf{R}^{(k+1)}(\mathbf{p}\mid\mathbf{p}_0)=\Delta\,\mathbf{R}^{(1)}(\mathbf{p}\mid\mathbf{p}_0)+\int \mathbf{D}^{++}_{21}(\mathbf{p}\mid\mathbf{q})\,\Delta\,\mathbf{R}^{(k)}(\mathbf{q}\mid\mathbf{p}_0)\ \frac{\mathrm{d}^2q}{(2\pi)^2}\,, \tag{2.75}$$

where we defined the *reflection scattering kernel*

$$\mathbf{D}^{++}_{21}(\mathbf{p}\mid\mathbf{q})=(\alpha_1(\mathbf{p})-\alpha_2(\mathbf{p}))\ \left[\mathbf{M}^{++}_{21}(\mathbf{p}\mid\mathbf{p})\right]^{-1}\mathcal{K}^{++}_{21}(\mathbf{p}\mid\mathbf{q})\ \mathbf{M}^{++}_{21}(\mathbf{p}\mid\mathbf{q})\,. \tag{2.76}$$

Let us now expand the first few iterates and express them as a function of $\Delta\,\mathbf{R}^{(1)}(\mathbf{p}\mid\mathbf{p}_0)$ only. We have

$$\Delta\,\mathbf{R}^{(2)}(\mathbf{p}\mid\mathbf{p}_0)=\Delta\,\mathbf{R}^{(1)}(\mathbf{p}\mid\mathbf{p}_0)+\int \mathbf{D}^{++}_{21}(\mathbf{p}\mid\mathbf{q})\,\Delta\,\mathbf{R}^{(1)}(\mathbf{q}\mid\mathbf{p}_0)\ \frac{\mathrm{d}^2q}{(2\pi)^2}\,, \tag{2.77a}$$

$$\begin{aligned}\Delta\,\mathbf{R}^{(3)}(\mathbf{p}\mid\mathbf{p}_0)=&\Delta\,\mathbf{R}^{(1)}(\mathbf{p}\mid\mathbf{p}_0)+\int \mathbf{D}^{++}_{21}(\mathbf{p}\mid\mathbf{q})\,\Delta\,\mathbf{R}^{(1)}(\mathbf{q}\mid\mathbf{p}_0)\ \frac{\mathrm{d}^2q}{(2\pi)^2}\\&+\int\int \mathbf{D}^{++}_{21}(\mathbf{p}\mid\mathbf{q})\,\mathbf{D}^{++}_{21}(\mathbf{q}\mid\mathbf{q}')\,\Delta\,\mathbf{R}^{(1)}(\mathbf{q}'\mid\mathbf{p}_0)\ \frac{\mathrm{d}^2q'}{(2\pi)^2}\,\frac{\mathrm{d}^2q}{(2\pi)^2}\,.\end{aligned} \tag{2.77b}$$

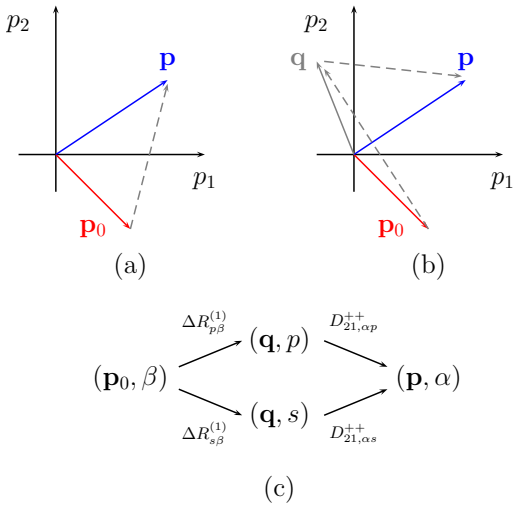


Figure 2.2: Sketches of (a) a single scattering leap from $\mathbf{p}_0$ to $\mathbf{p}$, (b) a two-scattering event diagram showing a first leap from $\mathbf{p}_0$ to an intermediate in-plane wave vector $\mathbf{q}$ and a second leap from $\mathbf{q}$ to $\mathbf{p}$. (c) Illustration of elementary transitions of probability amplitudes between the modes involved in the scattering events illustrated in (b).

We now start to see that by recursion we can show that as $k \to \infty$ the solution of the reduced Rayleigh equation for the deviation $\Delta\,\mathbf{R}$ can be written as

$$\Delta\,\mathbf{R}(\mathbf{p} \mid \mathbf{p}_0) = \sum_{k=0}^{\infty} \left\langle \boldsymbol{\mathcal{D}}_{21}^{++,k}(\mathbf{p}), \Delta\,\mathbf{R}^{(1)}(\cdot \mid \mathbf{p}_0) \right\rangle , \tag{2.78}$$

where the angle bracket denote the action of the linear functional $\boldsymbol{\mathcal{D}}_{21}^{++,k}(\mathbf{p})$ parametrized by $\mathbf{p}$ and which is the k^{th} power, in the sense of composition, of the functional $\boldsymbol{\mathcal{D}}_{21}^{++}(\mathbf{p})$ defined by the kernel $\mathbf{D}_{21}^{++}$. Explicitly, we have for any 2×2 matrix valued function $\mathbf{F} : \mathbf{q} \in \mathbb{R}^2 \mapsto \mathbf{F}(\mathbf{q}) \in \mathfrak{M}_2(\mathbb{C})$,

$$\left\langle \boldsymbol{\mathcal{D}}_{21}^{++,k}(\mathbf{p}), \mathbf{F} \right\rangle \stackrel{\text{def}}{=} \int \cdots \int \mathbf{D}_{21}^{++}(\mathbf{p} \mid \mathbf{p}_k)\, \mathbf{D}_{21}^{++}(\mathbf{p}_k \mid \mathbf{p}_{k-1}) \cdots \mathbf{D}_{21}^{++}(\mathbf{p}_2 \mid \mathbf{p}_1)\; \mathbf{F}(\mathbf{p}_1)\; \frac{\mathrm{d}^2 p_1}{(2\pi)^2} \cdots \frac{\mathrm{d}^2 p_k}{(2\pi)^2} . \tag{2.79}$$

By convention we set $\boldsymbol{\mathcal{D}}_{21}^{++,0}(\mathbf{p})$ to be the Dirac mass centered at $\mathbf{p}$, i.e. $\left\langle \boldsymbol{\mathcal{D}}_{21}^{++,0}(\mathbf{p}), \mathbf{F} \right\rangle \stackrel{\text{def}}{=} \mathbf{F}(\mathbf{p})$.

Remark 2.6. We have obtained in Eq. (2.79) what is known in the theory of integral equations as the Liouville-Neumann, or simply Neumann series expansion [57] of the solution of the Fredholm integral equation of the second kind Eq. (2.70). Physicists will certainly recognize what they call a Born series.

The physical interpretation of Eq. (2.78) is that the first iterate (Born approximation) corresponds to the reflection amplitude obtained by summing all possible single scattering path since Eq. (2.74) only involves the incident in-plane wave vector $\mathbf{p}_0$ and the outgoing in-plane wave

vector $\mathbf{p}$. This could be represented by a diagram where the in-plane wave vector is incoming with value $\mathbf{p}_0$ and makes a single leap to $\mathbf{p}$. Then the second iterate for $\Delta\,\mathbf{R}$, which reads

$$\Delta\,\mathbf{R}^{(2)}(\mathbf{p}\mid\mathbf{p}_0)=\Delta\,\mathbf{R}^{(1)}(\mathbf{p}\mid\mathbf{p}_0)+\int\,\mathbf{D}_{21}^{++}(\mathbf{p}\mid\mathbf{q})\,\Delta\,\mathbf{R}^{(1)}(\mathbf{q}\mid\mathbf{p}_0)\;\frac{\mathrm{d}^2q}{(2\pi)^2}\,, \tag{2.80}$$

is the sum of the single scattering diagram and a diagram representing the sum of all possible two-scattering events path starting from the in-plane wave vector $\mathbf{p}_0$ and outgoing with $\mathbf{p}$ but with any intermediate in-plane wave vector $\mathbf{q}$ as suggested by the integral over $\mathbf{q}$. The remarkable point here is that the two-scattering events term is constructed as applying the scattering kernel to the single scattering term. The scattering kernel can be interpreted as a *transition probability amplitude* for scattering from $\mathbf{q}$ to $\mathbf{p}$ (and for corresponding polarization coupling involved) and this independently of the prior history of the path. To be more precise, given an incident polarization state β with in-plane wave vector $\mathbf{p}_0$ and for a fixed outgoing polarization state $\alpha\in\{p,s\}$ with in-plane wave vector $\mathbf{p}$, the integrand in Eq. (2.80) reads

$$D_{21,\alpha p}^{++}(\mathbf{p}\mid\mathbf{q})\Delta R_{p\beta}^{(1)}(\mathbf{q}\mid\mathbf{p}_0)+D_{21,\alpha s}^{++}(\mathbf{p}\mid\mathbf{q})\Delta R_{s\beta}^{(1)}(\mathbf{q}\mid\mathbf{p}_0)\,,$$

which can be interpreted as an elementary flow of probability amplitude coming from states $(\mathbf{q},p)$ and $(\mathbf{q},s)$ into state $(\mathbf{p},\alpha)$. Indeed, for higher multiple scattering events, each term in Eq. (2.78) is of the form given by Eq. (2.79) which reminds of a Chapman-Kolmogorov equation with transition probabilities for Markovian processes [58]. Whether this is a general property of the scattering of light by any surfaces or if it is only a property of the reduced Rayleigh equations, we must keep in mind that the solution given in Eq. (2.78) is only valid under the assumption of convergence of the series of scattering diagrams. The analysis of the condition for the convergence of this series, probably requiring tools from functional analysis and would essentially analyze under which conditions the linear functional defined by the scattering kernel is contractant, would be an important result both theoretically and numerically. Indeed, such a result would decide whether the algorithm built around Eq. (2.70) would converge and how fast, if it does. The convergence of the series Eq. (2.78) will be studied in details for sinusoidal surfaces as a function of physical parameters such as the wavelength, the lattice constant and the surface profile amplitude in Chapter 6. The convergence of the method will also be illustrated in Chapter 8 for weakly rough dielectric surfaces. We will refer to the above method, based on the iteration of Eq. (2.70), as the method of *Fixed point Iterates of the reduced Rayleigh Equation of the Second kind*, or in short, method of FIRES. We have chosen this denomination in order to stress what the method does and that the starting point is the reduced Rayleigh equations of the second kind. In particular, we make this distinction so that no confusion is made with the conventional Born approximation applied directly on the volume integral representation, and to make a distinction with the similar method but starting from the coupled Rayleigh equations (see Section 2.4).

For the transmission amplitude, the first iterate (starting from $\Delta\,\mathbf{T}^{(0)}(\mathbf{p}\mid\mathbf{p}_0)=0$) reads

$$\begin{aligned}\Delta\,\mathbf{T}^{(1)}(\mathbf{p}\mid\mathbf{p}_0)&=(\alpha_1(\mathbf{p})-\alpha_2(\mathbf{p}))\;\left[\mathbf{M}_{12}^{-;-}(\mathbf{p}\mid\mathbf{p})\right]^{-1}\;\mathcal{K}_{12}^{-;-}(\mathbf{p}\mid\mathbf{p}_0)\;\mathbf{M}_{12}^{-;-}(\mathbf{p}\mid\mathbf{p}_0)\,\boldsymbol{\tau}^{(0)}(\mathbf{p}_0)\\&=\mathcal{K}_{12}^{-;-}(\mathbf{p}\mid\mathbf{p}_0)\,\boldsymbol{\tau}\,(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}})\;.\end{aligned} \tag{2.81}$$

The transmission scattering kernel reads

$$\mathbf{D}_{12}^{-;-}(\mathbf{p}\mid\mathbf{q})=(\alpha_1(\mathbf{p})-\alpha_2(\mathbf{p}))\;\left[\mathbf{M}_{12}^{-;-}(\mathbf{p}\mid\mathbf{p})\right]^{-1}\mathcal{K}_{12}^{-;-}(\mathbf{p}\mid\mathbf{q})\;\mathbf{M}_{12}^{-;-}(\mathbf{p}\mid\mathbf{q})\,, \tag{2.82}$$

and the solution for $\Delta\,\mathbf{T}(\mathbf{p} \mid \mathbf{p}_0)$ as a series of multiple scattering diagrams reads

$$\Delta\,\mathbf{T}(\mathbf{p} \mid \mathbf{p}_0) = \sum_{k=0}^{\infty} \left\langle \boldsymbol{\mathcal{D}}_{12}^{-,-,k}(\mathbf{p}), \Delta\,\mathbf{T}^{(1)}(\cdot \mid \mathbf{p}_0) \right\rangle \,, \tag{2.83}$$

with the obvious conventions analogous to the case of reflection.

2.6 Reduced Rayleigh equations for scalar waves

2.6.1 Motivation and context

In this section, we derive the reduced Rayleigh equations for scalar waves. The derivation of the reduced Rayleigh equations in this case is significantly simpler than that of electromagnetic waves but, apart from the technical issues related to polarization, the derivation is essentially identical. Why do we consider scalar waves now that we have the equations for electromagnetic waves? The motivation is two-fold. First, even though the equations for scalar waves do not fully represent the physics of electromagnetic waves, they give useful approximations and physical insights. In particular, the comparison between phenomena (or absence of phenomena) observed when studying scalar waves and electromagnetic waves can give important clues on the types of phenomena which fundamentally arise as a consequence of the polarization of electromagnetic waves or if they can be explained based on simpler scalar considerations. Such a distinction will be necessary when studying the Yoneda and Brewster scattering effect in Chapter 10 for example. Second, we would like to extend the framework to other types of waves that may be described by the Helmholtz equation, such as acoustic waves and quantum matter waves for example, within some assumptions. To fix the idea, consider the case of the scattering of a quantum particle of mass m, represented by a wave function $\psi : (\mathbf{r}, t) \in \mathbb{R}^3 \times \mathbb{R} \mapsto \psi(\mathbf{r}, t) \in \mathbb{C}$, by a surface separating two regions of constant and distinct potentials V_1 and V_2. We know that for a time independent Hamiltonian, the solution of the Schrödinger equation

$$i\hbar \frac{\partial \psi}{\partial t} = \left[-\frac{\hbar^2}{2m}\,\Delta + V \right] \psi \,, \tag{2.84}$$

with initial condition $\psi(\mathbf{r}, 0) = \psi_0(\mathbf{r})$, is given by

$$\psi(\mathbf{r}, t) = \sum_{\alpha} \langle \psi_\alpha, \psi_0 \rangle \, \exp\left(-i \frac{E_\alpha t}{\hbar} \right) \, \psi_\alpha(\mathbf{r}) \,, \tag{2.85}$$

where $\langle f, g \rangle \overset{\text{def}}{=} \int_{\mathbb{R}^3} f^*(\mathbf{r}) g(\mathbf{r}) \mathrm{d}^3 r$ is an inner-product in $\mathcal{L}^2(\mathbb{R}^3; \mathbb{C})$, and the (E_α, ψ_α) are solutions of the eigenvalue problem (counted with their degeneracy)

$$\left[-\frac{\hbar^2}{2m}\,\Delta + V \right] \psi_\alpha = E_\alpha \psi_\alpha \,. \tag{2.86}$$

Note that the sum over α may be discrete or continuous depending on the problem. The above problem is then solved by diagonalizing the Hamiltonian (i.e. solving Eq. (2.86)) which, under the assumption that we are looking for unbounded modes, or scattering modes, for which we have a continuous spectrum for $E_\alpha > V_1$ and $E_\alpha > V_2$, can be re-written as a Helmholtz equation

$$\Delta \psi + \epsilon \frac{\omega_{\text{eff}}^2}{c^2} \psi = 0 \,, \tag{2.87}$$

where we now drop the index α and define $\omega_{\text{eff}}^2 / c^2 = 2mE/\hbar^2$, and $\epsilon = 1 - V/E$.

Remark 2.7. Here we have simply made an arbitrary identification between the energy E and the effective angular frequency ω_{eff}. It should not be confused with the usual relation $E = \hbar\omega$ of a de Broglie wave for example. Indeed, it was expected that one should not get a linear relation between the effective frequency of the Helmholtz equation and the energy because the Schrödinger equation is a partial differential equation of order one in time while the classical wave equation is of order two in time. It is here only a mathematically convenient definition which helps us recover a familiar Helmholtz equation. In fact, we could have chosen another convention, for example $\omega_{\text{eff}}^2/c^2 = 2m(E - V_{\min})/\hbar^2$, and $\epsilon = 1 - (V - V_{\min})/(E - V_{\min})$, with $V_{\min} = \min(V_1, V_2)$, so that the effective dielectric constant is unity in the region of lowest potential for example. However, note that the consequence of such a definition for the effective angular frequency which is not consistent with a linear frequency-energy relation implies that a plane wave of frequency ω_{eff} is *not* a de Broglie wave with energy $E = \hbar\omega_{\text{eff}}$. One should then be careful about the physical interpretation and always go back to expressions with the energy in the end.

2.6.2 RRE for scalar waves

For the scattering of a quantum particle, the boundary conditions at the potential interface are given by the continuity of the wave function and its normal derivative with respect to the interface

$$\lim_{\eta\to 0} \left[\psi(\mathbf{s}(\mathbf{x}) + \eta\,\mathbf{n}(\mathbf{x})) - \psi(\mathbf{s}(\mathbf{x}) - \eta\,\mathbf{n}(\mathbf{x}))\right] = 0 \tag{2.88a}$$

$$\lim_{\eta\to 0}\ \mathbf{n}(\mathbf{x}) \cdot \left[\nabla\psi(\mathbf{s}(\mathbf{x}) + \eta\,\mathbf{n}(\mathbf{x})) - \nabla\psi(\mathbf{s}(\mathbf{x}) - \eta\,\mathbf{n}(\mathbf{x}))\right] = 0\,. \tag{2.88b}$$

Similarly to what was done in the case of electromagnetic wave the wave function ψ may be written as a sum of plane waves in each region of constant potential as

$$\psi_j(\mathbf{r}) = \sum_{a=\pm} \int_{\mathbb{R}^2} \psi_j^a(\mathbf{q})\, \exp(i\,\mathbf{k}_j^a(\mathbf{q}) \cdot \mathbf{r})\, \frac{\mathrm{d}^2 q}{(2\pi)^2} \tag{2.89}$$

Assuming that Eq. (2.89) is valid close to the interface of the step of potential, the continuity of the field and its normal derivative with respect to the interface yields

$$\sum_{a=\pm} \int \psi_1^a(\mathbf{q})\, \exp(i\,\mathbf{k}_1^a(\mathbf{q}) \cdot \mathbf{s}(\mathbf{x}))\, \frac{\mathrm{d}^2 q}{(2\pi)^2} = \sum_{a=\pm} \int \psi_2^a(\mathbf{q})\, \exp(i\,\mathbf{k}_2^a(\mathbf{q}) \cdot \mathbf{s}(\mathbf{x}))\, \frac{\mathrm{d}^2 q}{(2\pi)^2} \tag{2.90a}$$

$$\sum_{a=\pm} \int \psi_1^a(\mathbf{q})\, \exp(i\,\mathbf{k}_1^a(\mathbf{q}) \cdot \mathbf{s}(\mathbf{x}))\, \mathbf{n}(\mathbf{x}) \cdot \mathbf{k}_1^a(\mathbf{q})\, \frac{\mathrm{d}^2 q}{(2\pi)^2} = \sum_{a=\pm} \int \psi_2^a(\mathbf{q})\, \exp(i\,\mathbf{k}_2^a(\mathbf{q}) \cdot \mathbf{s}(\mathbf{x}))\, \mathbf{n}(\mathbf{x}) \cdot \mathbf{k}_2^a(\mathbf{q})\, \frac{\mathrm{d}^2 q}{(2\pi)^2} \tag{2.90b}$$

By multiplying Eq. (2.90a) by $\mathbf{n}(\mathbf{x}) \cdot \mathbf{k}_2^b(\mathbf{p}) \exp(-i\,\mathbf{k}_2^b(\mathbf{p}) \cdot \mathbf{s}(\mathbf{x}))$ and Eq. (2.90b) by $\exp(-i\,\mathbf{k}_2^b(\mathbf{p}) \cdot \mathbf{s}(\mathbf{x}))$, where $\mathbf{p} = p_1\,\hat{\mathbf{e}}_1 + p_2\,\hat{\mathbf{e}}_2$ is an arbitrary in-plane wave vector and $b \in \{\pm\}$ is arbitrarily chosen, and finally summing the two resulting equations and integrating over $\mathbf{x}$ we obtain

$$\sum_{a=\pm} \int \psi_1^a(\mathbf{q})\ \mathbf{w}_{21}^{ba}(\mathbf{p} \mid \mathbf{q}) \cdot [\mathbf{k}_2^b(\mathbf{p}) + \mathbf{k}_1^a(\mathbf{q})]\, \frac{\mathrm{d}^2 q}{(2\pi)^2} = \sum_{a=\pm} \int \psi_2^a(\mathbf{q})\ \mathbf{w}_{22}^{ba}(\mathbf{p} \mid \mathbf{q}) \cdot [\mathbf{k}_2^b(\mathbf{p}) + \mathbf{k}_2^a(\mathbf{q})]\, \frac{\mathrm{d}^2 q}{(2\pi)^2}\,. \tag{2.91}$$

The definition of $\mathbf{w}_{lm}^{ba}$ was given in Eq. (2.18) in the derivation of the reduced Rayleigh equations in the case of electromagnetic waves. Following the same steps as in the derivation of the transfer equations in the case of electromagnetic waves, we find without difficulty the *transfer equations* for scalar waves subjected to the continuity of the field and its normal derivative with respect to the interface

$$\sum_{a=\pm}\int \left(\epsilon_2-\epsilon_1\right)\frac{\omega^2}{c^2}\,\mathcal{J}_{2,1}^{b,a}\left(\mathbf{p}\,|\,\mathbf{q}\right)\,\psi_1^a\left(\mathbf{q}\right)\frac{\mathrm{d}^2q}{(2\pi)^2}=2\,b\,\alpha_2\left(\mathbf{p}\right)\,\psi_2^b\left(\mathbf{p}\right)\,, \tag{2.92a}$$

$$\sum_{a=\pm}\int \left(\epsilon_1-\epsilon_2\right)\frac{\omega^2}{c^2}\,\mathcal{J}_{1,2}^{b,a}\left(\mathbf{p}\,|\,\mathbf{q}\right)\,\psi_2^a\left(\mathbf{q}\right)\frac{\mathrm{d}^2q}{(2\pi)^2}=2\,b\,\alpha_1\left(\mathbf{p}\right)\,\psi_1^b\left(\mathbf{p}\right)\,. \tag{2.92b}$$

Here the definition of $\mathcal{J}_{l,m}^{b,a}$ was given in Eq. (2.21). The transfer equations for the scalar waves subjected to the aforementioned boundary conditions are similar to the transfer equations for electromagnetic waves but replacing the matrix of change of polarization basis $\mathbf{P}_{l,m}^{b,a}$ by the scalar 1 (compare e.g. Eqs. (2.43) and (2.92a)).

Remark 2.8. We would like to stress that the above transfer equations for scalar wave are only valid for the boundary conditions imposing the continuity of the field and its normal derivative when crossing the interface. Such boundary conditions are relevant for the scattering of a quantum particle by a potential, for example. However, one must keep in mind that for the problem of scattering of acoustic waves, the boundary conditions would be different. Consequently, for a problem of scattering of scalar wave with different boundary conditions than the one chosen above, the corresponding transfer equations would read in general as

$$\sum_{a=\pm}\int \mathcal{J}_{2,1}^{b,a}\left(\mathbf{p}\,|\,\mathbf{q}\right)\,F_{2,1}^{b,a}(\mathbf{p}\,\mid\,\mathbf{q})\,\psi_1^a\left(\mathbf{q}\right)\frac{\mathrm{d}^2q}{(2\pi)^2}=2\,b\,\alpha_2\left(\mathbf{p}\right)\,\psi_2^b\left(\mathbf{p}\right)\,, \tag{2.93a}$$

$$\sum_{a=\pm}\int \mathcal{J}_{1,2}^{b,a}\left(\mathbf{p}\,|\,\mathbf{q}\right)\,F_{1,2}^{b,a}(\mathbf{p}\,\mid\,\mathbf{q})\,\psi_2^a\left(\mathbf{q}\right)\frac{\mathrm{d}^2q}{(2\pi)^2}=2\,b\,\alpha_1\left(\mathbf{p}\right)\,\psi_1^b\left(\mathbf{p}\right)\,, \tag{2.93b}$$

where F_{lm}^{ba} is a scalar function which depends on the boundary conditions. In the concrete example of the acoustic scattering problem from and through a penetrable interface, the problem is modeled by the following equations[4],

$$\frac{\partial\,\mathbf{v}}{\partial t}+\frac{1}{\rho_j}\,\nabla\psi=0 \tag{2.94a}$$

$$\frac{\partial\psi}{\partial t}+c_j^2\rho_j\,\nabla\cdot\mathbf{v}=0\,, \tag{2.94b}$$

where $\mathbf{v}$, ψ, are respectively the velocity field and pressure variation of the fluid, and the constants ρ_j, c_j are respectively the average density and speed of sound in medium j. By taking the time derivative of Eq. (2.94b) and substituting Eq. (2.94a) into it, we obtain the wave equation

$$\frac{\partial^2\psi}{\partial t^2}-c_j^2\,\Delta\psi=0\,. \tag{2.95}$$

[4]The system of equations, Eq. (2.94), is obtained from a linearization of Euler's equations.

It is common to assume *the continuity of the pressure and the continuity of the normal velocity across the interface*, i.e.

$$\lim_{\eta\to 0}\left[\psi(\mathbf{s}(\mathbf{x})+\eta\,\mathbf{n}(\mathbf{x}))-\psi(\mathbf{s}(\mathbf{x})-\eta\,\mathbf{n}(\mathbf{x}))\right]=0 \tag{2.96a}$$

$$\lim_{\eta\to 0}\ \mathbf{n}(\mathbf{x})\cdot\left[\frac{1}{\rho_1}\,\nabla\psi(\mathbf{s}(\mathbf{x})+\eta\,\mathbf{n}(\mathbf{x}))-\frac{1}{\rho_2}\,\nabla\psi(\mathbf{s}(\mathbf{x})-\eta\,\mathbf{n}(\mathbf{x}))\right]=0\,. \tag{2.96b}$$

In such a case, we leave to the reader to show that for acoustic waves we have

$$\begin{aligned}F_{lm}^{ba}(\mathbf{p}\,|\,\mathbf{q}) &= \rho_2\,\left[\mathbf{k}_2^b(\mathbf{p})-\mathbf{k}_1^a(\mathbf{q})\right]\cdot\left[\rho_2^{-1}\,\mathbf{k}_2^b(\mathbf{p})-\rho_1^{-1}\,\mathbf{k}_1^a(\mathbf{q})\right]\\ &= \rho_2\,\left(\frac{\epsilon_2}{\rho_2}-\frac{\epsilon_1}{\rho_1}\right)\,\frac{\omega^2}{c^2}+\left(1-\frac{\rho_2}{\rho_1}\right)\mathbf{k}_2^b(\mathbf{p})\cdot\mathbf{k}_1^a(\mathbf{q})\,,\end{aligned} \tag{2.97}$$

where we have set an arbitrary reference speed $c>0$ and defined a effective dielectric constant $\epsilon_j=c^2/c_j^2$. Other types of boundary conditions are of course possible and the factor F_{lm}^{ba} must be adapted accordingly by re-deriving the reduced Rayleigh equations each time. Note that in the limit case $\rho_2=\rho_1$ the boundary conditions are those of continuity of the field and its normal derivative with respect to the surface and that we indeed recover the corresponding transfer equations derived in Eqs. (2.92a) and (2.92b) since

$$F_{lm}^{ba}(\mathbf{p}\,|\,\mathbf{q})\big|_{\rho_2=\rho_1}=\left(\epsilon_2-\epsilon_1\right)\,\frac{\omega^2}{c^2}\,. \tag{2.98}$$

Considering now the scattering of an plane wave incident in the potential region 1, $\psi_0(\mathbf{r})=\Psi_0\exp(i\,\mathbf{k}_1^-(\mathbf{p}_0)\cdot\mathbf{r})$, and assuming that the only waves allowed to be scattered are those of with wave vector directed upward ($a=+$) in the potential region 1 and downward ($a=-$) in the potential region 2, we can write

$$\psi_1^-(\mathbf{q})=\Psi_0\,(2\pi)^2\delta(\mathbf{q}-\mathbf{p}_0) \tag{2.99a}$$

$$\psi_2^+(\mathbf{q})=0 \tag{2.99b}$$

$$\psi_1^+(\mathbf{q})=R(\mathbf{q}\mid\mathbf{p}_0)\,\Psi_0 \tag{2.99c}$$

$$\psi_2^-(\mathbf{q})=T(\mathbf{q}\mid\mathbf{p}_0)\,\Psi_0\,. \tag{2.99d}$$

Plugging the above equations in Eqs. (2.92a) and (2.92b) respectively applied for $b=+$ and $b=-$, we obtain the following reduced Rayleigh equations

$$\int\,\mathcal{J}_{2,1}^{+,+}\,(\mathbf{p}\,|\,\mathbf{q})\;R(\mathbf{q}\mid\mathbf{p}_0)\,\frac{\mathrm{d}^2q}{(2\pi)^2}=-\,\mathcal{J}_{2,1}^{+,-}\,(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}})\;, \tag{2.100a}$$

$$\int\,\mathcal{J}_{1,2}^{-,-}\,(\mathbf{p}\,|\,\mathbf{q})\;T(\mathbf{q}\mid\mathbf{p}_0)\,\frac{\mathrm{d}^2q}{(2\pi)^2}=\frac{2\,\alpha_1\,(\mathbf{p})\,c^2}{(\epsilon_1-\epsilon_2)\,\omega^2}\,(2\pi)^2\delta(\mathbf{p}-\mathbf{p}_0)\,. \tag{2.100b}$$

The corresponding equations in the case of other types of boundary conditions are immediately obtained in view of Remark 2.8.

Remark 2.9. In the rest of this work, when deriving approximate solutions of the reduced Rayleigh equations, we will always treat the case of electromagnetic waves. In order to obtain the corresponding approximate solutions for scalar waves it will suffices to replace the matrices $\mathbf{M}_{l,m}^{b,a}$ by the scalar factor for the chosen boundary conditions, which may simply reduce to a constant for the case of scattering by a quantum particle.

2.7 Summary

It is time to summarize what has be done is the present chapter. We have derived the theory of the reduced Rayleigh equations for electromagnetic waves as well as for scalar waves subjected to different types of boundary conditions in a unified framework. The reduced Rayleigh equations are decoupled integral equations for the reflection and transmission amplitudes in which the right-hand side, or source term, encodes the incident field. Maybe more fundamental than the reduced Rayleigh equations are what we have called the transfer equations which are integral relations linking the Fourier amplitudes of the fields below and above the interfaces. The transfer equations will be particularly important to generalize the framework to multilayer systems. Furthermore, by considering the deviation of the solution of the reduced Rayleigh equations from the solution for a planar surface, we have obtained Fredholm integral equations of the second type, which we have named the reduced Rayleigh equations of the second type. The latter equations allow for a Born series expansion of the solution to the reduced Rayleigh equations in terms of multiple scattering events diagrams. We will see later that the iterative procedure for obtaining the series expansion will lead to a powerful numerical scheme for solving the reduced Rayleigh equations.

Chapter 3

Multilayer systems

In this chapter, we will generalize the formalism of the reduced Rayleigh equations to multilayer systems. The transfer equations Eqs. (2.44) and (2.47) derived in Chapter 2 will serve as building blocks to derive in a systematic way the reduced Rayleigh equations for a stack of arbitrary many rough layers. The system of sub- and superscripts introduced in Chapter 2, which might seem more pedantic than useful so far, will reveal here their power in composition rules.

We will derive *transfer equations* between any two layers in the stack which relate the field amplitudes in each of the two chosen layers without explicitly using the field amplitudes in intermediate layers. In particular, by applying the transfer equations to the whole stack, we will deduce the integral equations satisfied by the reflection and transmission amplitudes for the scattering by the whole system, namely the reduced Rayleigh equations. The transfer equations keep the same integral equation structure as for the case of a single interface but the corresponding *multi-interface transfer kernel* is now defined via multiple integral composition of intermediate single-interface transfer kernels.

Although the derivation of the reduced Rayleigh equations for a multilayer system is rather straightforward, two types of numerical issues are to be expected when it comes to solving them. First, it is quite common to encounter numerical instabilities for film systems if no care is taken. Indeed, we expect from a physical point of view the existence of exponentially growing modes inside the film. These modes should not, in theory, represent any problem since there are modes which are growing but are bounded within the film. Nevertheless, exponentially growing modes are known to lead to instabilities when it comes to numerics. Second, the composition rule to obtain the multi-interface transfer kernels significantly increases the complexity for setting up the numerical linear system associated with the reduced Rayleigh equations. We will see that already for a system containing two rough interfaces the computation of the transfer kernel represents a challenge in itself.

Taking the example of a system containing two interfaces, we will show that the complexity associated with a system where only one of the interface is rough and the other planar is essentially the same as a system made of a single rough interface. In such a case, we will also explain in details how the equations can be recast in a form which is numerically stable. Unfortunately, constructing a robust and low complexity algorithm for the general case where, for example, two interfaces are rough has not yet been achieved, and is an ongoing research topic. We will suggest potential tracks, based on the reduced Rayleigh equations of the second kind,

and the probability interpretation associated with it, which may cure the curse of dimensionality.

3.1 The system

We consider a system composed of $n+1$ layers separated by n interfaces ($n \in \mathbb{N}^*$). The media are indexed from 1 to $n+1$, where the first and $n+1$th media are semi-infinite in the x_3-direction. We assume that each interface admits an average plane parallel to $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2)$ and are bounded in the $\hat{\mathbf{e}}_3$-direction. The i^{th} interface ($i \in [\![1, n]\!]$), separating the media i and $i+1$, admits an equation of the form

$$x_3 = \zeta_i(\mathbf{x}) \ , \tag{3.1}$$

for $\mathbf{x}$ in the $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2)$ plane. We also assume that all interfaces are disjoint and ordered, i.e. $\forall\, i \in [\![1, n-1]\!], \zeta_{i+1} < \zeta_i$, as shown in Fig. 3.1.

We are interested in finding integral equations satisfied by the reflection and transmission amplitudes, $\mathbf{R}(\mathbf{p} \,|\, \mathbf{p}_0)$ and $\mathbf{T}(\mathbf{p} \,|\, \mathbf{p}_0)$, after interaction with the whole system. If we come back to Eq. (2.44), that we may write for the fields on both sides of the first interface for example, we now need to know the upwards propagating field in the first slab, which does not vanish anymore since a set of subsequent interfaces also scatter back the transmitted light. The idea is to successively eliminate all the fields in the intermediate interfaces and reach the last medium (which is assumed semi-infinite).

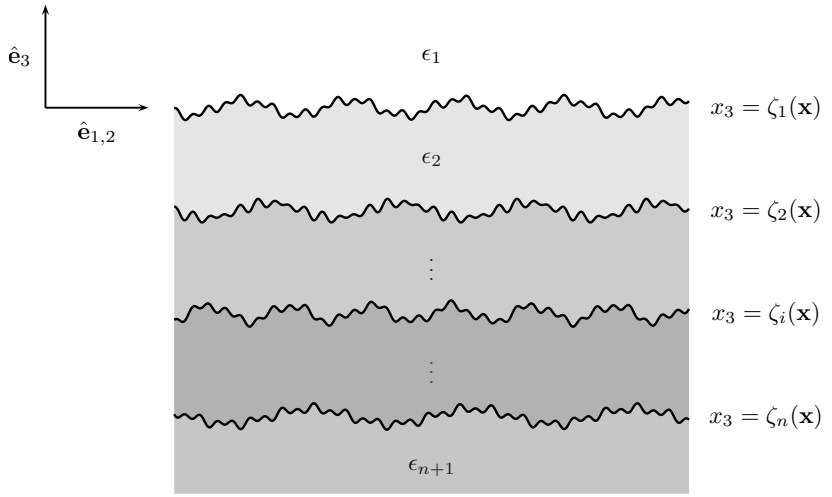


Figure 3.1: Illustration of a multilayer system and indexation of the media and the surface profiles.

3.2 The transfer equations

We assume that the derivation of the transfer equations carried out in Chapter 2 is still valid for each interface, i.e. we assume that the electric field in each layer may be expanded in plane

waves. For each interface we can write Eq. (2.44), which reads for $i \in [\![1, n]\!]$,

$$\sum_{a_i=\pm} \int \mathcal{J}_{i+1,i}^{a_{i+1},a_i}(\mathbf{p}_{i+1} \,|\, \mathbf{p}_i)\ \mathbf{M}_{i+1,i}^{a_{i+1},a_i}(\mathbf{p}_{i+1} \,|\, \mathbf{p}_i)\ \boldsymbol{\mathcal{E}}_i^{a_i}(\mathbf{p}_i)\, \frac{\mathrm{d}^2 p_i}{(2\pi)^2} = \frac{2\sqrt{\epsilon_i \epsilon_{i+1}}\, a_{i+1}\, \alpha_{i+1}(\mathbf{p}_{i+1})}{\epsilon_{i+1} - \epsilon_i}\ \boldsymbol{\mathcal{E}}_{i+1}^{a_{i+1}}(\mathbf{p}_{i+1})\,. \tag{3.2}$$

The reason for indexing a_i, a_{i+1} and $\mathbf{p}_i$, $\mathbf{p}_{i+1}$ in Eq. (3.2) will become clear in a few lines. We recall that a_{i+1} and $\mathbf{p}_{i+1}$ can be arbitrarily chosen. Moreover, the definition of $\mathcal{J}_{i+1,i}^{b,a}$ must be precised. In Eq. (2.21), the profile function is now naturally replaced by the one of the interface i. More generally, the kernel scalar factor $\mathcal{J}_{l,m}^{b,a}(\mathbf{p} \,|\, \mathbf{q})$ for $l, m \in \{i, i+1\}$ and $l \neq m$ is defined as

$$\mathcal{J}_{l,m}^{b,a}(\mathbf{p} \,|\, \mathbf{q}) = \left(b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q})\right)^{-1} \int \exp\left[-i(\mathbf{k}_l^b(\mathbf{p}) - \mathbf{k}_m^a(\mathbf{q})) \cdot (\mathbf{x} + \zeta_i(\mathbf{x})\, \hat{\mathbf{e}}_3)\right]\, \mathrm{d}^2 x\,. \tag{3.3}$$

One may argue that we do not strictly need two indices for $\mathcal{J}_{i+1,i}^{b,a}$, since it only appears for successive layers, and it is not defined for non-successive layers. However, we keep the two indices because we will interchange them when we deal with the backward relation later on, so *the order of the indices is important*. Furthermore, it will be convenient to define the *single interface kernels* $\boldsymbol{\Theta}_{l,m}^{b,a}$ defined for successive media, i.e. $l, m \in [\![1, n+1]\!]$ such that $|l - m| = 1$, $a, b \in \{\pm\}$, and $\mathbf{p}, \mathbf{q}$ in the vector plane $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2)$ by

$$\boldsymbol{\Theta}_{l,m}^{b,a}(\mathbf{p} \,|\, \mathbf{q}) = \alpha_l^{-1}(\mathbf{p})\, \mathcal{J}_{l,m}^{b,a}(\mathbf{p} \,|\, \mathbf{q})\ \mathbf{M}_{l,m}^{b,a}(\mathbf{p} \,|\, \mathbf{q})\ . \tag{3.4}$$

The transfer kernel is qualified as *forward* if $l > m$ and *backward* otherwise. In order to compact further notations, we define $\eta_{i+1,i} = \frac{2\sqrt{\epsilon_i \epsilon_{i+1}}}{\epsilon_{i+1} - \epsilon_i}$. With these new notations, Eq. (3.2) reads

$$\sum_{a_i=\pm} \int \boldsymbol{\Theta}_{i+1,i}^{a_{i+1},a_i}(\mathbf{p}_{i+1} \,|\, \mathbf{p}_i)\ \boldsymbol{\mathcal{E}}_i^{a_i}(\mathbf{p}_i)\, \frac{\mathrm{d}^2 p_i}{(2\pi)^2} = a_{i+1}\, \eta_{i+1,i}\ \boldsymbol{\mathcal{E}}_{i+1}^{a_{i+1}}(\mathbf{p}_{i+1})\,. \tag{3.5}$$

By multiplying to the left both sides of Eq. (3.5) by $a_{i+1}\ \boldsymbol{\Theta}_{i+2,i+1}^{a_{i+2},a_{i+1}}(\mathbf{p}_{i+2} \,|\, \mathbf{p}_{i+1})$, where $a_{i+2} = \pm$ and $\mathbf{p}_{i+2}$ can be arbitrarily chosen, and summing over $a_{i+1} = \pm$ and $\mathbf{p}_{i+1}$, one obtains

$$\begin{aligned} &\sum_{a_{i+1}=\pm} a_{i+1} \sum_{a_i=\pm} \int\int \boldsymbol{\Theta}_{i+2,i+1}^{a_{i+2},a_{i+1}}(\mathbf{p}_{i+2} \,|\, \mathbf{p}_{i+1})\, \boldsymbol{\Theta}_{i+1,i}^{a_{i+1},a_i}(\mathbf{p}_{i+1} \,|\, \mathbf{p}_i)\ \boldsymbol{\mathcal{E}}_i^{a_i}(\mathbf{p}_i)\, \frac{\mathrm{d}^2 p_i}{(2\pi)^2}\, \frac{\mathrm{d}^2 p_{i+1}}{(2\pi)^2} \\ &\quad = \eta_{i+1,i} \sum_{a_{i+1}=\pm} \int \boldsymbol{\Theta}_{i+2,i+1}^{a_{i+2},a_{i+1}}(\mathbf{p}_{i+2} \,|\, \mathbf{p}_{i+1})\, \boldsymbol{\mathcal{E}}_{i+1}^{a_{i+1}}(\mathbf{p}_{i+1})\, \frac{\mathrm{d}^2 p_{i+1}}{(2\pi)^2}\ . \end{aligned} \tag{3.6}$$

On the right hand side of the above equation, we recognize the left hand side of Eq. (3.5) taken at interface $i+2$, hence

$$\begin{aligned} &\sum_{a_{i+1}=\pm} a_{i+1} \sum_{a_i=\pm} \int\int \boldsymbol{\Theta}_{i+2,i+1}^{a_{i+2},a_{i+1}}(\mathbf{p}_{i+2} \,|\, \mathbf{p}_{i+1})\, \boldsymbol{\Theta}_{i+1,i}^{a_{i+1},a_i}(\mathbf{p}_{i+1} \,|\, \mathbf{p}_i)\ \boldsymbol{\mathcal{E}}_i^{a_i}(\mathbf{p}_i)\, \frac{\mathrm{d}^2 p_i}{(2\pi)^2}\, \frac{\mathrm{d}^2 p_{i+1}}{(2\pi)^2} \\ &\quad = a_{i+2}\, \eta_{i+2,i+1}\, \eta_{i+1,i}\ \boldsymbol{\mathcal{E}}_{i+2}^{a_{i+2}}(\mathbf{p}_{i+2})\,. \end{aligned} \tag{3.7}$$

By interchanging the order of the integrals we obtain

$$\sum_{a_i=\pm} \int \boldsymbol{\Theta}_{i+2,i}^{a_{i+2},a_i}(\mathbf{p}_{i+2} \,|\, \mathbf{p}_i)\ \boldsymbol{\mathcal{E}}_i^{a_i}(\mathbf{p}_i)\, \frac{\mathrm{d}^2 p_i}{(2\pi)^2} = a_{i+2}\, \eta_{i+2,i}\ \boldsymbol{\mathcal{E}}_{i+2}^{a_{i+2}}(\mathbf{p}_{i+2})\,. \tag{3.8}$$

The above integral equation has the same structure as the forward transfer equation for a single interface, Eq. (2.44), and relates the field amplitudes in media i and $i+2$ without explicitly referring to the field amplitudes in medium $i+1$. This comes with a cost since the new transfer kernel $\boldsymbol{\Theta}_{i+2,i}^{a_{i+2},a_i}$, which will be called the *forward two-interfaces transfer kernel* between media i and $i+2$, is defined by the following transfer kernels composition rule

$$\boldsymbol{\Theta}_{i+2,i}^{a_{i+2},a_i}(\mathbf{p}\,|\,\mathbf{q}) = \sum_{a_{i+1}=\pm} a_{i+1} \int \boldsymbol{\Theta}_{i+2,i+1}^{a_{i+2},a_{i+1}}(\mathbf{p}\,|\,\mathbf{p}')\,\boldsymbol{\Theta}_{i+1,i}^{a_{i+1},a_i}(\mathbf{p}'\,|\,\mathbf{q})\frac{\mathrm{d}^2p'}{(2\pi)^2}\,. \tag{3.9}$$

In addition, we have defined in Eq. (3.8) the right-hand side factor

$$\eta_{i+2,i} = \eta_{i+2,i+1}\,\eta_{i+1,i}\,. \tag{3.10}$$

By iterating the process, we can successively eliminate intermediate fields between *any* two arbitrary layers i and $i+j$, $j \in [\![2, n+1-i]\!]$, and we obtain the following *forward transfer equation*

$$\sum_{a_i=\pm} \int \boldsymbol{\Theta}_{i+j,i}^{a_{i+j},a_i}(\mathbf{p}_{i+j}\,|\,\mathbf{p}_i)\;\boldsymbol{\mathcal{E}}_i^{a_i}(\mathbf{p}_i)\frac{\mathrm{d}^2p_i}{(2\pi)^2} = a_{i+j}\,\eta_{i+j,i}\;\boldsymbol{\mathcal{E}}_{i+j}^{a_{i+j}}(\mathbf{p}_{i+j})\,, \tag{3.11}$$

where the *j-interfaces forward transfer kernel* from i to $i+j$ is defined from the composition rule for $k \in [\![2, j]\!]$

$$\begin{aligned}&\boldsymbol{\Theta}_{i+k,i}^{a_{i+k},a_i}(\mathbf{p}_{i+k}\,|\,\mathbf{p}_i)\\ &= \sum_{a_{i+k-1}=\pm} a_{i+k-1}\int \boldsymbol{\Theta}_{i+k,i+k-1}^{a_{i+k},a_{i+k-1}}(\mathbf{p}_{i+k}\,|\,\mathbf{p}_{i+k\text{-}1})\;\boldsymbol{\Theta}_{i+k-1,i}^{a_{i+k-1},a_i}(\mathbf{p}_{i+k\text{-}1}\,|\,\mathbf{p}_i)\frac{\mathrm{d}^2p_{i+k\text{-}1}}{(2\pi)^2}\,,\end{aligned} \tag{3.12}$$

and initialized for $k=1$ with the single-interface transfer kernel $\boldsymbol{\Theta}_{i+1,i}^{a_{i+1},a_i}(\mathbf{p}_{i+1}\,|\,\mathbf{p}_i)$. Equivalently one may prefer the following expression expanded in single-interfaces transfer kernels

$$\begin{aligned}&\boldsymbol{\Theta}_{i+j,i}^{a_{i+j},a_i}(\mathbf{p}_{i+j}\,|\,\mathbf{p}_i) =\\ &\sum_{a_{i+j-1}=\pm} a_{i+j-1}\int\cdots\sum_{a_{i+1}=\pm} a_{i+1}\int \boldsymbol{\Theta}_{i+j,i+j-1}^{a_{i+j},a_{i+j-1}}(\mathbf{p}_{i+j}\,|\,\mathbf{p}_{i+j\text{-}1})\cdots\boldsymbol{\Theta}_{i+1,i}^{a_{i+1},a_i}(\mathbf{p}_{i+1}\,|\,\mathbf{p}_i)\frac{\mathrm{d}^2p_{i+1}}{(2\pi)^2}\cdots\frac{\mathrm{d}^2p_{i+j\text{-}1}}{(2\pi)^2}\,.\end{aligned} \tag{3.13}$$

In addition, the right-hand side factor $\eta_{i+j,i}$ is also defined by recurrence and reads

$$\eta_{i+j,i} = \prod_{k=1}^{j}\eta_{i+k,i+k-1}\,. \tag{3.14}$$

Similarly, we can derive the *backward transfer equation*:

$$\sum_{a_i=\pm} \int \boldsymbol{\Theta}_{i-j,i}^{a_{i-j},a_i}(\mathbf{p}_{i\text{-}j}\,|\,\mathbf{p}_i)\;\boldsymbol{\mathcal{E}}_i^{a_i}(\mathbf{p}_i)\frac{\mathrm{d}^2p_i}{(2\pi)^2} = a_{i-j}\,\eta_{i-j,i}\;\boldsymbol{\mathcal{E}}_{i-j}^{a_{i-j}}(\mathbf{p}_{i-j})\,, \tag{3.15}$$

where *j-interfaces backward transfer kernel* from i to $i-j$ is defined from the composition rule for $k \in [\![2, j]\!]$

$$\begin{aligned}&\boldsymbol{\Theta}_{i-k,i}^{a_{i-k},a_i}(\mathbf{p}_{i\text{-}k}\,|\,\mathbf{p}_i)\\ &= \sum_{a_{i-k+1}=\pm} a_{i-k+1}\int \boldsymbol{\Theta}_{i-k,i-k+1}^{a_{i-k},a_{i-k+1}}(\mathbf{p}_{i\text{-}k}\,|\,\mathbf{p}_{i\text{-}k+1})\;\boldsymbol{\Theta}_{i-k+1,i}^{a_{i-k+1},a_i}(\mathbf{p}_{i\text{-}k+1}\,|\,\mathbf{p}_i)\frac{\mathrm{d}^2p_{i\text{-}k+1}}{(2\pi)^2}\,.\end{aligned} \tag{3.16}$$

initialized for $k = 1$ with the single-interface transfer kernel $\boldsymbol{\Theta}_{i-1,i}^{a_{i-1},a_i}(\mathbf{p}_{i\text{-}1} \,|\, \mathbf{p}_i)$. Equivalently one may prefer the following expression expanded in single-interfaces transfer kernels

$$\boldsymbol{\Theta}_{i-j,i}^{a_{i-j},a_i}(\mathbf{p}_{i\text{-}j} \,|\, \mathbf{p}_i) = \sum_{a_{i-j+1}=\pm} a_{i-j+1} \int \cdots \sum_{a_{i-1}=\pm} a_{i-1} \int \boldsymbol{\Theta}_{i-j,i-j+1}^{a_{i-j},a_{i-j+1}}(\mathbf{p}_{i\text{-}j} \,|\, \mathbf{p}_{i\text{-}j+1}) \cdots \boldsymbol{\Theta}_{i-1,i}^{a_{i-1},a_i}(\mathbf{p}_{i\text{-}1} \,|\, \mathbf{p}_i) \frac{\mathrm{d}^2 p_{i\text{-}1}}{(2\pi)^2} \cdots \frac{\mathrm{d}^2 p_{i\text{-}j+1}}{(2\pi)^2} \,. \tag{3.17}$$

In addition, the right-hand side factor $\eta_{i-j,i}$ is also defined by recurrence and reads

$$\eta_{i-j,i} = \prod_{k=1}^{j} \eta_{i-k,i-k+1} \,. \tag{3.18}$$

We have finally obtained the transfer equations generalized to any two slabs in a stack. These equations can be applied to the whole set of slabs to obtain the integral equations for the total reflection and transmission amplitudes as shown in the following section.

3.3 Reflection and transmission by a multilayer system

Assuming that we consider the reflection and transmission of an incident plane wave incoming in medium 1, whose amplitude is given by Eq. (2.48), and according to the definition of the reflection and transmission amplitudes Eqs. (2.49) and (2.50), and knowing that for $a_{n+1} = +$ we have $\boldsymbol{\mathcal{E}}_{n+1}^{+}(\mathbf{p}) = 0$ (i.e. no upwards propagating wave in the last medium), we have

$$\boldsymbol{\mathcal{E}}_1^{-}(\mathbf{q}) = (2\pi)^2 \, \delta(\mathbf{q} - \mathbf{p}_0) \; \boldsymbol{\mathcal{E}}_0 \,, \tag{3.19a}$$

$$\boldsymbol{\mathcal{E}}_1^{+}(\mathbf{q}) = \mathbf{R}(\mathbf{q} \,|\, \mathbf{p}_0) \, \boldsymbol{\mathcal{E}}_0, \tag{3.19b}$$

$$\boldsymbol{\mathcal{E}}_{n+1}^{+}(\mathbf{q}) = \mathbf{0} \,, \tag{3.19c}$$

$$\boldsymbol{\mathcal{E}}_{n+1}^{-}(\mathbf{q}) = \mathbf{T}(\mathbf{q} \,|\, \mathbf{p}_0) \, \boldsymbol{\mathcal{E}}_0 \,. \tag{3.19d}$$

By applying the transfer equation, Eq. (3.11), for $a_{n+1} = +$, between the first and last media (i.e. $i = 1$ and $j = n$), and inserting Eqs. (3.19a), (3.19c) and (3.19b), we obtain the following integral equation for the reflection amplitude

$$\int \boldsymbol{\Theta}_{n+1,1}^{+,+}(\mathbf{p} \,|\, \mathbf{q}) \; \mathbf{R}(\mathbf{q} \,|\, \mathbf{p}_{\mathrm{o}}) \, \frac{\mathrm{d}^2 q}{(2\pi)^2} = - \, \boldsymbol{\Theta}_{n+1,1}^{+,-}(\mathbf{p} \,|\, \mathbf{p}_{\mathrm{o}}) \;, \tag{3.20}$$

By applying the transfer equation, Eq. (3.15), for $a_1 = -$, between the first and last media (i.e. $i = n+1$ and $j = n$), inserting Eqs. (3.19a), (3.19c) and (3.19d), we obtain the following integral equation for the transmission amplitude

$$\int \boldsymbol{\Theta}_{1,n+1}^{-,-}(\mathbf{p} \,|\, \mathbf{q}) \; \mathbf{T}(\mathbf{q} \,|\, \mathbf{p}_{\mathrm{o}}) \, \frac{\mathrm{d}^2 q}{(2\pi)^2} = -\eta_{1,n+1} \, (2\pi)^2 \delta(\mathbf{p} - \mathbf{p}_0) \; \mathbf{I}_2 \;, \tag{3.21}$$

Remark 3.1. These equations have the same structure as in the single interface case. The only difference resides in the transfer kernel which is defined as multiple (integral) compositions of single interface transfer kernels.

3.4 Stack of flat interfaces – Fabry-Perot interferometer

When all interfaces are planar the profile equation of the j^{th} interface is given by,

$$x_3 = \zeta_j(\mathbf{x}) = -H_j \,. \tag{3.22}$$

where $H_j > 0$. For convenience we choose the origin on the first interface, hence $H_1 = 0$. Similarly to what we did in section 2.3, it is immediate to show that $\mathcal{J}_{j+1,j}^{a_{j+1},a_j}$ for a planar interface reads

$$\mathcal{J}_{j+1,j}^{a_{j+1},a_j}(\mathbf{p}_{j+1}\,|\,\mathbf{p}_j) = \frac{e^{i(a_{j+1}\alpha_{j+1}(\mathbf{p}_{j+1})-a_j\alpha_j(\mathbf{p}_j))H_j}}{a_{j+1}\alpha_{j+1}(\mathbf{p}_{j+1})-a_j\alpha_j(\mathbf{p}_j)}\,(2\pi)^2\,\delta(\mathbf{p}_{j+1}-\mathbf{p}_j)\,, \tag{3.23}$$

thus, all integrals in the multi-interface transfer kernels become trivial due to the Dirac masses and we are left with

$$\mathbf{\Theta}_{n+1,1}^{a_{n+1},a_1}(\mathbf{p}\,|\,\mathbf{q}) = \bar{\mathbf{\Theta}}_{n+1,1}^{a_{n+1},a_1}(\mathbf{p})\,(2\pi)^2\delta(\mathbf{p}-\mathbf{q})\,, \tag{3.24}$$

where we defined

$$\bar{\mathbf{\Theta}}_{n+1,1}^{a_{n+1},a_1}(\mathbf{p}) = \sum_{\substack{a_n=\pm\\ \vdots \\ a_2=\pm}} \frac{a_n\cdots a_2}{\alpha_n(\mathbf{p})\cdots\alpha_2(\mathbf{p})}\prod_{j=1}^{n}\frac{e^{i(a_{j+1}\alpha_{j+1}(\mathbf{p})-a_j\alpha_j(\mathbf{p}))H_j}}{a_{j+1}\alpha_{j+1}(\mathbf{p})-a_j\alpha_j(\mathbf{p})}\;\mathbf{M}_{j+1,j}^{a_{j+1},a_j}(\mathbf{p}\,|\,\mathbf{p})\,. \tag{3.25}$$

By inserting Eq. (3.24) into Eq. (3.20), we can solve for the reflection amplitude:

$$\mathbf{R}(\mathbf{p}\,|\,\mathbf{p}_0) = -\left[\bar{\mathbf{\Theta}}_{n+1,1}^{+,+}(\mathbf{p}_0)\right]^{-1}\;\bar{\mathbf{\Theta}}_{n+1,1}^{+,-}(\mathbf{p}_0)\,(2\pi)^2\delta(\mathbf{p}-\mathbf{p}_0)\,. \tag{3.26}$$

Remark 3.2. As expected from a system consisting of planar parallel interfaces, only the specular mode is present as indicated by the Dirac mass. Moreover, from the definition of $\bar{\mathbf{\Theta}}_{n+1,1}^{a_{n+1},a_1}(\mathbf{p})$, given by Eq. (3.25), it is clear that it is a diagonal matrix, and so is its inverse. Indeed, all factors $\mathbf{M}_{j+1,j}^{a_{j+1},a_j}(\mathbf{p}\,|\,\mathbf{p})$ are diagonal by definition of the $\mathbf{M}_{j+1,j}^{a_{j+1},a_j}$ matrix when evaluated at a pair of identical in-plane wave-vectors. Hence, $\mathbf{R}(\mathbf{p}\,|\,\mathbf{p}_0)$ is diagonal, which means that there is no polarization coupling.

Similarly, for the transmission amplitude we obtain:

$$\mathbf{T}(\mathbf{p}\,|\,\mathbf{p}_0) = -\eta_{1,n+1}\left[\bar{\mathbf{\Theta}}_{1,n+1}^{-,-}(\mathbf{p}_0)\right]^{-1}\,(2\pi)^2\delta(\mathbf{p}-\mathbf{p}_0)\,, \tag{3.27}$$

with

$$\bar{\mathbf{\Theta}}_{1,n+1}^{a_1,a_{n+1}}(\mathbf{p}) = \sum_{\substack{a_2=\pm\\ \vdots \\ a_n=\pm}} \frac{a_2\cdots a_n}{\alpha_2(\mathbf{p})\cdots\alpha_n(\mathbf{p})}\prod_{j=1}^{n}\frac{e^{i(a_j\alpha_j(\mathbf{p})-a_{j+1}\alpha_{j+1}(\mathbf{p}))H_{j+1}}}{a_j\alpha_j(\mathbf{p})-a_{j+1}\alpha_{j+1}(\mathbf{p})}\;\mathbf{M}_{j,j+1}^{a_j,a_{j+1}}(\mathbf{p}\,|\,\mathbf{p})\,. \tag{3.28}$$

We may verify that the expression given by Eq. (3.26) corresponds to a well known expression for the reflection amplitudes of a Fabry-Perot interferometer for the case of a system made of

three media separated by $n = 2$ interfaces. Denoting by d the thickness of the film (second medium), Eq. (3.26) yields for s-polarized wave $R_{ss}(\mathbf{p}\,|\,\mathbf{p}_0) = r_{ss}(\mathbf{p}_0)\,(2\pi)^2\delta(\mathbf{p}-\mathbf{p}_0)$ with

$$\begin{aligned} r_{ss} &= -\frac{e^{id(\alpha_2+\alpha_1)}(\alpha_3-\alpha_2)^{-1}(\alpha_2+\alpha_1)^{-1} - e^{id(-\alpha_2+\alpha_1)}(\alpha_3+\alpha_2)^{-1}(-\alpha_2+\alpha_1)^{-1}}{e^{id(\alpha_2-\alpha_1)}(\alpha_3-\alpha_2)^{-1}(\alpha_2-\alpha_1)^{-1} - e^{id(-\alpha_2-\alpha_1)}(\alpha_3+\alpha_2)^{-1}(-\alpha_2-\alpha_1)^{-1}} \\ &= e^{2id\alpha_1}\,\frac{e^{id\alpha_2}(\alpha_3-\alpha_2)^{-1}(\alpha_2+\alpha_1)^{-1} - e^{-id\alpha_2}(\alpha_3+\alpha_2)^{-1}(\alpha_1-\alpha_2)^{-1}}{e^{id\alpha_2}(\alpha_3-\alpha_2)^{-1}(\alpha_1-\alpha_2)^{-1} - e^{-id\alpha_2}(\alpha_3+\alpha_2)^{-1}(\alpha_2+\alpha_1)^{-1}} \\ &= e^{2id\alpha_1}\,\frac{\frac{\alpha_1-\alpha_2}{\alpha_2+\alpha_1} + e^{-2id\alpha_2}\frac{\alpha_2-\alpha_3}{\alpha_3+\alpha_2}}{1 + e^{-2id\alpha_2}\frac{\alpha_1-\alpha_2}{\alpha_2+\alpha_1}\frac{\alpha_2-\alpha_3}{\alpha_3+\alpha_2}} \\ &= e^{2id\alpha_1}\,\frac{r_{ss}^{(1,2)} + r_{ss}^{(2,3)}e^{-2id\alpha_2}}{1 + r_{ss}^{(1,2)}r_{ss}^{(2,3)}e^{-2id\alpha_2}}\,, \end{aligned} \tag{3.29}$$

where $r_{ss}^{(1,2)}$ and $r_{ss}^{(2,3)}$ are the Fresnel coefficients for s-polarized reflection for a single planar interface separating media 1 and 2, and 2 and 3 respectively. Equation (3.29) is a well known expression for the reflection amplitude of the Fabry-Perot interferometer for s-polarized light up to the phase factor $e^{2id\alpha_1}$ and the sign in the exponential factor $e^{-2id\alpha_2}$ which depend respectively of the choice of origin and the orientation of the x_3-axis (see e.g. [59]). The derivation of the Fabry-Perot amplitudes for the reflected p-polarized wave and for the case of transmission is similar and is left to the reader. The reduced Rayleigh equations for multilayer system are then consistent with well known results in the limit case where the interfaces are planar.

3.5 Numerical challenges

This section can be read superficially at first and we recommend the interested reader to read it in more depths after reading Part II.

In this last section, we anticipate on some numerical issues and challenges for the resolution of the reduced Rayleigh equations. The aim is not to discuss the details of the numerical resolution, which will be the object of Part II, but rather point at some properties of the equations which are likely to make the numerical resolution delicate. For the sake of clarity, we will restrict ourselves to the case of a system made of three media separated by two arbitrary interfaces. The corresponding reduced Rayleigh equations satisfied by the reflection and transmission amplitudes are respectively given by

$$\int \mathbf{\Theta}_{31}^{++}(\mathbf{p}\,|\,\mathbf{q})\;\mathbf{R}(\mathbf{q}\,|\,\mathbf{p}_0)\,\frac{\mathrm{d}^2q}{(2\pi)^2} = -\,\mathbf{\Theta}_{31}^{+-}(\mathbf{p}\,|\,\mathbf{p}_0) \tag{3.30a}$$

$$\int \mathbf{\Theta}_{13}^{--}(\mathbf{p}\,|\,\mathbf{q})\;\mathbf{T}(\mathbf{q}\,|\,\mathbf{p}_0)\,\frac{\mathrm{d}^2q}{(2\pi)^2} = -\eta_{13}\,(2\pi)^2\,\delta(\mathbf{p}-\mathbf{p}_0)\,. \tag{3.30b}$$

As a small digression, let us mention that these equations will be at the basis of the study of Selényi rings in Chapter 10 and in one of the model of the plasmonic photonic surface studied in Chapter 12.

3.5.1 Numerical stability and exponentially growing modes

The first issue we will be concerned with is that of the stability of the numerical scheme. Numerically, the integral equations (3.30) are discretized to obtain linear systems, and only certain finite set of discrete modes are taken into account. These modes may be propagating,

evanescent or even exponentially growing modes inside the film (medium 2). If the norm in-plane wave vector associated with such a evanescent or growing mode is large, the exponential decay or growth rate may become large. Physically, such rapid growth or decay will be attenuated by a corresponding amplitude which is expected to be weak. Due to numerical precision, making a small (and inevitable) error on the amplitudes of such modes may yield to dramatic instability. As a rule of thumb, the decay or growth rate is controlled by $|\alpha_2(\mathbf{p})|\, d$. It is then expected that the thicker the film or the larger the dielectric constant of the film, the more likely instability are to occur for a given set of modes. In other words, troubles are expected when the film becomes of the order of a few wavelengths thick. One may naively think that the modes in the film should not be an issue since we are not evaluating them by virtue of the transfer equations. However, the fact that the amplitudes of the modes in the film do not explicitly appear does not mean that they do not contribute, quite on the contrary. Their contribution is encoded in the transfer kernel. It is therefore expected that the source of the trouble will be found in analyzing the transfer kernels. By definition the transfer kernel $\mathbf{\Theta}_{31}^{++}(\mathbf{p}\,|\,\mathbf{q})$ in Eq. (3.30) is given

$$\mathbf{\Theta}_{31}^{++}(\mathbf{p}\,|\,\mathbf{q}) = \sum_{a_2=\pm} a_2 \int \mathbf{\Theta}_{32}^{+a_2}(\mathbf{p}\,|\mathbf{p}')\ \mathbf{\Theta}_{21}^{a_2+}(\mathbf{p}'|\,\mathbf{q})\ \frac{\mathrm{d}p'}{(2\pi)^2}\,. \tag{3.31}$$

Cases where one of the interfaces is planar.

Let us assume first that the second surface is planar and we set the origin of the coordinate system on the second surface, hence $\zeta_2(\mathbf{x}) = 0$ and the surface profile function of the first interface can be written as $\zeta_1(\mathbf{x}) = d + \zeta(\mathbf{x})$, where d is the average film thickness and $\int \zeta(\mathbf{x})\,\mathrm{d}^2x = 0$. The scalar kernel factor $\mathcal{J}_{32}^{a_3,a_2}$ associated with the second interface thus reduces to

$$\mathcal{J}_{32}^{a_3,a_2}(\mathbf{p}\,|\,\mathbf{q}) = [a_3\alpha_3(\mathbf{p}) - a_2\alpha_2(\mathbf{p})]^{-1}\,(2\pi)^2\,\delta(\mathbf{p}-\mathbf{q})\,, \tag{3.32}$$

and the scalar kernel factor $\mathcal{J}_{21}^{a_2,a_1}$ associated with the first interface can be factorized as

$$\begin{aligned}\mathcal{J}_{21}^{a_2,a_1}(\mathbf{p}\,|\,\mathbf{q}) &= \frac{\exp[-i(a_2\alpha_2(\mathbf{p}) - a_1\alpha_1(\mathbf{q}))d]}{a_2\alpha_2(\mathbf{p}) - a_1\alpha_1(\mathbf{q})} \int \exp\left[-i(\mathbf{k}_2^{a_2}(\mathbf{p}) - \mathbf{k}_1^{a_1}(\mathbf{q}))\cdot(\mathbf{x} + \zeta(\mathbf{x})\,\hat{\mathbf{e}}_3)\right]\mathrm{d}^2x \\ &\stackrel{\text{def}}{=} \exp[-i(a_2\alpha_2(\mathbf{p}) - a_1\alpha_1(\mathbf{q}))d]\ \hat{\mathcal{J}}_{21}^{a_2,a_1}(\mathbf{p}\,|\,\mathbf{q})\,.\end{aligned} \tag{3.33}$$

The last equality is a defining equality for $\hat{\mathcal{J}}_{21}^{a_2,a_1}(\mathbf{p}\,|\,\mathbf{q})$. Note that the profile function in the definition of $\hat{\mathcal{J}}_{21}^{a_2,a_1}(\mathbf{p}\,|\,\mathbf{q})$ is ζ and not ζ_1, thus $\hat{\mathcal{J}}_{21}^{a_2,a_1}(\mathbf{p}\,|\,\mathbf{q})$ represents the scalar kernel factor one would get if the origin was chosen to lie on the average plane associated with the first interface, hence the phase factor in the right-hand side of Eq. (3.33). Expansion of Eq. (3.31) thus reads

$$\begin{aligned}\mathbf{\Theta}_{31}^{++}(\mathbf{p}\,|\,\mathbf{q}) = \Bigg[& e^{-i\alpha_2(\mathbf{p})d}\,\frac{\mathbf{M}_{32}^{++}(\mathbf{p}\,|\,\mathbf{p})\,\hat{\mathcal{J}}_{21}^{++}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{M}_{21}^{++}(\mathbf{p}\,|\,\mathbf{q})}{\alpha_3(\mathbf{p}) - \alpha_2(\mathbf{p})} \\ & - e^{i\alpha_2(\mathbf{p})d}\,\frac{\mathbf{M}_{32}^{+-}(\mathbf{p}\,|\,\mathbf{p})\,\hat{\mathcal{J}}_{21}^{-+}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{M}_{21}^{-+}(\mathbf{p}\,|\,\mathbf{q})}{\alpha_3(\mathbf{p}) + \alpha_2(\mathbf{p})}\Bigg]\,\alpha_3^{-1}(\mathbf{p})\alpha_2^{-1}(\mathbf{p})\,e^{i\alpha_1(\mathbf{q})d}\,.\end{aligned} \tag{3.34}$$

Let us now analyze the behaviour of the exponential factors as $|\,\mathbf{p}\,|$ or $|\,\mathbf{q}\,|$ becomes large. For $|\,\mathbf{q}\,| > n_1\omega/c$, $\alpha_1(\mathbf{q}) = i\beta_1(\mathbf{q})$, where $\beta_1(\mathbf{q}) = (\mathbf{q}^2 - \epsilon_1\omega^2/c^2)^{1/2} > 0$. Hence $e^{i\alpha_1(\mathbf{q})d} = e^{-\beta_1(\mathbf{q})d} \to 0$ as $|\,\mathbf{q}\,| \to \infty$. This term is therefore not troublesome. Similarly, for $|\,\mathbf{p}\,| > n_2\omega/c$, $\alpha_2(\mathbf{p}) = i\beta_2(\mathbf{p})$, where $\beta_2(\mathbf{p}) = (\mathbf{p}^2 - \epsilon_2\omega^2/c^2)^{1/2} > 0$. The exponential factor in the first term of the square bracket in Eq. (3.34) thus becomes $e^{-i\alpha_2(\mathbf{p})d} = e^{\beta_2(\mathbf{p})d} \to \infty$ as $|\,\mathbf{p}\,| \to \infty$ while the exponential factor in the second term becomes $e^{i\alpha_2(\mathbf{p})d} = e^{-\beta_2(\mathbf{p})d} \to 0$ as $|\,\mathbf{p}\,| \to \infty$. It

is therefore recommended to take care of the factor $e^{-i\alpha_2(\mathbf{p})d}$. Let us factorize it out so that Eq. (3.34) reads

$$\begin{aligned}\boldsymbol{\Theta}_{31}^{++}(\mathbf{p}\,|\,\mathbf{q}) = & \left[\frac{\mathbf{M}_{32}^{++}(\mathbf{p}\,|\,\mathbf{p})\,\hat{\mathcal{J}}_{21}^{++}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{M}_{21}^{++}(\mathbf{p}\,|\,\mathbf{q})}{\alpha_3(\mathbf{p})-\alpha_2(\mathbf{p})}\right. \\ & \left. - e^{2i\alpha_2(\mathbf{p})d}\,\frac{\mathbf{M}_{32}^{+-}(\mathbf{p}\,|\,\mathbf{p})\,\hat{\mathcal{J}}_{21}^{-+}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{M}_{21}^{-+}(\mathbf{p}\,|\,\mathbf{q})}{\alpha_3(\mathbf{p})+\alpha_2(\mathbf{p})}\right]\alpha_3^{-1}(\mathbf{p})\alpha_2^{-1}(\mathbf{p})\,e^{i\alpha_1(\mathbf{q})d}\,e^{-i\alpha_2(\mathbf{p})d} \\ & \stackrel{\text{def}}{=} \mathbf{M}(\mathbf{p}\,|\,\mathbf{q})\,\alpha_3^{-1}(\mathbf{p})\alpha_2^{-1}(\mathbf{p})\,e^{-i\alpha_2(\mathbf{p})d}\,, \end{aligned} \tag{3.35}$$

where the last equality defines $\mathbf{M}(\mathbf{p}\,|\,\mathbf{q})$. So far we have only analyzed the kernel $\boldsymbol{\Theta}_{31}^{++}$ of the integral equation (3.30a). Consider now the right-hand side of Eq. (3.30a). Following the same steps as the one executed for expanding the kernel $\boldsymbol{\Theta}_{31}^{++}$, we obtain that

$$\begin{aligned}\boldsymbol{\Theta}_{31}^{+-}(\mathbf{p}\,|\,\mathbf{p}_0) = & \left[\frac{\mathbf{M}_{32}^{++}(\mathbf{p}\,|\,\mathbf{p})\,\hat{\mathcal{J}}_{21}^{+-}(\mathbf{p}\,|\,\mathbf{p}_0)\,\mathbf{M}_{21}^{+-}(\mathbf{p}\,|\,\mathbf{p}_0)}{\alpha_3(\mathbf{p})-\alpha_2(\mathbf{p})}\right. \\ & \left. - e^{2i\alpha_2(\mathbf{p})d}\,\frac{\mathbf{M}_{32}^{+-}(\mathbf{p}\,|\,\mathbf{p})\,\hat{\mathcal{J}}_{21}^{--}(\mathbf{p}\,|\,\mathbf{p}_0)\,\mathbf{M}_{21}^{--}(\mathbf{p}\,|\,\mathbf{p}_0)}{\alpha_3(\mathbf{p})+\alpha_2(\mathbf{p})}\right]\alpha_3^{-1}(\mathbf{p})\alpha_2^{-1}(\mathbf{p})\,e^{-i\alpha_1(\mathbf{p}_0)d}\,e^{-i\alpha_2(\mathbf{p})d} \\ & \stackrel{\text{def}}{=} \mathbf{N}(\mathbf{p}\,|\,\mathbf{p}_0)\,\alpha_3^{-1}(\mathbf{p})\alpha_2^{-1}(\mathbf{p})\,e^{-i\alpha_2(\mathbf{p})d}\,, \end{aligned} \tag{3.36}$$

where the last equality defines $\mathbf{N}(\mathbf{p}\,|\,\mathbf{p}_0)$. The exponential factor $e^{-i\alpha_2(\mathbf{p})d}$ can then be also factorized out of the right-hand side of Eq. (3.30a). Since this factor only depends on the variable $\mathbf{p}$, it can be taken out of the integral in Eq. (3.30a) and simplified on both side of the equation (and so can the factor $\alpha_3^{-1}(\mathbf{p})\alpha_2^{-1}(\mathbf{p})$). Consequently, we have removed the exponentially growing terms on both side of the equation, and we obtain a new formulation of Eq. (3.30a)

$$\int \mathbf{M}(\mathbf{p}\,|\,\mathbf{q})\;\mathbf{R}(\mathbf{q}\,|\,\mathbf{p}_0)\,\frac{\mathrm{d}^2q}{(2\pi)^2} = -\,\mathbf{N}(\mathbf{p}\,|\,\mathbf{p}_0)\,, \tag{3.37}$$

where the new kernel and right-hand side were defined above. Mathematically speaking Eq. (3.37) is equivalent to Eq. (3.30a), however, from a numerical point of view, we have experienced that Eq. (3.37) leads to a numerical scheme which seems stable for film of several wavelengths thick while a numerical scheme directly based on Eq. (3.30a) leads to instability for a film thickness of the order of the wavelength.

Let us now consider the case where the first interface is planar and the second is arbitrary instead. In contrast to the previous case, we will see that the present one contains an additional difficulty. In that case, we make the choice of origin on the the planar interface, i.e. that we have $\zeta_1 = 0$ and $\zeta_2 = \zeta - d$ with $d > 0$ still denotes the film thickness. Thus we have

$$\mathcal{J}_{21}^{a_2,a_1}(\mathbf{p}\,|\,\mathbf{q}) = [a_2\alpha_2(\mathbf{p}) - a_1\alpha_1(\mathbf{p})]^{-1}\,(2\pi)^2\,\delta(\mathbf{p}-\mathbf{q})\,, \tag{3.38}$$

and

$$\begin{aligned}\mathcal{J}_{32}^{a_3,a_2}(\mathbf{p}\,|\,\mathbf{q}) = & \frac{\exp[i(a_3\alpha_3(\mathbf{p}) - a_2\alpha_2(\mathbf{q}))d]}{a_3\alpha_3(\mathbf{p}) - a_2\alpha_2(\mathbf{q})}\int \exp\left[-i(\mathbf{k}_3^{a_3}(\mathbf{p}) - \mathbf{k}_2^{a_2}(\mathbf{q}))\cdot(\mathbf{x}+\zeta(\mathbf{x})\,\hat{\mathbf{e}}_3)\right]\mathrm{d}^2x \\ & \stackrel{\text{def}}{=} \exp[i(a_3\alpha_3(\mathbf{p}) - a_2\alpha_2(\mathbf{q}))d]\;\hat{\mathcal{J}}_{32}^{a_3,a_2}(\mathbf{p}\,|\,\mathbf{q})\,. \end{aligned} \tag{3.39}$$

Following the same steps as in the previous case we obtain

$$\begin{aligned}\boldsymbol{\Theta}_{31}^{++}(\mathbf{p}\,|\,\mathbf{q}) = \Bigg[&e^{-i\alpha_2(\mathbf{q})d}\,\frac{\hat{\mathcal{J}}_{32}^{++}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{M}_{32}^{++}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{M}_{21}^{++}(\mathbf{q}\,|\,\mathbf{q})}{\alpha_2(\mathbf{q})-\alpha_1(\mathbf{q})}\\ &+e^{i\alpha_2(\mathbf{q})d}\,\frac{\hat{\mathcal{J}}_{32}^{+-}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{M}_{32}^{+-}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{M}_{21}^{-+}(\mathbf{q}\,|\,\mathbf{q})}{\alpha_2(\mathbf{q})+\alpha_1(\mathbf{q})}\Bigg]\,\alpha_3^{-1}(\mathbf{p})\alpha_2^{-1}(\mathbf{q})\,e^{i\alpha_3(\mathbf{p})d}\,. \end{aligned}\tag{3.40}$$

As in the previous case the troublesome factor is identified to be $e^{-i\alpha_2(\mathbf{q})d}$ which we factorize out

$$\begin{aligned}\boldsymbol{\Theta}_{31}^{++}(\mathbf{p}\,|\,\mathbf{q}) = \Bigg[&\frac{\hat{\mathcal{J}}_{32}^{++}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{M}_{32}^{++}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{M}_{21}^{++}(\mathbf{q}\,|\,\mathbf{q})}{\alpha_2(\mathbf{q})-\alpha_1(\mathbf{q})}\\ &+e^{2i\alpha_2(\mathbf{q})d}\,\frac{\hat{\mathcal{J}}_{32}^{+-}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{M}_{32}^{+-}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{M}_{21}^{-+}(\mathbf{q}\,|\,\mathbf{q})}{\alpha_2(\mathbf{q})+\alpha_1(\mathbf{q})}\Bigg]\,\alpha_3^{-1}(\mathbf{p})\alpha_2^{-1}(\mathbf{q})\,e^{i\alpha_3(\mathbf{p})d}\,e^{-i\alpha_2(\mathbf{q})d}\\ &\stackrel{\text{def}}{=}\mathbf{M}(\mathbf{p}\,|\,\mathbf{q})\,\alpha_3^{-1}(\mathbf{p})\,e^{i\alpha_3(\mathbf{p})d}\,e^{-i\alpha_2(\mathbf{q})d}\,. \end{aligned}\tag{3.41}$$

Note that here we re-use the same notation as before for what will be become the new transfer kernel for simplicity, although the definition differs. Similarly, the right-hand side can be recast as

$$\begin{aligned}\boldsymbol{\Theta}_{31}^{+-}(\mathbf{p}\,|\,\mathbf{p}_0) = \Bigg[&\frac{\hat{\mathcal{J}}_{32}^{++}(\mathbf{p}\,|\,\mathbf{p}_0)\,\mathbf{M}_{32}^{++}(\mathbf{p}\,|\,\mathbf{p}_0)\,\mathbf{M}_{21}^{+-}(\mathbf{p}_0\,|\,\mathbf{p}_0)}{\alpha_2(\mathbf{p}_0)+\alpha_1(\mathbf{p}_0)}\\ &+e^{2i\alpha_2(\mathbf{p}_0)d}\,\frac{\hat{\mathcal{J}}_{32}^{+-}(\mathbf{p}\,|\,\mathbf{p}_0)\,\mathbf{M}_{32}^{+-}(\mathbf{p}\,|\,\mathbf{p}_0)\,\mathbf{M}_{21}^{--}(\mathbf{p}_0\,|\,\mathbf{p}_0)}{\alpha_2(\mathbf{p}_0)-\alpha_1(\mathbf{p}_0)}\Bigg]\,\alpha_3^{-1}(\mathbf{p})\alpha_2^{-1}(\mathbf{p}_0)\,e^{i\alpha_3(\mathbf{p})d}\,e^{-i\alpha_2(\mathbf{p}_0)d}\\ &\stackrel{\text{def}}{=}\mathbf{N}(\mathbf{p}\,|\,\mathbf{p}_0)\,\alpha_3^{-1}(\mathbf{p})\,e^{i\alpha_3(\mathbf{p})d}\,e^{-i\alpha_2(\mathbf{p}_0)d}\,. \end{aligned}\tag{3.42}$$

Note that here we do not need to factorized by $e^{-i\alpha_2(\mathbf{p}_0)d}$ since it is a constant, but we have done so for the sake of keeping some symmetry in the equations. The factor $\alpha_3^{-1}(\mathbf{p})\,e^{i\alpha_3(\mathbf{p})d}$ can be taken out of the integral in Eq. (3.30a) and be simplified on both side of the equation. However, this factor is not the troublesome one. Indeed, the exponentially growing factor $e^{-i\alpha_2(\mathbf{q})d}$ remains inside the integral in Eq. (3.30a) and cannot be simplified. We need a more subtle manipulation here, namely a *change of unknown*. We let the factor $e^{-i\alpha_2(\mathbf{q})d}$ to be "absorbed" by the unknown by defining

$$\tilde{\mathbf{R}}(\mathbf{q}\,|\,\mathbf{p}_0) = e^{-i(\alpha_2(\mathbf{q})-\alpha_2(\mathbf{p}_0))d}\;\mathbf{R}(\mathbf{q}\,|\,\mathbf{p}_0)\,. \tag{3.43}$$

The resulting integral equation reads

$$\int \mathbf{M}(\mathbf{p}\,|\,\mathbf{q})\;\tilde{\mathbf{R}}(\mathbf{q}\,|\,\mathbf{p}_0)\,\frac{\mathrm{d}^2q}{(2\pi)^2} = -\,\mathbf{N}(\mathbf{p}\,|\,\mathbf{p}_0)\,. \tag{3.44}$$

Equation (3.44) was experienced to lead to a more stable numerical scheme than that obtained from Eq. (3.30a) used as it reads. The strategy in that case is then to solve Eq. (3.44) for $\tilde{\mathbf{R}}$ and then deduce $\mathbf{R}$ from Eq. (3.43) by a simple point-wise multiplication by $e^{i(\alpha_2(\mathbf{q})-\alpha_2(\mathbf{p}_0))d}$.

Remark 3.3. We could choose to keep the origin on the average plane of the second interface instead and obtain slightly different expressions, Nevertheless, this choice would lead to the same issue with a factor $e^{-i\alpha_2(\mathbf{q})d}$ which cannot be factorized out of the integral.

Similar consideration are valid for the reduced Rayleigh equation for the transmission amplitude. The only difference is that the case involving a change of unknown is that where the first interface is arbitrary and the second is planar instead. We summarize below our recommendations for all the cases. We will commonly denote by $\mathbf{M}$, $\mathbf{N}$ and $\mathbf{X}$ respectively the kernel of the integral equation, its right-hand side and the unknown, i.e. we have

$$\int \mathbf{M}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{X}(\mathbf{q}\,|\,\mathbf{p}_0)\,\frac{\mathrm{d}^2 q}{(2\pi)^2} = -\,\mathbf{N}(\mathbf{p}\,|\,\mathbf{p}_0)\,. \tag{3.45}$$

The expressions of $\mathbf{M}$, $\mathbf{N}$ and $\mathbf{X}$ for the different cases are given below.

Reflection rough-planar ($\zeta_1 = d + \zeta$, $\zeta_2 = 0$)

$$\begin{aligned}
\mathbf{M}(\mathbf{p}\,|\,\mathbf{q}) &= \left[\frac{\mathbf{M}_{32}^{++}(\mathbf{p}\,|\,\mathbf{p})\,\hat{\mathcal{J}}_{21}^{++}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{M}_{21}^{++}(\mathbf{p}\,|\,\mathbf{q})}{\alpha_3(\mathbf{p}) - \alpha_2(\mathbf{p})} - e^{2i\alpha_2(\mathbf{p})d}\,\frac{\mathbf{M}_{32}^{+-}(\mathbf{p}\,|\,\mathbf{p})\,\hat{\mathcal{J}}_{21}^{-+}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{M}_{21}^{-+}(\mathbf{p}\,|\,\mathbf{q})}{\alpha_3(\mathbf{p}) + \alpha_2(\mathbf{p})}\right] \\
\mathbf{N}(\mathbf{p}\,|\,\mathbf{p}_0) &= \left[\frac{\mathbf{M}_{32}^{++}(\mathbf{p}\,|\,\mathbf{p})\,\hat{\mathcal{J}}_{21}^{+-}(\mathbf{p}\,|\,\mathbf{p}_0)\,\mathbf{M}_{21}^{+-}(\mathbf{p}\,|\,\mathbf{p}_0)}{\alpha_3(\mathbf{p}) - \alpha_2(\mathbf{p})} - e^{2i\alpha_2(\mathbf{p})d}\,\frac{\mathbf{M}_{32}^{+-}(\mathbf{p}\,|\,\mathbf{p})\,\hat{\mathcal{J}}_{21}^{--}(\mathbf{p}\,|\,\mathbf{p}_0)\,\mathbf{M}_{21}^{--}(\mathbf{p}\,|\,\mathbf{p}_0)}{\alpha_3(\mathbf{p}) + \alpha_2(\mathbf{p})}\right] \\
\mathbf{X}(\mathbf{q}\,|\,\mathbf{p}_0) &= \exp\left[i(\alpha_1(\mathbf{q}) + \alpha_1(\mathbf{p}_0))d\right]\,\mathbf{R}(\mathbf{q}\,|\,\mathbf{p}_0)\,.
\end{aligned}$$

Reflection planar-rough ($\zeta_1 = 0$, $\zeta_2 = \zeta - d$)

$$\begin{aligned}
\mathbf{M}(\mathbf{p}\,|\,\mathbf{q}) &= \left[\frac{\hat{\mathcal{J}}_{32}^{++}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{M}_{32}^{++}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{M}_{21}^{++}(\mathbf{q}\,|\,\mathbf{q})}{\alpha_2(\mathbf{q})[\alpha_2(\mathbf{q}) - \alpha_1(\mathbf{q})]} + e^{2i\alpha_2(\mathbf{q})d}\,\frac{\hat{\mathcal{J}}_{32}^{+-}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{M}_{32}^{+-}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{M}_{21}^{-+}(\mathbf{q}\,|\,\mathbf{q})}{\alpha_2(\mathbf{q})[\alpha_2(\mathbf{q}) + \alpha_1(\mathbf{q})]}\right] \\
\mathbf{N}(\mathbf{p}\,|\,\mathbf{p}_0) &= \left[\frac{\hat{\mathcal{J}}_{32}^{++}(\mathbf{p}\,|\,\mathbf{p}_0)\,\mathbf{M}_{32}^{++}(\mathbf{p}\,|\,\mathbf{p}_0)\,\mathbf{M}_{21}^{+-}(\mathbf{p}_0\,|\,\mathbf{p}_0)}{\alpha_2(\mathbf{p}_0)[\alpha_2(\mathbf{p}_0) + \alpha_1(\mathbf{p}_0)]} + e^{2i\alpha_2(\mathbf{p}_0)d}\,\frac{\hat{\mathcal{J}}_{32}^{+-}(\mathbf{p}\,|\,\mathbf{p}_0)\,\mathbf{M}_{32}^{+-}(\mathbf{p}\,|\,\mathbf{p}_0)\,\mathbf{M}_{21}^{--}(\mathbf{p}_0\,|\,\mathbf{p}_0)}{\alpha_2(\mathbf{p}_0)[\alpha_2(\mathbf{p}_0) - \alpha_1(\mathbf{p}_0)]}\right] \\
\mathbf{X}(\mathbf{q}\,|\,\mathbf{p}_0) &= \exp\left[-i(\alpha_2(\mathbf{q}) - \alpha_2(\mathbf{p}_0))d\right]\,\mathbf{R}(\mathbf{q}\,|\,\mathbf{p}_0)\,.
\end{aligned}$$

Transmission rough-planar ($\zeta_1 = d + \zeta$, $\zeta_2 = 0$)

$$\begin{aligned}
\mathbf{M}(\mathbf{p}\,|\,\mathbf{q}) &= \left[e^{2i\alpha_2(\mathbf{q})d}\,\frac{\hat{\mathcal{J}}_{12}^{-+}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{M}_{12}^{-+}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{M}_{23}^{+-}(\mathbf{q}\,|\,\mathbf{q})}{\alpha_2(\mathbf{q})[\alpha_2(\mathbf{q}) + \alpha_3(\mathbf{q})]} - \frac{\hat{\mathcal{J}}_{12}^{--}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{M}_{12}^{--}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{M}_{23}^{--}(\mathbf{q}\,|\,\mathbf{q})}{\alpha_2(\mathbf{q})[-\alpha_2(\mathbf{q}) + \alpha_3(\mathbf{q})]}\right] \\
\mathbf{N}(\mathbf{p}\,|\,\mathbf{p}_0) &= \frac{4\sqrt{\epsilon_3\epsilon_2^2\epsilon_1}}{(\epsilon_1 - \epsilon_2)(\epsilon_2 - \epsilon_3)}\,\alpha_1(\mathbf{p}_0)\,\exp\left[-i\alpha_1(\mathbf{p}_0)d\right]\,(2\pi)^2\,\delta(\mathbf{p} - \mathbf{p}_0) \\
\mathbf{X}(\mathbf{q}\,|\,\mathbf{p}_0) &= \exp\left[-i\alpha_2(\mathbf{q})d\right]\,\mathbf{T}(\mathbf{q}\,|\,\mathbf{p}_0)\,.
\end{aligned}$$

Transmission planar-rough ($\zeta_1 = 0$, $\zeta_2 = \zeta - d$)

$$\begin{aligned}
\mathbf{M}(\mathbf{p}\,|\,\mathbf{q}) &= \left[e^{2i\alpha_2(\mathbf{p})d}\,\frac{\mathbf{M}_{12}^{-+}(\mathbf{p}\,|\,\mathbf{p})\,\hat{\mathcal{J}}_{23}^{+-}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{M}_{23}^{+-}(\mathbf{p}\,|\,\mathbf{q})}{\alpha_2(\mathbf{p}) + \alpha_3(\mathbf{q})} - \frac{\mathbf{M}_{12}^{--}(\mathbf{p}\,|\,\mathbf{p})\,\hat{\mathcal{J}}_{23}^{--}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{M}_{23}^{--}(\mathbf{p}\,|\,\mathbf{q})}{-\alpha_2(\mathbf{p}) + \alpha_3(\mathbf{q})}\right] \\
\mathbf{N}(\mathbf{p}\,|\,\mathbf{p}_0) &= \frac{4\sqrt{\epsilon_3\epsilon_2^2\epsilon_1}}{(\epsilon_1 - \epsilon_2)(\epsilon_2 - \epsilon_3)}\,\alpha_1(\mathbf{p}_0)\alpha_2(\mathbf{p}_0)\,\exp\left[i\alpha_2(\mathbf{p}_0)d\right]\,(2\pi)^2\,\delta(\mathbf{p} - \mathbf{p}_0) \\
\mathbf{X}(\mathbf{q}\,|\,\mathbf{p}_0) &= \exp\left[i\alpha_3(\mathbf{q})d\right]\,\mathbf{T}(\mathbf{q}\,|\,\mathbf{p}_0)\,.
\end{aligned}$$

Remark 3.4. In the above expressions, we have chosen to always make a change of unknown even in cases where it is not strictly necessary in regards to the stability of the numerical scheme. This is simply an aesthetic choice, in order to keep some kind of symmetry on the expressions. Only for the cases presenting a factor $\exp[-i\alpha_2(\mathbf{q})d]$ in the change of unknown is the change of unknown indispensable.

Case of two arbitrary interfaces.

The case where two interfaces are both rough remains a challenge to be taken. Stabilizing the reduced Rayleigh equations in this case has not been achieved in this work unfortunately. We believe, nevertheless that this should be possible, and that inspiration can be drawn from methods dealing with stacks of planar interfaces. In addition, the sub-cases treated above when only one of the interfaces is rough may give good hints on what the general treatment should be, as one may expect it to yield the aforementioned particular cases in the limit when one of the interfaces is set to be planar.

3.5.2 Complexity of the transfer kernel

In addition to the stability issue, there is an issue of increasing complexity as the number of layer increases. To illustrate this issue, let us compare the reduced Rayleigh equations for the reflection amplitudes associated with a system having a single rough interface and a system having two rough interfaces. In both cases, we have seen that the reduced Rayleigh equation can be written in the form

$$\int \mathbf{M}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{R}(\mathbf{q}\,|\,\mathbf{p}_0)\,\frac{\mathrm{d}^2q}{(2\pi)^2} = -\,\mathbf{N}(\mathbf{p}\,|\,\mathbf{p}_0)\,.$$

The difference resides, of course, in the kernel and right-hand side. Let us focus on the kernel $\mathbf{M}(\mathbf{p}\,|\,\mathbf{q})$. For a single interface, the kernel reads

$$\mathbf{M}(\mathbf{p}\,|\,\mathbf{q}) = \mathbf{\Theta}_{21}^{++}(\mathbf{p}\,|\,\mathbf{q}) = \alpha_2^{-1}(\mathbf{p})\,\mathcal{J}_{21}^{++}(\mathbf{p}\,|\,\mathbf{q})\;\mathbf{M}_{21}^{++}(\mathbf{p}\,|\,\mathbf{q})\,,$$

while for two interfaces it reads

$$\mathbf{\Theta}_{31}^{++}(\mathbf{p}\,|\,\mathbf{q}) = \sum_{a_2=\pm} a_2 \int \mathbf{\Theta}_{32}^{+a_2}(\mathbf{p}\,|\mathbf{p}')\;\mathbf{\Theta}_{21}^{a_2+}(\mathbf{p}'|\,\mathbf{q})\,\frac{\mathrm{d}p'}{(2\pi)^2}\,.$$

We will see in Chapter 5 that numerically, an integral equation such as the reduced Rayleigh equation can be discretized into a linear system of the form

$$\mathbf{M}\,\mathbf{R} = -\,\mathbf{N}\,,$$

where here $\mathbf{M}$ is a square matrix and $\mathbf{R}$ and $\mathbf{N}$ are column vectors (technically two columns-matrices, one for each incident p and s-polarization, cf. Chapter 5). For a set number of discretized points (or modes) N, solving the resulting linear with a so-called direct method take of the order of N^3 operations. This is independent of the kernel and right-hand side. However, there is an additional complexity that must be taken into account, namely, that of setting up the linear system. Roughly speaking, for a single interface, the number of operation for setting the matrix $\mathbf{M}$ can be written as νN^2 assuming that each of the N^2 elements of the matrix requires ν operations for their computation. For two rough interfaces, the kernel leads to a matrix of the form

$$\mathbf{M} = \mathbf{M}_1\,\mathbf{M}_2 - \mathbf{M}_3\,\mathbf{M}_4\,,$$

where $\mathbf{M}_1$, $\mathbf{M}_2$, $\mathbf{M}_3$, and $\mathbf{M}_4$ all have the same complexity as the matrix $\mathbf{M}$ associated with the kernel of a single interface problem. Setting up the matrix $\mathbf{M}$ for a two-interface problem thus requires first to set up four matrices of the same complexity as that of the single interface problem, so $4\nu N^2$ which is not a big problem, but more problematic are the two matrix products which require of the order of N^3 operations each. Therefore for two-rough interfaces, the numerical bottleneck becomes setting up the matrix of the linear system, not solving it. This will be illustrated in more details in Chapter 5. The reader will convince oneself that the power in the complexity scaling increases by one each time one interface is added to the system. We are facing the so-called curse of dimensionality.

3.5.3 Reduction of the complexity with the RRE of the second kind

But there is hope. Intuitively, one would expect that the complexity of a scattering problem should scale linearly with the number of interfaces, and not increase the power of the asymptotic scaling each time an interface is introduced. Why is that? If we think of methods such as boundary elements methods, the interfaces are discretized and the numerical problem scales with the number of discrete points. Thus, the scaling of the complexity is linear with the number of interfaces. Can we solve the reduced Rayleigh equations numerically with a complexity which is linear with the number of interfaces? The answer is yes, in principle.

The main idea is the following. Since setting up the transfer kernel is costly, let us avoid setting it up. Of course, we still need to keep the whole information about the system, and we should expect to still have to set up *single interface transfer kernels* but we can avoid the costly matrix-matrix multiplications associated with the multi-interface transfer kernel.

Instead of the reduced Rayleigh equation for the reflection amplitude, let us consider the reduced Rayleigh equation of the second kind instead (see Section 2.5). Following the same step as in Section 2.5 we can derive the reduced Rayleigh equation of the second kind for a multi-layer system. The starting point is the reduced Rayleigh equation

$$\int \mathbf{\Theta}_{31}^{++}(\mathbf{p}\,|\,\mathbf{q})\;\mathbf{R}(\mathbf{q}\,|\,\mathbf{p}_0)\,\frac{\mathrm{d}^2q}{(2\pi)^2} = -\,\mathbf{\Theta}_{31}^{+-}(\mathbf{p}\,|\,\mathbf{p}_0)\,, \tag{3.46}$$

where the transfer kernel and right-hand side are defined by

$$\mathbf{\Theta}_{31}^{a_3a_1}(\mathbf{p}\,|\,\mathbf{q}) = \sum_{a_2=\pm} a_2 \int \mathbf{\Theta}_{32}^{a_3a_2}(\mathbf{p}\,|\,\mathbf{p}')\,\mathbf{\Theta}_{21}^{a_2a_1}(\mathbf{p}'\,|\,\mathbf{q})\frac{\mathrm{d}^2p'}{(2\pi)^2} \tag{3.47a}$$

$$\begin{aligned}\mathbf{\Theta}_{32}^{a_3a_2}(\mathbf{p}\,|\,\mathbf{p}') &= \alpha_3^{-1}(\mathbf{p})\,\mathcal{J}_{32}^{a_3a_2}(\mathbf{p}\,|\,\mathbf{p}')\;\mathbf{M}_{32}^{a_3a_2}(\mathbf{p}\,|\,\mathbf{p}') \\ &= \left[\frac{(2\pi)^2\,\delta(\mathbf{p}-\mathbf{p}')}{a_3\alpha_3(\mathbf{p}) - a_2\alpha_2(\mathbf{p}')} + \mathcal{K}_{32}^{a_3a_2}(\mathbf{p}\,|\,\mathbf{p}')\right]\,\alpha_3^{-1}(\mathbf{p})\;\mathbf{M}_{32}^{a_3a_2}(\mathbf{p}\,|\,\mathbf{p}')\end{aligned} \tag{3.47b}$$

$$\begin{aligned}\mathbf{\Theta}_{21}^{a_2a_1}(\mathbf{p}'\,|\,\mathbf{q}) &= \alpha_2^{-1}(\mathbf{p}')\,\mathcal{J}_{21}^{a_2a_1}(\mathbf{p}'\,|\,\mathbf{q})\;\mathbf{M}_{21}^{a_2a_1}(\mathbf{p}'\,|\,\mathbf{q}) \\ &= \left[\frac{(2\pi)^2\,\delta(\mathbf{p}'-\mathbf{q})}{a_2\alpha_2(\mathbf{p}') - a_1\alpha_1(\mathbf{q})} + \mathcal{K}_{21}^{a_2a_1}(\mathbf{p}'\,|\,\mathbf{q})\right]\,\alpha_2^{-1}(\mathbf{p}')\,e^{-i(a_2\alpha_2(\mathbf{p}')-a_1\alpha_1(\mathbf{q}))d}\;\mathbf{M}_{21}^{a_2a_1}(\mathbf{p}'\,|\,\mathbf{q})\,.\end{aligned} \tag{3.47c}$$

Here we have chosen to place the origin of the coordinate system on the average of the second surface so that we can write $\zeta_1 = d + h_1$ and $\zeta_2 = h_2$ where d is the average thickness of the film and h_1 and h_2 are function with zero average. The h_i are then replacing the ζ_i in the definition

of the $\mathcal{K}_{lm}^{ba}$. Thanks to the delta-$\mathcal{K}$ splitting, the two-interface transfer kernel can be written as

$$\boldsymbol{\Theta}_{31}^{a_3a_1}(\mathbf{p}\,|\,\mathbf{q}) = (2\pi)^2\,\delta(\mathbf{p}-\mathbf{q})\;\mathbf{K}_1^{a_3a_1}(\mathbf{p}) + \mathbf{K}_2^{a_3a_1}(\mathbf{p}\,|\,\mathbf{q}) + \mathbf{K}_3^{a_3a_1}(\mathbf{p}\,|\,\mathbf{q}) + \mathbf{K}_4^{a_3a_1}(\mathbf{p}\,|\,\mathbf{q})\,, \tag{3.48}$$

where we have defined

$$\mathbf{K}_1^{a_3a_1}(\mathbf{p}) \overset{\text{def}}{=} \sum_{a_2=\pm} a_2 \frac{e^{-i(a_2\alpha_2(\mathbf{p})-a_1\alpha_1(\mathbf{p}))d}\;\mathbf{M}_{32}^{a_3a_2}(\mathbf{p}\,|\,\mathbf{p})\,\mathbf{M}_{21}^{a_2a_1}(\mathbf{p}\,|\,\mathbf{p})}{\alpha_3(\mathbf{p})\alpha_2(\mathbf{p})(a_3\alpha_3(\mathbf{p})-a_2\alpha_2(\mathbf{p}))(a_2\alpha_2(\mathbf{p})-a_1\alpha_1(\mathbf{p}))} \tag{3.49a}$$

$$\mathbf{K}_2^{a_3a_1}(\mathbf{p}\,|\,\mathbf{q}) \overset{\text{def}}{=} \sum_{a_2=\pm} a_2 \frac{e^{-i(a_2\alpha_2(\mathbf{p})-a_1\alpha_1(\mathbf{q}))d}\;\mathcal{K}_{21}^{a_2a_1}(\mathbf{p}\,|\,\mathbf{q})\;\mathbf{M}_{32}^{a_3a_2}(\mathbf{p}\,|\,\mathbf{p})\,\mathbf{M}_{21}^{a_2a_1}(\mathbf{p}\,|\,\mathbf{q})}{\alpha_3(\mathbf{p})\alpha_2(\mathbf{p})(a_3\alpha_3(\mathbf{p})-a_2\alpha_2(\mathbf{p}))} \tag{3.49b}$$

$$\mathbf{K}_3^{a_3a_1}(\mathbf{p}\,|\,\mathbf{q}) \overset{\text{def}}{=} \sum_{a_2=\pm} a_2 \frac{e^{-i(a_2\alpha_2(\mathbf{q})-a_1\alpha_1(\mathbf{q}))d}\;\mathcal{K}_{32}^{a_3a_2}(\mathbf{p}\,|\,\mathbf{q})\;\mathbf{M}_{32}^{a_3a_2}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{M}_{21}^{a_2a_1}(\mathbf{q}\,|\,\mathbf{q})}{\alpha_3(\mathbf{p})\alpha_2(\mathbf{q})(a_2\alpha_2(\mathbf{q})-a_1\alpha_1(\mathbf{q}))} \tag{3.49c}$$

$$\begin{aligned}\mathbf{K}_4^{a_3a_1}(\mathbf{p}\,|\,\mathbf{q}) \overset{\text{def}}{=} \sum_{a_2=\pm} a_2 \int & \frac{e^{-i(a_2\alpha_2(\mathbf{p}')-a_1\alpha_1(\mathbf{q}))d}}{\alpha_3(\mathbf{p})\alpha_2(\mathbf{p}')}\,\mathcal{K}_{32}^{a_3a_2}(\mathbf{p}\,|\,\mathbf{p}')\,\mathbf{M}_{32}^{a_3a_2}(\mathbf{p}\,|\,\mathbf{p}') \\ & \times \mathcal{K}_{21}^{a_2a_1}(\mathbf{p}'\,|\,\mathbf{q})\,\mathbf{M}_{21}^{a_2a_1}(\mathbf{p}'\,|\,\mathbf{q})\,\frac{\mathrm{d}^2p'}{(2\pi)^2}\,.\end{aligned} \tag{3.49d}$$

By plugging Eq. (3.48) into Eq. (3.46) we obtain

$$\begin{aligned}\mathbf{K}_1^{++}(\mathbf{p})\;\mathbf{R}(\mathbf{p}\,|\,\mathbf{p}_0) + \int \sum_{j=2}^{4} \mathbf{K}_j^{++}(\mathbf{p}\,|\,\mathbf{q})\;\mathbf{R}(\mathbf{q}\,|\,\mathbf{p}_0)\;\frac{\mathrm{d}^2q}{(2\pi)^2} \\ = -(2\pi)^2\delta(\mathbf{p}-\mathbf{p}_0)\,\mathbf{K}_1^{+-}(\mathbf{p}_0) - \sum_{j=2}^{4}\mathbf{K}_j^{+-}(\mathbf{p}\,|\,\mathbf{p}_0)\,.\end{aligned} \tag{3.50}$$

By setting the change of unknown

$$\Delta\,\mathbf{R}(\mathbf{p}\,|\,\mathbf{p}_0) \overset{\text{def}}{=} \mathbf{R}(\mathbf{p}\,|\,\mathbf{p}_0) - \boldsymbol{\rho}^{(0)}(\mathbf{p}_0)\,(2\pi)^2\,\delta(\mathbf{p}-\mathbf{p}_0)\,, \tag{3.51}$$

where $\boldsymbol{\rho}^{(0)}(\mathbf{p}_0)$ is the amplitude of the associated Fabry-Perot system (i.e. the response for the similar system but with planar interfaces $h_j = 0$) given by

$$\boldsymbol{\rho}^{(0)}(\mathbf{p}_0) \overset{\text{def}}{=} \left[\mathbf{K}_1^{++}(\mathbf{p}_0)\right]^{-1}\,\mathbf{K}_1^{+-}(\mathbf{p}_0)\,, \tag{3.52}$$

we get the reduced Rayleigh equation of the second kind for $\Delta\,\mathbf{R}(\mathbf{p}\,|\,\mathbf{p}_0)$

$$\begin{aligned}\mathbf{K}_1^{++}(\mathbf{p})\,\Delta\,\mathbf{R}(\mathbf{p}\,|\,\mathbf{p}_0) + \int \sum_{j=2}^{4} \mathbf{K}_j^{++}(\mathbf{p}\,|\,\mathbf{q})\,\Delta\,\mathbf{R}(\mathbf{q}\,|\,\mathbf{p}_0)\;\frac{\mathrm{d}^2q}{(2\pi)^2} \\ = -\sum_{j=2}^{4}\mathbf{K}_j^{+-}(\mathbf{p}\,|\,\mathbf{p}_0) - \sum_{j=2}^{4}\mathbf{K}_j^{++}(\mathbf{p}\,|\,\mathbf{p}_0)\,\boldsymbol{\rho}^{(0)}(\mathbf{p}_0)\,.\end{aligned} \tag{3.53}$$

So what have we gained? The reader may argue that we still have matrix-matrix products to be evaluated in the discretized version Eq. (3.53) if we want to solve it. This seems clear from the terms in $\mathbf{K}_4^{a_3a_1}$ since $\mathbf{K}_4^{a_3a_1}$ is a integral composition of two modified single-interface transfer kernels, similar to $\boldsymbol{\Theta}_{31}^{a_3a_1}$. This would indeed be the case if one would use a direct solver, but the power of the reduced Rayleigh equation of the second kind resides in its iterative resolution. Indeed, starting with the initial guess $\Delta\,\mathbf{R}^{(0)} = \mathbf{0}$ and then taking successive iterates, no matrix-matrix multiplication is necessary any longer but only matrix-vector multiplications.

Indeed numerically there will be terms of the form $\mathbf{A}\,\mathbf{B}\,\Delta\mathbf{R}^{(k)}$, where $\mathbf{A}$ and $\mathbf{B}$ are square matrices, which can be evaluated by computing first $\mathbf{Y} = \mathbf{B}\,\Delta\mathbf{R}^{(k)}$ and then $\mathbf{A}\,\mathbf{Y}$, i.e. using only matrix-vector multiplications. This is a significant reduction of complexity, under the assumption that the fixed point iterative method converges quickly.

We have shown here a promising idea for reducing the complexity of associated with solving the reduced Rayleigh equations for multi-layer system with rough interfaces. This method, however, does not cure the inherent unstable feature of the reduced Rayleigh equations as given in their "usual" form as discussed earlier. The main challenge for future research will be to find a form of the reduced Rayleigh equation of the second kind which is numerically stable.

3.6 Summary

In the present chapter the reduced Rayleigh equations have been generalized to systems of stacks of homogeneous layers separated by arbitrary interfaces. The derivation is based on the composition of the elementary transfer equations derived in Chapter 2 for each successive pair of layer and yields transfer equations for the resulting system. The increase in complexity associated with the addition of layers with rough interfaces has been discussed and was shown to represent a computational challenge. A possible solution for this issue, based on the reduced Rayleigh equations of the second kind, have been suggested to avoid setting up the costly multi-layer transfer kernel. In addition, the numerical stability of the obtained reduced Rayleigh equations have been analyzed and numerically stable expressions have been suggested in the case of systems possessing one rough and one planar interface. Finding stable expressions to implement in the case where two rough interfaces are present still remains to be achieved.

Chapter 4

Periodic structures

The present chapter is devoted to the analysis of the consequence of the periodicity of the interfaces in the reduced Rayleigh equations. First, from the reduced Rayleigh equations derived in Chapter 2 in the case of a single interface separating two media, we will recover the grating formula or the Bloch-Floquet theorem. This result states that the only diffracted modes allowed in this system are the ones with an in-plane wave-vector that differs from the in-plane wave vector of the incident wave by a reciprocal lattice of the surface. Consequently, the integral equations for the reflection and transmission amplitudes will reduce to an infinite but countable set of linear equations. Therefore, from a numerical point of view, no discretization in the in-plane wave vector space is needed compared to the case of an arbitrary surface, as the unknown is the discrete set of reflection and transmission amplitudes and the resolution parameter of numerical method is simply a cut-off in the in-plane wave-vector space, or equivalently in the set of diffractive orders. We will make use of this result in Chapter 5.

Then, based on the theoretical framework developed in Chapter 3, we will consider the case of a system made of a stack of layers where all interfaces are periodic but may or may not share the same basis lattice vectors. We will restrict our analysis to the case of a system having two interfaces for the sake of clarity and illustrate that care must be taken when treating such systems by distinguishing degenerate and non-degenerate cases regarding the summation of reciprocal lattice vectors.

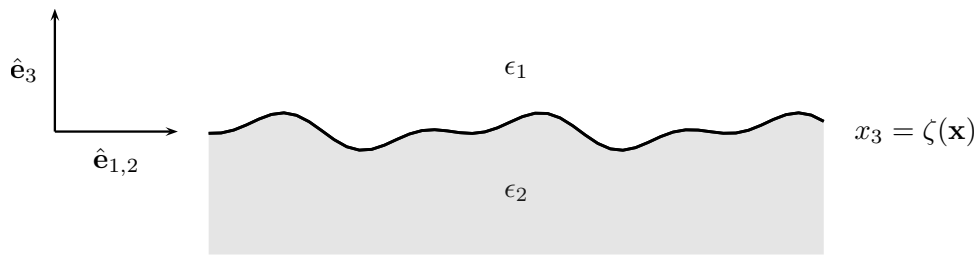


Figure 4.1: A system of two half spaces separated by a periodic interface.

4.1 Single periodic interface

We consider a single interface separating two media of dielectric constants ϵ_1 and ϵ_2 and whose profile is periodic, i.e. there exist $\mathbf{a}_1, \mathbf{a}_2 \in \mathbb{R}^2$, such that for all $\boldsymbol{\ell} \in \mathbb{Z}^2$ and for all $\mathbf{x} \in \mathbb{R}^2$, we have

$$\zeta(\mathbf{x} + \mathbf{a}^{(\boldsymbol{\ell})}) = \zeta(\mathbf{x}) \,, \tag{4.1}$$

where $\mathbf{a}^{(\boldsymbol{\ell})} = \ell_1 \mathbf{a}_1 + \ell_2 \mathbf{a}_2$ is a lattice vector. In particular, we will assume in the following that $\mathbf{a}_1, \mathbf{a}_2$ are vectors forming the primitive lattice cell a_c. Hence the profile at any point can be constructed from the profile in the unit cell. In mathematical terms, this means that for all $\mathbf{x} \in \mathbb{R}^2$, there exists $\boldsymbol{\ell} \in \mathbb{Z}^2$, and $\mathbf{x}_c \in a_c$, such that

$$\mathbf{x} = \mathbf{x}_c + \mathbf{a}^{(\boldsymbol{\ell})} \,, \tag{4.2}$$

and consequently,

$$\zeta(\mathbf{x}) = \zeta(\mathbf{x}_c) \,. \tag{4.3}$$

We recall that in the reduced Rayleigh equations for the reflection and transmission amplitudes Eqs. (2.51) and (2.52), the information about the interface is entirely encoded in the scalar kernel (and right-hand-side) factor $\mathcal{J}^{b,a}_{l,m}$. Let us analyze the consequences of the periodicity on this factor. We recall that for $\mathbf{p}$ and $\mathbf{q} \in \mathbb{R}^2$,

$$\mathcal{J}^{b,a}_{l,m}(\mathbf{p} \,|\, \mathbf{q}) = \begin{cases} I^{b,a}_{l,m}(\mathbf{p} \,|\, \mathbf{q}) \,, & \mathbf{q} \notin \mathcal{R}^{b,a}_{l,m}(\mathbf{p}) \\ -i\,\hat{\zeta}(\mathbf{p} - \mathbf{q}), & \mathbf{q} \in \mathcal{R}^{b,a}_{l,m}(\mathbf{p}) \end{cases} \tag{4.3}$$

For $\mathbf{q} \notin \mathcal{R}^{b,a}_{l,m}(\mathbf{p})$, we have

$$I^{b,a}_{l,m}(\mathbf{p} \,|\, \mathbf{q}) = (b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q}))^{-1} \int e^{-i(\mathbf{k}^b_l(\mathbf{p}) - \mathbf{k}^a_m(\mathbf{q}))\cdot\mathbf{s}(\mathbf{x})} \, \mathrm{d}^2x \,. \tag{4.4}$$

The integral becomes

$$\begin{aligned} \int e^{-i(\mathbf{k}^b_l(\mathbf{p}) - \mathbf{k}^a_m(\mathbf{q}))\cdot\mathbf{s}(\mathbf{x})} \, \mathrm{d}^2x &= \int_{\mathbb{R}^2} e^{-i(\mathbf{p}-\mathbf{q})\cdot\mathbf{x}} \, e^{-i(b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q}))\zeta(\mathbf{x})} \, \mathrm{d}^2x \\ &= \sum_{\boldsymbol{\ell}\in\mathbb{Z}^2} \int_{a_c} e^{-i(\mathbf{p}-\mathbf{q})\cdot(\mathbf{x}_c + \mathbf{a}^{(\boldsymbol{\ell})})} \, e^{-i(b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q}))\zeta(\mathbf{x}_c)} \, \mathrm{d}^2x_c \\ &= \sum_{\boldsymbol{\ell}\in\mathbb{Z}^2} e^{-i(\mathbf{p}-\mathbf{q})\cdot\mathbf{a}^{(\boldsymbol{\ell})}} \int_{a_c} e^{-i(\mathbf{p}-\mathbf{q})\cdot\mathbf{x}_c} \, e^{-i(b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q}))\zeta(\mathbf{x}_c)} \, \mathrm{d}^2x_c \,. \end{aligned} \tag{4.5}$$

Now using Poisson summation, we get

$$\sum_{\boldsymbol{\ell}\in\mathbb{Z}^2} e^{-i(\mathbf{p}-\mathbf{q})\cdot\mathbf{a}^{(\boldsymbol{\ell})}} = \frac{(2\pi)^2}{a_c} \sum_{\boldsymbol{\ell}\in\mathbb{Z}^2} \delta(\mathbf{p} - \mathbf{q} - \mathbf{G}^{(\boldsymbol{\ell})}) \,, \tag{4.6}$$

where $\mathbf{G}^{(\boldsymbol{\ell})}$ is a reciprocal lattice vector defined as

$$\mathbf{G}^{(\boldsymbol{\ell})} = \ell_1 \mathbf{b}_1 + \ell_2 \mathbf{b}_2 \,, \tag{4.7}$$

where the basis reciprocal lattice vector are defined, for $i, j \in \{1, 2\}$, by

$$\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\, \delta_{ij} \,. \tag{4.8}$$

Thus we obtain

$$I^{b,a}_{l,m}(\mathbf{p}\,|\,\mathbf{q}) = (2\pi)^2 \sum_{\boldsymbol{\ell}\in\mathbb{Z}^2} \delta(\mathbf{p}-\mathbf{q}-\mathbf{G}^{(\boldsymbol{\ell})})\, \tilde{I}^{b,a}_{l,m}(\mathbf{p}\,|\,\mathbf{q})\,, \tag{4.9}$$

where we have introduced

$$\tilde{I}^{b,a}_{l,m}(\mathbf{p}\,|\,\mathbf{q}) \stackrel{\text{def}}{=} (b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q}))^{-1}\, \frac{1}{a_c} \int_{a_c} e^{-i(\mathbf{k}^b_l(\mathbf{p}) - \mathbf{k}^a_m(\mathbf{q}))\cdot\mathbf{s}(\mathbf{x})}\, \mathrm{d}^2x\ . \tag{4.10}$$

For $\mathbf{q} \in \mathcal{R}^{b,a}_{l,m}(\mathbf{p})$, we have to consider

$$\begin{aligned}
\hat{\zeta}(\mathbf{p}-\mathbf{q}) &= \int \zeta(\mathbf{x})\, e^{-i(\mathbf{p}-\mathbf{q})\cdot\mathbf{x}}\, \mathrm{d}^2x \\
&= \sum_{\boldsymbol{\ell}\in\mathbb{Z}^2} \int_{a_c} \zeta(\mathbf{x}_c)\, e^{-i(\mathbf{p}-\mathbf{q})\cdot(\mathbf{x}_c + \mathbf{a}^{(l)})}\, \mathrm{d}^2x_c \\
&= \sum_{\boldsymbol{\ell}\in\mathbb{Z}^2} e^{-i(\mathbf{p}-\mathbf{q})\cdot\mathbf{a}^{(l)}} a_c\, \hat{\zeta}_{a_c}(\mathbf{p}-\mathbf{q}) \\
&= (2\pi)^2 \sum_{\boldsymbol{\ell}\in\mathbb{Z}^2} \delta(\mathbf{p}-\mathbf{q}-\mathbf{G}^{(\boldsymbol{\ell})})\, \hat{\zeta}_{a_c}(\mathbf{G}^{(\boldsymbol{\ell})})\,,
\end{aligned} \tag{4.11}$$

where we have introduced

$$\hat{\zeta}_{a_c}(\mathbf{q}) \stackrel{\text{def}}{=} \frac{1}{a_c} \int_{a_c} \zeta(\mathbf{x})\, e^{-i\,\mathbf{q}\,\cdot\,\mathbf{x}}\, \mathrm{d}^2x\ . \tag{4.12}$$

To sum up, we have

$$\mathcal{J}^{b,a}_{l,m}(\mathbf{p}\,|\,\mathbf{q}) = (2\pi)^2 \sum_{\boldsymbol{\ell}\in\mathbb{Z}^2} \delta(\mathbf{p}-\mathbf{q}-\mathbf{G}^{(\boldsymbol{\ell})})\, \tilde{\mathcal{J}}^{b,a}_{l,m}(\mathbf{p}\,|\,\mathbf{q})\,, \tag{4.13}$$

with

$$\tilde{\mathcal{J}}^{b,a}_{l,m}(\mathbf{p}\,|\,\mathbf{q}) \stackrel{\text{def}}{=} \begin{cases} \tilde{I}^{b,a}_{l,m}(\mathbf{p}\,|\,\mathbf{q}), & \mathbf{q} \notin \mathcal{R}^{b,a}_{l,m}(\mathbf{p}) \\ -i\,\hat{\zeta}_{a_c}(\mathbf{p}-\mathbf{q}), & \mathbf{q} \in \mathcal{R}^{b,a}_{l,m}(\mathbf{p})\ . \end{cases} \tag{4.14}$$

4.1.1 Reflection of a plane wave by a periodic surface

By plugging the expression of $\mathcal{J}^{b,a}_{l,m}$ in Eq. (4.13) into the reduced Rayleigh equation for the reflection amplitude, Eq. (2.51), we obtain

$$\begin{aligned}
&\sum_{\boldsymbol{\ell}\in\mathbb{Z}^2} \tilde{\mathcal{J}}^{+,+}_{2,1}(\mathbf{p}\,|\,\mathbf{p}-\mathbf{G}^{(\boldsymbol{\ell})})\, \mathbf{M}^{+,+}_{2,1}(\mathbf{p}\,|\,\mathbf{p}-\mathbf{G}^{(\boldsymbol{\ell})})\, \mathbf{R}(\mathbf{p}-\mathbf{G}^{(\boldsymbol{\ell})}\,|\mathbf{p}_0) \\
&= -(2\pi)^2 \sum_{\boldsymbol{\ell}\in\mathbb{Z}^2} \delta(\mathbf{p}-\mathbf{p}_0-\mathbf{G}^{(\boldsymbol{\ell})})\, \tilde{\mathcal{J}}^{+,-}_{2,1}(\mathbf{p}_0+\mathbf{G}^{(\boldsymbol{\ell})}\,|\,\mathbf{p}_0)\, \mathbf{M}^{+,-}_{2,1}(\mathbf{p}_0+\mathbf{G}^{(\boldsymbol{\ell})}\,|\mathbf{p}_0)\,,
\end{aligned} \tag{4.15}$$

where we have performed the integration on the left hand side using the fundamental property of the Dirac delta. The distribution on the right hand side being a Dirac comb, it is clear from the theory of distribution that a solution $\mathbf{R}$ must be itself a Dirac comb, with the same support as the Dirac masses on the right hand side. In other words, there exists a sequence of reflection amplitude matrices $(\mathbf{R}^{(\mathbf{m})}(\mathbf{p}_0))_{\mathbf{m}\in\mathbb{Z}^2}$, such that

$$\mathbf{R}(\mathbf{p}\,|\mathbf{p}_0) = (2\pi)^2 \sum_{\mathbf{m}\in\mathbb{Z}^2} \delta(\mathbf{p}-\mathbf{p}_\mathbf{m})\ \mathbf{R}^{(\mathbf{m})}(\mathbf{p}_0)\,, \tag{4.16}$$

with $\mathbf{p_m} \overset{\text{def}}{=} \mathbf{p}_0 + \mathbf{G}^{(\mathbf{m})}$. From a physical point of view, this means that the only waves allowed to be reflected are the ones having a transverse wave vector lying on the reciprocal lattice up to the incident transverse wave vector $\mathbf{p}_0$ (see Fig. 1.9 in Chapter 1). In other words, we have just shown that the reduced Rayleigh equation is consistent with the grating formula. Inserting Eq. (4.16) into Eq. (4.15) we get

$$\begin{aligned}&\sum_{\boldsymbol{\ell}\in\mathbb{Z}^2}\sum_{\mathbf{m}\in\mathbb{Z}^2}\delta(\mathbf{p}-\mathbf{p}_{\mathbf{m}+\boldsymbol{\ell}})\,\tilde{\mathcal{J}}^{+,+}_{2,1}(\mathbf{p}_{\mathbf{m}+\boldsymbol{\ell}}\,|\,\mathbf{p}_{\mathbf{m}})\;\mathbf{M}^{+,+}_{2,1}(\mathbf{p}_{\mathbf{m}+\boldsymbol{\ell}}\,|\,\mathbf{p}_{\mathbf{m}})\;\mathbf{R}^{(\mathbf{m})}(\mathbf{p}_0)\\&=-\sum_{\boldsymbol{\ell}\in\mathbb{Z}^2}\delta(\mathbf{p}-\mathbf{p}_{\boldsymbol{\ell}})\,\tilde{\mathcal{J}}^{+,-}_{2,1}(\mathbf{p}_{\boldsymbol{\ell}}\,|\,\mathbf{p}_0)\;\mathbf{M}^{+,-}_{2,1}(\mathbf{p}_{\boldsymbol{\ell}}\,|\mathbf{p}_0)\,.\end{aligned}\tag{4.17}$$

We now change the summation index $\boldsymbol{\ell}\leftrightarrow\boldsymbol{\ell}+\mathbf{m}$ in the left hand side to obtain

$$\begin{aligned}&\sum_{\boldsymbol{\ell}\in\mathbb{Z}^2}\delta(\mathbf{p}-\mathbf{p}_{\boldsymbol{\ell}})\sum_{\mathbf{m}\in\mathbb{Z}^2}\tilde{\mathcal{J}}^{+,+}_{2,1}(\mathbf{p}_{\boldsymbol{\ell}}\,|\,\mathbf{p}_{\mathbf{m}})\;\mathbf{M}^{+,+}_{2,1}(\mathbf{p}_{\boldsymbol{\ell}}\,|\,\mathbf{p}_{\mathbf{m}})\;\mathbf{R}^{(\mathbf{m})}(\mathbf{p}_0)\\&=-\sum_{\boldsymbol{\ell}\in\mathbb{Z}^2}\delta(\mathbf{p}-\mathbf{p}_{\boldsymbol{\ell}})\,\tilde{\mathcal{J}}^{+,-}_{2,1}(\mathbf{p}_{\boldsymbol{\ell}}\,|\,\mathbf{p}_0)\;\mathbf{M}^{+,-}_{2,1}(\mathbf{p}_{\boldsymbol{\ell}}\,|\mathbf{p}_0)\,.\end{aligned}\tag{4.18}$$

Now by integration against a smooth test function φ having a compact support included in a disc of radius $\min(|\,\mathbf{b}_1\,|,|\,\mathbf{b}_2\,|)/2$ centered on $\mathbf{p}_{\boldsymbol{\ell}}$ and such that $\varphi(\mathbf{p}_{\boldsymbol{\ell}})=1$, it follows that the coefficients of the Dirac delta centered on $\mathbf{p}_{\boldsymbol{\ell}}$ on the left and right hand side must be equal, and this for all $\boldsymbol{\ell}$. We obtain the following infinite linear system, namely that for all $\boldsymbol{\ell}\in\mathbb{Z}^2$,

$$\sum_{\mathbf{m}\in\mathbb{Z}^2}\tilde{\mathcal{J}}^{+,+}_{2,1}(\mathbf{p}_{\boldsymbol{\ell}}\,|\,\mathbf{p}_{\mathbf{m}})\;\mathbf{M}^{+,+}_{2,1}(\mathbf{p}_{\boldsymbol{\ell}}\,|\,\mathbf{p}_{\mathbf{m}})\;\mathbf{R}^{(\mathbf{m})}(\mathbf{p}_0)=-\tilde{\mathcal{J}}^{+,-}_{2,1}(\mathbf{p}_{\boldsymbol{\ell}}\,|\,\mathbf{p}_0)\;\mathbf{M}^{+,-}_{2,1}(\mathbf{p}_{\boldsymbol{\ell}}\,|\mathbf{p}_0)\,.\tag{4.19}$$

4.1.2 Transmission of a plane wave through a periodic surface

Similarly to the reasoning applied in the previous section, by plugging Eq. (4.13) into the reduced Rayleigh equation for the transmission amplitude, Eq. (2.52), we obtain

$$\sum_{\boldsymbol{\ell}\in\mathbb{Z}^2}\tilde{\mathcal{J}}^{-,-}_{1,2}(\mathbf{p}\,|\,\mathbf{p}-\mathbf{G}^{(\boldsymbol{\ell})})\,\mathbf{M}^{-,-}_{1,2}(\mathbf{p}\,|\,\mathbf{p}-\mathbf{G}^{(\boldsymbol{\ell})})\,\mathbf{T}(\mathbf{p}-\mathbf{G}^{(\boldsymbol{\ell})}\,|\mathbf{p}_0)=\frac{2\sqrt{\epsilon_1\epsilon_2}\,\alpha_1(\mathbf{p}_0)}{\epsilon_2-\epsilon_1}\,(2\pi)^2\,\delta(\mathbf{p}-\mathbf{p}_0)\,\mathbf{I}_2\,,\tag{4.20}$$

where we have performed the integration on the left hand side using the fundamental property of the Dirac delta. The distribution on the right hand side is now only a single Dirac delta. Since the transmission amplitude is taken at different point on a lattice $\mathbf{p}-\mathbf{G}^{(\boldsymbol{\ell})}$, the same argument as before applies and a solution $\mathbf{T}$ must be itself a Dirac comb. In other words, there exists a sequence of transmission amplitude matrices $(\mathbf{T}^{(\mathbf{m})}(\mathbf{p}_0))_{\mathbf{m}\in\mathbb{Z}^2}$ such that,

$$\mathbf{T}(\mathbf{p}\,|\mathbf{p}_0)=(2\pi)^2\sum_{\mathbf{m}\in\mathbb{Z}^2}\delta(\mathbf{p}-\mathbf{p}_{\mathbf{m}})\;\mathbf{T}^{(\mathbf{m})}(\mathbf{p}_0)\,.\tag{4.21}$$

Inserting the above equation into Eq. (4.20) we get

$$\begin{aligned}&\sum_{\boldsymbol{\ell}\in\mathbb{Z}^2}\sum_{\mathbf{m}\in\mathbb{Z}^2}\delta(\mathbf{p}-\mathbf{p}_{\mathbf{m}+\boldsymbol{\ell}})\,\tilde{\mathcal{J}}^{-,-}_{1,2}(\mathbf{p}_{\mathbf{m}+\boldsymbol{\ell}}\,|\,\mathbf{p}_{\mathbf{m}})\;\mathbf{M}^{-,-}_{1,2}(\mathbf{p}_{\mathbf{m}+\boldsymbol{\ell}}\,|\,\mathbf{p}_{\mathbf{m}})\;\mathbf{T}^{(\mathbf{m})}(\mathbf{p}_0)\\&\qquad=\frac{2\sqrt{\epsilon_1\epsilon_2}\,\alpha_1(\mathbf{p}_0)}{\epsilon_2-\epsilon_1}\,\delta(\mathbf{p}-\mathbf{p}_0)\,\mathbf{I}_2\,.\end{aligned}\tag{4.22}$$

We now change the summation index $\boldsymbol{\ell} \leftrightarrow \boldsymbol{\ell} + \mathbf{m}$ on the left hand side to obtain

$$\sum_{\boldsymbol{\ell}\in\mathbb{Z}^2} \delta(\mathbf{p}-\mathbf{p}_{\boldsymbol{\ell}}) \sum_{\mathbf{m}\in\mathbb{Z}^2} \tilde{\mathcal{J}}^{-;-}_{1,2}(\mathbf{p}_{\boldsymbol{\ell}}\,|\,\mathbf{p}_{\mathbf{m}})\ \mathbf{M}^{-;-}_{1,2}(\mathbf{p}_{\boldsymbol{\ell}}\,|\,\mathbf{p}_{\mathbf{m}})\ \mathbf{T}^{(\mathbf{m})}(\mathbf{p}_0) = \frac{2\sqrt{\epsilon_1\epsilon_2}\,\alpha_1(\mathbf{p}_0)}{\epsilon_2-\epsilon_1}\,\delta(\mathbf{p}-\mathbf{p}_0)\,\mathbf{I}_2\,. \quad (4.22)$$

Now by integration against a smooth test function φ having a compact support included in a disc of radius $\min(|\,\mathbf{b}_1\,|,|\,\mathbf{b}_2\,|)/2$ centered on $\mathbf{p}_{\boldsymbol{\ell}}$ and such that $\varphi(\mathbf{p}_{\boldsymbol{\ell}})=1$, it follows that for all $\boldsymbol{\ell}\in\mathbb{Z}^2$,

$$\sum_{\mathbf{m}\in\mathbb{Z}^2} \tilde{\mathcal{J}}^{-;-}_{1,2}(\mathbf{p}_{\boldsymbol{\ell}}\,|\,\mathbf{p}_{\mathbf{m}})\ \mathbf{M}^{-;-}_{1,2}(\mathbf{p}_{\boldsymbol{\ell}}\,|\,\mathbf{p}_{\mathbf{m}})\ \mathbf{T}^{(\mathbf{m})}(\mathbf{p}_0) = \frac{2\sqrt{\epsilon_1\epsilon_2}\,\alpha_1(\mathbf{p}_0)}{\epsilon_2-\epsilon_1}\,\delta_{\boldsymbol{\ell}}\,\mathbf{I}_2\,. \quad (4.23)$$

Here $\delta_{\boldsymbol{\ell}} \stackrel{\text{def}}{=} \delta_{\ell_1,0}\,\delta_{\ell_2,0}$ denotes a Kronecker delta, i.e. $\delta_{\boldsymbol{\ell}}=1$ if $\boldsymbol{\ell}=\mathbf{0}$ and 0 otherwise.

The reduced Rayleigh equations which were integral equations in the case of an arbitrary interface reduce, in the case of a periodic interface, to a infinite dimensional countable linear system of equations as can be seen from Eqs. (4.19) and (4.23).

4.2 Multilayer with periodic interfaces

In this section we combine what has been done in the previous section and in Chapter 3 to derive the reduced Rayleigh equations for the problem of scattering by a multilayer system whose interfaces are periodic. To stay as general as possible we treat the case where each surface has *a priori* a period and a profile in the unit cell different from each other.

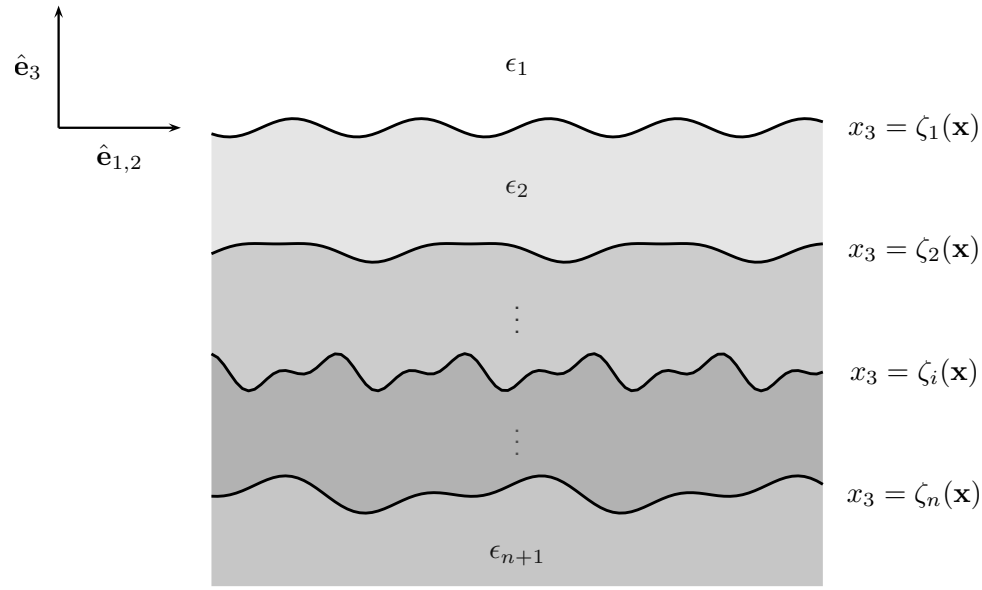


Figure 4.2: Multilayer system with periodic interfaces.

4.3 The system

The system under consideration is described by a stack of $n+1$ layers separated by n periodic interfaces, i.e. for all $i \in [\![1,n]\!]$, there exist $\mathbf{a}_{i,1}, \mathbf{a}_{i,2} \in \mathbb{R}^2$, such that for all $\boldsymbol{\ell} \in \mathbb{Z}^2$ we have

$$\zeta_i(\mathbf{x} + \mathbf{a}_i^{(\boldsymbol{\ell})}) = \zeta_i(\mathbf{x}) \ . \tag{4.24}$$

As in the previous section, we define $\mathbf{a}_i^{(\boldsymbol{\ell})} \stackrel{\text{def}}{=} \ell_1 \, \mathbf{a}_{i,1} + \ell_2 \, \mathbf{a}_{i,2}$, and we take $\mathbf{a}_{i,1}, \mathbf{a}_{i,2}$ as the primitive basis vectors of the unit cell of the i^{th} interface. Following section 4.1, each factor $\mathcal{J}_{i+1,i}^{a_{i+1},a_i}$ in the transfer kernel $\boldsymbol{\Theta}_{n+1,1}^{a_{n+1},a_1}$, can be written as

$$\mathcal{J}_{i+1,i}^{a_{i+1},a_i}(\mathbf{p}_{i+1} \,|\, \mathbf{p}_i) = (2\pi)^2 \sum_{\boldsymbol{\ell} \in \mathbb{Z}^2} \delta(\mathbf{p}_{i+1} - \mathbf{p}_i - \mathbf{G}_i^{(\boldsymbol{\ell})}) \ \tilde{\mathcal{J}}_{i+1,i}^{a_{i+1},a_i}(\mathbf{p}_{i+1} \,|\, \mathbf{p}_{i+1} - \mathbf{G}_i^{(\boldsymbol{\ell})}) \, . \tag{4.25}$$

Here we recall that the interface between media i, and $i+1$ is indexed as i, and $a_{c,i}$, $\mathbf{G}_i^{(\boldsymbol{\ell})}$ are respectively the area of the unit cell and a reciprocal lattice vector of the considered interface. In particular, we recall that the reciprocal lattice vectors are such that,

$$\mathbf{G}_i^{(\boldsymbol{\ell})} = \ell_1 \, \mathbf{b}_{i,1} + \ell_2 \, \mathbf{b}_{i,2} \tag{4.26}$$

where the primitive vectors of the reciprocal lattice $\mathbf{b}_{i,1}$, $\mathbf{b}_{i,2}$ are defined by

$$\mathbf{a}_{i,j} \cdot \mathbf{b}_{i,k} = 2\pi \, \delta_{jk} \, . \tag{4.27}$$

Consequently, the forward transfer kernel reads

$$\begin{aligned} \boldsymbol{\Theta}_{n+1,1}^{a_{n+1},a_1}(\mathbf{p} \,|\, \mathbf{q}) = (2\pi)^2 \sum_{\substack{a_n = \pm \\ \vdots \\ a_2 = \pm}} \sum_{\substack{\boldsymbol{\ell}_n \in \mathbb{Z}^2 \\ \vdots \\ \boldsymbol{\ell}_1 \in \mathbb{Z}^2}} a_n \cdots a_2 \, \delta \left(\mathbf{p} - \mathbf{q} - \sum_{i=1}^{n} \mathbf{G}_i^{(\boldsymbol{\ell}_i)} \right) \\ \times \ \tilde{\boldsymbol{\Theta}}_{n+1,n}^{a_{n+1},a_n} \left(\mathbf{q} + \sum_{i=1}^{n} \mathbf{G}_i^{(\boldsymbol{\ell}_i)} \,|\, \mathbf{q} + \sum_{i=1}^{n-1} \mathbf{G}_i^{(\boldsymbol{\ell}_i)} \right) \cdots \tilde{\boldsymbol{\Theta}}_{2,1}^{a_2,a_1} \left(\mathbf{q} + \mathbf{G}_1^{(\boldsymbol{\ell}_1)} \,|\, \mathbf{q} \right) \end{aligned} \tag{4.28}$$

Here we have introduced

$$\tilde{\boldsymbol{\Theta}}_{l,m}^{b,a}(\mathbf{p} \,|\, \mathbf{q}) \stackrel{\text{def}}{=} \alpha_l^{-1}(\mathbf{p}) \, \tilde{\mathcal{J}}_{l,m}^{b,a}(\mathbf{p} \,|\, \mathbf{q}) \ \mathbf{M}_{l,m}^{b,a}(\mathbf{p} \,|\, \mathbf{q}) \ . \tag{4.29}$$

4.4 Reflection and transmission

From now on, we treat the case of two interfaces, $n = 2$, for the sake of clarity as we are going to illustrate a fundamental difference between the case for a single interface and the case for several interfaces. As can already be seen in Eq. (4.28), sums of reciprocal lattice vectors of the two lattices are involved in the transfer kernel. We define as a *reciprocal lattice sum* the set of all vectors of the form $\mathbf{G} = \sum_{i=1}^{n} \mathbf{G}_i^{(\boldsymbol{\ell}_i)}$ with $\boldsymbol{\ell}_i \in \mathbb{Z}^2$ for all $i \in [\![1,n]\!]$ (we restrict ourselves to $n = 2$ hereafter). The lattice sum leads to two sub-cases to be considered: the case where the two reciprocal lattices differs only by a rational scaling factor, which leads to a problem similar to that of the case for a single interface; and the case where the two reciprocal lattices do not only differ by a rational scaling factor which leads to a different problem. The two cases will

be referred to respectively as degenerate and non-degenerate. In all cases, Eq. (4.28) yields the following equation when plugged into the reduced Rayleigh equation

$$\begin{aligned}&\sum_{\boldsymbol{\ell}_1,\boldsymbol{\ell}_2\in\mathbf{Z}^2}\sum_{a_2=\pm} a_2\;\tilde{\boldsymbol{\Theta}}_{32}^{+a_2}\left(\mathbf{p}\,|\,\mathbf{p}-\mathbf{G}_2^{(\boldsymbol{\ell}_2)}\right)\tilde{\boldsymbol{\Theta}}_{21}^{a_2+}\left(\mathbf{p}-\mathbf{G}_2^{(\boldsymbol{\ell}_2)}\,|\,\mathbf{p}-\mathbf{G}_2^{(\boldsymbol{\ell}_2)}-\mathbf{G}_1^{(\boldsymbol{\ell}_1)}\right)\mathbf{R}\left(\mathbf{p}-\mathbf{G}_2^{(\boldsymbol{\ell}_2)}-\mathbf{G}_1^{(\boldsymbol{\ell}_1)}\,|\,\mathbf{p}_0\right)\\ =-&\sum_{\boldsymbol{\ell}_1,\boldsymbol{\ell}_2\in\mathbf{Z}^2}\sum_{a_2=\pm} a_2\;\tilde{\boldsymbol{\Theta}}_{32}^{+a_2}\left(\mathbf{p}\,|\,\mathbf{p}_0+\mathbf{G}_1^{(\boldsymbol{\ell}_1)}\right)\tilde{\boldsymbol{\Theta}}_{21}^{a_2-}\left(\mathbf{p}_0+\mathbf{G}_1^{(\boldsymbol{\ell}_1)}\,|\,\mathbf{p}_0\right)\,(2\pi)^2\delta\left(\mathbf{p}-\mathbf{p}_0-\mathbf{G}_2^{(\boldsymbol{\ell}_2)}-\mathbf{G}_1^{(\boldsymbol{\ell}_1)}\right)\,.\end{aligned} \tag{4.30}$$

Note that here that compared to Eq. (4.28) we have chosen to use the Dirac delta to re-write some arguments of the single-interface transfer kernels in an equivalent way in the right-hand side for later convenience.

4.4.1 Degenerate case: rational lattice sum

We start by treating the degenerate case which corresponds to a system for which there exists a common lattice for two interfaces. In other words, there exists a unit cell, a priori bigger than that of the unit cells of each interface, which allows one to deduce the whole set of interfaces by translation of the two profiles. Taking the simplified case of a one-dimensional lattice, this corresponds to the case where the two lattice constants a_1 and a_2, for the first and second interface respectively, are such that

$$a_2 = \frac{p}{q}\,a_1\,, \tag{4.31}$$

with $p, q \in \mathbb{N}$ and we assume that the fraction $\frac{p}{q}$ is irreducible. Let us assume without loss of generality that $a_1 < a_2$. For such lattices, it is clear that the set of the two interfaces is itself periodic of period $a = p\,a_1 = q\,a_2$. The corresponding reciprocal lattice constants are then given by $b_j = 2\pi/a_j$ and we can define the reciprocal lattice constant for the overall lattice constant, a, as $b = 2\pi/a$. We have directly from Eq. (4.31) that

$$b_1 = p\,b \tag{4.32a}$$
$$b_2 = q\,b\,. \tag{4.32b}$$

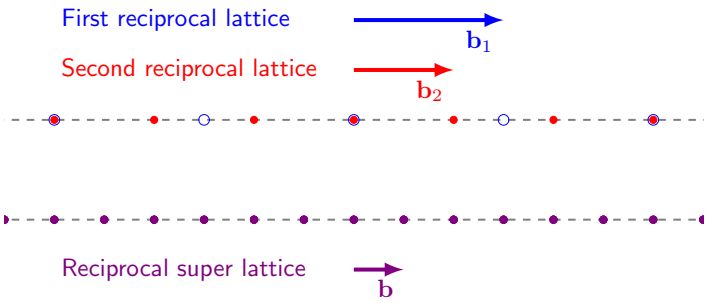
A vector of the reciprocal lattice sum hence reads

$$G = \ell_1 b_1 + \ell_2 b_2 = (\ell_1 p + \ell_2 q)b\,. \tag{4.33}$$

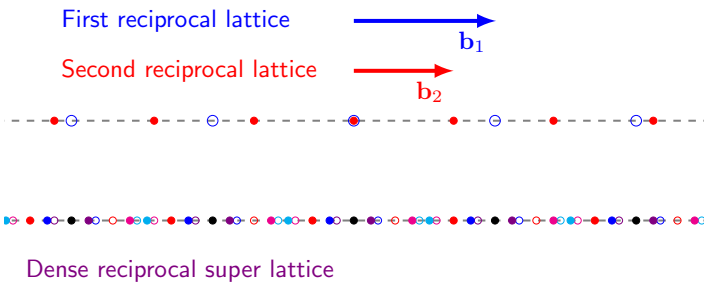
It is now clear that the reciprocal lattice sum is nothing but the reciprocal lattice of the latticed defined based on the lattice constant a. Indeed, $\mathbf{G}$ is an integer multiple of $b = 2\pi/a$ since p and q are integers and the smallest non-zero reciprocal lattice sum point is indeed b. To see this, it suffices to choose $\ell_1 = q$ and $\ell_2 = 1 - p$ which leads to

$$G = (qp + (1-p)q)b = b\,. \tag{4.34}$$

Up to some geometrical technicalities the same ideas can be generalized for two-dimensional lattices (this does not pose any problem for aligned rectangular lattices for example). The only thing to be taken from the previous analysis is that when there exists a common cell for both lattices which serves as a unit cell for the *whole* system, then the corresponding lattice vectors $\mathbf{a}_1^{(s)}$ and $\mathbf{a}_2^{(s)}$ (the subscript denoting the two independent directions and the superscript the



(a) Degenerate case: rational scaling factor.



(b) Non-degenerate case: irrational scaling factor.

Figure 4.3: Illustration of reciprocal lattice sums for (a) the degenerate case when the two lattices differs by a rational scaling factor ($b_2/b_1 = 2/3$ on the illustration) and (b) the non-degenerate case when the scaling factor is irrational ($b_2/b_1 = 1/\sqrt{2}$ on the illustration). Different colors in (b) correspond to different fixed value of $\ell_1 \in [\![0, 5]\!]$ (black, red, magenta, violet, blue, cyan) with open circles indicating negative values of ℓ_1 and filled circles indicating positive values of ℓ_1.

fact that there are primitive lattice vector of the whole system), can be used to construct the reciprocal lattice vectors of the lattice sum, namely $\mathbf{b}_1^{(s)}$ and $\mathbf{b}_2^{(s)}$ such that

$$\mathbf{a}_i^{(s)} \cdot \mathbf{b}_j^{(s)} = 2\pi\delta_{ij} \,. \tag{4.35}$$

The corresponding lattice, defined by the set $\{\mathbf{G}_s^{(\boldsymbol{\ell})} = \ell_1\, \mathbf{b}_1^{(s)} + \ell_2\, \mathbf{b}_2^{(s)} \,|\, \boldsymbol{\ell} = (\ell_1, \ell_2) \in \mathbb{Z}^2\}$ thus corresponds to reciprocal lattice sum. An illustration of the above considerations is given in Fig. 4.3(a).

What have we learned from this analysis? We have learned that the sums over $\boldsymbol{\ell}_1$ and $\boldsymbol{\ell}_2$ in Eq. (4.30), i.e. sums over the two individual reciprocal lattices, can in fact be re-written as a

single sum over the *super* reciprocal lattice as follows

$$\begin{aligned}&\sum_{\boldsymbol{\ell}\in\mathbf{Z}^2}\sum_{\boldsymbol{\ell}_2\in\mathbb{Z}^2}\sum_{a_2=\pm} a_2\;\tilde{\boldsymbol{\Theta}}_{32}^{+a_2}\left(\mathbf{p}\,|\,\mathbf{p}-\mathbf{G}_2^{(\boldsymbol{\ell}_2)}\right)\tilde{\boldsymbol{\Theta}}_{21}^{a_2+}\left(\mathbf{p}-\mathbf{G}_2^{(\boldsymbol{\ell}_2)}\,|\,\mathbf{p}-\mathbf{G}_s^{(\boldsymbol{\ell})}\right)\mathbf{R}\left(\mathbf{p}-\mathbf{G}_s^{(\boldsymbol{\ell})}\,|\,\mathbf{p}_0\right)\\ =&-\sum_{\boldsymbol{\ell}\in\mathbf{Z}^2}(2\pi)^2\delta\left(\mathbf{p}-\mathbf{p}_0-\mathbf{G}_s^{(\boldsymbol{\ell})}\right)\sum_{\boldsymbol{\ell}_1\in\mathbb{Z}^2}\sum_{a_2=\pm} a_2\;\tilde{\boldsymbol{\Theta}}_{32}^{+a_2}\left(\mathbf{p}_0+\mathbf{G}_s^{(\boldsymbol{\ell})}\,|\,\mathbf{p}_0+\mathbf{G}_1^{(\boldsymbol{\ell}_1)}\right)\tilde{\boldsymbol{\Theta}}_{21}^{a_2-}\left(\mathbf{p}_0+\mathbf{G}_1^{(\boldsymbol{\ell}_1)}\,|\,\mathbf{p}_0\right),\end{aligned}\tag{4.36}$$

which we recast as

$$\sum_{\boldsymbol{\ell}\in\mathbf{Z}^2}\mathbf{K}_L^{(r)}(\mathbf{p}\,|\,\mathbf{p}-\mathbf{G}_s^{(\boldsymbol{\ell})})\;\mathbf{R}\left(\mathbf{p}-\mathbf{G}_s^{(\boldsymbol{\ell})}\,|\,\mathbf{p}_0\right)=-(2\pi)^2\sum_{\boldsymbol{\ell}\in\mathbf{Z}^2}\delta\left(\mathbf{p}-\mathbf{p}_{\boldsymbol{\ell}}\right)\mathbf{K}_R^{(r)}(\mathbf{p}_{\boldsymbol{\ell}}\,|\,\mathbf{p}_0),\tag{4.37}$$

with

$$\mathbf{p}_{\boldsymbol{\ell}}\overset{\text{def}}{=}\mathbf{p}_0+\mathbf{G}_s^{(\boldsymbol{\ell})}\tag{4.38a}$$

$$\mathbf{K}_L^{(r)}(\mathbf{p}\,|\,\mathbf{q})\overset{\text{def}}{=}\sum_{\boldsymbol{\ell}_2\in\mathbb{Z}^2}\sum_{a_2=\pm} a_2\;\tilde{\boldsymbol{\Theta}}_{32}^{+a_2}\left(\mathbf{p}\,|\,\mathbf{p}-\mathbf{G}_2^{(\boldsymbol{\ell}_2)}\right)\tilde{\boldsymbol{\Theta}}_{21}^{a_2+}\left(\mathbf{p}-\mathbf{G}_2^{(\boldsymbol{\ell}_2)}\,|\,\mathbf{q}\right)\tag{4.38b}$$

$$\mathbf{K}_R^{(r)}(\mathbf{p}\,|\,\mathbf{q})\overset{\text{def}}{=}\sum_{\boldsymbol{\ell}_1\in\mathbb{Z}^2}\sum_{a_2=\pm} a_2\;\tilde{\boldsymbol{\Theta}}_{32}^{+a_2}\left(\mathbf{p}\,|\,\mathbf{q}+\mathbf{G}_1^{(\boldsymbol{\ell}_1)}\right)\tilde{\boldsymbol{\Theta}}_{21}^{a_2-}\left(\mathbf{q}+\mathbf{G}_1^{(\boldsymbol{\ell}_1)}\,|\,\mathbf{q}\right).\tag{4.38c}$$

In other words, we group terms such that $\mathbf{G}_1^{(\boldsymbol{\ell}_1)}+\mathbf{G}_2^{(\boldsymbol{\ell}_2)}$ is constant, equal to some $\mathbf{G}_s^{(\boldsymbol{\ell})}$. We can now repeat the arguments used in the case of a single interface. The term in the right-hand side is a Dirac comb on the lattice of the $\mathbf{p}_{\boldsymbol{\ell}}\overset{\text{def}}{=}\mathbf{p}_0+\mathbf{G}_s^{(\boldsymbol{\ell})}$ which imposes that $\mathbf{R}$ itself must be of the form Eq. (4.16). In other words, only plane waves with in-plane wave vectors differing by a super reciprocal lattice vector from the incident wave vector are allowed to exist in the system. Plugging Eq. (4.16) into Eq. (4.37) we obtain

$$\sum_{\boldsymbol{\ell}\in\mathbf{Z}^2}\sum_{\mathbf{m}\in\mathbf{Z}^2}\delta\left(\mathbf{p}-\mathbf{p}_{\boldsymbol{\ell}+\mathbf{m}}\right)\mathbf{K}_L^{(r)}(\mathbf{p}_{\boldsymbol{\ell}+\mathbf{m}}\,|\,\mathbf{p}_{\mathbf{m}})\;\mathbf{R}^{(\mathbf{m})}(\mathbf{p}_0)=-\sum_{\boldsymbol{\ell}\in\mathbf{Z}^2}\delta\left(\mathbf{p}-\mathbf{p}_{\boldsymbol{\ell}}\right)\mathbf{K}_R^{(r)}(\mathbf{p}_{\boldsymbol{\ell}}\,|\,\mathbf{p}_0),\tag{4.39}$$

which after the change of summation index $\boldsymbol{\ell}+\mathbf{m}\leftrightarrow\boldsymbol{\ell}$ and integration against a smooth test function φ having a compact support included in a disc of radius $\min(|\,\mathbf{b}_1^{(s)}\,|,|\,\mathbf{b}_2^{(s)}\,|)/2$ centered on $\mathbf{p}_{\boldsymbol{\ell}}$ and such that $\varphi(\mathbf{p}_{\boldsymbol{\ell}})=1$, yields the infinite countable linear system

$$\sum_{\mathbf{m}\in\mathbf{Z}^2}\mathbf{K}_L^{(r)}(\mathbf{p}_{\boldsymbol{\ell}}\,|\,\mathbf{p}_{\mathbf{m}})\;\mathbf{R}^{(\mathbf{m})}(\mathbf{p}_0)=-\,\mathbf{K}_R^{(r)}(\mathbf{p}_{\boldsymbol{\ell}}\,|\,\mathbf{p}_0),\tag{4.40}$$

for $\boldsymbol{\ell}\in\mathbb{Z}^2$.

Remark 4.1. The reduced Rayleigh equation (4.40) for the case of two periodic surfaces in the degenerate case takes the same form as the reduced Rayleigh equation for the case of a single periodic surface. The main difference resides in the complexity of the kernel of the linear system, which as can be seen in Eq. (4.38) is a discrete composition of periodic single-interface transfer kernels. In particular, such a composition is nothing but a matrix-matrix product (the sum of matrix-matrix products to be accurate taking into account the sum over $a_2=\pm$) as will be clear in Part II. It is then expected that solving the reduced Rayleigh equation for the case of two periodic interfaces in the degenerate case should not be different from the single interface case. However, we have foreseen in Chapter 3 some difficulties concerning the complexity of the kernel and on the stability of film system which are still present independently of whether the system is periodic or not.

The case of transmission is handled in a similar fashion without difficulty and we obtain the following reduced Rayleigh equation in the degenerate case

$$\sum_{\mathbf{m}\in\mathbf{Z}^2} \mathbf{K}_L^{(t)}(\mathbf{p}_{\boldsymbol{\ell}} \,|\, \mathbf{p}_{\mathbf{m}})\ \mathbf{T}^{(\mathbf{m})}(\mathbf{p}_0) = \eta_{13}\,\delta_{\boldsymbol{\ell}}\ \mathbf{I}_2\,, \tag{4.41}$$

where $\mathbf{K}_L^{(t)}$ now is defined as

$$\mathbf{K}_L^{(t)}(\mathbf{p}\,|\,\mathbf{q}) \overset{\text{def}}{=} \sum_{\boldsymbol{\ell}_1\in\mathbb{Z}^2}\sum_{a_2=\pm} a_2\ \tilde{\boldsymbol{\Theta}}_{12}^{-a_2}\left(\mathbf{p}\,|\,\mathbf{p}-\mathbf{G}_1^{(\boldsymbol{\ell}_1)}\right)\tilde{\boldsymbol{\Theta}}_{23}^{a_2-}\left(\mathbf{p}-\mathbf{G}_1^{(\boldsymbol{\ell}_1)}\,|\,\mathbf{q}\right)\,. \tag{4.42}$$

4.4.2 Non-degenerate case: irrational (dense) lattice sum

Let us now consider, what will be referred to as the non-degenerate case, for which each interface is periodic but the system as a whole is not periodic. In this case, we do not expect the Bloch-Floquet theorem to hold and therefore the scattering of an incident plane wave is not expected to yield a scattered field that can be written as a sum of plane waves whose wave vector obeys the grating formula. The system thus may simply be described as any arbitrary system without taking into account the fact that each surface is periodic. There is nevertheless something of interest to investigate to try to take advantage of the periodicity of each surface. Starting back to Eq. (4.30), which we have derived without making other assumptions than the two interfaces being periodic with different period, and when considering a non-degenerate system, we observe something which is both interesting mathematically (or strange depending on the opinion) and quite natural from a physics point of view. We would like of course to take the same steps as in the degenerate case, but we now cannot find a common super lattice to both lattices. As a consequence, we loses the grating formula as expected from physical consideration about the overall non-periodicity of the system. The presence of the Dirac deltas on the right-hand side nevertheless hints at the fact that the allowed wave vectors still may form a discrete, or countable set. Indeed, the set of the sum of elements of two countable sets is still a countable set. But this does not mean that each elements can be well separated from another, in other words, we may have to deal with a set of in-plane wave vector which is countable but dense in $\mathbb{R}^2$. Let us illustrate this idea by considering again a one-dimensional system where the lattice constants are related by a irrational scaling factor, say $\sqrt{2}$ for the sake of the example

$$a_2 = \sqrt{2}\,a_1\,. \tag{4.43}$$

The reciprocal lattice constants are then given by

$$b_1 = \frac{2\pi}{a_1} \tag{4.44a}$$

$$b_2 = \frac{2\pi}{a_2} = \frac{b_1}{\sqrt{2}}\,, \tag{4.44b}$$

and the sum of a reciprocal lattice vector of the first lattice and a reciprocal lattice vector of the second lattice hence reads

$$G = \ell_1\,b_1 + \ell_2\,b_2 = \left(\ell_1 + \frac{\ell_2}{\sqrt{2}}\right)b_1\,. \tag{4.45}$$

The set of points of the form Eq. (4.45) is illustrated in Fig. 4.3(b). The different colors for the points indicate the corresponding fixed value of ℓ_1 and varying value of ℓ_2. In contrast to Fig. 4.3(a), Fig. 4.3(b) illustrates the absence of an overall lattice and the densification of

the set of points on the real line when varying ℓ_1 and ℓ_2. It is clear that not all the points on the real line will be reached since, we have already indicated that the set of reciprocal points is countable. In particular, another way to see this, is that no rational number other than the integers are reachable. As a consequence of the countable property of the set, not all the irrational numbers are reachable either. In short, the set of points given by Eq. (4.45) has properties similar to $\mathbb{Q}$ with respect to $\mathbb{R}$, it is countable and dense.

This seems to make sense intuitively. Now, let us come back to the term on the right-hand side of Eq. (4.30). Taking a one-dimensional analogue for the sake of simplicity, the right-hand side is of the form

$$\text{RHS} = \sum_{G \in \mathcal{S}} a_G \, \delta(p - G) \, , \tag{4.46}$$

where $\mathcal{S}$ denotes some countable dense subset of $\mathbb{R}$, like our aforementioned lattice sum, and the a_G are some coefficients indexed by $G \in \mathcal{S}$ (which we can do without trouble since $\mathcal{S}$ is countable). Now comes the puzzling mathematical question. What does the above equation actually mean? We would like to attempt our good old definition of the Dirac delta and try to test it against a smooth test function with compact support φ, which would lead to

$$\int_{\mathbb{R}} \sum_{G \in \mathcal{S}} a_G \, \delta(p - G) \, \varphi(p) \, \mathrm{d}p = \sum_{G \in \mathcal{S} \, \cap \, \mathrm{supp}\varphi} a_G \, \varphi(G) \, . \tag{4.47}$$

The right-hand side in Eq. (4.47) would diverge if nothing is specified about the a_G. Indeed, taking $a_G = 1$ for all $G \in \mathcal{S}$ would lead to $\sum_{G \in \mathcal{S} \, \cap \, \mathrm{supp}\varphi} \varphi(G) = \infty$ for a positive test function, since we would sum an infinite number of terms of the same order. Intuitively, its seems that the coefficients a_G must then have a fast enough decay rate when ordered with decreasing modulus for the right-hand side of Eq. (4.47) to be defined. Coming back to the picture drawn in Fig. 4.3(b), we could expect that the weight a_G associated with somewhat increasing ℓ_1 and ell_2 decays as the set of points becomes denser and denser so that when summed with the corresponding weight, a finite result may be obtained.

We have done our best here to discuss intuitively the meaning of the Eq. (4.30) but it is clear that a rigorous treatment would require elements of measure theory and distribution theory. We are physicist after all, so let us be satisfied with the intuition for this work. What does it bring us then with respect to solving such a problem with a computer? As already mentioned, one way of dealing with the non-degenerate case could be to treat it as an arbitrary system without specific symmetry. One would somehow average the response over a volume of wave vectors to discretize the problem. Another way, which is similar to what would be expose in the next chapter for solving the reduced Rayleigh equation for a single periodic interface, is to assume that we approximate the system by a periodic system with a big period and treat it as if it were a degenerate case. For example, this would be the same as approximating the scaling factor $\sqrt{2}$ by a rational number in Eq. (4.45) and we would be back to the degenerate case. Of course, the closer the approximation the denser the number of reciprocal points, and the heavier the simulation. Nevertheless, we may expect that as the approximation of the scaling constant becomes better and better, the optical response would converge. One may hope then that a not too close approximation of the scaling constant is needed before obtaining a decent approximation of the optical response. Such systems are unfortunately not studied further in this work, but we hope that the above discussion will be the starting point of further research.

4.5 Summary

The present chapter has been devoted to analyzing the consequences of the periodicity of the interfaces for the reduced Rayleigh equations. First, we have seen that for a system containing a single periodic interface, the fields can be written as a discrete sum of plane waves with wave vectors satisfying the grating equation. The reduced Rayleigh equations hence become infinite countable linear systems of equations for the discrete set of reflection and transmission amplitudes. Then the cases of periodic multi-layer systems were analyzed in a similar way. We have seen that similar arguments as those used for a single periodic interface apply but with a distinction to be made between systems whose the periods of the different interfaces are rational multiples of one another or not. The difficulty to treat numerical the latter case was discussed with suggested solutions.

Part II

Numerical analysis and approximation methods

Chapter 5

Solving the reduced Rayleigh equations; direct method

"If my calculations are correct, when this baby hits 88 miles per hour...
you're gonna see some serious shit." – Dr. Emmett Brown
Back to the Future (1985).

In Part I, we have derived the reduced Rayleigh equations for a single arbitrary interface (Chapter 2) and presented a framework which allows us to generalize the reduced Rayleigh equations to multilayer system (Chapter 3). Also we have applied these equations in the case of periodic interfaces and recover the well known grating formula (Chapters 4). We have seen that for arbitrary rough interfaces the reduced Rayleigh equations yield a set of decoupled (but one can also consider their coupled counterpart) integral equations for the reflection and transmission amplitudes. We have shown that in the case of periodic interfaces, these integral equations become infinite countable linear systems of equations for the discrete propagative and evanescent modes.

In the present chapter, we present what will be called the direct method for solving the reduced Rayleigh equations we have encountered in Part I. We will see that arbitrary and periodic surfaces are numerically handled in a similar fashion in the sense that they lead numerically to solving linear systems. Indeed, on a digital computer, one may represent a surface only by a finite number of parameters, which usually translates into finite system size, discretization points, or a number of Fourier components. The transfer kernels in the reduced Rayleigh equations will be expanded in sums of Fourier moments of the surfaces profile. For this reason, we will start by describing the numerical procedures for a periodic system, and then explain how arbitrary surfaces eventually fall into the same procedures.

The present chapter has essentially two goals. First, the procedure to go from the reduced Rayleigh equations to an implementable numerical scheme will be described in details, and a complexity analysis will be made. Second, due to the finite size of the linear systems that can be handled numerically, the questions of truncation and convergence will arise naturally. A significant part of the chapter will be devoted to a numerical study of convergence with the number of modes kept in the truncation of the linear system. In particular, we will look for convergence criteria and study the convergence rate as a function of reduced physical parameters. It must be noted that these results have not been published yet as they must be complemented by further case studies and proper mathematical analysis. The results nevertheless illustrate clearly the main behaviors to be expected for the convergence of the

approximated solutions with respect to the physical parameters which, to our knowledge, have never been published before in the case of two-dimensional penetrable media and hence constitute a significant contribution of the present work in this area.

Note that in this chapter, our only goal is to solve numerically the reduced Rayleigh equations for given transfer kernels and to give a preliminary numerical study of convergence. There are two important points that are *not* discussed in this chapter. (i) We assume the knowledge of the transfer kernels, i.e. that we assume the existence of an algorithm to evaluate $\mathcal{J}_{l,m}^{b,a}$ or its expansion in Fourier moments. This is not a trivial task and an entire appendix is devoted to it (Appendix A). (ii) Assuming now that we obtain a solution to the reduced Rayleigh equations, i.e. that we obtain a Fourier expansion of the electric field in terms of reflection and transmission amplitudes. Is this a solution of the scattering problem given by the Helmholtz equation and the set of boundary conditions? We are facing the issue of the consistency of the reduced Rayleigh equations with respect to the initial scattering problem. This question will be addressed in Chapter 7 on the Rayleigh hypothesis and on consistency analysis.

5.1 Reduced units

Consider the reduced Rayleigh equation for the reflection amplitude Eq. (2.51)

$$\int \mathcal{J}_{2,1}^{+,+}(\mathbf{p}\,|\,\mathbf{q})\ \mathbf{M}_{2,1}^{+,+}(\mathbf{p}\,|\,\mathbf{q})\ \mathbf{R}(\mathbf{q}\,|\,\mathbf{p}_o)\,\frac{\mathrm{d}^2 q}{(2\pi)^2} = -\,\mathcal{J}_{2,1}^{+,-}(\mathbf{p}\,|\,\mathbf{p}_o)\ \mathbf{M}_{2,1}^{+,-}(\mathbf{p}\,|\,\mathbf{p}_o)\ ,$$

for a plane wave with in-plane wave vector $\mathbf{p}_0$ incident on the interface between medium 1 characterized by the dielectric constant ϵ_1 and medium 2 characterized by the dielectric constant ϵ_2. The wavelength in vacuum of the incident plane wave is given by $\lambda_0 = 2\pi\, c/\omega$, where ω is the angular frequency and c the speed of light in vacuum. If we assume medium 1 to have a real and positive dielectric constant, which is a rather general case if we wish the incident wave to propagate without being absorbed, we can choose that the wavelength in medium 1, $\lambda_1 = \lambda_0/\sqrt{\epsilon_1}$, will play the role of a length scale of reference. Hence all lengths may be expressed in unit of $\lambda_1/(2\pi)$. Equivalently, all wave numbers may be expressed in unit of $k_1 = \sqrt{\epsilon_1}\omega/c$. We introduce the following reduced quantities denoted with a bar $\bar{\cdot}$:

$$\mathbf{x} \overset{\text{def}}{=} \frac{\lambda_1}{2\pi}\ \bar{\mathbf{x}} \tag{5.1a}$$

$$\zeta(\mathbf{x}) \overset{\text{def}}{=} \frac{\lambda_1}{2\pi}\bar{\zeta}(\bar{\mathbf{x}}) \tag{5.1b}$$

$$z = \sqrt{\frac{\epsilon_2}{\epsilon_1}} \tag{5.1c}$$

$$\mathbf{p} \overset{\text{def}}{=} \sqrt{\epsilon_1}\frac{\omega}{c}\,\bar{\mathbf{p}} \tag{5.1d}$$

$$\alpha_1(\mathbf{p}) = \left(\epsilon_1\frac{\omega^2}{c^2} - \overset{2}{\mathbf{p}}\right)^{1/2} = \sqrt{\epsilon_1}\frac{\omega}{c}\ \left(1 - \overset{2}{\bar{\mathbf{p}}}\right)^{1/2} \overset{\text{def}}{=} \sqrt{\epsilon_1}\frac{\omega}{c}\ \bar{\alpha}_1(\bar{\mathbf{p}}) \tag{5.1e}$$

$$\alpha_2(\mathbf{p}) = \left(\epsilon_2\frac{\omega^2}{c^2} - \overset{2}{\mathbf{p}}\right)^{1/2} = \sqrt{\epsilon_1}\frac{\omega}{c}\ \left(z^2 - \overset{2}{\bar{\mathbf{p}}}\right)^{1/2} \overset{\text{def}}{=} \sqrt{\epsilon_1}\frac{\omega}{c}\ \bar{\alpha}_z(\bar{\mathbf{p}}) \tag{5.1f}$$

$$\mathbf{R}(\mathbf{p}\ |\ \mathbf{p}_0) \overset{\text{def}}{=} \left(\epsilon_1\frac{\omega^2}{c^2}\right)^{-1}\ \bar{\mathbf{R}}(\bar{\mathbf{p}}\ |\ \bar{\mathbf{p}}_0)\,. \tag{5.1g}$$

Using the definitions in Eq. (5.1) for the reduced parameters in the reduced Rayleigh equation we can express the polarization coupling matrix as

$$\begin{aligned}
\mathbf{M}_{2,1}^{b,a}(\mathbf{p}\,|\,\mathbf{q}) &= \begin{pmatrix} |\mathbf{p}||\mathbf{q}| + ab\,\alpha_2(\mathbf{p})\,\alpha_1(\mathbf{q})\,\hat{\mathbf{p}}\cdot\hat{\mathbf{q}} & -b\sqrt{\epsilon_1}\,\frac{\omega}{c}\alpha_2(\mathbf{p})[\hat{\mathbf{p}}\times\hat{\mathbf{q}}]\cdot\hat{\mathbf{e}}_3 \\ a\sqrt{\epsilon_2}\,\frac{\omega}{c}\,\alpha_1(\mathbf{q})[\hat{\mathbf{p}}\times\hat{\mathbf{q}}]\cdot\hat{\mathbf{e}}_3 & \sqrt{\epsilon_2\epsilon_1}\,\frac{\omega^2}{c^2}\hat{\mathbf{p}}\cdot\hat{\mathbf{q}} \end{pmatrix} \\
&= \epsilon_1\,\frac{\omega^2}{c^2}\begin{pmatrix} |\bar{\mathbf{p}}||\bar{\mathbf{q}}| + ab\,\bar{\alpha}_z(\bar{\mathbf{p}})\,\bar{\alpha}_1(\bar{\mathbf{q}})\;\hat{\mathbf{p}}\cdot\hat{\mathbf{q}} & -b\,\bar{\alpha}_z(\bar{\mathbf{p}})[\hat{\mathbf{p}}\times\hat{\mathbf{q}}]\cdot\hat{\mathbf{e}}_3 \\ a\,z\,\bar{\alpha}_1(\bar{\mathbf{q}})\,[\hat{\mathbf{p}}\times\hat{\mathbf{q}}]\cdot\hat{\mathbf{e}}_3 & z\,\hat{\mathbf{p}}\cdot\hat{\mathbf{q}} \end{pmatrix} \\
&\stackrel{\text{def}}{=} \epsilon_1\,\frac{\omega^2}{c^2}\;\bar{\mathbf{M}}_{z,1}^{b,a}(\bar{\mathbf{p}}\,|\,\bar{\mathbf{q}})\,, \qquad (5.2)
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{J}_{2,1}^{b,a}(\mathbf{p}\;|\;\mathbf{q}) &= (b\alpha_2(\mathbf{p}) - a\alpha_1(\mathbf{q}))^{-1}\int \exp\left(-i(\mathbf{p}-\mathbf{q})\cdot\mathbf{x}\right)\,\exp\left(-i(b\alpha_2(\mathbf{p}) - a\alpha_1(\mathbf{q}))\zeta(\mathbf{x})\right)\,\mathrm{d}^2x \\
&= \left(\sqrt{\epsilon_1}\frac{\omega}{c}\right)^{-1}(b\bar{\alpha}_z(\bar{\mathbf{p}}) - a\bar{\alpha}_1(\bar{\mathbf{q}}))^{-1}\int \exp\left(-i(\bar{\mathbf{p}}-\bar{\mathbf{q}})\cdot\frac{\sqrt{\epsilon_1}\omega\,\mathbf{x}}{c}\right) \\
&\quad \exp\left(-i(b\bar{\alpha}_z(\bar{\mathbf{p}}) - a\bar{\alpha}_1(\bar{\mathbf{q}}))\frac{\sqrt{\epsilon_1}\omega\,\zeta(\mathbf{x})}{c}\right)\,\mathrm{d}^2x \\
&= \left(\sqrt{\epsilon_1}\frac{\omega}{c}\right)^{-3}(b\bar{\alpha}_z(\bar{\mathbf{p}}) - a\bar{\alpha}_1(\bar{\mathbf{q}}))^{-1}\int \exp\left(-i(\bar{\mathbf{p}}-\bar{\mathbf{q}})\cdot\bar{\mathbf{x}}\right) \\
&\quad \exp\left(-i(b\bar{\alpha}_z(\bar{\mathbf{p}}) - a\bar{\alpha}_1(\bar{\mathbf{q}}))\,\bar{\zeta}(\bar{\mathbf{x}})\right)\,\mathrm{d}^2\bar{x} \\
&\stackrel{\text{def}}{=} \left(\sqrt{\epsilon_1}\frac{\omega}{c}\right)^{-3}\bar{\mathcal{J}}_{z,1}^{b,a}(\bar{\mathbf{p}}\;|\;\bar{\mathbf{q}})\,. \qquad (5.3)
\end{aligned}$$

By substituting Eqs. (5.2) and (5.3) into the reduced Rayleigh equation (2.51), making a change of variable $\mathbf{q} = \sqrt{\epsilon_1}\frac{\omega}{c}\;\bar{\mathbf{q}}$ and using the definition of the reduced reflection amplitudes Eq. (5.1g), we obtain the reduced Rayleigh equation in reduced units, or *dimensionless reduced Rayleigh equations*

$$\int \bar{\mathcal{J}}_{z,1}^{+,+}(\bar{\mathbf{p}}\,|\,\bar{\mathbf{q}})\;\bar{\mathbf{M}}_{z,1}^{+,+}(\bar{\mathbf{p}}\,|\,\bar{\mathbf{q}})\;\bar{\mathbf{R}}\,(\bar{\mathbf{q}}\,|\,\bar{\mathbf{p}}_{\mathrm{o}})\,\frac{\mathrm{d}^2\bar{q}}{(2\pi)^2} = -\,\bar{\mathcal{J}}_{z,1}^{+,-}(\bar{\mathbf{p}}\,|\,\bar{\mathbf{p}}_{\mathrm{o}})\;\bar{\mathbf{M}}_{z,1}^{+,-}(\bar{\mathbf{p}}\,|\,\bar{\mathbf{p}}_{\mathrm{o}})\;. \qquad (5.4)$$

What do we learn from the reduction of units procedure? By inspection of the dimensionless reduced Rayleigh equations and the definition of $\bar{\mathbf{M}}_{z,1}^{b,a}(\bar{\mathbf{p}}\,|\,\bar{\mathbf{q}})$ and $\bar{\mathcal{J}}_{z,1}^{b,a}(\bar{\mathbf{p}}\;|\;\bar{\mathbf{q}})$, Eqs. (5.2) and (5.3), we clearly see that the solution of the dimensionless reduced Rayleigh equation only depends on the impedance parameter z, and the function $\bar{\zeta}$. Physically, this means on the one hand that for a given surface profile and wavelength, the scattering of a plane wave will be equivalent for any choice of media as long as the impedance is kept constant. By *equivalent*, we mean that the reflection amplitudes would stay the same up to an *overall* scaling factor $(\epsilon_1\omega^2/c^2)^{-1}$ as can be seen from Eq. (5.1g). Alternatively, for a given choice of media, or fixed z, changing *simultaneously* the wavelength of the incident light by a factor β and scaling the surface profile by the same factor in the three directions would result in equivalent reflection amplitudes. To summarize, what matter are the dielectric contrast, or impedance, and the relative length scale between the wavelength of the incident light and the characteristic lengths of the surface profile.

Let us apply further the dimensionless reduced Rayleigh equation in the case of a periodic interface. We have seen in Chapter 4 that in the case of a periodic interface, the reduced Rayleigh

equation becomes a infinite dimensional (countable) linear system and that the reflection amplitudes is a sum of weighted Dirac masses on the lattice $(\mathbf{p}_{\boldsymbol{\ell}})_{\boldsymbol{\ell}\in\mathbb{Z}^2}$, Eqs. (4.19) and (4.16)

$$\mathbf{R}(\mathbf{p}\,|\mathbf{p}_0) = (2\pi)^2 \sum_{\mathbf{m}\in\mathbb{Z}^2} \delta(\mathbf{p}-\mathbf{p}_{\mathbf{m}})\ \mathbf{R}^{(\mathbf{m})}(\mathbf{p}_0)$$

and the reduced Rayeigh equation reads

$$\sum_{\mathbf{m}\in\mathbb{Z}^2} \tilde{\mathcal{J}}^{+,+}_{2,1}(\mathbf{p}_{\boldsymbol{\ell}}\,|\,\mathbf{p}_{\mathbf{m}})\ \mathbf{M}^{+,+}_{2,1}(\mathbf{p}_{\boldsymbol{\ell}}\,|\,\mathbf{p}_{\mathbf{m}})\ \mathbf{R}^{(\mathbf{m})}(\mathbf{p}_0) = -\tilde{\mathcal{J}}^{+,-}_{2,1}(\mathbf{p}_{\boldsymbol{\ell}}\,|\,\mathbf{p}_0)\ \mathbf{M}^{+,-}_{2,1}(\mathbf{p}_{\boldsymbol{\ell}}\,|\mathbf{p}_0)\,.$$

Following the reduction of unit procedure, we obtain

$$\bar{\mathbf{R}}(\bar{\mathbf{p}}\,|\bar{\mathbf{p}}_0) = (2\pi)^2 \sum_{\mathbf{m}\in\mathbb{Z}^2} \delta(\bar{\mathbf{p}}-\bar{\mathbf{p}}_{\mathbf{m}})\ \bar{\mathbf{R}}^{(\mathbf{m})}(\bar{\mathbf{p}}_0) \tag{5.5}$$

and

$$\sum_{\mathbf{m}\in\mathbb{Z}^2} \tilde{\bar{\mathcal{J}}}^{+,+}_{z,1}(\bar{\mathbf{p}}_{\boldsymbol{\ell}}\,|\,\bar{\mathbf{p}}_{\mathbf{m}})\ \bar{\mathbf{M}}^{+,+}_{z,1}(\bar{\mathbf{p}}_{\boldsymbol{\ell}}\,|\,\bar{\mathbf{p}}_{\mathbf{m}})\ \bar{\mathbf{R}}^{(\mathbf{m})}(\bar{\mathbf{p}}_0) = -\tilde{\bar{\mathcal{J}}}^{+,-}_{z,1}(\bar{\mathbf{p}}_{\boldsymbol{\ell}}\,|\,\bar{\mathbf{p}}_0)\ \bar{\mathbf{M}}^{+,-}_{z,1}(\bar{\mathbf{p}}_{\boldsymbol{\ell}}\,|\bar{\mathbf{p}}_0)\,, \tag{5.6}$$

where the reduced quantities read

$$\bar{\mathbf{p}}_{\boldsymbol{\ell}} = \bar{\mathbf{p}}_0 + \bar{\mathbf{G}}^{(\boldsymbol{\ell})} \stackrel{\text{def}}{=} \bar{\mathbf{p}}_0 + \left(\sqrt{\epsilon_1}\frac{\omega}{c}\right)^{-1} \mathbf{G}^{(\boldsymbol{\ell})} \tag{5.7a}$$

$$\bar{\mathbf{R}}^{(\mathbf{m})}(\bar{\mathbf{p}}_0) \stackrel{\text{def}}{=} \mathbf{R}^{(\mathbf{m})}(\mathbf{p}_0) \tag{5.7b}$$

$$\tilde{\mathcal{J}}^{b,a}_{2,1}(\mathbf{p}_{\boldsymbol{\ell}}\ |\ \mathbf{p}_{\mathbf{m}}) \stackrel{\text{def}}{=} \left(\sqrt{\epsilon_1}\frac{\omega}{c}\right)^{-3} \tilde{\bar{\mathcal{J}}}^{b,a}_{z,1}(\bar{\mathbf{p}}_{\boldsymbol{\ell}}\ |\ \bar{\mathbf{p}}_{\mathbf{m}})\,. \tag{5.7c}$$

In particular, this is equivalent to introducing the reduced lattice basis vector (for $i \in \{1,2\}$)

$$\bar{\mathbf{a}}_i = \frac{2\pi}{\lambda_1}\ \mathbf{a}_i, \tag{5.8}$$

whose reciprocal basis vectors are given by

$$\bar{\mathbf{b}}_i = \frac{\lambda_1}{2\pi}\ \mathbf{b}_i = \left(\sqrt{\epsilon_1}\frac{\omega}{c}\right)^{-1} \mathbf{b}_i\,, \tag{5.9}$$

so that the reciprocal lattice vectors read

$$\bar{\mathbf{G}}^{(\boldsymbol{\ell})} = \ell_1\ \bar{\mathbf{b}}_1 + \ell_2\ \bar{\mathbf{b}}_2\ . \tag{5.10}$$

In the case of a periodic system, and for a given normalized profile within the reduced unit cell, the solution of the dimensionless reduced Rayleigh equation *only depends on four independent parameters*: the impedance z, the reduced profile amplitude $2\pi H/\lambda_1$ (where H denotes a characteristic amplitude of the profile, e.g. $\max|\zeta|$), and the reduced lattice constants $2\pi a_i/\lambda_1$. This means that when we will study the convergence and consistency properties of the numerical schemes for solving the reduced Rayleigh equations in subsequent section and chapters, we will only need to consider how these properties behave as functions of these four free parameters and possibly for different normalized profiles.

For randomly rough surfaces, the surface profile is generally specified by a distribution of height characterized by a rms-roughness σ and an auto-correlation function characterized by a correlation length a. In that case, for given normalized probability density of height and normalized auto-correlation function, the free dimensionless parameters to consider are the impedance, the reduced rms-roughness $2\pi\sigma/\lambda_1$ and reduced correlation length $2\pi a/\lambda_1$.

Remark 5.1. We have illustrated the procedure of unit reduction in the case of the reduced Rayleigh equation for the reflection amplitudes but the same procedure applies for the transmission amplitudes and is left to the reader. Similarly, the procedure can be generalized easily to multilayer systems.

In the following, in order to keep the notations as light as possible, we will illustrate the implementation of the numerical schemes for solving the reduced Rayleigh equations in full units. The reader will have no difficulty in implementing the schemes for corresponding dimensionless equations following the procedure illustrated above. However, when dealing with the numerical analysis, we will come back to reduced units for the discussion since the scaling of quantities such as the convergence rate will involve only the few identified independent reduced parameters.

5.2 Periodic systems

We first start by presenting the direct method for solving numerically the reduced Rayleigh equations in the case of a periodic system made of a single interface as a prototypical system. The procedure can be easily extended to periodic (degenerate) multilayer system and we will briefly illustrate the increase in complexity from a single periodic interface to two periodic interfaces.

5.2.1 Implementation

Let us consider a reduced Rayleigh equation in the case of a system of periodic interfaces such that the transfer equation takes the form of an infinite countable liner system, such as for all $\boldsymbol{\ell} \in \mathbb{Z}^2$,

$$\sum_{\mathbf{m}\in\mathbb{Z}^2} \mathbf{M}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_{\mathbf{m}})\, \mathbf{X}^{(\mathbf{m})}(\mathbf{p}_0) = \mathbf{N}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_0)\,. \tag{5.11}$$

Here $\mathbf{M}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_{\mathbf{m}})$ and $\mathbf{N}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_0)$ are respectively a given transfer kernel and right hand side corresponding to the system of interest and $\mathbf{X}^{(\mathbf{m})}(\mathbf{p}_0)$ can denote either the reflection or transmission amplitude matrix of the mode indexed by $\mathbf{m} \in \mathbb{Z}^2$ for a given incident wave vector characterized by its projection $\mathbf{p}_0$ in the $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2)$ plane (see previous chapters). The transfer kernel and the right hand side are of course different whether we consider the reflection or transmission amplitudes. In the following, $\mathbf{X}^{(\mathbf{m})}(\mathbf{p}_0)$ will be referred to as *scattering amplitudes*, the term *scattering* needing to be understood as a general term for *reflection* or *transmission* depending on the context. We recall that $\mathbf{X}^{(\mathbf{m})}(\mathbf{p}_0) = (X^{(\mathbf{m})}_{\alpha,\beta}(\mathbf{p}_0))_{\alpha\beta\in\{p,s\}}$ is a two by two complex matrix containing the scattering amplitude from incident polarized state β to an outgoing polarization state α of the mode $\mathbf{m}$. Consequently, Eq. (5.11), for a given $\boldsymbol{\ell}$, contains four scalar equations which can be written in index form as

$$\sum_{\mathbf{m}\in\mathbb{Z}^2} \sum_{\gamma\in\{p,s\}} M_{\alpha\gamma}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_{\mathbf{m}})\, X^{(\mathbf{m})}_{\gamma\beta}(\mathbf{p}_0) = N_{\alpha\beta}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_0)\,, \tag{5.12}$$

for a given $\boldsymbol{\ell} \in \mathbb{Z}^2$ and $\alpha, \beta \in \{p, s\}$.

Remark 5.2. It is worth noting that the index β in Eq. (5.12) does not appear in the kernel term. This implies that Eq. (5.12) can be solved independently for $(X^{(\mathbf{m})}_{p})_{\mathbf{m}\in\mathbb{Z}^2}$ and for $(X^{(\mathbf{m})}_{s})_{\mathbf{m}\in\mathbb{Z}^2}$. Physically, this expresses the fact that one can measure *independently* the response from either a p- or s-polarized incident state, but the reciprocal statement is not true. It is *a priori* not

possible, to measure the separate contribution of the p and s- components of a mixed incident state on the response with a single measurement. The interesting fact from a numerical point of view though, is that the kernel stays the same for the two incident states, and only the right hand side changes. We will see that this allows for the *simultaneous* solution of the independent equations for $(X_{.p}^{(\mathbf{m})})_{\mathbf{m}\in\mathbb{Z}^2}$ and $(X_{.s}^{(\mathbf{m})})_{\mathbf{m}\in\mathbb{Z}^2}$.

Now that the remark on the separation of the problem into p and s incident states has been made, we need to introduce a cut off in the $\boldsymbol{\ell}$-plane. Indeed, only a finite number of scalar unknowns can be handled with a digital computer. Only a finite subset of modes indexed in $\mathbb{Z}^2$ can be handled, and a common (but not unique) way to choose this subset is to pick terms within a disc of a certain radius in the $\boldsymbol{\ell}$-plane or equivalently in the $\mathbf{p}$-plane of transverse wave vectors. We define such a disc for an index radius $N_c \geq 0$ by $\mathcal{D}_{N_c} = \{\mathbf{m} \in \mathbb{Z}^2 \mid m_1^2 + m_2^2 \leq N_c^2\}$. This has for consequences that (i) an approximate numerical solution can be computed only for $(\mathbf{X}^{(\mathbf{m})}(\mathbf{p}_0))_{\mathbf{m}\in\mathcal{D}_{N_c}}$ and (ii) that a *truncation error* is introduced in the sum in Eq. (5.11). The approximate numerical problem then satisfies, for all $\boldsymbol{\ell} \in \mathcal{D}_{N_c}$,

$$\sum_{\mathbf{m}\in\mathcal{D}_{N_c}} \mathbf{M}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_{\mathbf{m}})\, \mathbf{X}_{N_c}^{(\mathbf{m})}(\mathbf{p}_0) = \mathbf{N}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_0)\,. \tag{5.13}$$

Here we have stressed that the solution of Eq. (5.13) is an approximation of that of Eq. (5.11) as indicated by subscript N_c. It is intuitively expected that, provided stability, the approximate solution $(\mathbf{X}_{N_c}^{(\mathbf{m})}(\mathbf{p}_0))_{\mathbf{m}\in\mathcal{D}_{N_c}}$ converges in some sense to the exact solution $(\mathbf{X}^{(\mathbf{m})}(\mathbf{p}_0))_{\mathbf{m}\in\mathbb{Z}^2}$ of Eq. (5.11) as $N_c \to \infty$. This question will be addressed in more details when discussing the convergence rate in a later subsection. We will drop the subscript N_c to lighten notations if the context is not ambiguous.

The approximate numerical problem Eq. (5.13) is a finite linear system, or to be more accurate, two independent linear systems with the same kernel matrix but a different right hand side for the incident p or s-polarization state. In order to handle it numerically, it is convenient to re-write these systems in the conventional way in which the kernel matrix is a two dimensional array and the unknowns and right hand sides take the form of one dimensional arrays. So far, the unknown elements $(X_{.p}^{(\mathbf{m})}(\mathbf{p}_0))_{\mathbf{m}\in\mathcal{D}_{N_c}}$ and $(X_{.s}^{(\mathbf{m})}(\mathbf{p}_0))_{\mathbf{m}\in\mathcal{D}_{N_c}}$ are indexed by three indices, α, m_1 and m_2. We re-order the elements in the following way. First, with use a bijective mapping from $\mathcal{D}_{N_c}$ to $[\![1, D]\!]$, where $D = \mathrm{card}(\mathcal{D}_{N_c})$ denotes the cardinal of $\mathcal{D}_{N_c}$, i.e. the number of distinct elements in the set $\mathcal{D}_{N_c}$. In other words, we associate a single integer n from 1 to D to each element $\mathbf{m} = (m_1, m_2) \in \mathcal{D}_{N_c}$. There are several such bijections ($D!$ to be exact) but it is convenient[1] to choose a bijection such that the elements of $\mathcal{D}_{N_c}$ are ordered with increasing value of $m_1^2 + m_2^2$, i.e. that $|\mathbf{m}_{n+1}| \geq |\mathbf{m}_n|$. For an illustration of the mode indexing mapping see Fig. 5.1. Then, the elements of $(X_{.\beta}^{(\mathbf{m})}(\mathbf{p}_0))_{\mathbf{m}\in\mathcal{D}_{N_c}}$ (for $\beta \in \{p, s\}$) are ordered in a column vector $\mathbf{X}_\beta(\mathbf{p}_0)$ in such a way that the D first elements are of the form $X_{p\beta}^{(\mathbf{m}_n)}(\mathbf{p}_0)$ with increasing values of n and the D remaining elements are of the form $X_{s\beta}^{(\mathbf{m}_n)}(\mathbf{p}_0)$ with increasing values of n, i.e. that we write

$$\mathbf{X}_\beta(\mathbf{p}_0) = \left(X_{p\beta}^{(\mathbf{m}_1)}(\mathbf{p}_0), \cdots, X_{p\beta}^{(\mathbf{m}_D)}(\mathbf{p}_0), X_{s\beta}^{(\mathbf{m}_1)}(\mathbf{p}_0), \cdots, X_{s\beta}^{(\mathbf{m}_D)}(\mathbf{p}_0)\right)^{\mathrm{T}}. \tag{5.14}$$

Consistently, the kernel is re-ordered in a matrix $\mathbf{M}(\mathbf{p}_0)$ and the right hand side in a column

[1] Convenient but not necessary.

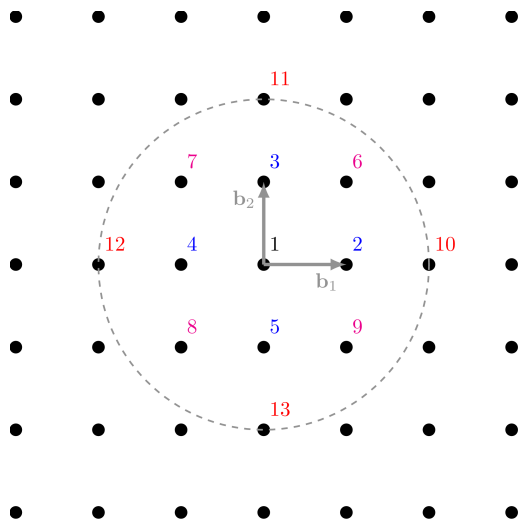


Figure 5.1: Vertices of the reciprocal lattice of a square lattice. The basis vectors $(\mathbf{b}_i)_{i\in\{1,2\}}$ are the primitive reciprocal lattice vectors. The domain $\mathcal{D}_{N_c}$ is delimited by the dashed circle. Here the index cutoff radius $N_c = 2$ and $\operatorname{card}\mathcal{D}_{N_c} = 13$. A possible indexing is suggested indicated as integers associated with each vertex in the domain delimited by the dashed circle. Integers indices printed in the same color indicates that their associated reciprocal lattice vectors share the same norm.

vector $\mathbf{N}_\beta(\mathbf{p}_0)$ such that for $(i,j) \in [\![1, D]\!]^2$,

$$M_{ij}(\mathbf{p}_0) = M_{pp}(\mathbf{p}_{\boldsymbol{\ell}_i} \mid \mathbf{p}_{\mathbf{m}_j}) \tag{5.15}$$
$$M_{D+i,j}(\mathbf{p}_0) = M_{sp}(\mathbf{p}_{\boldsymbol{\ell}_i} \mid \mathbf{p}_{\mathbf{m}_j}) \tag{5.16}$$
$$M_{i,D+j}(\mathbf{p}_0) = M_{ps}(\mathbf{p}_{\boldsymbol{\ell}_i} \mid \mathbf{p}_{\mathbf{m}_j}) \tag{5.17}$$
$$M_{D+i,D+j}(\mathbf{p}_0) = M_{ss}(\mathbf{p}_{\boldsymbol{\ell}_i} \mid \mathbf{p}_{\mathbf{m}_j})\,, \tag{5.18}$$

and

$$\mathbf{N}_\beta(\mathbf{p}_0) = \left(N_{p\beta}(\mathbf{p}_{\boldsymbol{\ell}_1} \mid \mathbf{p}_0), \cdots, N_{p\beta}(\mathbf{p}_{\boldsymbol{\ell}_D} \mid \mathbf{p}_0), N_{s\beta}(\mathbf{p}_{\boldsymbol{\ell}_1} \mid \mathbf{p}_0), \cdots, N_{s\beta}(\mathbf{p}_{\boldsymbol{\ell}_D} \mid \mathbf{p}_0)\right)^{\mathrm{T}}\,. \tag{5.19}$$

This re-ordering yields a linear system appropriate for numerical implementation of the form

$\mathbf{M}\,\mathbf{X} = \mathbf{N}$ which when expanded reads

$$\begin{pmatrix} M_{pp}(\mathbf{p}_{\mathbf{m}_1} \mid \mathbf{p}_{\mathbf{m}_1}) & \cdots & M_{pp}(\mathbf{p}_{\mathbf{m}_1} \mid \mathbf{p}_{\mathbf{m}_D}) & M_{ps}(\mathbf{p}_{\mathbf{m}_1} \mid \mathbf{p}_{\mathbf{m}_1}) & \cdots & M_{ps}(\mathbf{p}_{\mathbf{m}_1} \mid \mathbf{p}_{\mathbf{m}_D}) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ M_{pp}(\mathbf{p}_{\mathbf{m}_D} \mid \mathbf{p}_{\mathbf{m}_1}) & \cdots & M_{pp}(\mathbf{p}_{\mathbf{m}_D} \mid \mathbf{p}_{\mathbf{m}_D}) & M_{ps}(\mathbf{p}_{\mathbf{m}_D} \mid \mathbf{p}_{\mathbf{m}_1}) & \cdots & M_{ps}(\mathbf{p}_{\mathbf{m}_D} \mid \mathbf{p}_{\mathbf{m}_D}) \\ M_{sp}(\mathbf{p}_{\mathbf{m}_1} \mid \mathbf{p}_{\mathbf{m}_1}) & \cdots & M_{sp}(\mathbf{p}_{\mathbf{m}_1} \mid \mathbf{p}_{\mathbf{m}_D}) & M_{ss}(\mathbf{p}_{\mathbf{m}_1} \mid \mathbf{p}_{\mathbf{m}_1}) & \cdots & M_{ss}(\mathbf{p}_{\mathbf{m}_1} \mid \mathbf{p}_{\mathbf{m}_D}) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ M_{sp}(\mathbf{p}_{\mathbf{m}_D} \mid \mathbf{p}_{\mathbf{m}_1}) & \cdots & M_{sp}(\mathbf{p}_{\mathbf{m}_D} \mid \mathbf{p}_{\mathbf{m}_D}) & M_{ss}(\mathbf{p}_{\mathbf{m}_D} \mid \mathbf{p}_{\mathbf{m}_1}) & \cdots & M_{ss}(\mathbf{p}_{\mathbf{m}_D} \mid \mathbf{p}_{\mathbf{m}_D}) \end{pmatrix}$$

$$\begin{pmatrix} X_{pp}^{(\mathbf{m}_1)}(\mathbf{p}_0) & X_{ps}^{(\mathbf{m}_1)}(\mathbf{p}_0) \\ \vdots & \vdots \\ X_{pp}^{(\mathbf{m}_D)}(\mathbf{p}_0) & X_{ps}^{(\mathbf{m}_D)}(\mathbf{p}_0) \\ X_{sp}^{(\mathbf{m}_1)}(\mathbf{p}_0) & X_{ss}^{(\mathbf{m}_1)}(\mathbf{p}_0) \\ \vdots & \vdots \\ X_{sp}^{(\mathbf{m}_D)}(\mathbf{p}_0) & X_{ss}^{(\mathbf{m}_D)}(\mathbf{p}_0) \end{pmatrix} = \begin{pmatrix} N_{pp}(\mathbf{p}_{\mathbf{m}_1} \mid \mathbf{p}_0) & N_{ps}(\mathbf{p}_{\mathbf{m}_1} \mid \mathbf{p}_0) \\ \vdots & \vdots \\ N_{pp}(\mathbf{p}_{\mathbf{m}_D} \mid \mathbf{p}_0) & N_{ps}(\mathbf{p}_{\mathbf{m}_D} \mid \mathbf{p}_0) \\ N_{sp}(\mathbf{p}_{\mathbf{m}_1} \mid \mathbf{p}_0) & N_{ss}(\mathbf{p}_{\mathbf{m}_1} \mid \mathbf{p}_0) \\ \vdots & \vdots \\ N_{sp}(\mathbf{p}_{\mathbf{m}_D} \mid \mathbf{p}_0) & N_{ss}(\mathbf{p}_{\mathbf{m}_D} \mid \mathbf{p}_0) \end{pmatrix}.$$

Remark 5.3. The size of each numerical linear system is $2D$ where we recall that $D = \mathrm{card}(\mathcal{D}_{N_c})$ which physically corresponds to the numerical number of modes included in the expansion of the electric field. The cardinal of $\mathcal{D}_{N_c}$ increases with the index cut off radius N_c in a quadratic fashion. A rough estimate for D for large enough index cut off N_c is given by $D \sim \lfloor \pi N_c^2 \rfloor$. Thus the size of the linear systems scales roughly as $2\pi N_c^2$. The exact number of modes for a few integer values of N_c are reported in Table 5.1 for a square lattice.

N_c	D	$\lfloor \pi N_c^2 \rfloor$	N_c	D	$\lfloor \pi N_c^2 \rfloor$	N_c	D	$\lfloor \pi N_c^2 \rfloor$	N_c	D	$\lfloor \pi N_c^2 \rfloor$
0	1	0	6	113	113	12	441	452	18	1009	1017
1	5	3	7	149	153	13	529	530	19	1129	1134
2	13	12	8	197	201	14	613	615	20	1257	1256
3	29	28	9	253	254	15	709	706	21	1373	1385
4	49	50	10	317	314	16	797	804	22	1517	1520
5	81	78	11	377	380	17	901	907	23	1653	1661

Table 5.1: Number of modes $D = \mathrm{card}(\mathcal{D}_{N_c})$ for some values of the index cut off radius N_c for a square lattice. The rough estimate $D \sim \lfloor \pi N_c^2 \rfloor$ is also given.

5.2.2 Complexity analysis

Now that the truncation and re-ordering procedures have been described, it is clear that the resolution of a reduced Rayleigh equation consists in two main parts: (i) setting up the linear

system, and (ii) solving it. We analyze in the following the complexity associated with these two steps. We will see that solving the linear system, for a given number of modes D, always requires the same number of operations (if a direct solver is used) and this *independently* of the apparent complexity of the physical system, e.g. a system with a single interface compared to a system containing several interfaces. However *nothing* can be said on the number of modes required to achieve a good approximate solution. Indeed, if a system containing a single interface requires D modes for the solution to be approximated to some prescribed precision, it is not necessarily true that the same number of modes D will be sufficient if a second interface is added and hence an additional medium (say with a similar dielectric constant as the two original ones). We will rather focus on the complexity associated with setting up the linear system, i.e. in computing the elements of the kernel matrix $\mathbf{M}$ and the right hand side $\mathbf{N}$ for a given fixed number of modes D.

Solving the system

Assume the linear system is set up and that it is left to solve it. There are basically two classes of solvers for linear systems: the direct methods that follow an algorithm that gives the exact[2] solution in a finite number of operations, and iterative methods that give successive approximate solutions of the system. We will deal with iterative methods in a separate chapter, and focus on the LU decomposition method as a direct method. The reason for choosing specifically the LU decomposition here as a prototypical direct solver, rather than the Gauss elimination method for example, is that we found it to be practical since the linear system needs to be solved for two right hand sides. Indeed, once the LU factorization is achieved, solving for multiple right hand sides does not significantly increase the complexity [60].

The LU method consists in two steps: the LU factorization, which requires $(2D-1)2D(2D+1)/3$ operations and a back substitution step that requires $2D(2D+1)/2$ operations per right hand side [60]. The total number of operations for solving the linear system (for the two right hand sides) through LU factorization is then

$$N_{\mathrm{op,LU}} = \frac{(2D-1)2D(2D+1)}{3} + 2D(2D+1) = \frac{2D(2D+1)(2D+2)}{3} \sim \frac{(2D)^3}{3}\,. \tag{5.20}$$

Setting up the system

We have seen that for a given number of modes D the size of the linear system is $2D$. This means that they are $(2D)^2$ elements to compute to set up the matrix and $4D$ elements for the right hand sides corresponding to the p and s incident polarization states. Note that the matrix elements need only to be computed once for the two subsystems. At first sight, it seems that if an algorithm computing a single element requires K operations, then the number of operations needed to set up the whole system would be $N_{\mathrm{op}} = K((2D)^2 + 4D) = 4KD(D+1)$. This clearly shows that setting up the matrix system requires a number of operations that scales *quadratically* with the number of modes and *linearly* with the number of operations needed to compute a single element of the system entries. Let us analyze more in details the elements of $\mathbf{M}$ and $\mathbf{N}$ to see how expensive they are to compute, i.e. give an estimate of K and see whether some computation can be spared.

[2]Exact under the assumption that the entries are known exactly and exact arithmetic is used, which is never the case numerically due to round off errors.

Case 1: a single interface

Let us start by considering the case of a single periodic interface. In this case the kernel is of the form

$$\mathbf{M}\left(\mathbf{p}\,|\,\mathbf{q}\right) = \tilde{\mathcal{J}}^{b,a}_{l,m}\left(\mathbf{p}\,|\,\mathbf{q}\right)\,\mathbf{M}^{b,a}_{l,m}\left(\mathbf{p}\,|\,\mathbf{q}\right), \tag{5.21}$$

where it is reminded that $\tilde{\mathcal{J}}^{b,a}_{l,m}$ is essentially

$$\tilde{\mathcal{J}}^{b,a}_{l,m}\left(\mathbf{p}\,|\,\mathbf{q}\right) = \frac{1}{a_c}\,\frac{\int_{a_c} e^{-i\left(\mathbf{k}^b_l\left(\mathbf{p}\right)-\mathbf{k}^a_m\left(\mathbf{q}\right)\right)\cdot\mathbf{s}(\mathbf{x})}\,\mathrm{d}^2x}{b\alpha_l\left(\mathbf{p}\right)-a\alpha_m\left(\mathbf{q}\right)},$$

and

$$\mathbf{M}^{b,a}_{l,m}\left(\mathbf{p}\,|\,\mathbf{q}\right) = \begin{pmatrix} |\mathbf{p}||\mathbf{q}| + ab\alpha_l\left(\mathbf{p}\right)\alpha_m\left(\mathbf{q}\right)\hat{\mathbf{p}}\cdot\hat{\mathbf{q}} & -b\sqrt{\epsilon_m}\frac{\omega}{c}\alpha_l\left(\mathbf{p}\right)[\hat{\mathbf{p}}\times\hat{\mathbf{q}}]_3 \\ a\sqrt{\epsilon_l}\frac{\omega}{c}\alpha_m\left(\mathbf{q}\right)[\hat{\mathbf{p}}\times\hat{\mathbf{q}}]_3 & \sqrt{\epsilon_l\epsilon_m}\frac{\omega^2}{c^2}\hat{\mathbf{p}}\cdot\hat{\mathbf{q}} \end{pmatrix}.$$

The right hand side has a similar form for the equation for the reflection amplitude and a rather simpler form for that of the transmission amplitude (see Chapter 4). We thus focus on the kernel terms. According to our numerically convenient re-ordering, setting up the $\mathbf{M}$ matrix, which is of size $2D\times 2D$, consists in computing D^2 2×2 sub-matrices of the form given in Eq. (5.21) evaluated at different pairs of in-plane wave-vectors. Then each element of the sub-matrices can be placed in $\mathbf{M}$ according to (i) the polarization coupling and (ii) the pair of wave-vectors or equivalently mode indices. The pair of polarization indices gives in which quadrant they must be placed, and the pair of mode index set the position within the quadrant.

Remark 5.4. A remarkable property of Eq. (5.21) is that the factor $\tilde{\mathcal{J}}^{b,a}_{l,m}\left(\mathbf{p}\,|\,\mathbf{q}\right)$ *does not* depend on the polarization indices, which means that it needs to be computed only once for each pair of mode indices.

The previous remark is of great interest as the factor $\tilde{\mathcal{J}}^{b,a}_{l,m}\left(\mathbf{p}\,|\,\mathbf{q}\right)$ is the one that requires most of the computational power for setting up the matrix and the remark spares us roughly a factor 4 in computational time compared to a naive element by element evaluation of the matrix. To be more specific, let us now count the number of operations κ needed to set up a set of four elements associated with a pair of wave-vectors. Hence, the number of operations needed to set up the whole matrix $\mathbf{M}$ will be $D^2\kappa$. Assume that there exists an algorithm that gives an accurate evaluation of $\tilde{\mathcal{J}}^{b,a}_{l,m}\left(\mathbf{p}\,|\,\mathbf{q}\right)$ in J number of operations. From the definition of $\mathbf{M}^{b,a}_{l,m}\left(\mathbf{p}\,|\,\mathbf{q}\right)$ one needs about 4 evaluations of square roots, 20 multiplications and 6 additions (we have assumed that terms that do not depend on the wave vector are computed once and for all and are not counted). As it is usual to only keep multiplications when counting operations but taking into account that square roots are a bit more expensive than multiplications, let us simply say that they are roughly 26 operations then, i.e. $\kappa\approx 30+J$. How large can J be? This depends greatly on the method to evaluate $\tilde{\mathcal{J}}^{b,a}_{l,m}\left(\mathbf{p}\,|\,\mathbf{q}\right)$. Some surface profiles, like sinusoidal profiles, give closed form expressions for $\tilde{\mathcal{J}}^{b,a}_{l,m}\left(\mathbf{p}\,|\,\mathbf{q}\right)$ in terms of a Bessel function (see Chapter A). Efficient algorithm for evaluating Bessel functions can then be used. However it is in general not possible to obtain closed form expression for arbitrary profiles. One method that we have used consists in expanding $\tilde{\mathcal{J}}^{b,a}_{l,m}\left(\mathbf{p}\,|\,\mathbf{q}\right)$ is a power series in the following way. Let us define

$$\xi^{(\mathbf{m})}(\gamma) = \frac{1}{a_c}\int_{a_c} e^{-i\,\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{x}}\,e^{-i\gamma\zeta(\mathbf{x})}\,\mathrm{d}^2x\,. \tag{5.22}$$

With this definition we have

$$\tilde{\mathcal{J}}^{b,a}_{l,m}\left(\mathbf{p}_0^{(\boldsymbol{\ell})} \mid \mathbf{p}_0^{(\mathbf{m})}\right) = \frac{\xi^{(\boldsymbol{\ell}-\mathbf{m})}\left(b\alpha_l(\mathbf{p}_0^{(\boldsymbol{\ell})}) - a\alpha_m(\mathbf{p}_0^{(\mathbf{m})})\right)}{b\alpha_l(\mathbf{p}_0^{(\boldsymbol{\ell})}) - a\alpha_m(\mathbf{p}_0^{(\mathbf{m})})}\,. \tag{5.23}$$

We then focusing on evaluating $\xi^{(\mathbf{m})}(\gamma)$. By expanding the exponential factor $e^{-i\gamma\zeta(\mathbf{x})}$ in the integrand in Eq. (5.22) we obtain an expression for $\xi^{(\mathbf{m})}(\gamma)$ as a power series

$$\xi^{(\mathbf{m})}(\gamma) = \sum_{n=0}^{\infty} \frac{(-i\gamma)^n}{n!}\hat{\zeta}^{(n)}_{a_c}(\mathbf{G}^{(\mathbf{m})}) = \sum_{n=0}^{\infty} a_n^{(\mathbf{m})}(-i\gamma)^n\,, \tag{5.24}$$

where we define the *Fourier moment* of order n of the surface profile ζ as

$$\hat{\zeta}^{(n)}_{a_c}(\mathbf{G}^{(\mathbf{m})}) = \frac{1}{a_c}\int_{a_c} e^{-i\,\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{x}}\,\zeta^n(\mathbf{x})\,\mathrm{d}^2x\,. \tag{5.25}$$

Appendix A is devoted to the computation of $\xi^{(\mathbf{m})}(\gamma)$ and the Fourier moments for a collection of surface profiles. Here it is enough to notice that the coefficients $(a_n^{(\mathbf{m})})_{n\in\mathbb{N}}$ of the power series are *independent* of γ. The coefficients can therefore be pre-computed for a given set of mode indices.

Remark 5.5. One may be doubtful about the gain of pre-computed these coefficients before setting up the linear system as they still depend on mode indices and hence are different for the different elements of the linear system. Even though there may be indeed little gain if one is only interested to study the scattering from a given surface and for a given specific incident wave vector and wavelength, there is a significant gain if one is interested to loop over the parameters defining the incident excitation. Indeed, for a given surface profile, the coefficients $(a_n^{(\mathbf{m})})_{n\in\mathbb{N}}$ *only depend on the surface parameters* and can be re-used if either the wavelength, the dielectric properties of the media, and/or the incident wave-vector are changed. We will therefore disregard here the complexity associated with the computation of these coefficients.

By truncating the series so that only the N_s+1 first terms are kept, we then need to evaluate a polynomial of degree N_s. The naive approach consisting in summing term by term leads to $N_s(N_s+1)/2$ multiplications. This can however be improved by the use of the Hörner method [61, 62] for the evaluation of polynomials which applied to our example reads as follow

Hörner's algorithm:

Initialize $\xi^{(\mathbf{m})}(\gamma) \leftarrow a_{N_s}^{(\mathbf{m})}$.
Loop for $n = N_s$ to 1
  $\xi^{(\mathbf{m})}(\gamma) \leftarrow a_{n-1}^{(\mathbf{m})} + (-i\gamma)\,\xi^{(\mathbf{m})}(\gamma)$
End of the loop.

Hence only N_s multiplications are needed, which gives a significant speed up compared to the naive approach when N_s becomes large. Hence $J \approx N_s$.

The next question now is the following. How many terms must be kept in the power series to ensure an accurate evaluation of $\xi^{(\mathbf{m})}(\gamma)$? This is a delicate question that depends on both the surface profile parameters and the argument γ. Indeed, the convergence rate of the sequence of partial sums depends both on γ and on how fast the coefficients $(a_n^{(\mathbf{m})})_{n\in\mathbb{N}}$ vanish when $n\to\infty$.

Moreover, since the argument γ depends in practice of pairs of in-plane wave-vectors $\mathbf{p}_0^{(\boldsymbol{\ell})}$ and $\mathbf{p}_0^{(\mathbf{m})}$, the number of terms to achieved a prescribed accuracy can changed from matrix element to matrix element. In practice, either a fixed large enough number of terms N_s is imposed based on experience and then the accuracy is verified *a posteriori*, or an adaptive scheme can be used in which N_s is gradually increased and the outcomes for the approximation of $\xi^{(\mathbf{m})}(\gamma)$ for successive values of N_s are compared until the desired accuracy is reached. Thus, assuming that N_s is fixed for all elements in the linear system, the *number of operations to set up the linear system* is estimated to be

$$N_{\text{op,set}} \approx (N_s + 30)D(D+1)\,, \tag{5.26}$$

which yields the follows asymptotic scaling in terms of the index radius cutoff N_c

$$N_{\text{op,set}} \sim \begin{cases} 4(N_s+30)N_c^2\,, & \text{for 1D systems} \\ \pi^2(N_s+30)N_c^4\,, & \text{for 2D systems} \end{cases}. \tag{5.27}$$

Here we recall that N_s is the number of terms kept in the truncation of $\xi^{(\mathbf{m})}(\gamma)$, and we have used the rough scaling $D \sim \pi N_c^2$ for the number of modes D for two-dimensional surfaces and $D = 2N_c + 1$ for one-dimensional surfaces. In the case of a two-dimensional surface, the number of operations for setting up the linear system associated with a truncation index N_c scales as the fourth power of N_c. Note that the scaling coefficient is non-negligible. Indeed, a lower bound would be about 300 if $N_s = 0$, but in practice, $N_s \leq 120$, which gives an upper scaling constant around 1500.

The total number of operations associated with setting and solving the matrix system, for a two-dimensional surface, is then

$$N_{\text{op,tot}} \sim \pi^2(N_s+30)N_c^4 + \frac{(2\pi)^3}{3}N_c^6\,. \tag{5.28}$$

Here we voluntarily keep both the contribution from the setting and solving steps in the asymptotic scaling, which is cumbersome but helps illustrating that there will be *two regimes of dominant complexity* depending on N_c. For small values of N_c the computational burden will essentially resides in setting up the linear system due to the rather large scaling coefficient $\pi^2(N_s + 30)$ whereas for large values of N_c the term associated with solving the linear systems dominates due to its larger exponent compared to that of the term associated with setting up the linear systems. This effect is illustrated in Fig. (5.2) where the CPU time spent to set up and solve the system is reported as a function of N_c in the cases of a two-dimensional sinusoidal surface and of a supported hemiellipsoidal profile. In the case of the sinusoidal surface, a standard algorithm is used to evaluated Bessel functions [63, 64] whereas a power series expansion has been used in the case of the hemiellipsoidal profile with $N_s = 120$ (see Appendix A). In both cases the LAPACK solver GESV is used[3]. The two regimes are clearly visible, with the setting-up dominated regime for low values of N_c where the total time in black is essentially that spent in setting the system in red, and the solving dominated regime where the total time is essentially spent in the linear solver in blue. Note that the time spent in the solver is independent of the surface profile and that the critical values N_c for which equal time is spent for setting up and solving the linear system varies slightly with the surface profile as two different methods are chosen for computing $\xi^{(\mathbf{m})}(\gamma)$. For the sinusoidal surface this critical values is found to be around $N_c = 16$ and for the hemiellipsoidal profile it is found around $N_c = 11$.

[3]This general solver uses the LU decomposition, see http://www.netlib.org/lapack/

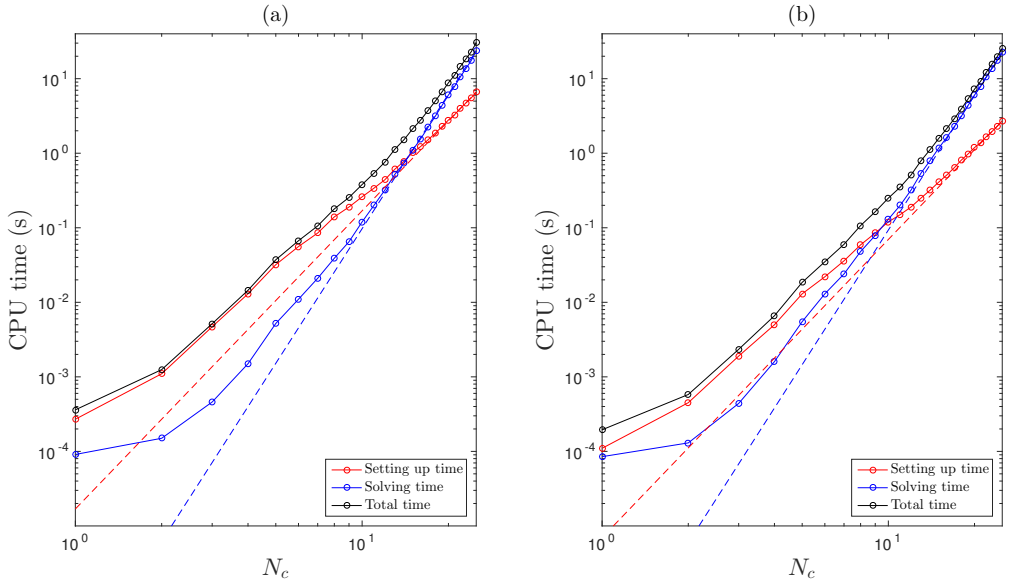


Figure 5.2: CPU time required for setting up the linear system and solving it as a function of the mode index cut off N_c. The dashed lines correspond to fit in $\mathcal{O}(N_c^4)$ and $\mathcal{O}(N_c^6)$ to the set up time and solving time respectively. (a) Two-dimensional sinusoidal surface. (b) Hemiellipsoidal profile.

Case 2: several interfaces

We have seen in Section 4.4.1 that for a periodic multi-layer system consisting of two interfaces periodic interfaces (in the degenerate case), the transfer kernel for the equation for the reflection amplitudes is of the of the form (see Eqs. (4.40) and (4.38))

$$\mathbf{K}_L^{(r)}(\mathbf{p}_{\boldsymbol{\ell}} \,|\, \mathbf{p}_{\mathbf{m}}) = \sum_{\boldsymbol{\ell}_2 \in \mathbb{Z}^2} \sum_{a_2 = \pm} a_2 \, \tilde{\boldsymbol{\Theta}}_{32}^{+a_2} \left(\mathbf{p}_{\boldsymbol{\ell}} \,|\, \mathbf{p}_{\boldsymbol{\ell}} - \mathbf{G}_2^{(\boldsymbol{\ell}_2)} \right) \tilde{\boldsymbol{\Theta}}_{21}^{a_2 +} \left(\mathbf{p}_{\boldsymbol{\ell}} - \mathbf{G}_2^{(\boldsymbol{\ell}_2)} \,|\, \mathbf{p}_{\mathbf{m}} \right) . \tag{5.29}$$

Once truncated, the sum over $\boldsymbol{\ell}_2$ will run over $\mathcal{D}_{N_c}$ instead of $\mathbb{Z}^2$ and it is clear that after the re-indexing procedure is applied, the kernel matrix can be written as a sum of two matrix products of the form

$$\mathbf{K}_L^{(r)} = \mathbf{M}_1 \, \mathbf{M}_2 - \mathbf{M}_3 \, \mathbf{M}_4 \, , \tag{5.30}$$

where the matrices $(\mathbf{M}_i)_{1 \leq i \leq 4}$ are readily obtained from the truncated versions the single interface kernels present in Eq. (5.29). As in the case for a single interface, setting up each matrix $\mathbf{M}_i$ takes about $\pi^2(N_s + 30)N_c^4$ operations[4], so about $4\pi^2(N_s + 30)N_c^4$ in total. But this is not the problematic part. What costs even more, when N_c becomes large, is simply to perform the matrix-matrix multiplications, which consists in $(2D)^3 = (2\pi)^3 N_c^6$ scalar multiplications for

[4]Some computation can be saved by setting the matrices simultaneously, but this is not important for the overall scaling.

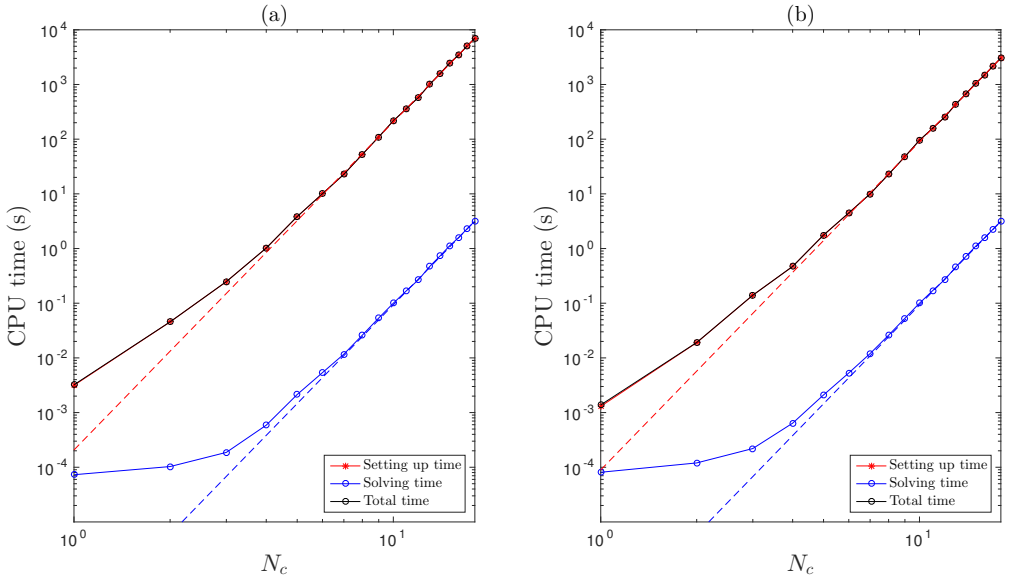


Figure 5.3: CPU time required for setting up the linear system and solving it as a function of the mode index cut off N_c. The dashed lines correspond to fits in $\mathcal{O}(N_c^6)$ to the set up time and solving time respectively. (a) Three layers system with two two-dimensional sinusoidal surfaces. (b) Three layers system with two hemiellipsoidal profiles.

each product. The total operation count for setting up the kernel matrix $\mathbf{K}_L^{(r)}$ is then

$$N_{\text{op},set} \approx 2^4\pi^3 N_c^6 + 4\pi^2(N_s+30)N_c^4\,. \tag{5.31}$$

This is larger than the number of operations required to solve the system ($\sim (2\pi)^3 N_c^6/3$), although of the same power scale with N_c, and hence always dominates the complexity of the whole algorithm. Figure 5.3 illustrate this fact very clearly. It shows that the total CPU time used to solve the scattering problem (black circles) is essentially the same as the CPU time used to set up the system. One can also appreciate the same power law scaling in N_c^6 both for the set up time and the solving time. We believe Fig. 5.3 comes as a clear illustration for the need for alternative ways to solve the problem of scattering in the case of two and more rough interfaces, hence justifying the discussions in Sections 3.5.2 and 3.5.3.

5.2.3 Truncation error and convergence

We now turn to the question of truncation error and the rate of convergence at which an approximate solution, $\mathbf{X}_{N_c}(\mathbf{p}_0) = (\mathbf{X}_{N_c}^{(\mathbf{m})}(\mathbf{p}_0))_{\mathbf{m}\in\mathbb{Z}^2}$, converges to the exact solution, $\mathbf{X}(\mathbf{p}_0) = (\mathbf{X}^{(\mathbf{m})}(\mathbf{p}_0))_{\mathbf{m}\in\mathbb{Z}^2}$ of the reduced Rayleigh equation. First, we need to define the sense to be given to convergence, or in other words we need to define a distance between the two solutions. The solutions will be viewed as elements in $\mathbb{C}^{\mathbb{N}}$, i.e. as a sequence of complex numbers. Indeed, as already mentioned, a re-ordering can be made so that the four indices required to identify a component of a solution (two indices for polarization coupling and two for the mode index)

only become one index in $\mathbb{N}$. We will see, that it is not necessary to explicitly give the bijection between $\{p,s\}^2\times\mathbb{Z}^2$ and $\mathbb{N}$ to make calculations, but we wanted to point out that such a bijection exists and that everything works as if we compared sequences of complex numbers due to the countable property of the set of elements required to describe a solution in the case of periodic systems. Thus we will use standard distances between two complex sequences associated with the conventional ℓ^p-norms, namely the distance between $\mathbf{X}_{N_c}(\mathbf{p}_0)$ and $\mathbf{X}(\mathbf{p}_0)$, that we will refer to as *truncation error*, associated with the ℓ^p-norm reads

$$\left\|\mathbf{X}_{N_c}-\mathbf{X}\right\|_p=\begin{cases}\left(\displaystyle\sum_{\mathbf{m}\in\mathbb{Z}^2}\sum_{\alpha,\beta\in\{p,s\}}\left|X^{(\mathbf{m})}_{\alpha\beta,N_c}-X^{(\mathbf{m})}_{\alpha\beta}\right|^p\right)^{1/p}, & p\in[1,\infty[\\ \displaystyle\sup_{\substack{\mathbf{m}\in\mathbb{Z}^2\\ \alpha,\beta\in\{p,s\}}}\left|X^{(\mathbf{m})}_{\alpha\beta,N_c}-X^{(\mathbf{m})}_{\alpha\beta}\right|, & p=\infty\end{cases}. \tag{5.32}$$

Here we have dropped the dependence on $\mathbf{p}_0$ to lighten notations. We would like to remind the reader, that $X^{(\mathbf{m})}_{\alpha\beta,N_c}(\mathbf{p}_0)=0$ for all polarization coupling and all mode index $\mathbf{m}$ such that $|\mathbf{m}|>N_c$. We then say that the approximate solution converges to the exact solution, in the sense of the ℓ^p-norm, if $\left\|\mathbf{X}_{N_c}-\mathbf{X}\right\|_p\to 0$ as $N_c\to\infty$. In practice, we will restrict ourselves to the norms corresponding to $p=1,2$ and ∞.

In fact, since the exact solution is usually not known, we will rather compare numerically consecutive approximate solutions, say $\mathbf{X}_{N_c}$ and $\mathbf{X}_{N_c+1}$, hence the sum over $\mathbf{m}\in\mathbb{Z}^2$ reduces to a sum over $\mathbf{m}\in\mathcal{D}_{N_c+1}$. Furthermore, in order to remove the sensitivity of the distance with respect to the overall scale of the full solution, we normalize the distance and define a *relative distance*, $\varepsilon_{N_c,p}$, as

$$\varepsilon_{N_c,p}=\frac{\left\|\mathbf{X}_{N_c+1}-\mathbf{X}_{N_c}\right\|_p}{\left\|\mathbf{X}_{N_c}\right\|_p}. \tag{5.33}$$

If $\mathbf{X}_{N_c}$ converges to $\mathbf{X}\neq 0$ with respect to the ℓ^p-norm then $\varepsilon_{N_c,p}$ converges to 0, but the reciprocal statement is not true in general. In fact, we should rather verify that $\left\|\mathbf{X}_{N_c+m}-\mathbf{X}_{N_c}\right\|_p$ goes to zero as $N_c\to\infty$ for any integer $m>0$, in order to show that $\mathbf{X}_{N_c}$ is a Cauchy sequence in $\ell^p(\mathbb{C})$. Nevertheless, for the sake of simplicity and because we are primarily interested in searching numerically some conjectures, we will abusively call $\varepsilon_{N_c,p}$ a *relative truncation error*, and say that the method converges (with respect to a given norm) if $\varepsilon_{N_c,p}\to 0$ as N_c is increased. We will specifically be interested in the speed at which $\varepsilon_{N_c,p}\to 0$ and study how the convergence rate scales with reduced physical parameters identified earlier such as $2\pi a_i/\lambda_1$, $2\pi\max|\zeta|/\lambda_1$, and $z=\sqrt{\epsilon_2/\epsilon_1}$.

Convergence rate

We now give some vocabulary and a definition of the convergence rate that we have mentioned previously as a quantity measuring how fast the sequence of numerically approximated solutions converges (given that it does). For clarity here the index denoting the norm used for the definition of the error will be dropped. When the approximate solution converges to the exact solution in the sense given above, i.e. $\lim_{N_c\to\infty}\epsilon_{N_c}=0$, we say that the approximate solution converges *linearly* if there exists a constant $\eta\in[0,1[$ such that for all index cutoff N_c large

enough, $N_c > \kappa \geq 0$, we have

$$\varepsilon_{N_c+1} \leq \eta \, \varepsilon_{N_c} \, . \tag{5.34}$$

We say that the convergence is *superlinear* of order $q > 1$ if there exists a constant C such that for all index cutoff large enough, $N_c > \kappa \geq 0$, we have

$$\varepsilon_{N_c+1} \leq C \, \varepsilon_{N_c}^{q} \, . \tag{5.35}$$

We may say that the convergence is quadratic if $q = 2$, cubic if $q = 3$, etc ... In the case of linear convergence, if we assume that $\varepsilon_{N_c} > 0$ and define $\eta_{N_c} = \varepsilon_{N_c+1}/\varepsilon_{N_c}$, linear convergence states that the sequence $(\eta_{N_c})_{N_c \in \mathbb{N}}$ is bounded by a constant smaller than 1 after a large enough value of N_c. Since by definition $\eta_{N_c} > 0$, the sequence $(\eta_{N_c})_{N_c \in \mathbb{N}}$ is a bounded sequence of real numbers in the case of linear (and also superlinear) convergence, and we can therefore extract a converging subsequence (Bolzano-Weierstrass theorem). Moreover, and more relevant for our definition of the convergence rate, the sequence $(\eta_{N_c})_{N_c \in \mathbb{N}}$ admits both a lim inf and a lim sup. We may define the two following *convergence rates* as

$$\eta_{\text{inf}} = \liminf_{N_c \to \infty} \eta_{N_c} \, , \tag{5.36}$$

$$\eta_{\text{sup}} = \limsup_{N_c \to \infty} \eta_{N_c} \, . \tag{5.37}$$

Intuitively, we could say that η_{inf} represents asymptotically the best improvement from one iteration to the next one can expect, whereas η_{sup} rather represents the worst. Asymptotically the actual improvement from an iterate to the next lies in between η_{inf} and η_{sup}. In practice, we will numerically take the average of the computed η_{N_c} after the first few approximated solutions (e.g. $N_c \leq \kappa \approx 3$) if a clear linear convergence is observed, i.e.

$$\eta = \frac{1}{N_c^{\max} + 1 - \kappa} \sum_{N_c=\kappa}^{N_c^{\max}} \eta_{N_c} \, , \tag{5.38}$$

where κ is some arbitrarily chosen threshold and $N_c^{\max}$ is the largest index cutoff for which a simulation is achieved.

Linear convergence is also known in the literature as *exponential* convergence although the denomination may sound overselling. What is meant by the term *exponential* is the following. Assume that for N_c large enough the ratio $\varepsilon_{N_c+1}/\varepsilon_{N_c}$ is constant equal to some $0 < \eta < 1$. Then we can write by taking the logarithm of $\varepsilon_{N_c+1}/\varepsilon_{N_c}$ that

$$\ln \varepsilon_{N_c+1} = \ln \eta + \ln \varepsilon_{N_c} \, , \tag{5.39}$$

which can be iterated to yield

$$\ln \varepsilon_{N_c} = N_c \ln \eta + \ln \varepsilon_0 \, . \tag{5.40}$$

or equivalently, by defining the decay rate $\beta = -\ln \eta > 0$

$$\varepsilon_{N_c} = \varepsilon_0 \exp(-\beta N_c) \, . \tag{5.41}$$

We have thus shown that linear convergence yields a model of the form of an exponential decay for the error with a decay rate $\beta = -\ln \eta > 0$, hence the term exponential convergence. We will see that this trend is indeed what is observed for the convergence of the numerical solution of the reduced Raleigh equations as the number of modes is increased, and we will pay particular

attention to the variation of the convergence rate η, or equivalently the decay rate β , with the reduced parameters identified in Section 5.1.

Dependence of the convergence rate with characteristic lengths

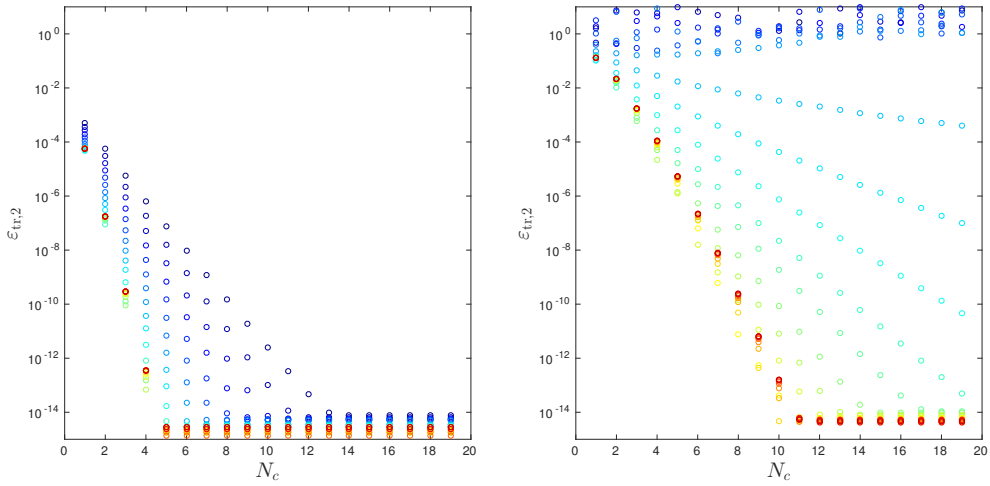
We start our numerical analysis of the convergence properties of the numerical solution of the reduced Rayleigh equations with respect to the number of modes, or rather the index cutoff N_c, with some numerical experiments. We insist here that our aim is not to give mathematically solid proofs but rather search some conjectures for convergence criterion and scaling of the convergence rates. The numerical conjectures may then be considered as a starting point for further rigorous mathematical analysis.

Our first step consists in studying how the convergence is affected with the different regimes of lengths scales one may encounter. The reduction of units in Section 5.1 hinted at the fact that there are essentially two types of length parameters which are expected to play a role in the physics of wave scattering by a surface; these are the characteristic transverse and out-of-plane lengths of the surface with respect to the wavelength. For a periodic surface, these are the lattice constants and the typical amplitude of the surface profile. To fix the ideas, we consider a two-dimensional sinusoidal profile with equal lattice constants along the two orthogonal directions ($a_1 = a_2 = a$), i.e. the surface profile is given by

$$\zeta(\mathbf{x}) = \frac{H}{2}\left[\sin\left(\frac{2\pi}{a}x_1\right) + \sin\left(\frac{2\pi}{a}x_2\right)\right], \tag{5.42}$$

where a denotes the lattice constant and H the amplitude. As mentioned previously in the present chapter, the $\xi^{(\mathbf{m})}(\gamma)$ integral associated with the above sinusoidal surface can be determined analytically in terms of Bessel functions of the first kind (see Appendix A) for which efficient and accurate numerical algorithms can be used [63, 64]. We also fix the dielectric constant of the media to $\epsilon_1 = 1$ and $\epsilon_2 = 2.25$ for now, i.e. that we fix the reduced parameter $z = \sqrt{\epsilon_2/\epsilon_1} = 1.5$, so that only the influence of the length scales are investigated. We will allow the dielectric contrast to vary in a second step later.

Figures 5.4(a) and 5.4(b) show the relative truncation error, using the euclidean norm, as a function of the index cutoff N_c for different values of the reduced lattice constant $2\pi a/\lambda_1$ respectively for a small reduced amplitude $2\pi H/\lambda_1 = 0.01$ (Fig. 5.4(a)) and a larger one $2\pi H/\lambda_1 = 0.5$ (Fig. 5.4(b)). The colors indicate different values of the reduced lattice constant $2\pi a/\lambda_1$ increasing from blue to red, i.e. blue represents a regime where $a \ll \lambda_1$ ($a/\lambda_1 = 0.02$) and red a regime where $a \gg \lambda_1$ ($a/\lambda_1 = 100$). We would like to mention, that all numerical experiments in Figs. 5.4(a) and 5.4(b) are achieved for normal incidence for simplicity. In particular, due to the obvious symmetry of the problem, we will not distinguish between the incident p- and s-polarized excitation for the error, since the solutions are the same up to rotation. It is clear from Fig. 5.4(a) that the numerical solution converges as the index cutoff N_c is increased for all the considered values of the reduced lattice constant. As indicated by the logarithmic scale for the relative truncation error, the decay of the error for each value of the reduced lattice constant is exponential in the sense given above, as points follow a straight line. The convergence can hence be characterized as linear and we can estimate the corresponding convergence rate η (or equivalently the corresponding decay rate β). This seems to be true for all the tested values of the reduced lattice constant but with different convergence rate. Indeed, Fig. 5.4(a) shows that the convergence is faster for larger values of the reduced lattice



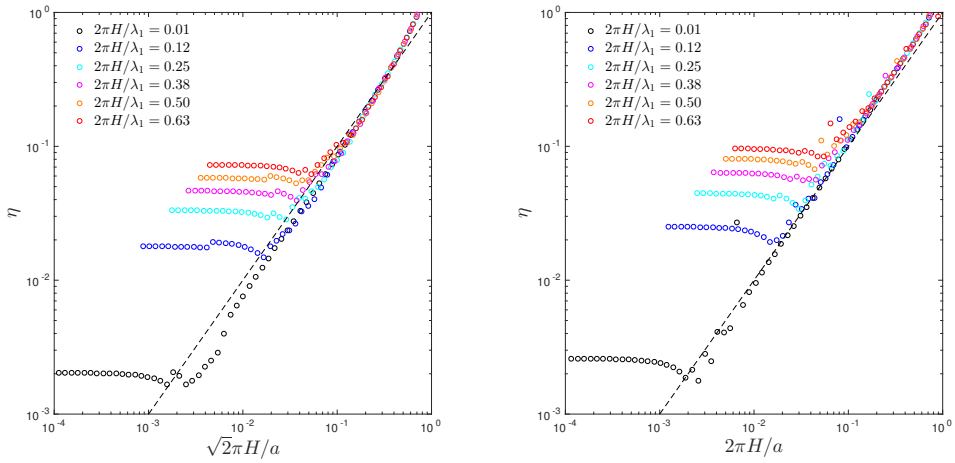
(a) Fixed $2\pi H/\lambda_1 = 0.01$ and variable a/λ_1 (logarithmic scale) between 0.02 and 100.

(b) Fixed $2\pi H/\lambda_1 = 0.5$ and variable a/λ_1 (logarithmic scale) between 0.02 and 100.

Figure 5.4: Relative truncation error using the euclidean norm as a function of the index cutoff N_c. The considered system is made of vacuum and glass ($\epsilon_1 = 1$, $\epsilon_2 = 2.25$ hence $z = 1.5$) separated by a two-dimensional sinusoidal surface. Each color represents a different value of the reduced lattice constant $2\pi a/\lambda_1$ (increasing value from blue to red). The reduced amplitude of the sinusoidal profile is fixed to $2\pi H/\lambda_1 = 0.01$ in (a) and to $2\pi H/\lambda_1 = 0.5$ in (b). The numerical experiments are achieved for normal incidence.

constant. The convergence rate hence depends on this parameter. Figure 5.4(b) illustrates the same conclusion as Fig. 5.4(a) but complement it with two important observations. First, for the same values of the reduced lattice constant as the ones from Fig. 5.4(a) the convergence rate differs if one increases the reduced amplitude $2\pi H/\lambda_1$. In other words, the convergence is slower the larger the reduced amplitude for a given reduced lattice constant. Second, convergence is not always guaranteed. Indeed, for small enough reduced lattice constants, the numerical solution are found to be unbounded which can be seen in Fig. 5.4(b) with the cluster of points oscillating around an error of one. Note that due to our definition of the relative truncation error, if the norm of the solution grows without bound, the relative error may stay bounded due to the normalization. We have indeed verified that the solutions grows without bound for the values of the reduced lattice constant which do not exhibit convergence in Fig. 5.4(b). These two new pieces of information tell us that: (i) there probably exist some conditions on the reduced lattice constant and the reduced amplitude to ensure convergence of the numerical solution; (ii) given that the numerical solution converges, the linear convergence rate η depends both on the reduced lattice constant and the reduced amplitude. In particular, we have observed that the larger the amplitude and the smaller the lattice constant, the slower the convergence. Could it be that convergence is controlled by a typical slope?

We show in Fig. 5.5 a graph of the convergence rate η as a function of $s = \sqrt{2}\pi H/a$ which



(a) 2D (b) 1D

Figure 5.5: Convergence rate η as a function of the maximum slope s of the sinusoidal profile. (a) Two-dimensional sinusoidal profile (maximum slope equal to $s = \sqrt{2}\pi H/a$), i.e. same system studied in Fig. 5.4. (b) One-dimensional sinusoidal profile (maximum slope equal to $2\pi H/a$). The different data sets are obtained for different values of the reduced amplitudes $s = 2\pi H/\lambda_1 \in \{0.01, 0.12, 0.24, 0.37, 0.50, 0.63\}$. The dashed line is the line of equation $\eta = s$ as a guide for the eyes.

is the maximum slope found on the profile. For a two-dimensional profile, we understand the term *slope* as a slope for a one dimensional cut of the surface along an arbitrary direction. Figure 5.5 is made by evaluating the convergence rate in similar simulations as the ones illustrated in Fig. 5.4 for different values of the reduced amplitude linearly space ($2\pi H/\lambda_1 \in \{0.01, 0.12, 0.25, 0.38, 0.50, 0.63\}$) as indicated by the different colors in the legend. For a given reduced amplitude (a given color), the points are determined by scanning over the reduced lattice constant (i.e. similar to the different curves in Fig. 5.4). Recall that given that convergence is guaranteed, the linear convergence rate η is between 0 and 1, and the smaller η the faster the convergence. Figure 5.5(a) reveals two interesting features of the convergence rate. First, for small enough surface slopes, the convergence rate is nearly constant, the value being essentially given by the reduced amplitude. Second, for large enough surface slopes, all the data points collapse on the same line (in a loglog plot) and the convergence rate η only depends on the characteristic slope s and not on the individual values of the reduced amplitude and lattice constant. In other words, Fig. 5.5(a) reveals two regimes of convergence rate which we can qualify as the *amplitude dominated regime* and the *slope dominated regime*. In the amplitude dominated regime the convergence rate η only depends on the reduced amplitude, while in the slope dominated regime the convergence rate η only depends on the surface slope in a power law fashion. We can see that η is roughly linear with s in the slope dominated regime although there could be some deviation as the data points do not follow exactly the dashed line in Fig. 5.5(a). The discrepancy may be caused either by some uncertainty in determining η and/or the behavior of η as a function of s may truly be non-linear. Only a finer mathematical

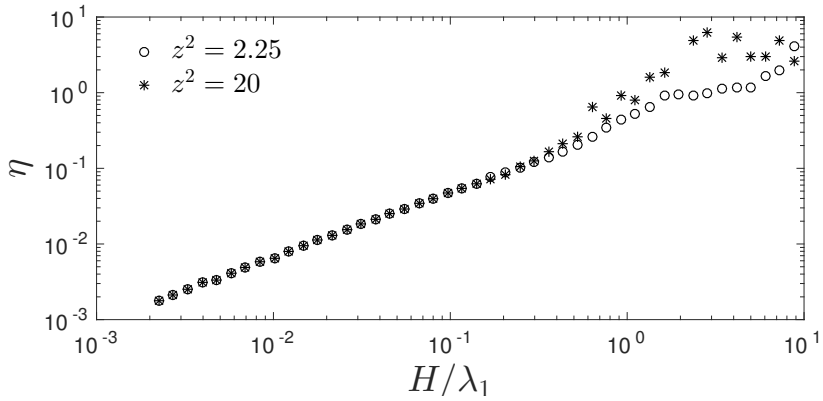


Figure 5.6: Convergence rate η as a function of the reduced amplitude H/λ_1 for a two-dimensional sinusoidal profile and fixed maximum slope $s = 10^{-4}$. Open circle correspond to a dielectric contrast of $z^2 = 2.25$ and stars correspond to a dielectric contrast of $z^2 = 20$.

analysis will resolve the issue. Figure 5.5(b) shows similar results as Fig. 5.5(a) but for a one dimensional sinusoidal profile of equation

$$\zeta(\mathbf{x}) = H \sin\left(\frac{2\pi}{a} x_1\right) , \tag{5.43}$$

for which the maximum slope is $s = 2\pi H/a$. The one and two-dimensional cases exhibit very similar behaviors when it comes to the convergence rate. In addition, it is quite remarkable that Fig. 5.5 also provides us with a simple convergence criterion, namely that we must have

$$s < C \approx 1 \tag{5.44}$$

where C is some constant close to unity, to guarantee convergence. But is it enough? What would happen if we continued to increase further the reduced amplitude while keeping the slope constant? How does the convergence rate scale with $2\pi H/\lambda_1$ in the amplitude dominated regime? Figure 5.6 shows the dependence of the convergence rate η on the normalized amplitude[5] H/λ_1 in the amplitude dominated regime for a fixed value of the maximum slope $s = 10^{-4}$ and this for two rather different values of the dielectric contrast $z^2 = 2.25$ and $z^2 = 20$. Focusing on the dielectric contrast $z^2 = 2.25$, which corresponds to the value fixed in our previous numerical experiments, we can see that in the amplitude dominated regime $\eta \approx H/\lambda_1$ as long as H/λ_1 is smaller than a value close to unity. In addition, changing the dielectric contrast seems to have no effect on the convergence rate in this regime.

To sum up, there seems to be two criterion to be satisfied to guarantee convergence. There is an amplitude criterion which roughly reads as $H/\lambda_1 < 1$ and a slope criterion which reads $s < 1$. The value of unity in the inequalities is not to be taken as exact as we have not found exactly unity in our simulations but rather some values close to unity. In the slope dominated regime, the convergence rate is roughly linear with the slope while in the amplitude dominated regime the convergence rate is roughly linear with the normalized amplitude. All these features are well

[5]The reason for removing the 2π is that we have found that $\eta \approx H/\lambda_1$ in this regime.

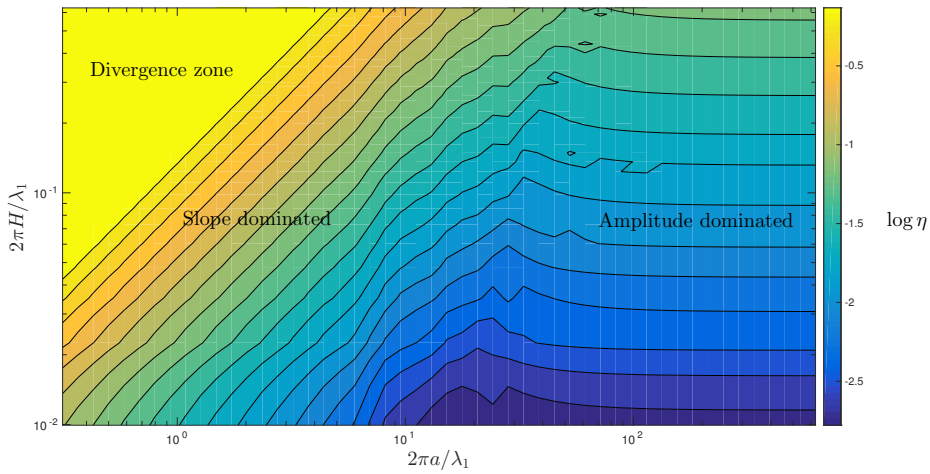


Figure 5.7: Contour diagram of $\log \eta$ in the $(2\pi a/\lambda_1, 2\pi H/\lambda_1)$-plane for a two-dimensional sinusoidal surface separating two dielectrics of dielectric contrast $z = 1.5$. For divergent results, η is forced to be saturated to unity for clarity.

captured by Fig. 5.7, which shows a contour diagram of the convergence rate (in a logarithmic scale) in the $(2\pi a/\lambda_1, 2\pi H/\lambda_1)$-plane. The diagram illustrates the zone of divergence where η has been forced to be saturated to 1, the slope dominated zone where lines of constant η are oblique straight lines of constant maximum slope s and the amplitude dominated regime where the lines of constant η are horizontal line of constant amplitude. Furthermore, as noted from Fig. 5.6, it seems that the dielectric contrast z does not influence significantly the convergence rate. However, we will see now that it somewhat influences the starting point in the sequence of relative error $(\varepsilon_{N_c})_{N_c>0}$.

Influence of the dielectric contrast on the convergence

Figure 5.8(a) shows the relative truncation error as a function of the index cutoff N_c for different values of the dielectric contrast z varying fro 0.01 (blue) to 100 (red) and for fixed values of the reduced lattice constant $(a/\lambda_1 = 1)$ and amplitudes $(H/\lambda_1 = 0.1)$. We can observe that for small enough values of z the decay rate of the error seems to be independent of z as all the data points asymptotically form parallel lines, i.e. after large enough values of N_c. The main difference between different values of z seems to reside in the starting point, ε_1. We can observe, roughly, that for small values of z the starting point ε_1 is independent of z while when z becomes large enough the starting point lifts towards larger values of the error. This means that there can be significant difference in the consecutive solutions for small values of N_c. In particular, the relative error can be so large as to somewhat saturate the definition due to the normalization as mentioned previously, and can as can be seen from the cluster of points around an error of unity in Fig. 5.8(a) for large values of z. Then after enough modes have been taken into account the error decays with a decay rate independent of z. These features are well illustrated in Figs. 5.9(a) and (b) in the case of a dielectric-dielectric system (red circles). Figure 5.9(a)

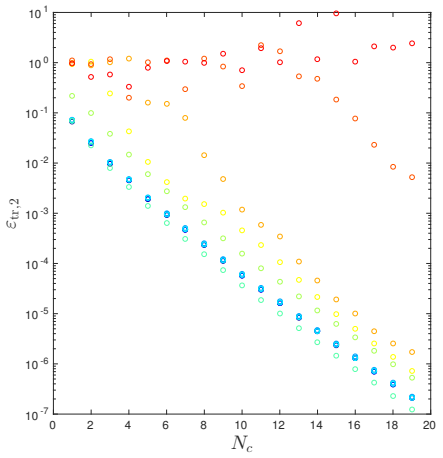
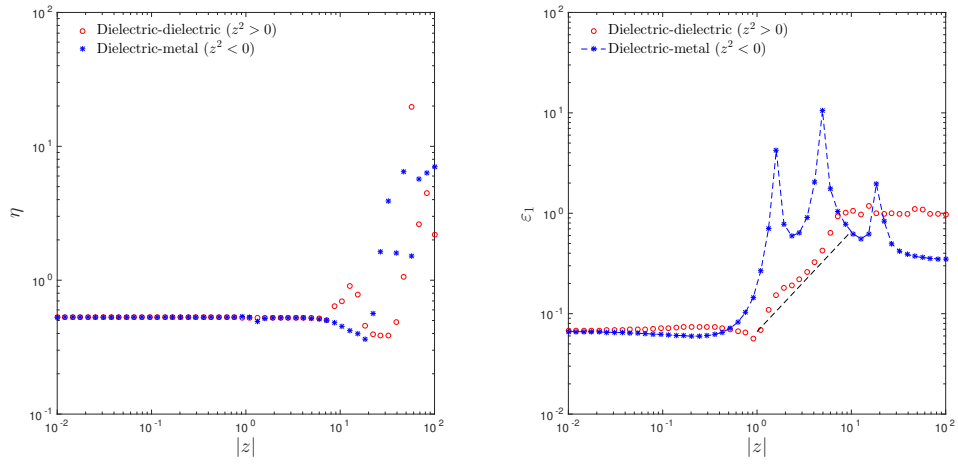


Figure 5.8: Relative truncation error using the euclidean norm as a function of the index cutoff N_c. The considered system is made of vacuum ($\epsilon_1 = 1$) and a dielectric for different values of the dielectric constant (z varies on a logarithmic scale between 0.01 (blue) and 100 (red)). The surface profile is a two-dimensional sinusoidal surface characterized by $a/\lambda_1 = 1$ and $H/\lambda_1 = 0.1$. The numerical experiments are achieved for normal incidence.

shows the convergence rate η measured for a two-dimensional sinusoidal profile characterized by the aforementioned values of the reduced lattice constant and amplitude, as a function of z. The fact that η is independent of z is then readily observed until some rather extreme value of the dielectric contrast $z \approx 10$ where the convergence rate varies wildly. We argue that these variations are not significantly due to true dependence of the decay rate on the dielectric contrast but rather an growing inaccuracy and even impossibility of determining the decay rate as the simulation were probably not pushed far enough in N_c to observe a clear enough convergence (or to not observe convergence at all within the range of N_c considered here, see the red data points in Fig. 5.9(a)). Figure 5.9(b) shows the behavior of the starting relative error as a function z for the same parameters used in Fig. 5.9(a). It is interesting to notice to regimes. First for dielectric contrasts z smaller than unity (i.e. medium of incidence optically denser that the medium of transmission) the starting relative error is constant with z. Then, for $z > 1$, the starting relative error increases linearly with z. The plateau for $z > 10$ corresponds to a case where the error is saturated due to the normalization and may not represent the true behavior of the absolute error as function of z. We speculate that the absolute error still increases linearly with z then.

Metallic systems

We have considered so far dielectric systems, and we are now exploring the convergence of the method for metallic systems. Fixing a dielectric contrast of $z^2 = -2.25$ and repeating the same numerical experiments as to obtain Fig. 5.7, we obtain the convergence diagram shown in Fig. 5.10(a). We can see that the convergence diagram for a dielectric-metallic system is overall very similar to that of a dielectric-dielectric system. The main feature that differs is now the presence of an island of divergence or slow convergence roughly between the slope dominated



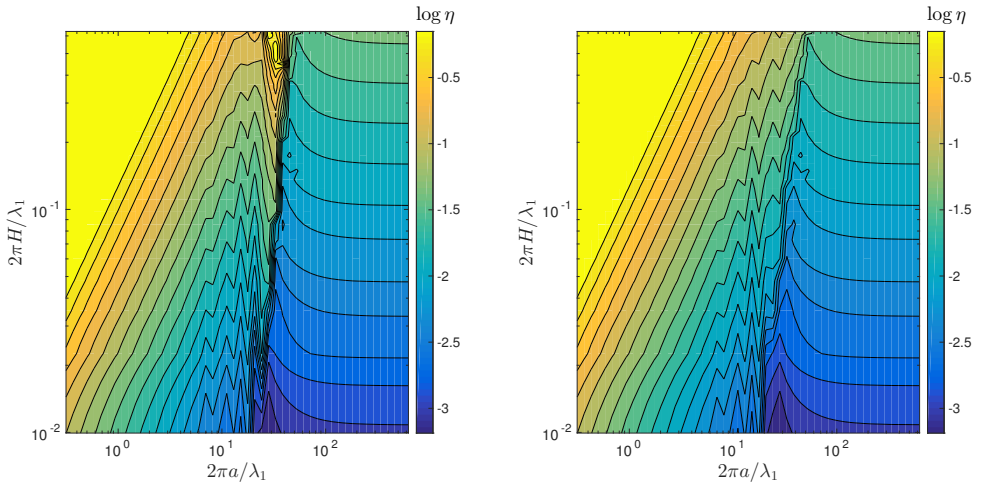
(a) (b)

Figure 5.9: (a) Convergence rate η and (b) starting error ϵ_0 as a functions of the modulus of the dielectric contrast $|z| = |\sqrt{\epsilon_2/\epsilon_1}|$ for a sinusoidal profile characterized by $a/\lambda_1 = 1$ and $H/\lambda_1 = 0.1$. The red open circles are results for dielectric-dielectric systems ($z^2 > 0$) while the blue stars are results for dielectric-metallic systems ($z^2 < 0$). The dashed line is the line of equation $\eta = z$ as a guide for the eyes.

regime and the amplitude dominated regime. Although a deeper investigation is needed to understand this feature, we believe this is due to the excitation of surface plasmon polaritons. Such resonances are expected to yield poorly convergent results, if not divergent. Figures 5.9(a) and 5.9(b) show the dependence of η and ε_1 on the absolute dielectric contrast in the amplitude dominated regime. It can be observed that, as for dielectric-dielectric systems, η is constant as a function of $|z|$ in the regime, as it can also be observed in Fig. 5.10, but signatures of a potential effect of resonances seems to be present in ε_1. A deeper and more detailed study should be carried out in future research on the topic. If we now consider a metal with loss by adding a small imaginary part to the dielectric constant, $z^2 = -2.25 + 0.1\,i$, the island of divergence vanishes as can be seen in Fig. 5.10(b). The imaginary part of the dielectric constant for the metal seems to damp the resonance and hence helps the method to converge. We also believe that the smaller the imaginary part of the dielectric constant, the slower the convergence will be in the region near a resonance.

5.3 Randomly rough systems

We will now explain briefly how to handle arbitrary, i.e. non-periodic, surfaces. First, we must keep in mind some characteristic length scales of the physical system in order to introduce as few artifacts as possible. What are the length scales at play in the case of scattering by a randomly rough interface? For a periodic interface, these were the wavelength of the incident light, the lattice constants and the characteristic amplitude of the profile. For a randomly rough surface,



(a) $z^2 = -2.25$ (b) $z^2 = -2.25 + 0.1\,i$

Figure 5.10: Contour diagram of $\log \eta$ in the $(2\pi a/\lambda_1, 2\pi H/\lambda_1)$-plane for a two-dimensional sinusoidal surface separating a dielectric medium and a metallic medium with dielectric contrast (a) $z^2 = -2.25$ and (b) $z^2 = -2.25 + 0.1\,i$. For divergent results, η is forced to be saturated to unity for clarity.

the so-called *correlation lengths* and the *rms-roughness* of the surface will be the characteristic lengths of interest together with the wavelength. Indeed, we have seen in Section 1.6.1 that the statistical properties of random surfaces, which are realizations of a stochastic process are essentially determined by the probability density of heights and the auto-correlation function. Two characteristic length scales associated with these two data, are the rms roughness and the correlation length. Since its is not possible to represent random surfaces of infinite size on a computer, only a certain portion of it will be represented (in a sense that will be made clear soon). Intuitively, it is natural to require that the portion of surface which is represented should have a characteristic size which is both larger than the wavelength and larger than the correlation length. Indeed, having for example a size of the surface smaller than the correlation length may induce a undesired correlation due to the truncation of the surface. In the following, we illustrate two methods one may apply to treat the case of random surfaces.

5.3.1 Method 1 - Periodizing the system

The first method we suggest here consists in periodizing the surface. Assume that a portion of the surface, say a square of size $L \times L$, is large enough to encode the statistical properties of the surface. In practice, this means that the length of the truncated surface is much larger that the typical correlation length, $L \gg a$. Then one can hope to sample densely enough the continuous diffusely scattered light, or to be more accurate the reflection and transmission amplitudes, $X_{\alpha\beta}(\mathbf{p} \,|\, \mathbf{p}_0)$, if one also choose $L \gg \lambda$ and repeat the same surface periodically in both directions. Indeed, we can now view the system as a periodic surface with a period much larger than the wavelength of the incident light such that, in virtue of the grating theorem, the in-plane wave vectors of diffracted modes $(\mathbf{p}_{\boldsymbol{\ell}})_{\boldsymbol{\ell}\in\mathbb{Z}^2}$ are densely packed and effectively sample the

continuous response of the infinite system. Then intuitively, the larger the lattice constant L, the better the approximation will be. In practice, this means that from a numerical point of view, it suffices to use what has been done for a periodic system in the previous sections of the present chapter. The continuous scattering amplitude can then be estimated at the sampling points, $\mathbf{p}_{\boldsymbol{\ell}}$, roughly as

$$X_{\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_0) \approx X^{(\boldsymbol{\ell})}_{\alpha\beta}(\mathbf{p}_0)\,\Delta p^{-2}\,, \tag{5.45}$$

by assuming that the scattering amplitude does not vary much in a small box $[\mathbf{p}_{\boldsymbol{\ell}}\cdot\hat{\mathbf{e}}_1 - \Delta p/2, \mathbf{p}_{\boldsymbol{\ell}}\cdot\hat{\mathbf{e}}_1 + \Delta p/2] \times [\mathbf{p}_{\boldsymbol{\ell}}\cdot\hat{\mathbf{e}}_2 - \Delta p/2, \mathbf{p}_{\boldsymbol{\ell}}\cdot\hat{\mathbf{e}}_2 + \Delta p/2]$, where $\Delta p = 2\pi/L$ is the step between two reciprocal lattice vectors.

5.3.2 Method 2 - Finite size system and numerical quadrature

A second method for solving the reduced Rayleigh equations numerically, or more generally, Fredholms integral equation of the first kind, consists in the following approximation steps leading to a finite size linear system. Consider a typical reduced Rayleigh equation of the form

$$\int \mathbf{M}(\mathbf{p}\,|\,\mathbf{q})\;\mathbf{X}(\mathbf{q}\,|\,\mathbf{p}_0)\,\frac{\mathrm{d}^2 q}{(2\pi)^2} = -\,\mathbf{N}(\mathbf{p}\,|\,\mathbf{p}_0)\,. \tag{5.46}$$

The first step consists in approximating the unknown solution by its projection onto some finite dimensional functional space, say a space of function defined by a basis $(\phi_j)_{1\le j\le N}$. This basis can be arbitrarily chosen, and can be for example a truncated Fourier basis, Lagrange interpolation polynomials, or whatever family of functions one may believe suited to approximate the unknown solution with as few terms as possible as

$$\mathbf{X}(\mathbf{q}\,|\,\mathbf{p}_0) \approx \sum_{j=1}^{N} \mathbf{X}_j(\mathbf{p}_0)\,\phi_j(\mathbf{q})\,. \tag{5.47}$$

The (matrix valued) coefficients $(\mathbf{X}_j(\mathbf{p}_0))_{1\le j\le N}$ are then the new unknown of the problem. By substitution of Eq. (5.47) into Eq. (5.46) we obtain

$$\sum_{j=1}^{N} \int \mathbf{M}(\mathbf{p}\,|\,\mathbf{q})\,\phi_j(\mathbf{q})\,\frac{\mathrm{d}^2 q}{(2\pi)^2}\,\mathbf{X}_j(\mathbf{p}_0) \approx -\,\mathbf{N}(\mathbf{p}\,|\,\mathbf{p}_0)\,. \tag{5.48}$$

The next step consists in choosing a numerical quadrature to approximate the integral. Such a quadrature can generally be written as

$$\int f(\mathbf{q})\,\frac{\mathrm{d}^2 q}{(2\pi)^2} \approx \sum_{k=1}^{\kappa} w_k\, f(\mathbf{q}_k)\,, \tag{5.49}$$

where κ weight coefficients, $(w_k)_{1\le k\le\kappa}$, and sampling points, $(\mathbf{q}_k)_{1\le k\le\kappa}$, are defined by the quadrature. The sampling points may be evenly spaced (Newton-Cotes quadratures) or unevenly spaced (Gauss quadratures) [61]. By approximation of the integral in Eq. (5.48) we obtain

$$\sum_{j=1}^{N}\sum_{k=1}^{\kappa} \mathbf{M}(\mathbf{p}\,|\,\mathbf{q}_k)\,w_k\,\phi_j(\mathbf{q}_k)\;\mathbf{X}_j(\mathbf{p}_0) \approx -\,\mathbf{N}(\mathbf{p}\,|\,\mathbf{p}_0)\,, \tag{5.50}$$

which we recast as

$$\sum_{j=1}^{N} \mathbf{K}_j(\mathbf{p})\;\mathbf{X}_j(\mathbf{p}_0) \approx -\,\mathbf{N}(\mathbf{p}\,|\,\mathbf{p}_0)\,, \tag{5.51}$$

with the definition

$$\mathbf{K}_j(\mathbf{p}) \stackrel{\text{def}}{=} \sum_{k=1}^{\kappa} w_k \, \phi_j(\mathbf{q}_k) \, \mathbf{M}(\mathbf{p} \,|\, \mathbf{q}_k) \,. \tag{5.52}$$

Equation (5.51) resemble an algebraic linear system of equations with finitely many unknowns, $(\mathbf{X}_j(\mathbf{p}_0))_{1 \le j \le N}$, but we now have infinitely many equations, namely one for each value of $\mathbf{p}$. We can now choose a set of N sample points, $(\mathbf{p}_i)_{1 \le i \le N}$, which may or may not be identical to the sampled $\mathbf{q}$-points, to obtain the following finite dimensional linear system

$$\sum_{j=1}^{N} \mathbf{K}_{ij} \, \mathbf{X}_j(\mathbf{p}_0) = -\,\mathbf{N}_i(\mathbf{p}_0) \,, \tag{5.53}$$

with

$$\mathbf{K}_{ij} \stackrel{\text{def}}{=} \mathbf{K}_j(\mathbf{p}_i) = \sum_{k=1}^{\kappa} w_k \, \phi_j(\mathbf{q}_k) \, \mathbf{M}(\mathbf{p}_i \,|\, \mathbf{q}_k) \tag{5.54}$$

$$\mathbf{N}_i(\mathbf{p}_0) \stackrel{\text{def}}{=} \mathbf{N}(\mathbf{p}_i \,|\, \mathbf{p}_0) \,. \tag{5.55}$$

5.4 Summary

Let us summarize the main results from the present chapter. First we have shown in details how to solve numerically the reduced Rayleigh equations for periodic systems. For a periodic system, the reduced Rayleigh equations reduce to infinite countable linear systems of equations for the unknown scattering amplitudes for the set of discrete modes. A finite linear system was then obtained by truncation of the set of discrete modes kept in the system and the resulting linear system of equations was solved numerically by using direct methods such as Gauss elimination or LU factorization. Then we have explored the convergence property of the numerical solution with increasing number of modes kept in the linear system. We have found that the convergence rate depends mainly on surface parameters such as the typical slope of the surface and the maximum amplitude of the surface in unit of wavelength. In particular, two regimes have been identified depending on the ratio of the lattice constant and the wavelength. For lattice constants smaller than the wavelength, the convergence rate is determined by the characteristic slope of the surface profile, while for lattice constants larger than the wavelength, the convergence rate is determined by the characteristic amplitude of the surface normalized by the wavelength. Finally, some techniques for treating numerically non-periodic surfaces have been briefly illustrated.

The next two chapters will be devoted to the studies of iterative methods and of the consistency of the reduced Rayleigh equations method. In particular, we will study how the convergence rate of the iterative methods is affected by the surface parameters and how to quantify the quality of the numerical solutions. We will see that the two identified regimes of convergence rate found in the present chapter will again be present in these two studies. We would like to note in passing that previous study assessing the accuracy of approximate methods [65] and the validity of the reduced Rayleigh equations based on a conservation of energy criterion [66] have also reported similar diagrams of accuracy showing the slope and amplitude dominated regimes.

Chapter 6

Solving the reduced Rayleigh equations; iterative methods

"Pourquoi faire des Mathématiques ? Parce que les Mathématiques, ça sert à faire de la Physique. La Physique, ça sert à faire des frigidaires. Les frigidaires, ça sert à y mettre des langoustes, et les langoustes, ça sert aux mathématiciens, qui les mangent et sont alors dans de bonnes dispositions pour faire des Mathématiques, qui servent à la Physique, qui sert à faire des frigidaires qui ...".

Laurent Schwartz.

In Chapter 5, we have presented how the reduced Rayleigh equations can be solved numerically, and that the two major steps in this process are (i) the construction of a linear system that represents a truncated approximation of a reduced Rayleigh equation, and (ii) solving the obtained linear system. In particular, we have seen that independently of the physical system considered, solving the linear system with so-called direct methods, such as Gauss elimination or LU factorization, always requires a number of operations that scales as the cube of the number of modes kept in the truncation of the reduced Rayleigh equation. Moreover in the cases of layered systems in which only one of the interface is rough and the others are flat, the complexity associated with solving the system will take over that of setting up the linear system for large enough number of modes kept in the truncation. It is then of interest to consider iterative methods to solve the system in order to decrease the complexity associated with it. It must be kept in mind though that for layered systems involving more than two rough interfaces, the complexity resides essentially in constructing the linear system, and improving the solving step would not help to significantly speed up the computer program. A potential alternative way, based on the reduced Rayleigh equations of the second kind has been discussed in Chapter 3 already to tackle this issue for layered system with more than one rough interface. Nevertheless, systems involving only one rough interface cover a non-negligible class of physical systems of interest from an application point of view, and it is the prior aim of this chapter to present iterative methods to give reasonably good numerical approximated solutions of the truncated reduced Rayleigh equations in these cases.

We will first present standard textbook iterative solvers such Jacobi, Gauss-Seidel, and successive over-relaxation (SOR) methods before dealing with iterative methods specifically designed for solving the reduced Rayleigh equations. We will then present the so-called *small amplitude perturbation theory* (SAPT) which consists in expanding the solution in powers of the surface profile function, in a sense that will be made clear soon, and approximating the

solution by computing each order of the expansion by the use of a recursive formula. We will see that this method can be viewed as an iterative method using several iteration matrices. The derivation of the small amplitude perturbation theory will be given first here for a system involving a single interface to fix the ideas, and then generalized to a stack of layers with arbitrarily many interfaces in Chapter 9. Finally, the fixed point method based on the *reduced Rayleigh equation of the second kind* will prove to be a powerful method which can be viewed as a more efficient perturbation method than small amplitude perturbation theory. Finally, we will consider a fixed point algorithm based on the *coupled Rayleigh equations*, and discuss its practicality or rather impracticality, due to stability issues.

The different methods will be compared with respect to their complexity, their convergence and stability for different class of physical systems, and their convergence rate towards the solution given by the direct method. Note that here we *do not* consider whether the truncation error is small enough or not, which is an issue that has already been discussed in Chapter 5 and that is a sole property of the truncated linear system, and is therefore independent of the method chosen for solving it (provided that the method manages to give the solution). We rather analyze the *convergence of the iterated approximated solutions of the truncated linear system towards the solution given by a direct method* that will serve as a reference.

6.1 Simple iterative methods

We consider here simple iterative methods to solve the linear system $\mathbf{M}\,\mathbf{X} = \mathbf{N}$ of size $2D$ associated with a truncation of a reduced Rayleigh equation that we assumed has already been set as explained in Chapter 5, and $\mathbf{M}$ is assumed non-singular. The term *simple* refers to the fact that these methods are found in most textbooks on numerical linear algebra but also to the fact that the construction of the sequence of approximate solutions, denoted $(\mathbf{X}_k)_{k\in\mathbb{N}}$, consists in applying operations on the current approximate solution. In other words, the computation of an approximate solution $\mathbf{X}_{k+1}$ only requires the knowledge of the approximate solution $\mathbf{X}_k$ and not that of any other prior approximate solutions. The following is adapted from Ref. [60] where more details can be found. Here we only give the minimum required for practical implementation, but the interested reader will consult Ref. [60] for more analysis.

6.1.1 Implementation

We first define the *splitting* of the matrix $\mathbf{M}$ as a pair of matrices $(\mathbf{M}_1, \mathbf{M}_2)$ such that

$$\mathbf{M} = \mathbf{M}_1 - \mathbf{M}_2\,, \tag{6.1}$$

with $\mathbf{M}_1$ constrained to be non-singular. The iterative method associated with the splitting $(\mathbf{M}_1, \mathbf{M}_2)$ is then given by

Initialize $\mathbf{X}_0$.

Loop until a desired precision, or a maximum number of iterations is reached

$$\mathbf{M}_1\,\mathbf{X}_{k+1} = \mathbf{M}_2\,\mathbf{X}_k + \mathbf{N} \tag{6.2}$$

End of the loop.

Remark 6.1. Note that *if* the sequence $(\mathbf{X}_k)_{k\in\mathbb{N}}$ converges to some limit $\mathbf{X}$ *then* by plugging the limit into Eq. (6.2), we find that the limit is necessarily the solution of the linear system $\mathbf{M}\,\mathbf{X} = \mathbf{N}$. However, nothing guarantees *a priori* that such a sequence converges in general.

The above iterative method thus requires to solve a linear system at each iteration, and is therefore advantageous only if $\mathbf{M}_1$ is somewhat easier to invert than $\mathbf{M}$. Systems which are easily invertible are typically diagonal or triangular. It is thus natural to consider splittings that yield such matrices for $\mathbf{M}_1$.

Jacobi's splitting

The so-called Jacobi iterative method is that associated with the splitting

$$\mathbf{M}_1 = \mathrm{diag}\,(\mathbf{M}) \qquad \text{and} \qquad \mathbf{M}_2 = \mathrm{diag}\,(\mathbf{M}) - \mathbf{M}\,, \tag{6.2}$$

where $\mathrm{diag}\,(\mathbf{M})$ denotes the matrix that consists in the diagonal of the matrix $\mathbf{M}$. Note that in order to invert the system in Eq. (6.2) with this splitting, $\mathbf{M}$ must have *all* its diagonal elements nonzero.

Remark 6.2. To invert Eq. (6.2) it is only needed to divide each row in the right hand side by the corresponding diagonal element of $\mathbf{M}$.

Gauss-Seidel's splitting

Let us denote by $\mathbf{T}_{\mathrm{up}}$ and $\mathbf{T}_{\mathrm{low}}$ the strictly (i.e. without the diagonal) upper and lower triangular part of $\mathbf{M}$ respectively. Namely, for $i, j \in [\![1, 2D]\!]$,

$$\begin{aligned} T_{\mathrm{up},ij} &= M_{ij} \qquad \text{if } j > i, \text{and 0 otherwise.} \\ T_{\mathrm{low},ij} &= M_{ij} \qquad \text{if } j < i, \text{and 0 otherwise.} \end{aligned}$$

The splitting associated with the Gauss-Seidel method is given by

$$\mathbf{M}_1 = \mathrm{diag}\,(\mathbf{M}) + \mathbf{T}_{\mathrm{low}} \qquad \text{and} \qquad \mathbf{M}_2 = -\,\mathbf{T}_{\mathrm{up}}\ . \tag{6.2}$$

Note that similarly to the Jacobi method, $\mathbf{M}_1 = \mathrm{diag}\,(\mathbf{M}) + \mathbf{T}_{\mathrm{low}}$ must be non-singular, which is equivalent to $\mathbf{M}$ having *all* its diagonal elements nonzero, since $\mathbf{M}_1$ is a triangular matrix.

Remark 6.3. Equation (6.2) is easy to solve because $\mathbf{M}_1$ is triangular, and it is only necessary to apply a forward substitution.

Successive over-relaxation (SOR)

The splitting associated with the SOR method is given by

$$\mathbf{M}_1 = \frac{1}{\omega}\,\mathrm{diag}\,(\mathbf{M}) + \mathbf{T}_{\mathrm{low}} \qquad \text{and} \qquad \mathbf{M}_2 = \frac{1-\omega}{\omega}\,\mathrm{diag}\,(\mathbf{M}) - \mathbf{T}_{\mathrm{up}}\,, \tag{6.2}$$

where $\omega \in \mathbb{R}_+^*$ is a *relaxation parameter*. Here again the matrix $\mathbf{M}_1$ being lower triangular, Eq. (6.2) can be inverted by forward substitution. The SOR method can be viewed as the extrapolated Gauss-Seidel method, as it yields the Gauss-Seidel method for $\omega = 1$. For $\omega < 1$ one talks about under-relaxation and for $\omega > 1$ one talks about over-relaxation. In practice, the relaxation parameter must be optimized to yields the best convergence rate.

6.1.2 Complexity analysis

In Chapter 5, we have seen that solving the linear system associated with a truncated reduced Rayleigh equation containing D modes (i.e. system of size $2D$) by using the LU factorization method yields an operation counts of $N_{\text{op,LU}} = 2D(2D+1)(2D+2)/3 \sim (2D)^3/3$. Let us analyze the number of operations required for each of the aforementioned iterative method. It is clear, almost by definition of the iterative method, that the total number of operations will depend on the convergence rate (under the assumption that convergence is given). The faster the convergence, the fewer the number of iterations needed to reach a desired accuracy to the solution and the fewer the number of operations. The convergence rate depends in general on the physical system (see Section 6.5). Therefore we focus here on the *number of operations per iteration*, $N_{\text{op/iter}}$, and if, after analysis of the convergence rate, it is found that k iterations are needed to reach a desired accuracy, the complexity will be $kN_{\text{op/iter}}$. In order to conclude whether an iterative method is worth using compared to a direct method, one needs to analyze (i) whether the iterative method converges, and (ii) if it converges fast enough so that k remains small. We will see that the number of operations per iteration scales as $8D^2$, therefore the iterative method will be beneficial if $k \ll D/3$.

Jacobi: For the Jacobi method, one iteration requires (i) two[1] matrix-vector multiplications by an *a priori* full matrix $\mathbf{M}_2$ with 0 on the diagonal, (ii) two vector additions that we neglect, and (iii) two matrix-vector multiplication but by the diagonal matrix $\text{diag}(\mathbf{M})^{-1}$. This makes a total of $N_{\text{op/iter}} = 8D^2$ scalar multiplications.

Gauss-Seidel and SOR: For the Gauss-Seidel method, one iteration requires (i) two matrix-vector multiplications by strictly upper triangular matrix $\mathbf{M}_2$, (ii) two vector additions that we neglect, and (iii) two forward substitution with a lower triangular matrix $\mathbf{M}_1$. This makes a total of $N_{\text{op/iter}} = 8D^2$ scalar multiplications, i.e exactly the same as for the Jacobi method. For the SOR method we need to add step multiplying the diagonal elements of $\mathbf{M}_1$ and $\mathbf{M}_2$ respectively by ω^{-1} and $(1-\omega)/\omega$ which add $4D$ scalar multiplications, which is insignificant for the asymptotic scaling.

6.2 Small amplitude perturbation theory

The theory of perturbation refers to a set of mathematical methods which aim at solving approximately and iteratively a problem which can be considered close (in some sense) to a simpler problem which can be solve more easily. The full problem being a perturbed problem of the simpler one, where the perturbation is parametrized by a small parameter. The technique was introduced first by astronomers, in particular Laplace and Lagrange in the 18th century in the context of the study of planetary motions and developed further by Poincaré in the 19th century and before leading to the theory KAM (Kolmogorov, Arnold, Moser) for the stability of systems of Hamiltonian mechanics subject to small perturbations in the half of the 20th century. The theory perturbation has quickly spreads in all fields of physics, in particular in quantum mechanics. Electromagnetic scattering is no exception, and Rayleigh himself solved approximately the problem scattering by a sinusoidal perfectly conducting surface by the use of a perturbative treatment to lowest non-trivial order in 1907. Fano used a perturbative solution up to second order in the surface profile to explain the origin of the Wood anomalies in 1940. We present here

[1] Remember that $\mathbf{X}_k$ contains two column vectors.

the derivation of what is known as small amplitude perturbation theory (SAPT), or simply small perturbation theory, in the literature on wave scattering. In the context of the reduced Rayleigh equation, this pertubative technique was used already by Toigo *et al.* in Ref. [37] where the two first order were computed. It was then developed, among other, by Maradudin and Greffet [56, 67, 68] as a relatively practical numerical iterative technique to solve the reduced Rayleigh equations. The first few terms in a perturbative expansion may often lead important insights on the mechanism involved in some scattering phenomena. The Wood anomaly is an example of a phenomenon which was explained in this way by identifying the excitation of a surface plasmon polariton. In the context of this thesis, we will see in Chapter 10 that the approximation to first order in the surface profile will lead to the physical understanding of some phenomena such as the Yoneda and Brewster scattering phenomena. In the present chapter, we are only interested in using the perturbative method as an iterative method which can numerically be implemented to any desired order.

6.2.1 Derivation

The basic principle of this method is to expand the kernel factor $\mathcal{J}_{l,m}^{b,a}$ in a series of Fourier transforms of the power of the surface profile function ζ and also to expand the unknown reflection or transmission amplitude in a similar series and matching terms of the same order. The expansions read as follows

$$\begin{aligned}
\mathcal{J}_{l,m}^{b,a}(\mathbf{p}\,|\,\mathbf{q}) &= (b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q}))^{-1} \int e^{-i(\mathbf{p}-\mathbf{q})\cdot\mathbf{x}}\, e^{-i(b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q}))\,\zeta(\mathbf{x})}\, \mathrm{d}^2x \\
&= \sum_{n=0}^{\infty} \frac{(-i)^n}{n!} (b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q}))^{n-1} \int e^{-i(\mathbf{p}-\mathbf{q})\cdot\mathbf{x}}\, \zeta^n(\mathbf{x})\, \mathrm{d}^2x \\
&= \sum_{n=0}^{\infty} \frac{(-i)^n}{n!} (b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q}))^{n-1}\, \hat{\zeta}^{(n)}(\mathbf{p}-\mathbf{q}) \qquad (6.3a) \\
\mathbf{R}(\mathbf{q}\,|\,\mathbf{p}_\mathrm{o}) &= \sum_{j=0}^{\infty} \frac{(-i)^j}{j!}\, \mathbf{R}^{(j)}(\mathbf{q}\,|\,\mathbf{p}_\mathrm{o}) \qquad (6.3b) \\
\mathbf{T}(\mathbf{q}\,|\,\mathbf{p}_\mathrm{o}) &= \sum_{j=0}^{\infty} \frac{(-i)^j}{j!}\, \mathbf{T}^{(j)}(\mathbf{q}\,|\,\mathbf{p}_\mathrm{o})\ . \qquad (6.3c)
\end{aligned}$$

Here we have defined the Fourier transform of the n^{th} power of ζ, or n^{th} *Fourier moment of the surface profile*, as

$$\hat{\zeta}^{(n)}(\mathbf{q}) = \int e^{-i\,\mathbf{q}\cdot\mathbf{x}}\, \zeta^n(\mathbf{x})\, \mathrm{d}^2x\ . \qquad (6.4)$$

We are now ready to proceed with the perturbative method.

Reflection

We start by substituting Eqs. (6.3a) and (6.3b) into Eq. (2.51) and get

$$\begin{aligned}
&\sum_{n=0}^{\infty}\sum_{j=0}^{\infty} \frac{(-i)^{n+j}}{n!\,j!} \int (\alpha_2(\mathbf{p}) - \alpha_1(\mathbf{q}))^{n-1}\, \hat{\zeta}^{(n)}(\mathbf{p}-\mathbf{q})\, \mathbf{M}_{2,1}^{+,+}(\mathbf{p}\,|\,\mathbf{q})\, \mathbf{R}^{(j)}(\mathbf{q}\,|\,\mathbf{p}_\mathrm{o})\, \frac{\mathrm{d}^2q}{(2\pi)^2} \\
&\qquad = -\sum_{m=0}^{\infty} \frac{(-i)^m}{m!} (\alpha_2(\mathbf{p}) + \alpha_1(\mathbf{p}_0))^{m-1}\, \hat{\zeta}^{(m)}(\mathbf{p}-\mathbf{p}_0)\, \mathbf{M}_{2,1}^{+,-}(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o})\ . \qquad (6.5)
\end{aligned}$$

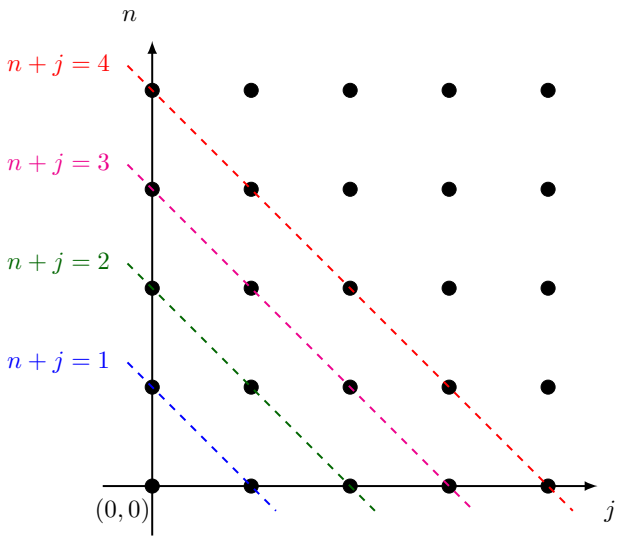


Figure 6.1: Illustration of the change of indices $m = n + j$. Summing over $(j, n) \in \mathbb{N}^2$ is equivalent to summing over sub-sets with elements of constant value $n + j = m$.

As illustrated in Fig. 6.1, summing over all $(n, j) \in \mathbb{N}^2$ is equivalent to summing over subsets $\mathcal{S}_m = \{(n, j) \in \mathbb{N}^2 \mid n + j = m\}$ of pairs of constant sum $m = n + j$, i.e. that we have $\sum_{n,j=0}^{\infty} \equiv \sum_{m=0}^{\infty} \sum_{(n,j)\in\mathcal{S}_m}$, therefore the previous equation can be re-written as

$$\sum_{m=0}^{\infty} \frac{(-i)^m}{m!} \sum_{n=0}^{m} \binom{m}{n} \int (\alpha_2(\mathbf{p}) - \alpha_1(\mathbf{q}))^{n-1}\, \hat{\zeta}^{(n)}(\mathbf{p}-\mathbf{q})\, \mathbf{M}_{2,1}^{+,+}(\mathbf{p}\,|\,\mathbf{q})\, \mathbf{R}^{(m-n)}(\mathbf{q}\,|\,\mathbf{p}_o)\, \frac{\mathrm{d}^2 q}{(2\pi)^2}$$
$$= -\sum_{m=0}^{\infty} \frac{(-i)^m}{m!} (\alpha_2(\mathbf{p}) + \alpha_1(\mathbf{p}_0))^{m-1}\, \hat{\zeta}^{(m)}(\mathbf{p}-\mathbf{p}_0)\, \mathbf{M}_{2,1}^{+,-}(\mathbf{p}\,|\,\mathbf{p}_o)\,. \tag{6.6}$$

Also note that we used that $\frac{1}{n!\,(m-n)!} = \frac{1}{m!}\binom{m}{n}$ by definition of the binomial coefficients. The procedure consists in matching terms of the same order on both sides of the equation. The zero order only consists of one term $n = m = 0$ and gives

$$\int (\alpha_2(\mathbf{p}) - \alpha_1(\mathbf{q}))^{-1}\, \hat{\zeta}^{(0)}(\mathbf{p}-\mathbf{q})\, \mathbf{M}_{2,1}^{+,+}(\mathbf{p}\,|\,\mathbf{q})\, \mathbf{R}^{(0)}(\mathbf{q}\,|\,\mathbf{p}_o)\, \frac{\mathrm{d}^2 q}{(2\pi)^2}$$
$$= -(\alpha_2(\mathbf{p}) + \alpha_1(\mathbf{p}_0))^{-1}\, \hat{\zeta}^{(0)}(\mathbf{p}-\mathbf{p}_0)\, \mathbf{M}_{2,1}^{+,-}(\mathbf{p}\,|\,\mathbf{p}_o)\,. \tag{6.7}$$

By using that $\hat{\zeta}^{(0)}(\mathbf{p}-\mathbf{q}) = (2\pi)^2\, \delta(\mathbf{p}-\mathbf{q})$, we finally obtain the zero order reflection amplitude

$$\mathbf{R}^{(0)}(\mathbf{p}\,|\,\mathbf{p}_o) = \frac{\alpha_1(\mathbf{p}_0) - \alpha_2(\mathbf{p}_0)}{\alpha_2(\mathbf{p}_0) + \alpha_1(\mathbf{p}_0)} \left[\mathbf{M}_{2,1}^{+,+}(\mathbf{p}_o\,|\,\mathbf{p}_o)\right]^{-1} \mathbf{M}_{2,1}^{+,-}(\mathbf{p}_o\,|\,\mathbf{p}_o)\,(2\pi)^2\,\delta(\mathbf{p}-\mathbf{p}_0)$$
$$\stackrel{\text{def}}{=} (2\pi)^2\,\delta(\mathbf{p}-\mathbf{p}_0)\,\boldsymbol{\rho}^{(0)}(\mathbf{p}_0)\,, \tag{6.8}$$

where the last equation is a defining equation for $\boldsymbol{\rho}^{(0)}(\mathbf{p}_0)$.

Remark 6.4. We have just obtained that the zero order of the reflection amplitude corresponds exactly to the reflection amplitude for a flat surface we have derived in the Section 2.3. This was to be expected in the sense that the zero order of the surface profile corresponds to its average plane.

For orders $m \geq 1$, we have

$$\sum_{n=0}^{m} \binom{m}{n} \int (\alpha_2(\mathbf{p}) - \alpha_1(\mathbf{q}))^{n-1} \hat{\zeta}^{(n)}(\mathbf{p}-\mathbf{q})\ \mathbf{M}_{2,1}^{+,+}(\mathbf{p}\,|\,\mathbf{q})\ \mathbf{R}^{(m-n)}(\mathbf{q}\,|\,\mathbf{p}_o)\ \frac{\mathrm{d}^2 q}{(2\pi)^2}$$
$$= -(\alpha_2(\mathbf{p}) + \alpha_1(\mathbf{p}_0))^{m-1} \hat{\zeta}^{(m)}(\mathbf{p}-\mathbf{p}_0)\ \mathbf{M}_{2,1}^{+,-}(\mathbf{p}\,|\,\mathbf{p}_o)\,, \tag{6.9}$$

which by isolating the term of interest, $n = 0$, gives

$$\begin{aligned}
&-(\alpha_2(\mathbf{p}) + \alpha_1(\mathbf{p}_0))^{m-1} \hat{\zeta}^{(m)}(\mathbf{p}-\mathbf{p}_0)\ \mathbf{M}_{2,1}^{+,-}(\mathbf{p}\,|\,\mathbf{p}_o) \\
&-\sum_{n=1}^{m} \binom{m}{n} \int (\alpha_2(\mathbf{p}) - \alpha_1(\mathbf{q}))^{n-1} \hat{\zeta}^{(n)}(\mathbf{p}-\mathbf{q})\ \mathbf{M}_{2,1}^{+,+}(\mathbf{p}\,|\,\mathbf{q})\ \mathbf{R}^{(m-n)}(\mathbf{q}\,|\,\mathbf{p}_o)\ \frac{\mathrm{d}^2 q}{(2\pi)^2} \\
&= \int (\alpha_2(\mathbf{p}) - \alpha_1(\mathbf{q}))^{-1} \hat{\zeta}^{(0)}(\mathbf{p}-\mathbf{q})\ \mathbf{M}_{2,1}^{+,+}(\mathbf{p}\,|\,\mathbf{q})\ \mathbf{R}^{(m)}(\mathbf{q}\,|\,\mathbf{p}_o)\ \frac{\mathrm{d}^2 q}{(2\pi)^2} \\
&= \int (\alpha_2(\mathbf{p}) - \alpha_1(\mathbf{q}))^{-1} (2\pi)^2 \delta(\mathbf{p}-\mathbf{q})\ \mathbf{M}_{2,1}^{+,+}(\mathbf{p}\,|\,\mathbf{q})\ \mathbf{R}^{(m)}(\mathbf{q}\,|\,\mathbf{p}_o)\ \frac{\mathrm{d}^2 q}{(2\pi)^2} \\
&= (\alpha_2(\mathbf{p}) - \alpha_1(\mathbf{p}))^{-1}\ \mathbf{M}_{2,1}^{+,+}(\mathbf{p}\,|\,\mathbf{p})\ \mathbf{R}^{(m)}(\mathbf{p}\,|\,\mathbf{p}_o)\,.
\end{aligned} \tag{6.10}$$

Thus one can express $\mathbf{R}^{(m)}$ as a function of $\mathbf{R}^{(m-1)} \cdots \mathbf{R}^{(0)}$, in other words we have a recursive relation for determining all orders,

$$\mathbf{R}^{(m)}(\mathbf{p}\,|\,\mathbf{p}_o) = (\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p}))\left[\mathbf{M}_{2,1}^{+,+}(\mathbf{p}\,|\,\mathbf{p})\right]^{-1} \Bigg[(\alpha_2(\mathbf{p}) + \alpha_1(\mathbf{p}_0))^{m-1} \hat{\zeta}^{(m)}(\mathbf{p}-\mathbf{p}_0)\ \mathbf{M}_{2,1}^{+,-}(\mathbf{p}\,|\,\mathbf{p}_o)$$
$$+\sum_{n=1}^{m} \binom{m}{n} \int (\alpha_2(\mathbf{p}) - \alpha_1(\mathbf{q}))^{n-1} \hat{\zeta}^{(n)}(\mathbf{p}-\mathbf{q})\ \mathbf{M}_{2,1}^{+,+}(\mathbf{p}\,|\,\mathbf{q})\ \mathbf{R}^{(m-n)}(\mathbf{q}\,|\,\mathbf{p}_o)\ \frac{\mathrm{d}^2 q}{(2\pi)^2}\Bigg]\,. \tag{6.11}$$

Note that in general, high orders require the evaluation of as many integrals as the order to be determined and can become costly. For the first order, only one such integral is to be evaluated and is straightforward to perform thanks to the fact that $\mathbf{R}^{(0)}(\mathbf{q}\,|\,\mathbf{p}_0) \propto \delta(\mathbf{q}-\mathbf{p}_0)$. Indeed,

applying the above equation for $m = 1$ gives

$$\begin{aligned}
\mathbf{R}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}}) &= (\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p}))\left[\mathbf{M}_{2,1}^{+,+}(\mathbf{p}\,|\,\mathbf{p})\right]^{-1}\left[\hat{\zeta}^{(1)}(\mathbf{p}-\mathbf{p}_0)\;\mathbf{M}_{2,1}^{+,-}(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}})\right.\\
&\quad\left. + \int \hat{\zeta}^{(1)}(\mathbf{p}-\mathbf{q})\;\mathbf{M}_{2,1}^{+,+}(\mathbf{p}\,|\,\mathbf{q})\;\mathbf{R}^{(0)}(\mathbf{q}\,|\,\mathbf{p}_{\mathrm{o}})\;\frac{\mathrm{d}^2 q}{(2\pi)^2}\right]\\
&= (\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p}))\left[\mathbf{M}_{2,1}^{+,+}(\mathbf{p}\,|\,\mathbf{p})\right]^{-1}\left[\hat{\zeta}^{(1)}(\mathbf{p}-\mathbf{p}_0)\;\mathbf{M}_{2,1}^{+,-}(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}})\right.\\
&\quad\left. + \frac{\alpha_1(\mathbf{p}_0) - \alpha_2(\mathbf{p}_0)}{\alpha_2(\mathbf{p}_0) + \alpha_1(\mathbf{p}_0)}\,\hat{\zeta}^{(1)}(\mathbf{p}-\mathbf{p}_0)\;\mathbf{M}_{2,1}^{+,+}(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}})\;\left[\mathbf{M}_{2,1}^{+,+}(\mathbf{p}_{\mathrm{o}}\,|\,\mathbf{p}_{\mathrm{o}})\right]^{-1}\;\mathbf{M}_{2,1}^{+,-}(\mathbf{p}_{\mathrm{o}}\,|\,\mathbf{p}_{\mathrm{o}})\right]\\
&= (\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p}))\,\hat{\zeta}^{(1)}(\mathbf{p}-\mathbf{p}_0)\;\left[\mathbf{M}_{2,1}^{+,+}(\mathbf{p}\,|\,\mathbf{p})\right]^{-1}\left[\mathbf{M}_{2,1}^{+,-}(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}})\right.\\
&\quad\left. + \frac{\alpha_1(\mathbf{p}_0) - \alpha_2(\mathbf{p}_0)}{\alpha_1(\mathbf{p}_0) + \alpha_2(\mathbf{p}_0)}\;\mathbf{M}_{2,1}^{+,+}(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}})\;\left[\mathbf{M}_{2,1}^{+,+}(\mathbf{p}_{\mathrm{o}}\,|\,\mathbf{p}_{\mathrm{o}})\right]^{-1}\;\mathbf{M}_{2,1}^{+,-}(\mathbf{p}_{\mathrm{o}}\,|\,\mathbf{p}_{\mathrm{o}})\right]\\
&\overset{\mathrm{def}}{=} \hat{\zeta}^{(1)}(\mathbf{p}-\mathbf{p}_0)\,\boldsymbol{\rho}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}}) \overset{\mathrm{def}}{=} (\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p}))\,\hat{\zeta}^{(1)}(\mathbf{p}-\mathbf{p}_0)\,\hat{\boldsymbol{\rho}}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}})\;. \qquad (6.12)
\end{aligned}$$

Here we have given two alternative way to write $\mathbf{R}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}})$ introducing either the amplitude $\boldsymbol{\rho}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}})$ or $\hat{\boldsymbol{\rho}}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}})$ that only differs by a factor $(\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p}))$. The reason for introducing these definitions will become clear in Chapter 10 when we will give a clear physical interpretation of the first order reflection amplitude. For now, what matters is to convince oneself that it is relatively simple to compute $\mathbf{R}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}})$.

Transmission

Repeating the reasoning for the transmission amplitudes, we start by substituting Eqs. (6.3a) and (6.3c) into Eq. (2.52) and get

$$\begin{aligned}
&\sum_{n=0}^{\infty}\sum_{j=0}^{\infty}\frac{(-i)^{n+j}}{n!\,j!}\int(-\alpha_1(\mathbf{p}) + \alpha_2(\mathbf{q}))^{n-1}\,\hat{\zeta}^{(n)}(\mathbf{p}-\mathbf{q})\;\mathbf{M}_{1,2}^{-,-}(\mathbf{p}\,|\,\mathbf{q})\;\mathbf{T}^{(j)}(\mathbf{q}\,|\,\mathbf{p}_{\mathrm{o}})\;\frac{\mathrm{d}^2 q}{(2\pi)^2}\\
&\qquad = \frac{2\sqrt{\epsilon_1\epsilon_2}\,\alpha_1(\mathbf{p}_0)}{\epsilon_2 - \epsilon_1}\,(2\pi)^2\,\delta(\mathbf{p}-\mathbf{p}_0)\,\mathbf{I}_2\;. \qquad (6.13)
\end{aligned}$$

By using the same re-summation argument as in the case of the reflection amplitude, the previous equation thus becomes

$$\begin{aligned}
&\sum_{m=0}^{\infty}\sum_{n=0}^{m}\frac{(-i)^m}{m!}\binom{m}{n}\int(-\alpha_1(\mathbf{p}) + \alpha_2(\mathbf{q}))^{n-1}\,\hat{\zeta}^{(n)}(\mathbf{p}-\mathbf{q})\;\mathbf{M}_{1,2}^{-,-}(\mathbf{p}\,|\,\mathbf{q})\;\mathbf{T}^{(m-n)}(\mathbf{q}\,|\,\mathbf{p}_{\mathrm{o}})\;\frac{\mathrm{d}^2 q}{(2\pi)^2}\\
&\qquad = \frac{2\sqrt{\epsilon_1\epsilon_2}\,\alpha_1(\mathbf{p}_0)}{\epsilon_2 - \epsilon_1}\,(2\pi)^2\,\delta(\mathbf{p}-\mathbf{p}_0)\,\mathbf{I}_2\;. \qquad (6.14)
\end{aligned}$$

Next we match the zero order to the right hand side and the other orders to zero. The zero order only consists of one term, $n = m = 0$, and gives

$$\begin{aligned}
\mathbf{T}^{(0)}(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}}) &= \frac{2\sqrt{\epsilon_1\epsilon_2}\alpha_1(\mathbf{p}_0)}{\epsilon_2 - \epsilon_1}\,(2\pi)^2\,\delta(\mathbf{p}-\mathbf{p}_0)\,(\alpha_2(\mathbf{p}_0) - \alpha_1(\mathbf{p}_0))\;\left[\mathbf{M}_{1,2}^{-,-}(\mathbf{p}_{\mathrm{o}}\,|\,\mathbf{p}_{\mathrm{o}})\right]^{-1}\\
&\overset{\mathrm{def}}{=} (2\pi)^2\,\delta(\mathbf{p}-\mathbf{p}_0)\,\boldsymbol{\tau}^{(0)}(\mathbf{p}_0) \qquad (6.15)
\end{aligned}$$

Here we have used that $\hat{\zeta}^{(0)}(\mathbf{p}-\mathbf{q}) = (2\pi)^2\,\delta(\mathbf{p}-\mathbf{q})$ and the last equation defines $\boldsymbol{\tau}^{(0)}(\mathbf{p}_0)$.

Remark 6.5. As observed for the reflection amplitudes, we have just obtained that the zero order of the transmission amplitudes corresponds exactly to the transmission amplitudes for a planar surface derived in the Section 2.3.

For orders $m \geq 1$, we have

$$\sum_{n=0}^{m} \binom{m}{n} \int (\alpha_2(\mathbf{q}) - \alpha_1(\mathbf{p}))^{n-1}\, \hat{\zeta}^{(n)}(\mathbf{p}-\mathbf{q})\; \mathbf{M}_{1,2}^{-;-}(\mathbf{p}\,|\,\mathbf{q})\; \mathbf{T}^{(m-n)}(\mathbf{q}\,|\,\mathbf{p}_o)\; \frac{\mathrm{d}^2 q}{(2\pi)^2} = \mathbf{0}\,, \tag{6.16}$$

which by isolating the term of interest, $n = 0$, gives

$$\begin{aligned}
&-\sum_{n=1}^{m} \binom{m}{n} \int (\alpha_2(\mathbf{q}) - \alpha_1(\mathbf{p}))^{n-1}\, \hat{\zeta}^{(n)}(\mathbf{p}-\mathbf{q})\; \mathbf{M}_{1,2}^{-;-}(\mathbf{p}\,|\,\mathbf{q})\; \mathbf{T}^{(m-n)}(\mathbf{q}\,|\,\mathbf{p}_o)\; \frac{\mathrm{d}^2 q}{(2\pi)^2} \\
&\qquad = \int (\alpha_2(\mathbf{q}) - \alpha_1(\mathbf{p}))^{-1}\, \hat{\zeta}^{(0)}(\mathbf{p}-\mathbf{q})\; \mathbf{M}_{1,2}^{-;-}(\mathbf{p}\,|\,\mathbf{q})\; \mathbf{T}^{(m)}(\mathbf{q}\,|\,\mathbf{p}_o)\; \frac{\mathrm{d}^2 q}{(2\pi)^2} \\
&\qquad = \int (\alpha_2(\mathbf{q}) - \alpha_1(\mathbf{p}))^{-1}\, (2\pi)^2 \delta(\mathbf{p}-\mathbf{q})\; \mathbf{M}_{1,2}^{-;-}(\mathbf{p}\,|\,\mathbf{q})\; \mathbf{T}^{(m)}(\mathbf{q}\,|\,\mathbf{p}_o)\; \frac{\mathrm{d}^2 q}{(2\pi)^2} \\
&\qquad = (\alpha_2(\mathbf{p}) - \alpha_1(\mathbf{p}))^{-1}\; \mathbf{M}_{1,2}^{-;-}(\mathbf{p}\,|\,\mathbf{p})\; \mathbf{T}^{(m)}(\mathbf{p}\,|\,\mathbf{p}_o)\,.
\end{aligned} \tag{6.17}$$

Thus one can express $\mathbf{T}^{(m)}$ as a function of $\mathbf{T}^{(m-1)} \cdots \mathbf{T}^{(0)}$. We thus have a recursive relation for determining all orders,

$$\begin{aligned}
&\mathbf{T}^{(m)}(\mathbf{p}\,|\,\mathbf{p}_o) = (\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p}))\; \left[\mathbf{M}_{1,2}^{-;-}(\mathbf{p}\,|\,\mathbf{p})\right]^{-1} \\
&\times \sum_{n=1}^{m} \binom{m}{n} \int (\alpha_2(\mathbf{q}) - \alpha_1(\mathbf{p}))^{n-1}\, \hat{\zeta}^{(n)}(\mathbf{p}-\mathbf{q})\; \mathbf{M}_{1,2}^{-;-}(\mathbf{p}\,|\,\mathbf{q})\; \mathbf{T}^{(m-n)}(\mathbf{q}\,|\,\mathbf{p}_o)\; \frac{\mathrm{d}^2 q}{(2\pi)^2}\,.
\end{aligned} \tag{6.18}$$

Note that in general, high orders require the evaluation of as many integrals as the order to be determined and can become costly. For the first order we take advantage of the fact that $\mathbf{T}^{(0)}(\mathbf{q}\,|\,\mathbf{p}_o) \propto \delta(\mathbf{q}-\mathbf{p}_0)$. Applying the above equation for $m = 1$ gives

$$\begin{aligned}
\mathbf{T}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_o) =& \frac{2\sqrt{\epsilon_1\epsilon_2}\,\alpha_1(\mathbf{p}_0)}{\epsilon_2 - \epsilon_1}\, (\alpha_2(\mathbf{p}_0) - \alpha_1(\mathbf{p}_0))(\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p})) \\
&\times \hat{\zeta}^{(1)}(\mathbf{p}-\mathbf{p}_0)\; \left[\mathbf{M}_{1,2}^{-;-}(\mathbf{p}\,|\,\mathbf{p})\right]^{-1}\; \mathbf{M}_{1,2}^{-;-}(\mathbf{p}\,|\,\mathbf{p}_o)\; \left[\mathbf{M}_{1,2}^{-;-}(\mathbf{p}_o\,|\,\mathbf{p}_o)\right]^{-1} \\
\overset{\text{def}}{=}& \hat{\zeta}^{(1)}(\mathbf{p}-\mathbf{p}_0)\, \boldsymbol{\tau}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_o) \overset{\text{def}}{=} (\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p}))\, \hat{\zeta}^{(1)}(\mathbf{p}-\mathbf{p}_0)\, \hat{\boldsymbol{\tau}}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_o)\,.
\end{aligned} \tag{6.19}$$

Here the last equation defines the amplitudes $\boldsymbol{\rho}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_o)$ and $\hat{\boldsymbol{\rho}}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_o)$, which will be of particular interest in Chapter 10 when discussing the physical interpretation of the first order transmission amplitude.

Periodic surfaces

We apply the method presented above for a periodic surface. In that case, the expansions read

$$\tilde{\mathcal{J}}^{b,a}_{l,m}(\mathbf{p}\,|\,\mathbf{q}) = \sum_{n=0}^{\infty} \frac{(-i)^n}{n!}\,(b\alpha_l\,(\mathbf{p}) - a\alpha_m\,(\mathbf{q}))^{n-1}\,\hat{\zeta}^{(n)}_{a_c}(\mathbf{p}-\mathbf{q}) \tag{6.20a}$$

$$\mathbf{R}^{(\mathbf{m})}(\mathbf{p}_0) = \sum_{j=0}^{\infty} \frac{(-i)^j}{j!}\;\mathbf{R}^{(j)}_{\mathbf{m}}(\mathbf{p}_0) \tag{6.20b}$$

$$\mathbf{T}^{(\mathbf{m})}(\mathbf{p}_0) = \sum_{j=0}^{\infty} \frac{(-i)^j}{j!}\;\mathbf{T}^{(j)}_{\mathbf{m}}(\mathbf{p}_0)\,. \tag{6.20c}$$

Here we have defined the n^{th} *Fourier moment* as

$$\hat{\zeta}^{(n)}_{a_c}(\mathbf{q}) = \frac{1}{a_c}\int_{a_c} e^{-i\,\mathbf{q}\cdot\mathbf{x}}\,\zeta^n\,(\mathbf{x})\,\mathrm{d}^2x\;. \tag{6.21}$$

Note that we have expanded the component of each diffractive order $\mathbf{m}$, which we already know form a countable set, and $\mathbf{R}^{(j)}_{\mathbf{m}}(\mathbf{p}_0)$ (resp. $\mathbf{T}^{(j)}_{\mathbf{m}}(\mathbf{p}_0)$) must be understood as the j^{th} order in the power of the surface profile for the $\mathbf{m}^{\text{th}}$ diffractive order of the reflection (resp. transmission) amplitude. Then following the same derivation as in the general case, we obtain that the terms of the expansion of the reflection amplitudes are given by

$$\mathbf{R}^{(0)}_{\boldsymbol{\ell}}(\mathbf{p}_0) = \frac{\alpha_1(\mathbf{p}_0)-\alpha_2(\mathbf{p}_0)}{\alpha_2(\mathbf{p}_0)+\alpha_1(\mathbf{p}_0)}\;\left[\mathbf{M}^{+,+}_{2,1}(\mathbf{p}_0\,|\,\mathbf{p}_0)\right]^{-1}\;\mathbf{M}^{+,-}_{2,1}(\mathbf{p}_0\,|\,\mathbf{p}_0)\,\delta_{\boldsymbol{\ell},\mathbf{0}}\,, \tag{6.22a}$$

$$\begin{aligned}\mathbf{R}^{(k)}_{\boldsymbol{\ell}}(\mathbf{p}_0) = {}& (\alpha_1(\mathbf{p}_{\boldsymbol{\ell}})-\alpha_2(\mathbf{p}_{\boldsymbol{\ell}}))\,\left[\mathbf{M}^{+,+}_{2,1}(\mathbf{p}_{\boldsymbol{\ell}}\,|\,\mathbf{p}_{\boldsymbol{\ell}})\right]^{-1}\;\Bigg[(\alpha_2(\mathbf{p}_{\boldsymbol{\ell}})+\alpha_1(\mathbf{p}_0))^{k-1}\;\hat{\zeta}^{(k)}_{a_c}(\mathbf{G}^{(\boldsymbol{\ell})})\;\mathbf{M}^{+,-}_{2,1}(\mathbf{p}_{\boldsymbol{\ell}}\,|\,\mathbf{p}_0)\\ &+\sum_{n=1}^{k}\binom{k}{n}\sum_{\mathbf{m}\in\mathbb{Z}^2}(\alpha_2(\mathbf{p}_{\boldsymbol{\ell}})-\alpha_1(\mathbf{p}_{\mathbf{m}}))^{n-1}\;\hat{\zeta}^{(n)}_{a_c}(\mathbf{G}^{(\boldsymbol{\ell}-\mathbf{m})})\;\mathbf{M}^{+,+}_{2,1}(\mathbf{p}_{\boldsymbol{\ell}}\,|\,\mathbf{p}_{\mathbf{m}})\;\mathbf{R}^{(k-n)}_{\mathbf{m}}(\mathbf{p}_0)\Bigg]\,,\end{aligned} \tag{6.22b}$$

for $k>0$ and that of the transmission amplitudes are given by

$$\mathbf{T}^{(0)}_{\boldsymbol{\ell}}(\mathbf{p}_0) = \frac{2\sqrt{\epsilon_1\epsilon_2}\alpha_1(\mathbf{p}_0)}{\epsilon_2-\epsilon_1}\;(\alpha_2(\mathbf{p}_0)-\alpha_1(\mathbf{p}_0))\;\left[\mathbf{M}^{-,-}_{1,2}(\mathbf{p}_0\,|\,\mathbf{p}_0)\right]^{-1}\;\delta_{\boldsymbol{\ell},\mathbf{0}}\,, \tag{6.23a}$$

$$\begin{aligned}\mathbf{T}^{(k)}_{\boldsymbol{\ell}}(\mathbf{p}_0) = {}& (\alpha_1(\mathbf{p}_{\boldsymbol{\ell}})-\alpha_2(\mathbf{p}_{\boldsymbol{\ell}}))\;\left[\mathbf{M}^{-,-}_{1,2}(\mathbf{p}_{\boldsymbol{\ell}}\,|\,\mathbf{p}_{\boldsymbol{\ell}})\right]^{-1}\\ &\times\sum_{n=1}^{k}\binom{k}{n}\sum_{\mathbf{m}\in\mathbb{Z}^2}(-\alpha_1(\mathbf{p}_{\boldsymbol{\ell}})+\alpha_2(\mathbf{p}_{\mathbf{m}}))^{n-1}\;\hat{\zeta}^{(n)}_{a_c}(\mathbf{G}^{(\boldsymbol{\ell}-\mathbf{m})})\;\mathbf{M}^{-,-}_{1,2}(\mathbf{p}_{\boldsymbol{\ell}}\,|\,\mathbf{p}_{\mathbf{m}})\;\mathbf{T}^{(k-n)}_{\mathbf{m}}(\mathbf{p}_0)\,.\end{aligned} \tag{6.23b}$$

We recall in passing the notation $\mathbf{p}_{\boldsymbol{\ell}} \overset{\text{def}}{=} \mathbf{p}_0 + \mathbf{G}^{(\boldsymbol{\ell})}$ with $\mathbf{G}^{(\boldsymbol{\ell})}$ a reciprocal lattice vector. Note that here we have chosen to denote the perturbation order k instead of m, used previously in the general case, in order to avoid confusion with the mode indices $\mathbf{m}$ for the diffractive orders.

Remark 6.6. These expressions can be found either by following the same derivation as that described in the general case and starting from the reduced Rayleigh equations for a periodic surface derived Chapter 4, or equivalently applying the reasoning from Chapter 4 to the SAPT results obtained in the general case.

6.2.2 Implementation

We can see from Eqs. (6.22) and (6.23) that the recursive relation giving $\mathbf{X}_{\boldsymbol{\ell}}^{(k)}$, where $\mathbf{X} = \mathbf{R}$ or $\mathbf{T}$ can be written in the form, for $k > 1$

$$\mathbf{X}_{\boldsymbol{\ell}}^{(k)}(\mathbf{p}_0) = \mathbf{Y}_{\boldsymbol{\ell}}^{(k)}(\mathbf{p}_0) + \sum_{n=1}^{k} \binom{k}{n} \sum_{\mathbf{m}\in\mathbb{Z}^2} \mathbf{M}_{\boldsymbol{\ell},\mathbf{m}}^{(n)}(\mathbf{p}_0)\ \mathbf{X}_{\mathbf{m}}^{(k-n)}(\mathbf{p}_0)\,. \tag{6.24}$$

Indeed, we have in the case of the reflection amplitudes

$$\begin{aligned}
\mathbf{Y}_{\boldsymbol{\ell}}^{(k)}(\mathbf{p}_0) &\overset{\text{def}}{=} \hat{\zeta}_{a_c}^{(k)}(\mathbf{G}^{(\boldsymbol{\ell})})\ (\alpha_1(\mathbf{p}_{\boldsymbol{\ell}}) - \alpha_2(\mathbf{p}_{\boldsymbol{\ell}}))\left[\mathbf{M}_{2,1}^{+,+}(\mathbf{p}_{\boldsymbol{\ell}}\,|\,\mathbf{p}_{\boldsymbol{\ell}})\right]^{-1} \\
&\quad\times (\alpha_2(\mathbf{p}_{\boldsymbol{\ell}}) + \alpha_1(\mathbf{p}_0))^{k-1}\ \mathbf{M}_{2,1}^{+,-}(\mathbf{p}_{\boldsymbol{\ell}}\,|\,\mathbf{p}_0) &\text{(6.25a)}\\
\mathbf{M}_{\boldsymbol{\ell},\mathbf{m}}^{(n)}(\mathbf{p}_0) &\overset{\text{def}}{=} \hat{\zeta}_{a_c}^{(n)}(\mathbf{G}^{(\boldsymbol{\ell}-\mathbf{m})})\ (\alpha_1(\mathbf{p}_{\boldsymbol{\ell}}) - \alpha_2(\mathbf{p}_{\boldsymbol{\ell}}))\left[\mathbf{M}_{2,1}^{+,+}(\mathbf{p}_{\boldsymbol{\ell}}\,|\,\mathbf{p}_{\boldsymbol{\ell}})\right]^{-1} \\
&\quad\times (\alpha_2(\mathbf{p}_{\boldsymbol{\ell}}) - \alpha_1(\mathbf{p}_{\mathbf{m}}))^{n-1}\ \mathbf{M}_{2,1}^{+,+}(\mathbf{p}_{\boldsymbol{\ell}}\,|\,\mathbf{p}_{\mathbf{m}})\,, &\text{(6.25b)}
\end{aligned}$$

and for the transmission amplitudes

$$\begin{aligned}
\mathbf{Y}_{\boldsymbol{\ell}}^{(k)}(\mathbf{p}_0) &\overset{\text{def}}{=} \mathbf{0} &\text{(6.26a)}\\
\mathbf{M}_{\boldsymbol{\ell},\mathbf{m}}^{(n)}(\mathbf{p}_0) &\overset{\text{def}}{=} \hat{\zeta}_{a_c}^{(n)}(\mathbf{G}^{(\boldsymbol{\ell}-\mathbf{m})})\ (\alpha_1(\mathbf{p}_{\boldsymbol{\ell}}) - \alpha_2(\mathbf{p}_{\boldsymbol{\ell}}))\left[\mathbf{M}_{1,2}^{-,-}(\mathbf{p}_{\boldsymbol{\ell}}\,|\,\mathbf{p}_{\boldsymbol{\ell}})\right]^{-1} \\
&\quad\times (-\alpha_1(\mathbf{p}_{\boldsymbol{\ell}}) + \alpha_2(\mathbf{p}_{\mathbf{m}}))^{n-1}\ \mathbf{M}_{1,2}^{-,-}(\mathbf{p}_{\boldsymbol{\ell}}\,|\,\mathbf{p}_{\mathbf{m}})\,, &\text{(6.26b)}
\end{aligned}$$

This means that for the truncated system, when applying the re-indexing described in Section 5.2.1, Eq. (6.24) yields

$$\mathbf{X}_{\beta}^{(k)}(\mathbf{p}_0) = \mathbf{Y}_{\beta}^{(k)}(\mathbf{p}_0) + \sum_{n=1}^{k} \binom{k}{n}\ \mathbf{M}^{(n)}(\mathbf{p}_0)\ \mathbf{X}_{\beta}^{(k-n)}(\mathbf{p}_0)\,, \tag{6.27}$$

where

$$\mathbf{X}_{\beta}^{(k)}(\mathbf{p}_0) = \left(X_{\mathbf{m}_1,p\beta}^{(k)}(\mathbf{p}_0),\cdots,X_{\mathbf{m}_D,p\beta}^{(k)}(\mathbf{p}_0),X_{\mathbf{m}_1,s\beta}^{(k)}(\mathbf{p}_0),\cdots,X_{\mathbf{m}_D,s\beta}^{(k)}(\mathbf{p}_0)\right)^{\mathrm{T}}\,, \tag{6.28}$$

and for $(i,j) \in [\![1,D]\!]^2$,

$$\begin{aligned}
M_{ij}^{(n)}(\mathbf{p}_0) &= M_{\boldsymbol{\ell}_i,\mathbf{m}_j,pp}^{(n)}(\mathbf{p}_0) &\text{(6.29a)}\\
M_{D+i,j}^{(n)}(\mathbf{p}_0) &= M_{\boldsymbol{\ell}_i,\mathbf{m}_j,sp}^{(n)}(\mathbf{p}_0) &\text{(6.29b)}\\
M_{i,D+j}^{(n)}(\mathbf{p}_0) &= M_{\boldsymbol{\ell}_i,\mathbf{m}_j,ps}^{(n)}(\mathbf{p}_0) &\text{(6.29c)}\\
M_{D+i,D+j}^{(n)}(\mathbf{p}_0) &= M_{\boldsymbol{\ell}_i,\mathbf{m}_j,ss}^{(n)}(\mathbf{p}_0)\,, &\text{(6.29d)}
\end{aligned}$$

and

$$\mathbf{Y}_{\beta}^{(k)}(\mathbf{p}_0) = \left(Y_{\mathbf{m}_1,p\beta}^{(k)}(\mathbf{p}_0),\cdots,Y_{\mathbf{m}_D,p\beta}^{(k)}(\mathbf{p}_0),Y_{\mathbf{m}_1,s\beta}^{(k)}(\mathbf{p}_0),\cdots,Y_{\mathbf{m}_D,s\beta}^{(k)}(\mathbf{p}_0)\right)^{\mathrm{T}}\,. \tag{6.30}$$

In order to compute the terms in the expansion of the reflection, or transmission amplitudes, up to order k, it is thus necessary to set up k matrices $(\mathbf{M}^{(n)})_{1\le n\le k}$ and $2k$ vectors (k for each incident polarization) $(\mathbf{Y}_{\beta}^{(n)})_{1\le n\le k}$. These matrices and vectors are then used and re-used to successively compute the $(\mathbf{X}_{\beta}^{(n)})_{1\le n\le k}$.

6.2.3 Complexity analysis

Setting up matrices

Let us count the number of operations required to set up the matrices $(\mathbf{M}^{(n)})_{1\leq n\leq k}$ and $(\mathbf{Y}^{(n)})_{1\leq n\leq k}$. To fix the ideas we will consider the case of the reflection amplitudes and we assume that the Fourier moments have already computed and are available for use. Consider the following algorithm to set up all the $(\mathbf{M}^{(n)})_{1\leq n\leq k}$ and $(\mathbf{Y}^{(n)})_{1\leq n\leq k}$ simultaneously.

- Loop over the set of modes $i \in [\![1, D]\!]$

  1 Compute $\alpha_1(\mathbf{p}_{\boldsymbol{\ell}_i})$ and $\alpha_2(\mathbf{p}_{\boldsymbol{\ell}_i})$ (≈ 10 op.).

  2 Set $A \leftarrow \alpha_1(\mathbf{p}_{\boldsymbol{\ell}_i}) + \alpha_2(\mathbf{p}_{\boldsymbol{\ell}_i})$ (1 op.) and $B \leftarrow 1$.

  3 Compute $\mathbf{M}^{+,-}_{2,1}(\mathbf{p}_{\boldsymbol{\ell}_i} \,|\, \mathbf{p}_0)$ (≈ 20 op.).

  4 Compute $\mathbf{M}^{+,+}_{2,1}(\mathbf{p}_{\boldsymbol{\ell}_i} \,|\, \mathbf{p}_{\boldsymbol{\ell}_i})$ (note diagonal matrix ≈ 5 op.).

  5 Compute $\tilde{\mathbf{Y}} \leftarrow (\alpha_1(\mathbf{p}_{\boldsymbol{\ell}_i}) - \alpha_2(\mathbf{p}_{\boldsymbol{\ell}_i})) \left[\mathbf{M}^{+,+}_{2,1}(\mathbf{p}_{\boldsymbol{\ell}_i} \,|\, \mathbf{p}_{\boldsymbol{\ell}_i})\right]^{-1} \mathbf{M}^{+,-}_{2,1}(\mathbf{p}_{\boldsymbol{\ell}_i} \,|\, \mathbf{p}_0)$ (7 op.).

  * *Setting up the $\mathbf{Y}^{(n)}_\beta(\mathbf{p}_0)$-vectors.*

  6 Loop over perturbative order $n \in [\![1, k]\!]$

    6.1 Get $\hat{\zeta}^{(n)}_{a_c}(\mathbf{G}^{(\boldsymbol{\ell}_i)})$ (0 op.).

    6.2 Set $C \leftarrow B \times \hat{\zeta}^{(n)}_{a_c}(\mathbf{G}^{(\boldsymbol{\ell}_i)})$ (1 op.).

    6.3 Compute $C \times \tilde{\mathbf{Y}}$ (4 op.) and set $Y^{(n)}_{\boldsymbol{\ell}_i,\alpha\beta}(\mathbf{p}_0) \leftarrow \tilde{Y}_{\alpha\beta}$ in the $\mathbf{Y}^{(n)}_\beta(\mathbf{p}_0)$-vectors ($\beta \in \{p, s\}$).

    6.4 Increment power $B \leftarrow B \times A$ (1 op.).

  * *Setting up the $\mathbf{M}^{(n)}$-matrices.*

  7 Loop over the set modes $j \in [\![1, D]\!]$

    7.1 Compute $\alpha_1(\mathbf{p}_{\mathbf{m}_j})$ (≈ 5 op.).

    7.2 Compute $\mathbf{M}^{+,+}_{2,1}(\mathbf{p}_{\boldsymbol{\ell}_i} \,|\, \mathbf{p}_{\boldsymbol{\ell}_j})$ (≈ 20 op.).

    7.3 Compute $\tilde{\mathbf{M}} \leftarrow (\alpha_1(\mathbf{p}_{\boldsymbol{\ell}_i}) - \alpha_2(\mathbf{p}_{\boldsymbol{\ell}_i})) \left[\mathbf{M}^{+,+}_{2,1}(\mathbf{p}_{\boldsymbol{\ell}_i} \,|\, \mathbf{p}_{\boldsymbol{\ell}_i})\right]^{-1} \mathbf{M}^{+,+}_{2,1}(\mathbf{p}_{\boldsymbol{\ell}_i} \,|\, \mathbf{p}_{\mathbf{m}_j})$ (7 op.).

    7.4 Set $A' \leftarrow \alpha_2(\mathbf{p}_{\boldsymbol{\ell}_i}) - \alpha_1(\mathbf{p}_{\boldsymbol{\ell}_j})$ and $B' \leftarrow 1$ (1 op.).

    7.5 Loop over perturbative order $n \in [\![1, k]\!]$

      7.5.1 Get $\hat{\zeta}^{(n)}_{a_c}(\mathbf{G}^{(\boldsymbol{\ell}_i)} - \mathbf{G}^{(\mathbf{m}_j)})$ (0 op.).

      7.5.2 Set $C' \leftarrow B' \times \hat{\zeta}^{(n)}_{a_c}(\mathbf{G}^{(\boldsymbol{\ell}_i)} - \mathbf{G}^{(\mathbf{m}_j)})$ (1 op.).

      7.5.3 Compute $C' \times \tilde{\mathbf{M}}$ (4 op.) and set $M^{(n)}_{\boldsymbol{\ell}_i,\mathbf{m}_j,\alpha,\beta}(\mathbf{p}_0) \leftarrow \tilde{M}_{\alpha\beta}$ according to Eq. (6.29).

      7.5.4 Increment power $B' \leftarrow B' \times A'$ (1 op.).

The number of operations indicated are rough estimates which may vary if one consider that some constants such as $\sqrt{\epsilon_j}$, ω^2/c^2 and their products are tabulated before hand or not, and we have treated equally multiplications, additions and square roots, which in practice should not really be. Nevertheless, we just aim at a rough estimation, since it is the scaling with the number of modes D and maximum perturbative order k which is of particular interest. Following the

algorithm, the total number of operations for setting up the $\mathbf{Y}_\beta^{(n)}$-vectors and $\mathbf{M}^{(n)}$-matrices is roughly

$$N_{op,setup}^{\text{SAPT}} \approx 43\, D + 6kD + (33 + 6k)D^2\ (6k + \kappa)D^2\,, \tag{6.31}$$

where the assymptotic scaling is taken for large D and κ is a constant which may vary depending on the details of the implementation, in our case we have estimated $\kappa = 33$. What is important is that the cost associated with setting up the iteration matrices scales quadratically with the number modes and linearly with the number of perturbative order kept in the expansion k.

Iterations

Now that the matrices are set up, it is time to successively determine each order of the scattering amplitudes. Let us count the number of operations required for each iteration for a fixed incident polarization β. Neglecting the number of operations that is required to set up $\mathbf{X}_\beta^{(0)2}$ the first iteration requires one matrix vector multiplication and a vector addition, giving $(2D)^2$ scalar multiplications and $(2D)(2D-1)+n = (2D)^2$ scalar additions, the second iteration requires two matrix vector multiplications and two vector additions (let's neglect multiplying by the binomial coefficient), giving $2(2D)^2$ scalar multiplications and $2(2D)(2D-1)+2(2D) = 2(2D)^2$ scalar additions, ... , and the kth iteration requires k matrix vector multiplications and k vector additions, giving $k(2D)^2$ scalar multiplications and $k(2D)^2$ scalar additions. In total, to compute the k first terms in the scattering amplitude expansions requires $(2D)^2+2(2D)^2+\cdots+k(2D)^2 = k(k+1)(2D)^2/2$ multiplications and as many additions. In addition, once we have the $k+1$ first terms of the expansion, we need to sum them up. Hence we roughly need

$$N_{\text{op,iter}}^{\text{SAPT}} \approx k^2(2D)^2 \tag{6.32}$$

operations to compute the scattering amplitudes approximated to kth order in power of the surface profile function, given the iteration matrices and vectors.

The total operation cost associated with this method is then

$$N_{\text{op,tot}}^{\text{SAPT}} = N_{\text{op,setup}}^{\text{SAPT}} + N_{\text{op,iter}}^{\text{SAPT}} \sim (6k+\kappa)D^2 + k^2(2D)^2\,. \tag{6.33}$$

In conclusion, the complexity associated with solving a reduced Rayleigh equation with SAPT as an iterative method scales *quadratically* with the number of modes and *quadratically* with the perturbation order k. Note that the other iterative methods studied in this chapter all scales linearly with the iteration number, and are simpler to implement. The iterative method based on SAPT may or may not be worth its slightly higher complexity depending on whether its convergence rate is advantageous. This will be explored in Section 6.5.

6.3 Fixed point iterates of the RRE of the second kind

6.3.1 Implementation

In Section 2.5, we have seen that reduced Rayleigh equations can be recast as integral equations of the second kind. It was suggested in Section 2.5 that one may solve the reduced Rayleigh equations of the second kind by interpreting it as a fixed point problem. We will refer to the following algorithm as the method of *Fixed point Iterates of the reduced Rayleigh Equations of*

[2] which contains only one non-zero element, that corresponding to the co-polarized zero diffractive order, being a Fresnel amplitude

the Second kind, or shorten with acronym FIRES. For a periodic system, the reduced Rayleigh equations of the second kind Eqs. (2.70, 2.73) read, for $\boldsymbol{\ell} \in \mathbb{Z}^2$,

$$(\alpha_2(\mathbf{p}_{\boldsymbol{\ell}}) - \alpha_1(\mathbf{p}_{\boldsymbol{\ell}}))^{-1}\ \mathbf{M}_{21}^{++}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_{\boldsymbol{\ell}})\, \Delta\mathbf{R}^{(\boldsymbol{\ell})}(\mathbf{p}_0) + \sum_{\mathbf{m}\in\mathbb{Z}^2} \tilde{\mathcal{K}}_{21}^{++}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_{\mathbf{m}})\ \mathbf{M}_{21}^{++}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_{\mathbf{m}})\, \Delta\mathbf{R}^{(\mathbf{m})}(\mathbf{p}_0)$$
$$= -\tilde{\mathcal{K}}_{21}^{++}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_0)\ \mathbf{M}_{21}^{++}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_0)\, \boldsymbol{\rho}^{(0)}(\mathbf{p}_0) - \tilde{\mathcal{K}}_{21}^{+-}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_0)\ \mathbf{M}_{21}^{+-}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_0)\,. \tag{6.34}$$

and

$$(\alpha_2(\mathbf{p}_{\boldsymbol{\ell}}) - \alpha_1(\mathbf{p}_{\boldsymbol{\ell}}))^{-1}\ \mathbf{M}_{12}^{-,-}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_{\boldsymbol{\ell}})\, \Delta\mathbf{T}^{(\boldsymbol{\ell})}(\mathbf{p}_0) + \sum_{\mathbf{m}\in\mathbb{Z}^2} \tilde{\mathcal{K}}_{12}^{-,-}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_{\mathbf{m}})\ \mathbf{M}_{12}^{-,-}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_{\mathbf{m}})\, \Delta\mathbf{T}^{(\mathbf{m})}(\mathbf{p}_0)$$
$$= -\tilde{\mathcal{K}}_{12}^{-,-}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_0)\ \mathbf{M}_{12}^{-,-}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_0)\, \boldsymbol{\tau}^{(0)}(\mathbf{p}_0)\,, \tag{6.35}$$

with

$$\tilde{\mathcal{K}}_{l,m}^{b,a}(\mathbf{p} \mid \mathbf{q}) \stackrel{\text{def}}{=} \frac{1}{a_c} \int_{a_c} \frac{\exp[-i(b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q}))\zeta(\mathbf{x})] - 1}{b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q})} \exp[-i(\mathbf{p}-\mathbf{q})\cdot\mathbf{x}]\ \mathrm{d}^2x\,, \tag{6.36}$$
$$\Delta\mathbf{R}^{(\boldsymbol{\ell})}(\mathbf{p}_0) \stackrel{\text{def}}{=} \mathbf{R}^{(\boldsymbol{\ell})}(\mathbf{p}_0) - \boldsymbol{\rho}^{(0)}(\mathbf{p}_0)\,, \tag{6.37}$$
$$\Delta\mathbf{T}^{(\boldsymbol{\ell})}(\mathbf{p}_0) \stackrel{\text{def}}{=} \mathbf{T}^{(\boldsymbol{\ell})}(\mathbf{p}_0) - \boldsymbol{\tau}^{(0)}(\mathbf{p}_0)\,. \tag{6.38}$$

Equations (6.34) and (6.35) yield the following recursive scheme

$$\Delta\mathbf{X}_{\boldsymbol{\ell}}^{(0)}(\mathbf{p}_0) = \mathbf{0} \tag{6.39a}$$
$$\Delta\mathbf{X}_{\boldsymbol{\ell}}^{(k+1)}(\mathbf{p}_0) = \mathbf{Y}_{\boldsymbol{\ell}}(\mathbf{p}_0) + \sum_{\mathbf{m}\in\mathbb{Z}^2} \mathbf{M}_{\boldsymbol{\ell},\mathbf{m}}(\mathbf{p}_0)\, \Delta\mathbf{X}_{\mathbf{m}}^{(k)}(\mathbf{p}_0) \qquad \text{for } k \in \mathbb{N}\,, \tag{6.39b}$$

where we denote by $\Delta\mathbf{X}_{\boldsymbol{\ell}}^{(k)}$ the k^{th} iterate approximation of $\Delta\mathbf{R}^{(\boldsymbol{\ell})}(\mathbf{p}_0)$ or $\Delta\mathbf{T}^{(\boldsymbol{\ell})}(\mathbf{p}_0)$ and the iteration matrices $\mathbf{M}_{\boldsymbol{\ell},\mathbf{m}}(\mathbf{p}_0)$ and $\mathbf{Y}_{\boldsymbol{\ell}}(\mathbf{p}_0)$ are given by

$$\mathbf{Y}_{\boldsymbol{\ell}}(\mathbf{p}_0) \stackrel{\text{def}}{=} (\alpha_1(\mathbf{p}_{\boldsymbol{\ell}}) - \alpha_2(\mathbf{p}_{\boldsymbol{\ell}}))\ \left[\mathbf{M}_{21}^{++}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_{\boldsymbol{\ell}})\right]^{-1}$$
$$\times \left[\tilde{\mathcal{K}}_{21}^{++}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_0)\ \mathbf{M}_{21}^{++}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_0)\, \boldsymbol{\rho}^{(0)}(\mathbf{p}_0) + \tilde{\mathcal{K}}_{21}^{+-}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_0)\ \mathbf{M}_{21}^{+-}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_0)\right] \tag{6.40a}$$
$$\mathbf{M}_{\boldsymbol{\ell},\mathbf{m}}(\mathbf{p}_0) \stackrel{\text{def}}{=} (\alpha_1(\mathbf{p}_{\boldsymbol{\ell}}) - \alpha_2(\mathbf{p}_{\boldsymbol{\ell}}))\ \left[\mathbf{M}_{21}^{++}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_{\boldsymbol{\ell}})\right]^{-1}\ \tilde{\mathcal{K}}_{21}^{++}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_{\mathbf{m}})\ \mathbf{M}_{21}^{++}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_{\mathbf{m}})\,, \tag{6.40b}$$

for the equation for the reflection amplitudes and by

$$\mathbf{Y}_{\boldsymbol{\ell}}(\mathbf{p}_0) \stackrel{\text{def}}{=} (\alpha_1(\mathbf{p}_{\boldsymbol{\ell}}) - \alpha_2(\mathbf{p}_{\boldsymbol{\ell}}))\ \left[\mathbf{M}_{12}^{-,-}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_{\boldsymbol{\ell}})\right]^{-1}\ \tilde{\mathcal{K}}_{12}^{-,-}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_0)\ \mathbf{M}_{12}^{-,-}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_0)\, \boldsymbol{\tau}^{(0)}(\mathbf{p}_0) \tag{6.41a}$$
$$\mathbf{M}_{\boldsymbol{\ell},\mathbf{m}}(\mathbf{p}_0) \stackrel{\text{def}}{=} (\alpha_1(\mathbf{p}_{\boldsymbol{\ell}}) - \alpha_2(\mathbf{p}_{\boldsymbol{\ell}}))\ \left[\mathbf{M}_{12}^{-,-}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_{\boldsymbol{\ell}})\right]^{-1}\ \tilde{\mathcal{K}}_{12}^{-,-}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_{\mathbf{m}})\ \mathbf{M}_{12}^{-,-}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_{\mathbf{m}})\,, \tag{6.41b}$$

for the equation of for the transmission amplitudes. This means that for the truncated system, when applying the re-indexing described in Section 5.2.1, Eq. (6.39) yields

$$\mathbf{X}_{\beta}^{(0)}(\mathbf{p}_0) = \mathbf{0} \tag{6.42a}$$
$$\mathbf{X}_{\beta}^{(k+1)}(\mathbf{p}_0) = \mathbf{Y}_{\beta}(\mathbf{p}_0) + \mathbf{M}(\mathbf{p}_0)\ \mathbf{X}_{\beta}^{(k)}(\mathbf{p}_0)\,, \tag{6.42b}$$

where

$$\Delta \mathbf{X}_{\beta}^{(k)}(\mathbf{p}_0) = \left(\Delta X_{\mathbf{m}_1,p\beta}^{(k)}(\mathbf{p}_0), \cdots, \Delta X_{\mathbf{m}_D,p\beta}^{(k)}(\mathbf{p}_0), \Delta X_{\mathbf{m}_1,s\beta}^{(k)}(\mathbf{p}_0), \cdots, \Delta X_{\mathbf{m}_D,s\beta}^{(k)}(\mathbf{p}_0)\right)^{\mathrm{T}}, \tag{6.43}$$

and for $(i,j) \in [\![1,D]\!]^2$,

$$M_{ij}(\mathbf{p}_0) = M_{\boldsymbol{\ell}_i,\mathbf{m}_j,pp}(\mathbf{p}_0) \tag{6.44}$$
$$M_{D+i,j}(\mathbf{p}_0) = M_{\boldsymbol{\ell}_i,\mathbf{m}_j,sp}(\mathbf{p}_0) \tag{6.45}$$
$$M_{i,D+j}(\mathbf{p}_0) = M_{\boldsymbol{\ell}_i,\mathbf{m}_j,ps}(\mathbf{p}_0) \tag{6.46}$$
$$M_{D+i,D+j}(\mathbf{p}_0) = M_{\boldsymbol{\ell}_i,\mathbf{m}_j,ss}(\mathbf{p}_0)\,, \tag{6.47}$$

and

$$\mathbf{Y}_{\beta}(\mathbf{p}_0) = \left(Y_{\mathbf{m}_1,p\beta}(\mathbf{p}_0), \cdots, Y_{\mathbf{m}_D,p\beta}(\mathbf{p}_0), Y_{\mathbf{m}_1,s\beta}(\mathbf{p}_0), \cdots, Y_{\mathbf{m}_D,s\beta}(\mathbf{p}_0)\right)^{\mathrm{T}}. \tag{6.48}$$

6.3.2 Complexity analysis

The complexity associated with setting up the iteration matrix $\mathbf{M}(\mathbf{p}_0)$ and the two column vectors $\mathbf{Y}_{\beta}(\mathbf{p}_0)$ is essentially the same as that associated with setting the linear system for the direct method. Then, each iteration requires, two matrix-vector multiplications and two vector additions (one per incident polarization state β), i.e. $8D^2$ scalar multiplications and $4D^2$ scalar additions. The complexity is essentially that of the simple iterative methods, such as the Jacobi method.

6.4 Fixed point iterates of the coupled Rayleigh equations

In Section 2.4, we have shown that one could obtain coupled integral equations for the reflection and transmission amplitudes, which we have called the coupled Rayleigh equations. It was suggested, based on the form of the equations, that one may gather the reflection and transmission amplitudes as one unknown and interpret the coupled Rayleigh equations as a fixed point problem.

6.4.1 Implementation

The implementation of the algorithm for obtaining the successive iterates for the reflection and transmission amplitudes is similar to that of the methods presented previously. We need to keep track simultaneously of the successive iterates of both the vector of reflection amplitudes, $\mathbf{R}_{\beta}^{(k)}(\mathbf{p}_0)$, and of the transmission amplitudes, $\mathbf{T}_{\beta}^{(k)}(\mathbf{p}_0)$, for which the indexing has been explained in details for the previous methods. The iterative scheme hence reads as follows.

- Initialize $\mathbf{R}_{\beta}^{(0)}(\mathbf{p}_0)$ and $\mathbf{T}_{\beta}^{(0)}(\mathbf{p}_0)$, for example with the Fresnel amplitudes.
- Loop until a desired accuracy or a set number of iteration is reached

$$\mathbf{R}_{\beta}^{(k+1)}(\mathbf{p}_0) = \mathbf{M}_1(\mathbf{p}_0)\ \mathbf{T}_{\beta}^{(k)}(\mathbf{p}_0) \tag{6.49a}$$
$$\mathbf{T}_{\beta}^{(k+1)}(\mathbf{p}_0) = \mathbf{M}_2(\mathbf{p}_0)\ \mathbf{R}_{\beta}^{(k)}(\mathbf{p}_0) + \mathbf{Y}_{\beta}(\mathbf{p}_0)\,. \tag{6.49b}$$

The iteration matrices $\mathbf{M}_1(\mathbf{p}_0)$, $\mathbf{M}_2(\mathbf{p}_0)$, and the vector $\mathbf{Y}_\beta(\mathbf{p}_0)$ are readily obtained from Eqs. (2.63a) and (2.63b), and read for a periodic surface, in terms of the $\boldsymbol{\ell}, \mathbf{m}$ indices,

$$\mathbf{M}_{1,\boldsymbol{\ell},\mathbf{m}}(\mathbf{p}_0) \stackrel{\text{def}}{=} \frac{\epsilon_1 - \epsilon_2}{2\sqrt{\epsilon_1\epsilon_2}\,\alpha_1(\mathbf{p}_{\boldsymbol{\ell}})}\; \tilde{\mathcal{J}}_{12}^{+,-}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_{\mathbf{m}})\; \mathbf{M}_{12}^{+,-}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_{\mathbf{m}}) \tag{6.50a}$$

$$\mathbf{M}_{2,\boldsymbol{\ell},\mathbf{m}}(\mathbf{p}_0) \stackrel{\text{def}}{=} \frac{\epsilon_1 - \epsilon_2}{2\sqrt{\epsilon_1\epsilon_2}\,\alpha_2(\mathbf{p}_{\boldsymbol{\ell}})}\; \tilde{\mathcal{J}}_{21}^{-,+}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_{\mathbf{m}})\; \mathbf{M}_{21}^{-,+}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_{\mathbf{m}}) \tag{6.50b}$$

$$\mathbf{Y}_{\boldsymbol{\ell}}(\mathbf{p}_0) \stackrel{\text{def}}{=} \frac{\epsilon_1 - \epsilon_2}{2\sqrt{\epsilon_1\epsilon_2}\,\alpha_1(\mathbf{p}_{\boldsymbol{\ell}})}\; \tilde{\mathcal{J}}_{21}^{-,-}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_0)\; \mathbf{M}_{21}^{-,-}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_0) \tag{6.50c}$$

The re-indexing procedure is the same as for the other methods.

Remark 6.7. In the algorithm described above we construct in fact two sequences of solutions in parallel. One may instead jump between the two sequences and obtain a factor two speed up. Indeed, one may, for example, use the freshly computed $\mathbf{R}_\beta^{(k+1)}(\mathbf{p}_0)$ from Eq. (6.49a) directly in Eq. (6.49b), and so on, in such a way that the $\mathbf{R}_\beta^{(k)}(\mathbf{p}_0)$ are only computed for odd k and the $\mathbf{T}_\beta^{(k)}(\mathbf{p}_0)$ for even k. One may also start the other way around and compute only $\mathbf{R}_\beta^{(k)}(\mathbf{p}_0)$ for even k and the $\mathbf{T}_\beta^{(k)}(\mathbf{p}_0)$ for odd k. For the study of the convergence rate, we will for convention use the naive algorithm (Eqs.(6.49a) and (6.49b)) where all terms are computed.

6.4.2 Complexity analysis

The complexity analysis is rather straightforward here. It consists essentially in twice that of the direct method[3] for setting up the iteration matrices plus a couple of matrix-vector for each full *R-T*-iteration for a given incident polarization state β. The complexity is hence twice that of the method of FIRES for example.

6.5 Comparison of the methods

We now proceed with some numerical comparison of the aforementioned methods. We have already analyzed the methods complexity and discussed iteration thresholds for a speed-up of the iterative methods compared to the direct method. It remains to study whether these methods are stable and estimate their rate of convergence towards the solution given by the direct method. We will see that the rate of convergence depends on the system parameters in a similar fashion as the convergence rate depended on the system parameters for the direct method when the convergence was studied as a function of the number of modes in Chapter 5. We will fix the index cutoff radius N_c of the reciprocal lattice to be large enough ($N_c = 20$) to obtain a *a priori* small enough truncation error with the direct method (see Chapter 5). Consequently, we remind that this fixes a set of scattered wave vectors, both for propagating and evanescent modes. For simplicity, we will only consider a two-dimensional sinusoidal surface between two dielectric, where the periods in the two orthogonal directions are equal $a_1 = a_2 = a$ and the only consider normal incidence. Thus we places ourselves in a similar framework as in Chapter 5.

The notion of convergence used in the present chapter is similar to that used in Chapter 5 and the vocabulary will remain essentially the same. There is, however, an important difference that we want to stress. Since, for a given set of parameters, the scattering problem can be solved

[3]Some computation can, however, be saved by setting up both $\mathbf{M}_1$ and $\mathbf{M}_2$ at once.

with the direct method, we can use its solution as a reference. The relative error will hence measure the distance between an approximate solution obtained for a given number of iteration k of an iterative method and the reference solution given by the direct method. If we denote the solution given by the direct method by $\mathbf{X}_{\mathrm{d}} = \left(\mathbf{X}_{\mathrm{d}}^{(\mathbf{m})}(\mathbf{p}_0,\omega)\right)_{\mathbf{m}\in\mathcal{D}_{N_c}}$ and the solution given by an indirect method after k iterations by $\mathbf{X}_{\mathrm{i},k} = \left(\mathbf{X}_{\mathrm{i},k}^{(\mathbf{m})}(\mathbf{p}_0,\omega)\right)_{\mathbf{m}\in\mathcal{D}_{N_c}}$ ($\mathbf{X} = \mathbf{R}$ or, $\mathbf{T}$), then we define the relative error in the Euclidean norm ε_k as

$$\varepsilon_k = \frac{|\mathbf{X}_{\mathrm{i},k} - \mathbf{X}_{\mathrm{d}}|}{|\mathbf{X}_{\mathrm{d}}|}\,. \tag{6.51}$$

Here, we define $|\mathbf{X}|$ as

$$|\mathbf{X}| = \left(\sum_{\mathbf{m}\in\mathcal{D}_{N_c}}\sum_{\alpha,\beta\in\{p,s\}} |x_{\alpha,\beta}^{(\mathbf{m})}(\mathbf{p}_0,\omega)|^2\right)^{1/2}, \tag{6.52}$$

where we remind that the notation $\alpha,\beta \in \{p,s\}$ denotes pairs of polarization states. Other norms could be used but we restrict ourselves to the Euclidean norm for the sake of simplicity.

We are now interested in studying how quickly the relative error ε decays (if it does) with the number of iterations of iterative methods and also how the corresponding convergence rate scales with the reduced parameters. The linear convergence rate determined numerically is defined in a similar fashion as in Chapter 5,

$$\eta = \frac{1}{K+1-\kappa}\sum_{k=\kappa}^{K}\eta_k\,, \tag{6.53}$$

where K is the maximum number of iterations considered before reaching machine precision and with

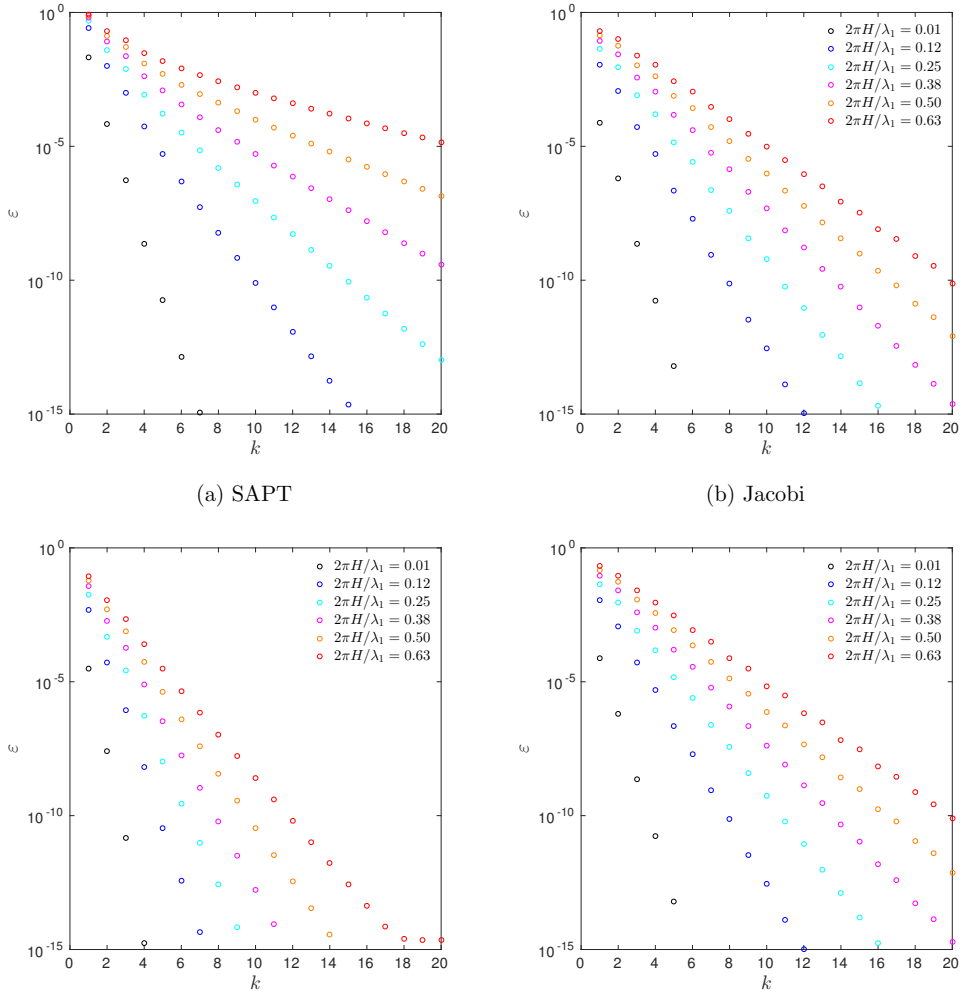
$$\eta_k = \frac{\varepsilon_{k+1}}{\varepsilon_k}\,. \tag{6.54}$$

6.5.1 Convergence rates

Consider the relative error ε as a function of the iteration number k obtained for the different methods, SAPT, Jacobi, Gauss-Seidel, and FIRES in Fig, 6.2 for a fixed reduced lattice constant ($a/\lambda_1 = 0.95$)[4] and for a set of reduced amplitudes. Moreover, the dielectric contrast is held fixed to $z = 1.5$.

First, it is clear that for each method taken separately, the convergence is slower for larger values of the reduced amplitude. It seems that the convergence is linear and the linear convergence rate, η, hence increases with the reduced amplitude. Furthermore, comparing now the methods for a fixed reduced amplitude, we can rank them with respect to their convergence rate from slowest to fastest: $\eta_{\mathrm{SAPT}} > \eta_{\mathrm{Jacobi}} \approx \eta_{\mathrm{FIRES}} > \eta_{\mathrm{Gauss-Seidel}}$. The reader will notice that the Jacobi method and the method of FIRES yield very similar results. They are not identical even though the difference cannot be appreciated from Figs. 6.2(b) and 6.2(d). The reason for this similarity can be explained quite simply. The method of FIRES consists in splitting the

[4]There is nothing special about the value of 0.95. It was arbitrary chosen as the closest to 1 in a bigger set of simulations, and for which the results illustrate well our discussion.



(a) SAPT

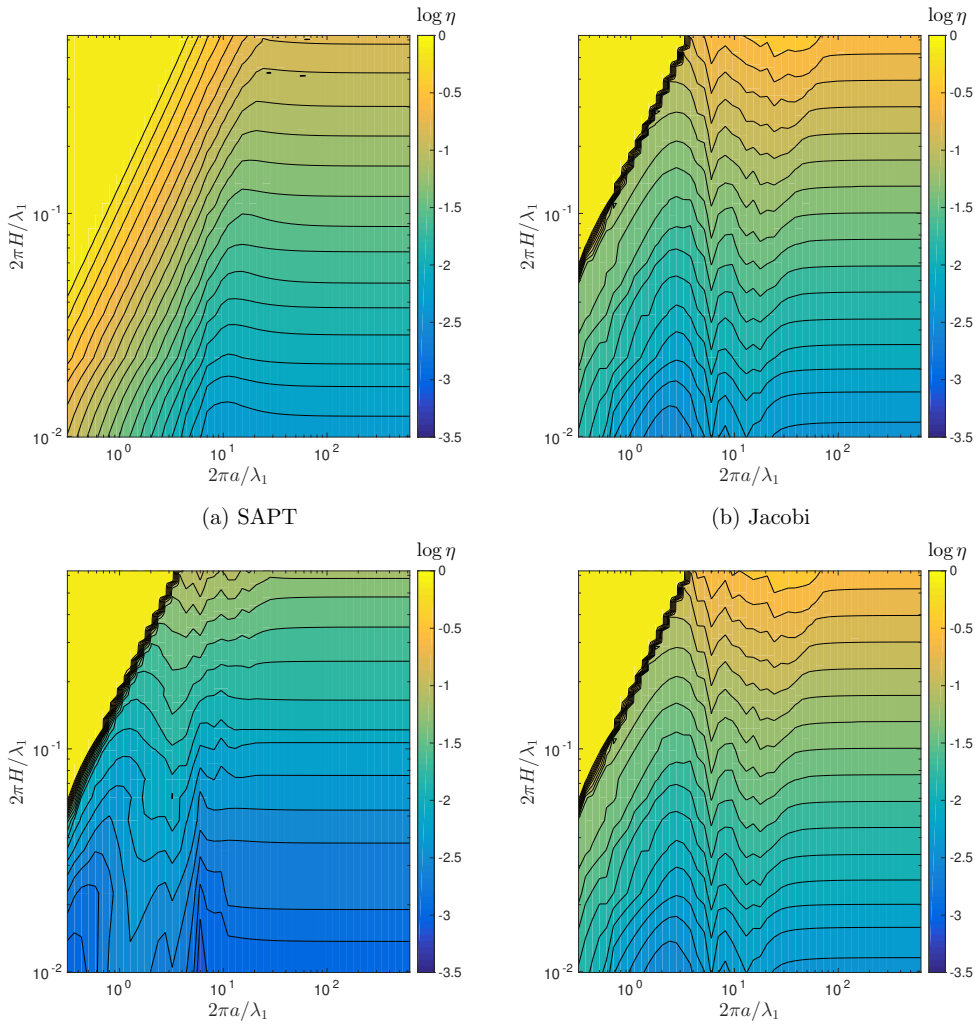
(b) Jacobi

(c) Gauss-Seidel

(d) FIRES

Figure 6.2: Error ε in the euclidean norm for the methods of SAPT (a), Jacobi (b), Gauss-Seidel (c) and FIRES (d) as a function of the iteration number k for two-dimensional sinusoidal surface characterized by $a/\lambda_1 = 0.95$ and for different amplitudes. The dielectric contrast is fixed to $z = 1.5$.

solution as a sum of the planar contribution and the roughness contribution of the surface, i.e. Fresnel plus scattering roughly speaking. This idea is very close to the splitting used in the Jacobi method, where the matrix of the linear system is split as its diagonal and what remains. One may hence expect similar results for the two methods for cases where the Fresnel contribution dominates in the diagonal.

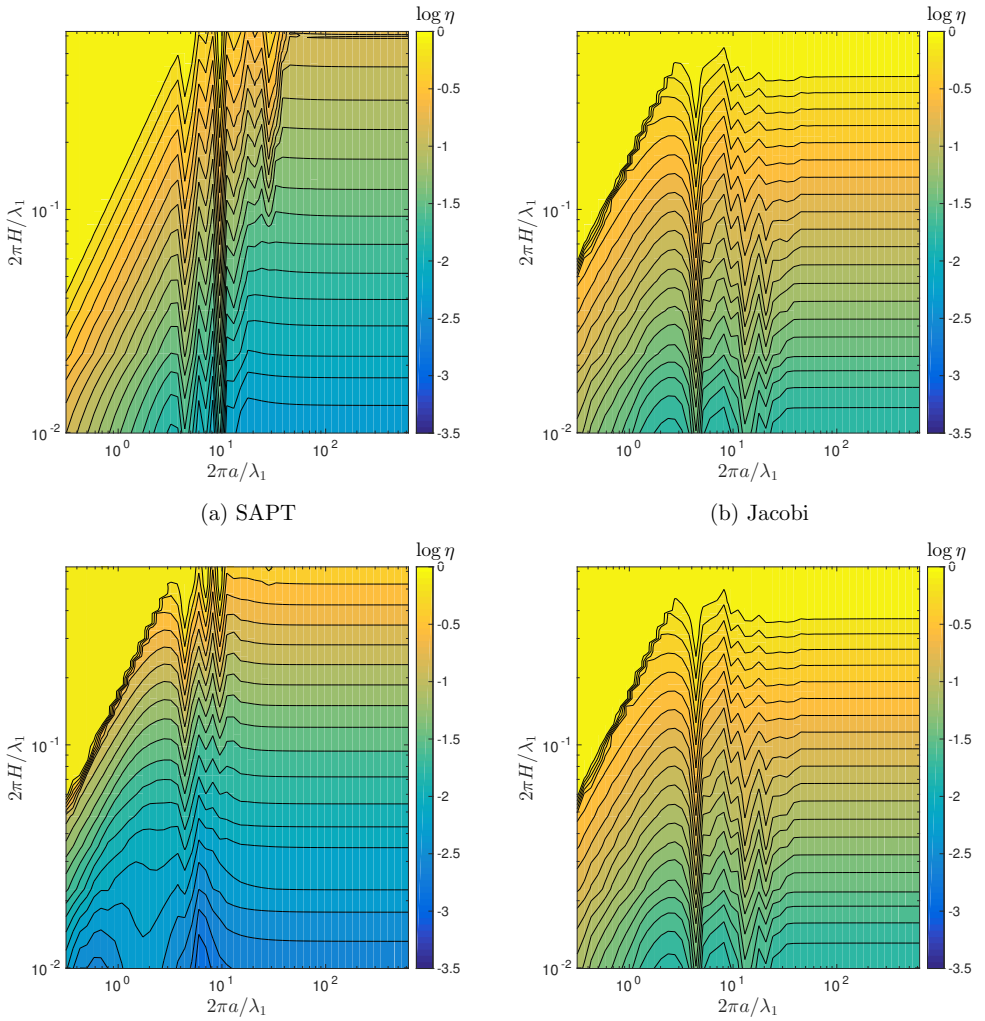


(a) SAPT (b) Jacobi

(c) Gauss-Seidel (d) FIRES

Figure 6.3: Diagram of the convergence rate $\log \eta$ in the $(2\pi a/\lambda_1, 2\pi H/\lambda_1)$-plane for the methods of (a) SAPT, (b) Jacobi, (c) Gauss-Seidel, and (d) FIRES. The results were obtained for a two-dimensional sinusoidal surface and the dielectric contrast is fixed to $z = 1.5$.

In a similar fashion as what was done in Chapter 5, Fig. 6.3 present diagrams of convergence rate in the $(2\pi a/\lambda_1, 2\pi H/\lambda_1)$ parameter space. The overall map of contour levels of constant convergence rate are reminiscent of that obtained in Chapter 5, see Fig. 5.7. This similarity indicates that solving the reduced Rayleigh equations using the presented methods will also be more and more challenging, in terms of computations, near the boundary of the convergence



(a) SAPT

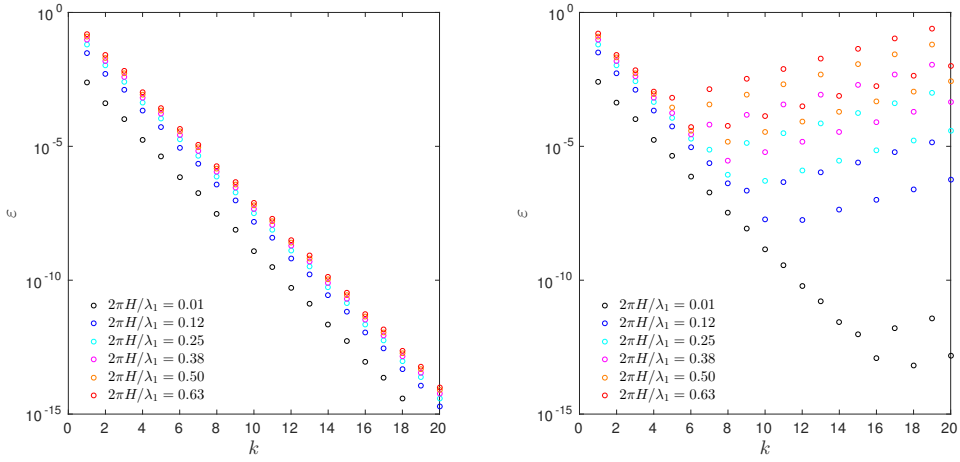
(b) Jacobi

(c) Gauss-Seidel

(d) FIRES

Figure 6.4: Diagram of the convergence rate $\log \eta$ in the $(2\pi a/\lambda_1, 2\pi H/\lambda_1)$-plane for the methods of (a) SAPT, (b) Jacobi, (c) Gauss-Seidel, and (d) FIRES. The results were obtained for a two-dimensional sinusoidal surface and the dielectric contrast is fixed to $z^2 = -2.25$.

domain identified in Chapter 5, as k will be required to be large. From Fig. 6.3, it also apparent that the Gauss-Seidel method overall performs best since it overall has smaller values of $\log \eta$ compared to the other methods. Moreover, SAPT seems to suffer from slow convergence in the slope dominate regime while it is comparable to the Jacobi and FIRES methods in the amplitude dominated regime.



(a) $a/\lambda_1 = 100$ (b) $a/\lambda_1 = 4.5$

Figure 6.5: Error ε in the euclidean norm for the method of Fixed Point Iterates of the Coupled Rayleigh Equations (FICRE) as a function of the iteration number k for two-dimensional sinusoidal surface characterized by (a) $a/\lambda_1 = 100$ and (b) $a/\lambda_1 = 4.5$ and different amplitudes. The dielectric contrast is fixed to $z = 1.5$.

Figure 6.4 presents the same study as for Fig. 6.3 but for a metallic system ($z^2 = -2.25$). We can see that all the iterative methods are affected by resonances. The Jacobi and FIRES method, which again are very similar are the most affected by the metallic nature of the substrate. It seems that the overall convergence diagram is shifted to lower critical values of the reduced amplitudes in addition to the resonances causing problem (see Figs. 6.4(b) and 6.4(d)). The Gauss-Seidel method performs best away from resonances (Fig. 6.4(c)) and SAPT seems to be less affected than the Jacobi and FIRES methods away from resonances (Fig. 6.4(a)).

6.5.2 The case of the FICRE

The attentive reader has probably already noticed that we have left aside the method of fixed point iterates of the coupled Rayleigh equations (FICRE) detailed in Section 6.4. The reason for this choice is that we devote here a specific discussion due to its peculiar stability property. Figures 6.5(a) and 6.5(b) show respectively the relative error ε as a function of the iteration number k for the FICRE method for $a/\lambda_1 = 100$ and $a/\lambda_1 = 4.5$, and a set of reduced amplitudes in each case. By observing Fig. 6.5(a), for which the lattice constant is much larger than the wavelength, convergence seems guaranteed for all the considered reduced amplitudes. What is particularly striking, however, is that the convergence rate is *independent* of the chosen amplitude. The initial error point of course increases with increasing amplitude, since the starting guess of the iterative scheme is chosen to be the Fresnel solution which is expected to deviate more and more from the true solution when the amplitude is increased. Nevertheless, the slope of the line of decay of ε in the semilog-plot in Fig. 6.5(a) has the same slope for all the reduced amplitudes. This is a quite remarkable feature. The second remarkable feature is that the scheme becomes unstable if the reduced lattice constant is decreased. This is readily

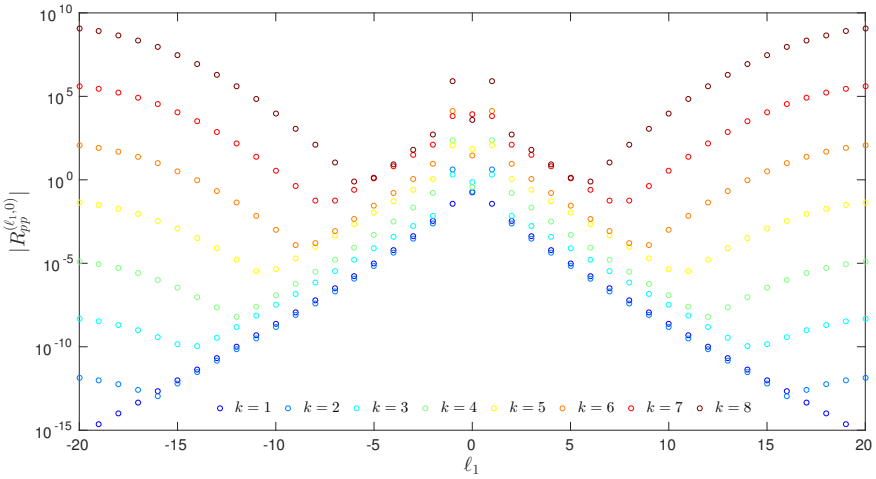


Figure 6.6: Modulus of the reflection amplitudes $R_{pp}^{(\ell_1,0)}$ as a function of ℓ_1 for different values of the iteration number k obtained with the method of FICRE for two-dimensional sinusoidal surface characterized by $a/\lambda_1 = 1$ and $H/\lambda_1 = 1/2$. The dielectric contrast is fixed to $z = 1.5$.

seen in Fig. 6.5(b) where for each the considered reduced amplitudes there exists a critical iteration number k_c for which the error starts to increase even though it has been decaying for $k < k_c$. The larger is the reduced amplitude, the smaller is the critical iteration number k_c. This behavior seems to indicate that the scheme may either be fundamentally unstable and that k_c decreases with increasing surface slope. The use of the FICRE method may then be guided by a trade off between fast enough convergence before the instability takes over.

Figure 6.6 illustrates well the instability of the numerical scheme. It shows the modulus of the reflection amplitudes of the co-polarized $p \to p$ diffracted modes, $|R_{pp}^{(\ell_1,0)}|$ scattered in the plane of incidence (i.e. $\ell_2 = 0$ in our configuration). As the iterations go, the initially, seemingly, reasonable distribution with decaying amplitudes for larger values of $|\ell_1|$, starts to lift up from the larger values of $|\ell_1|$ and progressively pollute the whole solution.

Remark 6.8. The convergence rate associated with the method of FICRE may seem advantageous compared to the other presented methods in some region of the parameter space, although it seems at first sight not so attractive. One must keep in mind that the results presented here were obtained with the naive version of the method of FICRE (see Remark 6.7). Using the algorithm described in Remark 6.7, the convergence rate η is then halved.

6.6 Summary

In the present chapter, a few iterative methods have been analyzed in terms of complexity and convergence rate with respect to the number of iterations. Standard linear algebra iterative methods such as the Jacobi and Gauss-Seidel methods have been compared with the iterative methods issued from small amplitude perturbation theory, the method of fixed point iterates of the reduced Rayleigh equations of the second kind and the fixed point iterates of the coupled

Rayleigh equations. We have observed that for all methods, the convergence rate with the number of iterations follows a similar behavior as the convergence rate with respect to the number of modes found in Chapter 5 as a function of the surface profile parameters. The slope dominated regime and the amplitude dominated regime are also present for the convergence rate of iterative methods. We have found that the Gauss-Seidel method converges fastest in the convergence region of the parameter space and that the other methods have similar convergence rate. Small amplitude perturbation theory has a rather poor convergence rate compared with other methods in the slope dominated regime. The Jacobi method and the method of fixed point iterates of the reduced Rayleigh equations of the second kind have a very similar convergence rate, which can be explained by their very close resemble in terms of implementation. The method of fixed point iterates of the coupled Rayleigh equations was treated as special case since it was found to be unstable. The instability develops quicker for small values of a/λ_1 than for larger ones. The overall conclusion to be drawn from this study is that simple linear algebra methods such as the Jacobi and Gauss-Seidel may be used in practice to yield a significant speed-up for solving the reduced Rayleigh equations for dielectric systems. In terms of complexity, the method based on small amplitude perturbation theory is not advantageous, as it is rather cumbersome to implement, has an additional increasing complexity with the number of iterations as compared with the other methods and does not yield a particularly larger speed-up. The method of fixed point iterates of the reduced Rayleigh equations of the second kind is particularly interesting in terms of convergence and simplicity of implementation.

Now that we have filled up our tool box for solving numerically the reduced Rayleigh equations, an important question remains to be addressed. How good is the obtained numerical solution?

Chapter 7

On the Rayleigh hypothesis and the validity of the RRE

"That's all well and good in practice, but how does it work in theory?"
Shmuel Weinberger.

7.1 The Rayleigh hypothesis

In 1907, Lord Rayleigh proposed a theory for the scattering of a plane wave by an impenetrable periodic grating in which he assumed that the scattered field could be decomposed as a sum of out-going plane waves[1] [15]. The unknown amplitudes of the expansion of the scattered field were then determined by the boundary conditions. In addition, Rayleigh also derived an approximate solution in the case of a sinusoidal perfectly conducting surface whose amplitude is small compared with the wavelength. The assumption on the representation of the scattered field was left unchallenged for almost half a century, until Lippmann noted that such a representation of the scattered field seemed, at first sight, not physical [69]. Based on a physical intuition, Lippmann suggested that the scattered field inside the grooves should not only contain out-going plane waves but also waves propagating towards the surface. Indeed, it seems reasonable to assume that the incoming wave excites secondary sources at points situated near the maxima of the surface profile which radiate as spherical waves in view of the Huygens-Fresnel principle, and therefore, points well inside the grooves should receive these downwards propagating secondary waves. Lippmann's note started further numerical and mathematical developments which aimed at justifying or delimiting the range of validity of the now so-called Rayleigh hypothesis. This has lead, quite surprisingly, to diverging opinions and conclusions while, as we will see, all the results obtained by the different authors were consistent between themselves. Only their interpretation has lead to confusion. We can identify, essentially two school of thought. On one side, some authors took as a criterion for correct results the conservation of energy, and claimed that the Rayleigh hypothesis is valid for a relatively wide range of surface parameters, in the sense that the reflection amplitudes converge (as more and more modes are kept in the system) and the energy is conserved. On the other side, others considered the conservation of energy as a necessary condition for the correctness of the result but not a sufficient one. Since the Rayleigh plane wave expansion of the field satisfies the Helmholtz equation, and an out-going wave radiation condition by construction, once the

[1] The terms "out-going plane waves" will implicitly refer to both progressive out-going waves and evanescent waves decaying exponentially away from the surface.

reflection amplitudes are determined, one should verify whether the boundary conditions are satisfied in order to conclude whether the field is indeed solution of the scattering problem. The latter approach yields a more restricted range of validity of the Rayleigh hypothesis. It is mainly the difference of what is to be considered as a correct solution which was the source of vivid discussions. We briefly summarize in the following paragraphs the main results and conclusions obtained by a few authors who, we believe, have contributed most to the discussion on the range of validity of the Rayleigh hypothesis without the intention of being exhaustive. We would like to stress, however, that the point here is not to judge whether some authors are right or wrong, but rather clarify certain confusion and point out the interest of the different approaches.

The first, and maybe one of the most important result concerning the range of validity of the Rayleigh hypothesis was obtained by Petit and Cadilhac in 1966 [70]. They considered the scattering of a scalar plane wave by a one-dimensional sinusoidal surface, of equation $x_3 = H \sin\big(2\pi x_1/a\big)$, subjected to the Dirichlet boundary condition on the surface. The problem being a two-dimensional one in the (x_1, x_3)-plane, they could use tools of complex analysis to show that: if the characteristic slope $2\pi H/a > 0.448$, the expansion in out-going plane waves for the scattered field, valid above the highest point on the surface (i.e. for $x_3 > H$), admits an analytic continuation with singularities located above the lowest point on the surface (i.e. $x_3 < -H$). Consequently, the Rayleigh hypothesis which states the validity of the out-going plane wave expansion of the scattered field inside the grooves breaks down, for the considered problem, for slopes $2\pi H/a > 0.448$. But does it means that it holds for slopes $2\pi H/a < 0.448$?

Millar completed the analysis of Petit and Cadilhac in 1969 and 1971 [54, 55], by showing, also using tools of complex analysis, that the Rayleigh hypothesis is valid *if and only if* the characteristic slope $2\pi H/a < 0.448$, or stated in another way that the analytic continuation of the field has no singularity above the lowest point of the surface under this condition. It must be noted that Millar's criterion for a sinusoidal surface has been obtained as a special case of a derivation valid for any type of one-dimensional surface profile provided it is periodic and represent an analytic arch in the complex plane. Millar showed in [54] that, for a general periodic analytic profile ζ, there always exists a constant γ such that for

$$2\pi\|\zeta\|_\infty/a < \gamma\,, \tag{7.1}$$

the Rayleigh hypothesis holds. However, Millar noted that the derivation for obtaining the existence of such a threshold γ is of little practical importance since determining its value or a lower bound is in general a difficult task. Nevertheless, the value[2] of $\gamma \approx 0.448$ could be proved for the sinusoidal surface [55].

Despite Millar's proof, a number of publications claimed the validity of the Rayleigh hypothesis well above the $2\pi H/a = 0.448$ threshold based on numerical work including (see e.g. [71, 72, 73]). Before following the discussion, it must be mentioned, as correctly pointed out by Wirgin [72] already in 1982 and more recently by Wauer and Rother [74], that there exists a whole family of numerical methods based on the Rayleigh expansion which differ by the sense given to the boundary conditions, and hence, may yield different results in terms of convergence. In the following, we restrict ourselves to those that seem numerically closer to satisfying the boundary conditions at every points of the surface, specifically the point

[2] 0.448 is an approximation of the solution of the equation $\operatorname{argsh}(1/x) = x + (1+x^2)^{1/2}$, solution which, is the exact value of the threshold.

matching method and the Rayleigh-Fourier method (which is exactly a version of the reduced Rayleigh equation for a perfectly conducting surface). The former discretizes the surface and directly writes the point-wise condition for each discretized point on the surface while the latter considered projections of the field onto the set of plane waves (see derivation of the RRE). When such procedures are applied, it is often shown numerically that convergent results (with increasing number of modes kept in the truncated expansion) are obtained for values of $2\pi H/a$ beyond 0.448 and that the energy balance can satisfied with very good accuracy. The authors usually then conclude by claiming the apparent numerical validity of the Rayleigh hypothesis beyond its analytical range of validity. To our knowledge, the first clear numerical answer to such numerical studies was given by van den Berg [75] (see also the interesting answers [76] to the equally interesting comments by Wirgin [72]). In Ref. [75], van den Berg considered the behavior of both the energy balance and the integrated mismatch of the boundary conditions along the surface as a function of the number of modes kept in the simulation, and this for different values of $2\pi H/a$. It was shown quite clearly that the error in the satisfaction of the boundary conditions decays with increasing number of modes for roughly $2\pi H/a < 0.448$ while it increases for $2\pi H/a > 0.448$. However, the deviation to the energy balance decays with increasing number of modes for parameters $2\pi H/a > 0.448$. The analytical limit of validity of the Rayleigh hypothesis was then illustrated numerically, and the distinction between a condition on satisfaction of the boundary conditions and of the energy balanced clearly made. If a field does not satisfy the boundary conditions, one should strictly speaking accept that the obtained field is *not* a solution of the scattering problem, and that the Rayleigh hypothesis used to represent the solution was an erroneous assumption. Note, however, that if one relaxes the boundary conditions (say that it should be satisfied in a least square sense instead of point wise), the apparently incorrect solution might become correct for the new relaxed problem. This has yield to modified methods for which the boundary conditions are required to be satisfied in a least square fashion for example (see e.g. [75, 72] and references therein). Moreover, the fact that for $2\pi H/a > 0.448$ some convergent results satisfying the conservation of energy are obtained may be of practical interest even though the solution does not satisfies the boundary conditions. Such results should of course be discarded for near-field studies. Nevertheless, the obtained solution could be a good approximation of the exact solution in the far-field. This resonates with what is known for asymptotic techniques for which the solutions are adequate only in the far-field, for example, and of limit use (or completely unphysical) in the near-field.

Finally, we would like to mention that the anaylis of Millar for the two-dimensional problem has later been generealized by van den Berg and Fokkema for the three-dimensional case [77, 78]. We are not aware, however, of a generalization for penetrable media. In the following sections, we will, inspired by the work of van den Berg, define the figures of merit to assess the quality of the numerical solutions. These will be the unitarity (conservation of energy) and consistency errors defined as some measures of the deviation to the satisfaction of the boundary conditions by the field. Using these tools, we will carry out numerically a parameter study of the validity of the Rayleigh hypothesis, or more safely, that of the reduced Rayleigh equations, first in the case of perfectly conducting sinusoidal surfaces and in the case of dielectric media. We will provide some analyitcal arguments to explain the observed behavior, and we will show that one may recover analytically Millar's threshold based on very simple notions, almost physically intuitive, which do not require the elaborate theory based on complex analysis developed by Petit, Millar, van den Berg and co-workers. The present work hence represents both a generalization of the numerical results of van den Berg's paper [75] for two-dimensional surfaces and penetrable surfaces, and suggests a new analytical path towards the derivation of simple criteria for assessing the validity

of the reduced Rayleigh equations for any surface.

7.2 Conservation of energy

Perhaps the simplest and the most commonly used a posteriori check for assessing the quality of a wave scattering simulation is that of the conservation of energy, or unitarity. Assuming, for the sake of simplicity, the case of the scattering of a monochromatic β-polarized plane wave by a periodic interface between two loss-less media, unitarity is expected to hold, i.e. that the sum of the efficiencies of all the reflected and transmitted propagating modes should equal unity,

$$U_\beta(\mathbf{p}_0) \overset{\text{def}}{=} \sum_{\substack{\boldsymbol{\ell}\in\mathbb{Z}^2 \\ \alpha\in\{p,s\}}} e_{\alpha\beta}^{(R,\boldsymbol{\ell})}(\mathbf{p}_0) + \sum_{\substack{\boldsymbol{\ell}\in\mathbb{Z}^2 \\ \alpha\in\{p,s\}}} e_{\alpha\beta}^{(T,\boldsymbol{\ell})}(\mathbf{p}_0) = 1\,. \tag{7.2}$$

We recall that by definition of the efficiencies, the sums are finite, since the efficiency for reflected (resp. transmitted) modes such that $|\,\mathbf{p}_{\boldsymbol{\ell}}\,| > n_1\,\omega/c$ (resp. $|\,\mathbf{p}_{\boldsymbol{\ell}}\,| > n_2\,\omega/c$) vanish. For reflection on a loss-less metallic surface, only the reflected modes would contribute. We must keep in mind, however, that unitarity, although it is a necessary condition for the correctness of the results, is not a sufficient one. Moreover, what does it mean numerically to satisfy the unitarity condition? What deviation from unity is acceptable? Are a few digits meaningful? Can we expect to go down to machine precision? In the literature, it is a common practice to verify that unitarity is satisfied within a few digits to assess the relative accuracy of the efficiencies. This is indeed a very reasonable thing to do when one is interested to compare with experimental results for which the accuracy is also limited by the instruments. Here, however, we will see that unitarity can be obtained up to machine precision, and can give in most cases a good assessment of the validity and accuracy of the results. One must nevertheless be careful with checking the convergence with the number of modes before assessing the validity or non validity of the results based solely on the unitarity. Verifying other conditions such as reciprocity can comfort ones opinion on the results but maybe the best verification is to check to which extent the fields satisfy the initial problem, that is the Helmholtz equation and the boundary conditions.

7.3 Consistency

We have seen in Section 2.3 that the resolution of the reduced Rayleigh equations in the limit case of a planar interface yields the correct Fresnel amplitudes. This is to be expected since a key hypothesis used to derive the integral equations for the reflection and transmission amplitudes is the so-called Rayleigh hypothesis, which states that the expression of the fields given by Eq. (1.31) can be taken down to the surface. It is not *a priori* obvious to know, for a given surface, whether the representation of the fields as a sum of plane waves described by Eq. (1.31) is valid in the whole space, included within the groves of the profile. In the case of a planar interface this hypothesis is verified but for an arbitrary rough surface, it is *a priori* not obvious to predict whether this hypothesis is valid or not, and if not, what is the error made by the method. To this end, we introduce the notion of *consistency errors*.

7.3.1 Definitions

The consistency errors are defined as deviation from the satisfaction of the boundary conditions by the fields obtained after resolution of the reduced Rayleigh equations. Indeed, since the plane wave expansion Eq. (1.31) satisfies the Helmholtz equation by construction, it is left to show

that the boundary conditions are satisfied. The *local consistency errors* associated with the electric and magnetic fields at a surface point $\mathbf{s}(\mathbf{x}) = \mathbf{x} + \zeta(\mathbf{x})\,\hat{\mathbf{e}}_3$ are defined as,

$$\varepsilon_{\mathrm{con,\,F}}(\mathbf{x}) = \left| \mathbf{n}(\mathbf{x}) \times \left[\tilde{\mathbf{F}}_1(\mathbf{s}(\mathbf{x})) - \tilde{\mathbf{F}}_2(\mathbf{s}(\mathbf{x})) \right] \right| , \tag{7.3}$$

where the field $\mathbf{F} = \mathbf{E}$ or $\mathbf{H}$ and we recall that $|\cdot|$ denotes the Euclidean norm. Here the tilde denote fields normalized by the norm of the incident field amplitude, i.e. $\tilde{\mathbf{E}} \stackrel{\mathrm{def}}{=} \mathbf{E}/E_0$ and $\tilde{\mathbf{H}} \stackrel{\mathrm{def}}{=} \mathbf{H}/H_0$. In other words, the local consistency errors, defined as Eq. (7.3), read as deviation from the boundary conditions being satisfied in the sense that a vanishing error corresponds to the boundary condition being satisfied locally. Although it will be of interest to analyze locally where the local consistency errors are strongest on the surface, we will study *integrated consistency errors* defined as,

$$\|\varepsilon_{\mathrm{con,\,F}}\|_p = \begin{cases} \left(\displaystyle\int_{\mathbb{R}^2} |\varepsilon_{\mathrm{con,\,F}}(\mathbf{x})|^p \, \mathrm{d}^2x \right)^{1/p}, & p \in [1, \infty[\\ \displaystyle\sup_{\mathbf{x} \in \mathbb{R}^2} |\varepsilon_{\mathrm{con,\,F}}(\mathbf{x})|, & p = \infty \end{cases} , \tag{7.4}$$

i.e. the L^p-norm of the function $\varepsilon_{\mathrm{con,\,F}}$. In practice, we will use $p = 1$, 2 or ∞. Also, since we will only consider the case of a periodic surface for our numerical study, it will be enough to concentrate on the unit cell and the more appropriate definition is then

$$\|\varepsilon_{\mathrm{con,\,F}}\|_p = \begin{cases} \left(\dfrac{1}{a_c} \displaystyle\int_{a_c} |\varepsilon_{\mathrm{con,\,F}}(\mathbf{x})|^p \, \mathrm{d}^2x \right)^{1/p}, & p \in [1, \infty[\\ \displaystyle\sup_{\mathbf{x} \in a_c} |\varepsilon_{\mathrm{con,\,F}}(\mathbf{x})|, & p = \infty \end{cases} , \tag{7.5}$$

where a_c denotes both the unit cell of the periodic lattice and its area.

As it stands now, the definition of the integrated consistency errors is somewhat of an absolute measure that is not bounded hence making the interpretation of its value and comparison for different surfaces or simulation parameters difficult. For example, we could be in a situation where the field is locally enhanced in a small region of the surface and for a fixed relative deviation of the boundary conditions the region where the field is enhanced would contribute more to the integrated consistency errors than a region of low field. We will therefore work with the *relative integrated consistency errors* defined as

$$\|\varepsilon_{\mathrm{con,\,F}}\|_{p,r} = \frac{\|\varepsilon_{\mathrm{con,\,F}}\|_p}{\| |\mathbf{n} \times \tilde{\mathbf{F}}_1| \|_p + \| |\mathbf{n} \times \tilde{\mathbf{F}}_2| \|_p} = \frac{\| |\mathbf{n} \times \tilde{\mathbf{F}}_1 - \mathbf{n} \times \tilde{\mathbf{F}}_2| \|_p}{\| |\mathbf{n} \times \tilde{\mathbf{F}}_1| \|_p + \| |\mathbf{n} \times \tilde{\mathbf{F}}_2| \|_p} , \qquad p \in [1, \infty] . \tag{7.6}$$

With this definition, the triangular inequality for the L^p-norm ensures unity as an upper bound for the relative integrated consistency errors, hence

$$0 \;\leq\; \|\varepsilon_{\mathrm{con,\,F}}\|_{p,r} \;\leq\; 1 . \tag{7.7}$$

A particularity of this definition, which can be seen both as a drawback or an advantage, is the non-linearity of the scaling of the relative integrated consistency errors. On the one hand, one may not appreciate well enough that values close to unity corresponds essentially to huge errors. One the other hand, as surface parameters are varied, it will therefore be of interest to see how the relative integrated consistency error undergoes a transition from values close to 0 to values close to 1 within a small region of parameter space. We may expect sharp step-like variations around critical values of the surface parameters, which may be helpful to conjecture

criteria for the validity of the reduced Rayleigh equations.

Note that this definition for the relative integrated consistency errors has an obvious practical limitation when dealing with scattering from a perfect electric conductor. Indeed, in that case, the electric field inside the perfect conductor vanishes identically and so does the corresponding contribution $|||\,\mathbf{n}\times\tilde{\mathbf{E}}_2\,|||_p$. Consequently, the previous definition of the relative integrated consistency errors becomes useless in the sense that any error profile would be normalized by itself hence giving a unit relative integrated error. In order to avoid this issue, we modify our definition as

$$\|\varepsilon_{\mathrm{con,\,E}}\|_{p,r} = \frac{|||\,\mathbf{n}\times\tilde{\mathbf{E}}_1\,|||_p}{|||\,\mathbf{n}\times\tilde{\mathbf{E}}_1\,|||_p + 1}\,, \qquad p\in[1,\infty],\ \text{for a perfect conductor.} \tag{7.8}$$

The interpretation of this definition is a shift of all electric fields by 1 (meaning an overall shift by the amplitude of the incident field since $\mathbf{E}_1$ is normalized by the incident field) both above and below the surface and then the use of the aforementioned definition. The error is thus given with respect to a reference uniform error having the incident field amplitude. In other words, when the absolute error is small compared to the incident field, $|||\,\mathbf{n}\times\tilde{\mathbf{E}}_1\,|||_p \ll 1$, the relative error is small, $\|\varepsilon_{\mathrm{con,\,E}}\|_{p,r} \ll 1$, whereas for an absolute error much larger than the incident field, $|||\,\mathbf{n}\times\tilde{\mathbf{E}}_1\,|||_p \gg 1$, the relative error is close to unity, i.e. $\|\varepsilon_{\mathrm{con,\,E}}\|_{p,r} \approx 1$. The relative error $\|\varepsilon_{\mathrm{con,\,E}}\|_{p,r} = 1/2$ corresponds to the threshold $|||\,\mathbf{n}\times\tilde{\mathbf{E}}_1\,|||_p = 1$, which means that the absolute error is of the order of the incident field amplitude in the sense of the L^p-norm.

Remark 7.1. Note that contrarily to what the notation may suggest, $\|\cdot\|_{p,r}$ is not a norm. Indeed, the axiom of homogeneity is obviously not satisfied.

Remark 7.2. The consistency errors could in principle depend on the incident field, both on the angle of incidence and polarization. In the following, we will only consider normal incidence for the sake of simplicity and the consistency errors for incidence p or s incident polarization will be indicated by a superscript, e.g. $\varepsilon^{(p)}_{\mathrm{con,F}}$ or $\varepsilon^{(s)}_{\mathrm{con,F}}$.

7.3.2 Near-field

The estimation of the consistency errors requires the knowledge of the near-field, or to be more accurate, the knowledge of the surface limit of the electromagnetic field on both sides of the interface. The near-field it-self will also be of interest for applications other than computing the consistency errors. We thus expose here how the near-field can be computed from the reflection and transmission amplitudes, $\mathbf{R}\,(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}})$ and $\mathbf{T}\,(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}})$. For this, it suffices to recall the field expansion Eq. (2.2) taken at the surface and substitute the definition of the reflection and transmission amplitudes Eqs. (2.49) and (2.50). The fields are then evaluated along the surface.

Remark 7.3. Numerically the expansion Eq. (1.31) will be replace by a numerical integral where the wave vectors are sampled for an arbitrary rough interface or a finite sum for a periodic interface where the set of the wave vectors is countable (see Chapter 4) but cut off. The resolution or cut off in the wave vector space will induce an error by-itself compared to what the exact solution of the reduced Rayleigh equations would be. Thus the computed consistency errors will contain both an error term due to the numerical approximations and the theoretical consistency error. Therefore one has to be careful with the interpretation of the consistency errors, and to first study how the error evolves by increasing the resolution parameters of the numerical

method so that when this error stabilizes (if it converges to a constant error) one can reasonably assume that one is left with the pure consistency error. In other words, one should first verify the the numerical solution has converged with respect to the number of mode (see Chapter 5).

7.4 Probing numerically the validity of the RRE

7.4.1 Perfect electric conductor

We start our numerical investigation into the validity of the reduced Rayleigh equations by treating the case of a perfect electric conductor. Although we have chosen in the remaining of this thesis to only discuss the scattering by systems made of penetrable media (dielectrics, metals) there exists an equation which is the sister of the reduced Rayleigh equations we have studied so far for penetrable media, which is often called simply the Rayleigh equation and reads [42]

$$\int \mathcal{J}^{+}(\mathbf{p}\,|\,\mathbf{q})\;\mathbf{M}(\mathbf{p}\,|\,\mathbf{q})\;\mathbf{R}(\mathbf{q}\,|\,\mathbf{p}_0)\;\frac{\mathrm{d}^2q}{(2\pi)^2} = -\,\mathcal{J}^{-}(\mathbf{p}\,|\,\mathbf{p}_0)\,\mathbf{M}(\mathbf{p}\,|\,\mathbf{p}_0) \tag{7.9a}$$

$$\mathcal{J}^{\pm}(\mathbf{p}\,|\,\mathbf{q}) \stackrel{\text{def}}{=} \frac{1}{\alpha(\mathbf{q})}\int \exp\big[-i(\mathbf{p}-\mathbf{q})\cdot\mathbf{x} \mp i\alpha(\mathbf{q})\zeta(\mathbf{x})\big]\,\mathrm{d}^2x \tag{7.9b}$$

$$\mathbf{M}(\mathbf{p}\,|\,\mathbf{q}) \stackrel{\text{def}}{=} \begin{pmatrix} |\,\mathbf{p}\,||\,\mathbf{q}\,| - \frac{\omega^2}{c^2}\;\hat{\mathbf{p}}\cdot\hat{\mathbf{q}} & \frac{\omega}{c}\alpha(\mathbf{q})\,[\hat{\mathbf{p}}\times\hat{\mathbf{q}}]\cdot\hat{\mathbf{e}}_3 \\ \frac{\omega^2}{c^2}\,[\hat{\mathbf{p}}\times\hat{\mathbf{q}}]\cdot\hat{\mathbf{e}}_3 & \frac{\omega}{c}\,\alpha(\mathbf{q})\;\hat{\mathbf{p}}\cdot\hat{\mathbf{q}} \end{pmatrix}. \tag{7.9c}$$

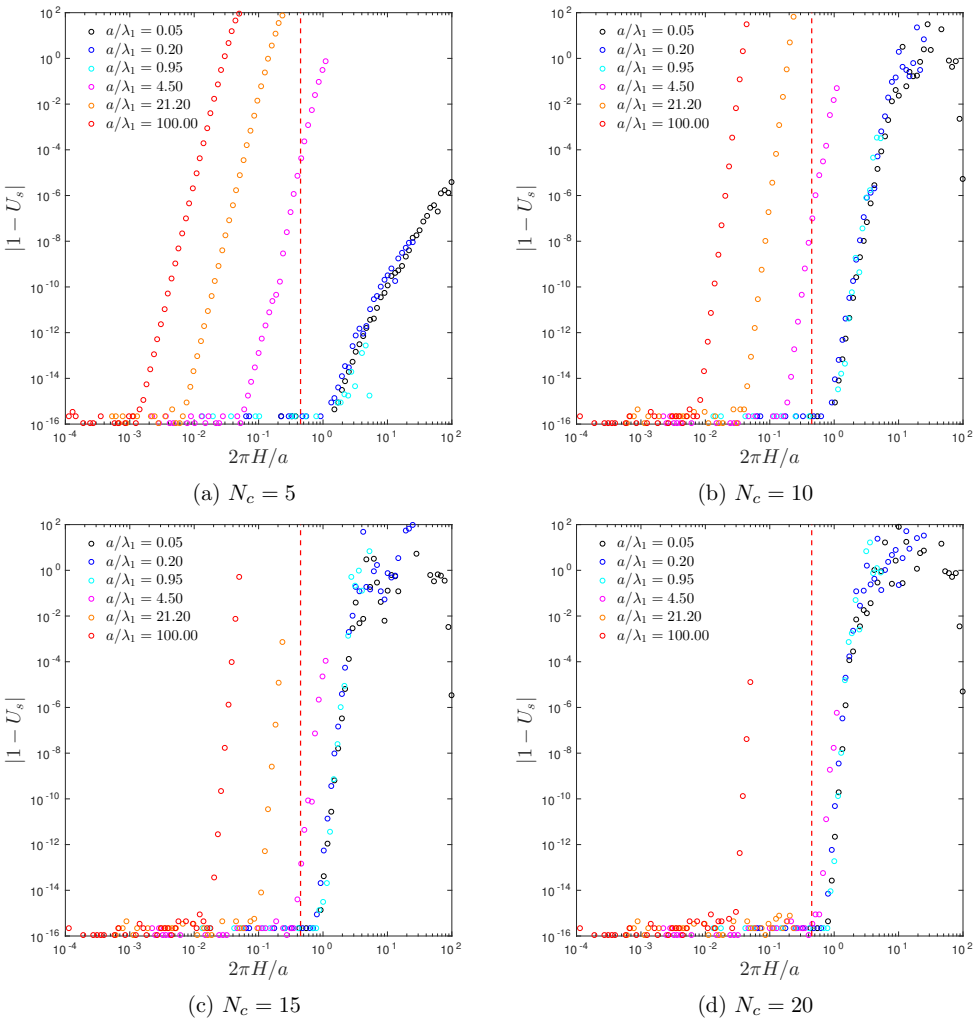
The term *reduced* is omitted for perfect electric conductor since there is no transmitted field, so there is no distinction to be made between coupled and decoupled equations. Note that vacuum is assumed here as the medium is incidence. Adjusting for a dielectric medium is straightforward by using similar scaling laws as derived in Section 5.1.

Remark 7.4. An interesting feature of this equation, is that it is rather simple to derive from scratch (see e.g. [42]) but it seems to be a rather difficult task to obtain it as a limit of the reduced Rayleigh equations as $\epsilon_2 \to -\infty$. This would be a interesting analysis problem to resolve, which to our knowledge has not been achieved yet.

The reason for studying the case of a perfect electric conductor here is that there exists a mathematical theorem which validates the Rayleigh hypothesis in the case of a one-dimensional sinusoidal surface whose maximum slope does not exceed a certain threshold. This is Millar's theorem [54, 55] that we have already discussed in Section 7.1. We will see that we can recover reasonably well Millar's theorem numerically and that some nuances should be made when it comes to interpreting the boundary conditions, as already suggested by Millar himself in [55] and that one should always remember that the hypotheses of a theorem are as important as its implications.

We will thus begin by studying the case of a one dimensional sinusoidal surface, $x_3 = H\sin(2\pi x_1/a)$ and discuss the results in light of Millar's theorem before treating the case of a two-dimensional sinusoidal surface $x_3 = H[\sin(2\pi x_1/a) + \sin(2\pi x_2/a)]/2$, and finally we will consider the case of a non-differential profile (a saw-tooth profile).

One-dimensional sinusoidal surface



(a) $N_c = 5$ (b) $N_c = 10$

(c) $N_c = 15$ (d) $N_c = 20$

Figure 7.1: Diagram of the deviation from unitarity, $|1 - U_s|$ as a function of the maximum slope $2\pi H/a$ of a one-dimensional sinusoidal perfect electric conducting surface and for different values of the reduced lattice constant. The index cutoff is (a) $N_c = 5$, (b) $N_c = 10$, (c) $N_c = 15$, and (d) $N_c = 20$. The red dashed line is the line of equation $2\pi H/a = 0.448$.

On unitarity – Consider the deviation from unity of U_s obtained for a set of one-dimensional surfaces with various amplitudes and periods. Figure 7.1 shows the deviation from unity of U_s, $|1 - U_s|$, as a function of the maximum slope of the surface $s = 2\pi H/a$, and this for different reduced period. The subfigures Figs. 7.1(a)-(d) correspond to different values of the cutoff index N_c chosen for the simulation, in order to illustrate how the unitarity converges as N_c is increased. Focusing on Fig. 7.1(a), the first observation one can make is that for surfaces having a small enough slope, unitarity can be satisfied to machine precision with a rather

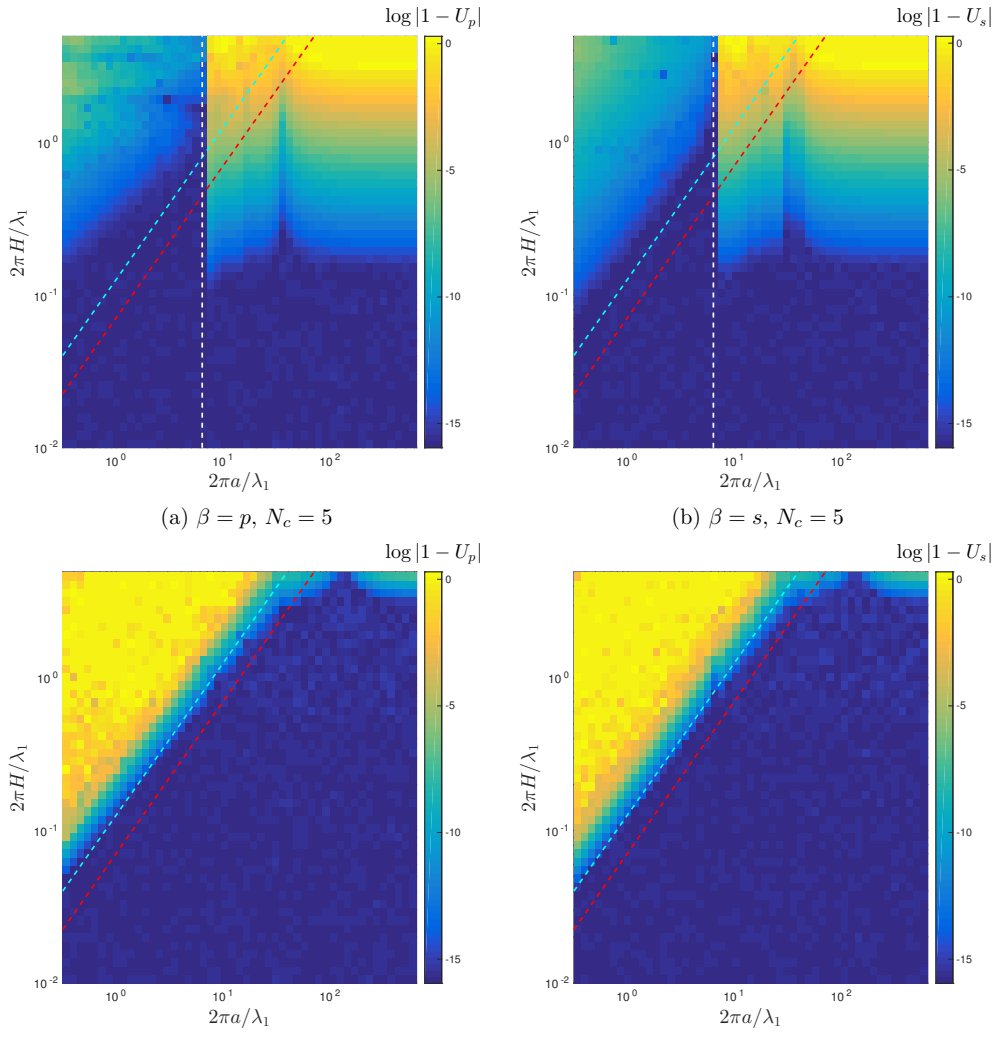
low value of the cutoff index. This is not surprising since in the limit of a planar surface $N_c = 0$ would suffice to give the exact result. As the slope is increased, the curves for the different reduced periods are well separated for a/λ_1 large enough, typically $a/\lambda_1 \gg 1$, while for $a/\lambda_1 \ll 1$ the curves overlap. Moreover, there seems to be a slope threshold for each value of a/λ_1 above which $|1 - U_s|$ increases monotonically with the slope in a power law fashion. These observations can be interpreted as follows. First, in the long wavelength regime, only one propagating mode is allowed and hence carries away the totality of the power. This means that the other modes taken into account in the simulation are evanescent and thus, a small cutoff in index space $|\boldsymbol{\ell}| \leq N_c$, may correspond to a large cutoff in k-space. In other words, one could expect a faster convergence in the long wavelength regime $a/\lambda_1 \ll 1$ than in the short wavelength regime $a/\lambda_1 \gg 1$. When N_c is increased, the slope threshold corresponding to the different values of a/λ_1 in the short wavelength regime shift towards larger slope values but the deviation of $|1 - U_s|$ above the threshold increases more and more rapidly with N_c (see Figs. 7.1)(a-d)). This change of increasing rate also occurs in the long wavelength regime, although the threshold slope remains stable. In the limit for $N_c \to \infty$, we can conjecture that all cases of a/λ_1 collapse to a unique curve as can be observed for $N_c = 20$ in Fig. 7.1(d). The value of a/λ_1 essentially dictates how far one must go with N_c before reaching this curve. Consequently, when convergence is obtained, there exists a slope threshold, independent of the wavelength, below which unitarity is satisfied (here up to machine precision), and above which the deviation from unitarity increases very quickly. This sharp transition seems to occur close to the Millar's threshold for the validity of the Rayleigh hypothesis for a one-dimensional perfectly conducting sinusoidal surface as indicated by the red vertical dashed lines in Fig. 7.1. The logarithmic scale is not ideal to appreciate whether the transition occurs at Millar's slope $s = 0.448$ or at a slightly larger value. We have found that for the values $a/\lambda_1 \ll 1$, the threshold seems to rather be at $s \approx 0.8$. Whether this threshold will move towards $s = 0.448$ by further increasing N_c is still an open question, but it does not seem to be the case by looking at simulations we have achieved for the largest value of N_c.

Under the assumption that the slope threshold obtained numerically here is indeed larger than that obtained by Millar, how should this result be interpreted. One may say, Millar's theorem must be questioned, the numerical experiment proves it wrong and we can in fact obtain correct results for slopes larger than 0.448. However, one must keep in mind that, a theorem is a theorem, hence it is mathematically true if its proof is irrefutable. To the author's knowledge, no-one has ever questioned Millar's proof, which should be taken as an indication of its validity. Is the numerical experiment incorrect then? No, not necessarily. The answer we suggest here to resolve this apparent paradox is the following: both Millar's theorem and the numerical experiment are correct, they are simply not answering the same questions. First of all, we have solely measured the deviation of U_s from unity. Unitarity being satisfied does not necessary mean that the result is correct in the sense that the fields solve the scattering problem. Second, even if the efficiencies are assumed to be corrected, this only means that the far-field is correct and nothing can be said about the near-field a priori. Millar's theorem deals with the validity of the Rayleigh hypothesis in the sense that the plane wave expansion is valid not only away from the surface but also in the grooves. This second observation raises a particularly interesting question. Could a method obtain a correct far-field with a wrong near-field? Third, it must be recalled that a key working hypothesis in Millar's theorem is that the boundary conditions should be satisfied point-wise at all points of the surface. As briefly mentioned in the derivation of the reduced Rayleigh equations in Chapter 2, the reduced Rayleigh equations do not assume that the boundary conditions should be satisfied point-wise but rather *weakly*, i.e. that we

consider the boundary condition as tested against plane waves by integration over the surface. Relaxing the boundary conditions may be the source of this apparent enhancement of validity of the Rayleigh hypothesis. Millar himself pointed out that this could in principle be the case if one considered that the boundary conditions should be satisfied for the L^2-norm for example instead of point-wise. This last point then even question the previous question: what is meant by a wrong near-field then? What type of boundary condition should the near-field satisfy to be physical?

Remark 7.5. We have presented results for an incident s-polarized wave but similar results are observed for an incident p-polarized wave. This can be seen in the diagrams of $|1 - U_\beta|$ in the $(2\pi H/\lambda_1, 2\pi a/\lambda_1)$-plane in Fig. 7.2 for $N_c = 5$ and for $N_c = 20$. We can observe the two different regimes of convergence for a/λ_1 being either larger or lower than unity. This is reminiscent of the slope vs amplitude dominated regimes described in Chapter 5 for the convergence rate. By comparing the diagrams of Fig. 7.2 for $N_c = 5$ and $N_c = 20$, we understand the importance of verifying first the convergence before studying the unitarity. Indeed, assuming one contents oneself with $N_c = 5$, one would be tempted to accept a solution as correct in the upper left corner (i.e. large amplitude and small period) where $|1 - U_\beta|$ is relatively low and discard a solution in the upper right corner (i.e. large amplitude and large period) where $|1 - U_\beta|$ is large. Such a practice would lead to conclusion in complete contradiction to what one would obtain by considering larger values of N_c as can be seen for $N_c = 20$. This observation raises an interesting question: Knowing that a solution obtained in the upper left corner of the diagram will diverge with increasing N_c, but which satisfies reasonably well the condition of unitarity for small values of N_c, can the solution obtained for a low truncation cutoff be considered as a good approximation of the exact solution for the far-field? This resemble a typical behavior of divergent asymptotic series which, although they are divergent, can yield good approximation to a function when limited to a small finite number of terms. If this is the case, how do we find a criterion for finding the optimal value of the index cutoff N_c? These are questions that may be partly answered by studying the consistency errors in the next section, but a more detailed study with a comparison with other methods should be encouraged.

On consistency and boundary conditions – It is now time to study the relative integrated consistency error, which for short we will simply designate as consistency error when the context is clear. We repeat the numerical experiment which we have exposed for the deviation from unitarity, with the same set of parameters but we now measure the consistency error. Figures 7.3(a) and (b) show the consistency error in the L^∞-norm for an incident s-polarized wave, $\|\varepsilon^{(s)}_{\mathrm{con,\,E}}\|_{\infty,r}$, as a function of the maximum slope of the profile $s = 2\pi H/a$. Several values of the reduced period are considered and the results were obtained with either a rather low index cutoff, $N_c = 5$ (a), or a relative large one $N_c = 20$ (b). Comparing Figs. 7.3(a) and (b), one can observe, as already pointed out during the study of unitarity, that more modes are required before obtaining reasonably well converged results in the regime $a/\lambda_1 > 1$ than in the regime $a/\lambda_1 < 1$. All the s-shaped curves exhibit a sharp variation from negligible error to maximal error around a critical slope s. As N_c is increased, all the error curves collapse to a single characteristic curve for which the transition occurs around, if not exactly at, the slope $s = 0.448$, i.e. Millar's threshold (see Fig. 7.3(b) and Millar's critical slope indicated by the vertical red dashed line). This is an interesting feature. This numerical experiment based on the consistency error that we have defined in this work seems to be a convincing illustration of Millar's result. Note that, as pointed out earlier, the fact that one obtains Millar's threshold here was not necessarily to be expected since it is not obvious whether the weak boundary condition of the Rayleigh equation would be equivalent to the point-wise boundary condition. Nevertheless,



(a) $\beta = p$, $N_c = 5$

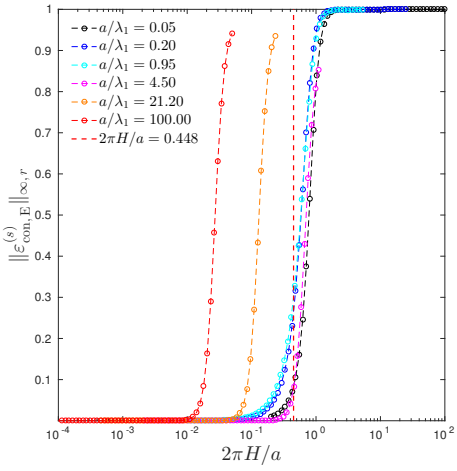
(b) $\beta = s$, $N_c = 5$

(c) $\beta = p$, $N_c = 20$

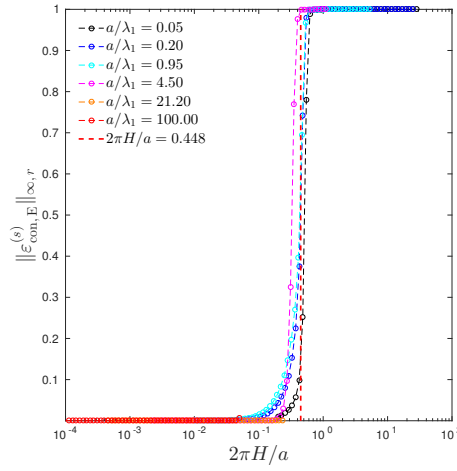
(d) $\beta = s$, $N_c = 20$

Figure 7.2: Diagram of deviation from unitarity, $\log|1 - U_\beta|$ in the $(2\pi a/\lambda_1, 2\pi H/\lambda_1)$-plane for a one-dimensional perfectly conducting sinusoidal surface obtained with $N_c = 5$ and $N_c = 20$. The red (resp. cyan) dashed line is the line of equation $2\pi H/a = 0.448$ (resp. $2\pi H/a = 0.8$). The white vertical dashed line is the line of equation $a/\lambda_1 = 1$. Note: U_β is saturated to 2 for clarity.

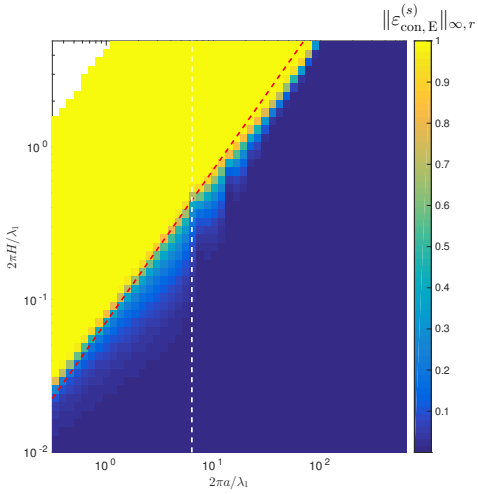
the numerical experiment seems to indicate that Millar's result holds for the Rayleigh equation, and this sheds some light on some of the questions raised after our study of the unitarity. We have demonstrated numerically here that *the satisfaction of unitarity does not necessarily mean the satisfaction of the boundary conditions*. We have indeed seen previously that unitarity holds



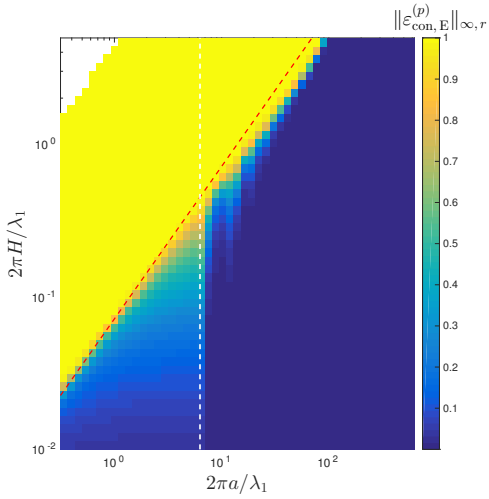
(a) $\beta = s$, $N_c = 5$



(b) $\beta = s$, $N_c = 20$



(c) $\beta = s$, $N_c = 20$



(d) $\beta = p$, $N_c = 20$

Figure 7.3: (a,b) Relative integrated consistency error, $\|\varepsilon^{(s)}_{\mathrm{con,E}}\|_{\infty,r}$, as a function of the maximum slope $s = 2\pi H/a$ for different values of a/λ_1 obtained for (a) $N_c = 5$ and (b) $N_c = 20$. (c,d) Diagrams of relative integrated consistency error, $\|\varepsilon^{(\beta)}_{\mathrm{con,E}}\|_{\infty,r}$, in the $(2\pi a/\lambda_1, 2\pi H/\lambda_1)$-plane for a one-dimensional perfectly conducting sinusoidal surface obtained with $N_c = 20$. The red dashed line is the line of equation $2\pi H/a = 0.448$. The white vertical dashed line is the line of equation $a/\lambda_1 = 1$.

for slopes up to about twice Millar's critical slope while the boundary conditions would not be satisfied according to our calculations. To our knowledge, this is a new result which illustrates well the necessity for having better check than unitarity to assess the quality of a simulation.

Furthermore, we would like to point out a detail which we have not resolved yet. The attentive reader may have noticed that the transition region, at the left of the threshold in Fig. 7.3(b), is not an abrupt one. There is a smooth increase of error before the sharp jump. This is particularly visible for p-polarized incident waves as illustrated in the consistency error diagram in Fig. 7.3(d) compared to s-polarized light in Fig. 7.3(c). This smooth slow transition occurs in the long wavelength regime, $a/\lambda_1 < 1$, and we can see that a residual error is present even below Millar's threshold. The residual error vanishes when the reduced amplitude tends to zero but we have experienced that it does not seem to vanish when more modes are taken into account. We do not understand at this stage the origin of this residual error. Note that the largest residual error occurs for $a/\lambda_1 \approx 1$ (see cyan curve in Fig. 7.3(b)). This means that the first diffracted modes $\ell_1 = \pm 1$ are at the limit of propagation and evanescence. We have seen in Section 1.4.2 for a metallic surface, that the in-plane component of the wave vector of a surface plasmon polariton has a slightly larger norm than $\sqrt{\epsilon_1}\,\omega/c$. In the limit $\epsilon_2 \to -\infty$, $|\,\mathbf{p}_{\mathrm{SPP}}\,|$, given by Eq. (1.98), approaches $\sqrt{\epsilon_1}\,\omega/c$. This observation could indicate that the Rayleigh equation has issues representing correctly surface plasmon polariton or modes which would resemble surface plasmon polariton for a perfectly conducting surface. This is, however, only a speculation and we would like to stress that we are not certain at this stage what is the origin of the residual error.

Two-dimensional sinusoidal surface

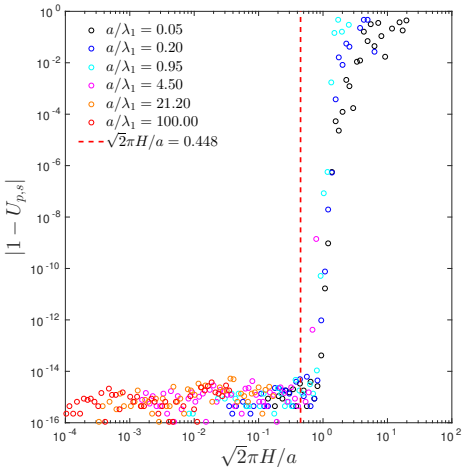
We now repeat our numerical experiment for a two-dimensional perfectly conducting sinusoidal surface of equation

$$x_3 = \frac{H}{2}\left[\sin\left(\frac{2\pi x_1}{a}\right) + \sin\left(\frac{2\pi x_2}{a}\right)\right]. \tag{7.10}$$

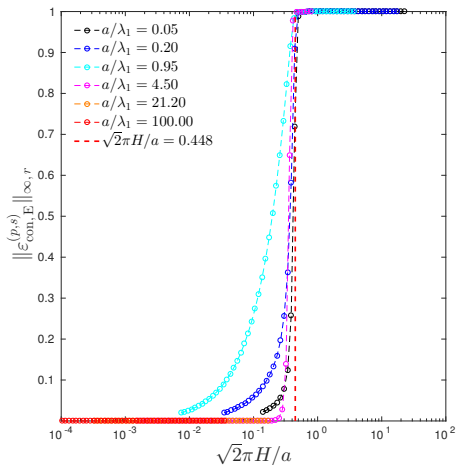
Since the results are very similar to those obtained for a one-dimensional surface, we skip the discussion about the convergence with N_c and consider only the case $N_c = 20$, i.e. large enough for our purpose. Note that the maximum slope for this two-dimensional surface is $s = \sqrt{2}\pi H/a$ and not $2\pi H/a$ as for the one-dimensional surface, and that due to the symmetry of the problem the incident s and p-polarized wave will yield the same results (at normal incidence). The unitarity and consistency error are hence the same for s and p-polarized waves and the corresponding subscript or superscript will then be denoted p, s here. As can be seen from Fig. 7.4, all the observations made for the one-dimensional sinusoidal surface holds for the two-dimensional one. In particular, it is remarkable that the slope threshold $s = 0.448$ still holds for the consistency error and that the threshold slope of the unitarity is still higher than the one for the consistency error. The fact that Millar's slope threshold holds independently of the dimension is not an obvious feature one may *a priori* expect, but it seems true. Not that for the two-dimensional surface, both consistency errors for s and p incident wave exhibit a significant residual error for $a/\lambda_1 < 1$. Indeed, now s and p polarizations have symmetric role at normal incidence. It is worth noting that the residual error is somewhat larger in the two-dimensional case than in the one-dimensional case as can be seen from Fig. 7.4(b) for $a/\lambda_1 \approx 1$.

7.4.2 Penetrable systems

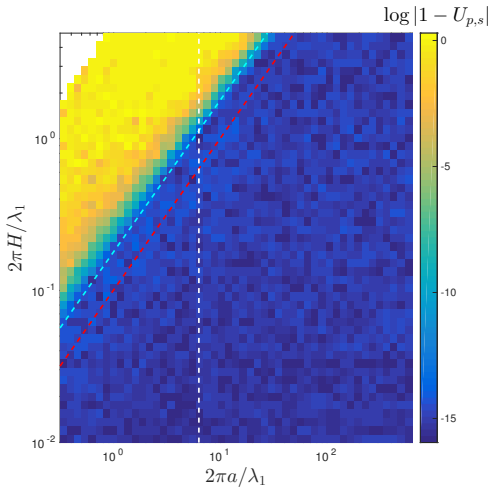
Now that we are familiar with the case of a perfect electric conductor, let us consider the case of an interface between two dielectric media. We repeat similar numerical experiments



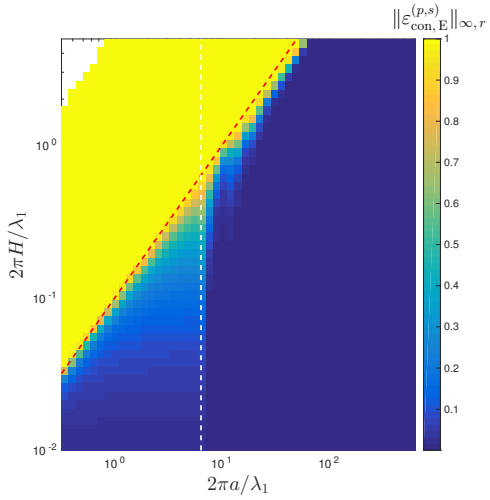
(a) Unitarity, $N_c = 20$



(b) Consistency error, $N_c = 20$

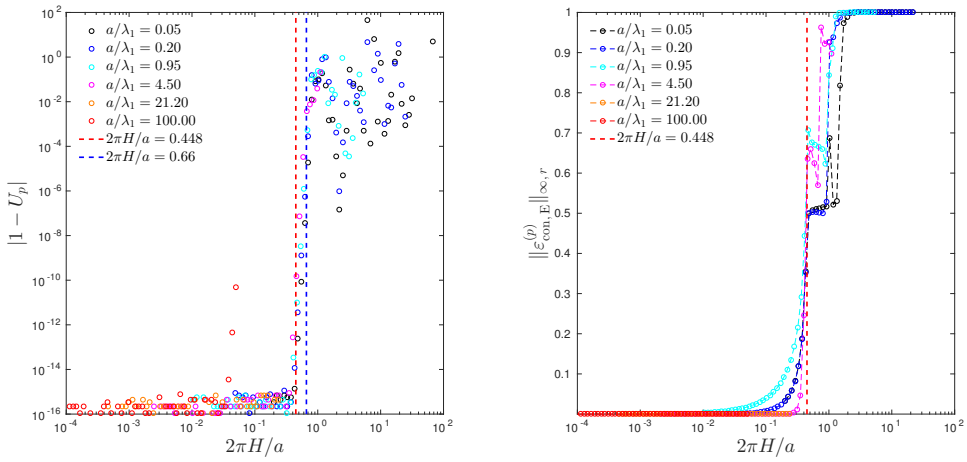


(c) Unitarity, $N_c = 20$



(d) Consistency error, $N_c = 20$

Figure 7.4: (a) Deviation from unitarity, $|1 - U_{p,s}|$, and (b) relative integrated consistency error, $\|\varepsilon^{(p,s)}_{\mathrm{con,E}}\|_{\infty,r}$, as a function of the maximum slope $s = \sqrt{2}\pi H/a$ for different values of a/λ_1. Diagrams of (c) $\log|1 - U_{p,s}|$, and (d) relative integrated consistency error, $\|\varepsilon^{(p,s)}_{\mathrm{con,E}}\|_{\infty,r}$, in the $(2\pi a/\lambda_1, 2\pi H/\lambda_1)$-plane. All results presented in these figures were obtained for a two-dimensional perfectly conducting sinusoidal surface obtained with $N_c = 20$. The red (resp. cyan) dashed line is the line of equation $\sqrt{2}\pi H/a = 0.448$, (resp. $\sqrt{2}\pi H/a = 0.8$). The white vertical dashed line is the line of equation $a/\lambda_1 = 1$. Note that both p and s-polarized light give the same results here due to the symmetry of the system.



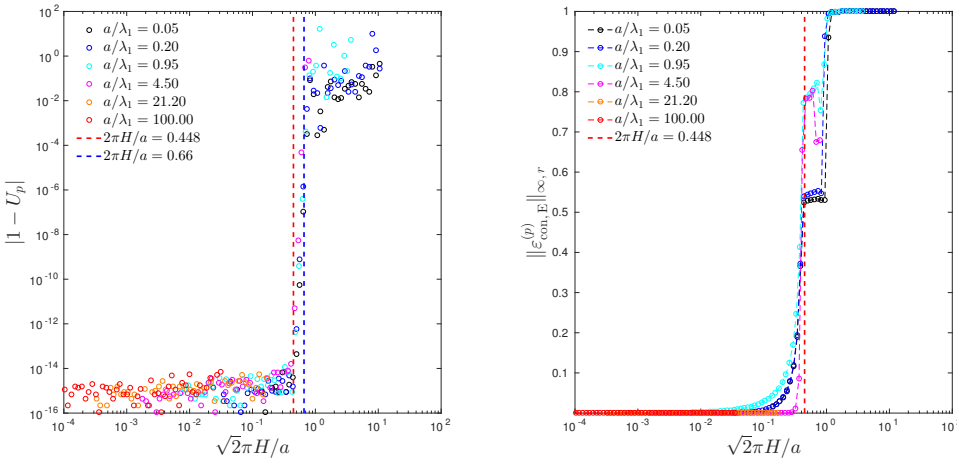
(a) (b)

Figure 7.5: (a) Deviation from unitarity $|1 - U_p|$ and (b) relative integrated consistency error $\|\varepsilon^{(p)}_{\mathrm{con,E}}\|_{\infty,r}$ as functions of the maximum slope of a one-dimensional sinusoidal surface $2\pi H/a$ for different reduced lattice constants. The dielectric contrast was held fixed at $z^2 = 2.25$ and the index radius cutoff was $N_c = 20$.

as the one achieved for a perfect electric conductor in the previous section but for a dielectric contrast $z^2 = 2.25$. Figures 7.5 and 7.6 present the deviation from unitarity, $|1 - U_p|$, and the relative consistency error, $\|\varepsilon^{(p)}_{\mathrm{con,E}}\|_{\infty,r}$, for a one- and two-dimensional sinusoidal profiles as a function of the maximum slope, and this for several values of a/λ_1. It is remarkable to observe that the Millar's slope threshold still seems to hold for penetrable sinusoidal profile (see Figs. 7.5(b) and 7.6(b)). It is not a priori obvious that the exact same threshold $s = 0.448$ should hold independently of the dielectric contrast. Indeed, Millar's criterion was derived under the assumption of a one-dimensional sinusoidal perfectly conducting profile. We now see that that this criterion can be extended to two-dimensional profiles too and for dielectric systems. At least this is what we conclude from the numerical experiments. The deviation from unitarity, Figs. 7.5(a) and Figs. 7.6(a), seems to also roughly have the same slope threshold as the consistency error in contrast to the case of a perfect electric conductor. By closer inspection, we estimate this threshold to be rather between $s = 0.448$ and $s \approx 0.66$.

7.4.3 Local consistency error and asymptotic amplitude decay

Surface map of local consistency error – We have studied in previous sections the integrated consistency error as a function of the geometrical parameters of the surface. It is instructive to have a closer look at the local consistency error in order to get insights into what may be at the origin of the non-satisfaction of the boundary conditions for some parameters. To fix the ideas, we will only consider a two-dimensional sinusoidal surface between two dielectric media here, with contrast $z^2 = 2.25$. Consider first the three cases depicted in Fig. 7.7. Figure 7.7(a) is a contour plot of the height of the surface within a unit cell. Note that the length scales are



(a) (b)

Figure 7.6: (a) Deviation from unitarity $|1 - U_p|$ and (b) relative integrated consistency error $\|\varepsilon^{(p)}_{\mathrm{con,E}}\|_{\infty,r}$ as functions of the maximum slope of a two-dimensional sinusoidal surface $\sqrt{2}\pi H/a$ for different reduced lattice constants. The dielectric contrast was held fixed at $z^2 = 2.25$ and the index radius cutoff was $N_c = 20$.

given in units of the period a. The maximum slope of the surface is fixed to be $\sqrt{2}\pi H/a = 0.1$, in such a way that we are below Millar's slope threshold. Figures 7.7(b)-(d) show the local consistency error, $\varepsilon^{(p)}_{\mathrm{con,E}}$, for three regimes of a/λ_1, namely, $a/\lambda_1 = 0.1$, $a/\lambda_1 = 1$, and $a/\lambda_1 = 10$ respectively. Figures 7.7(b) and 7.7(c) present similar local error maps, with maximum values of the order of 10^{-2} and we observe that the maxima of the local error occur both at the highest and lowest point on the surface. Figure 7.7(d), however, is very different from Figs. 7.7(b) and 7.7(c). The local error is insignificant, of the order of 10^{-13}, although it also has a tendency to be highest near the extrema of the surface profile. Note the high spatial frequency pattern of the local error in Fig. 7.7(d) compared to Fig. 7.7(b) and 7.7(c). The fact that the case presented in Fig. 7.7(b) and 7.7(c) have a non-negligible error is reminiscent of what we have observed for the integrated consistency error in the regime $a/\lambda_1 < 1$. We have seen that in this regime, there seems to exist a residual error. It seems that this residual error is due to a component of the field of a rather low spatial frequency which, for a reason which remains unclear to the author, is not compensated for by higher spatial frequency terms. In the regime $a/\lambda_1 > 1$, however, as one add more and more modes, i.e. terms of higher and higher spatial frequency in the plane wave expansion, the local error becomes smaller and smaller. At a finite mode cutoff, the error is then of the order of the amplitudes of the last few terms kept in the expansion. This interpretation is supported by the fact that the local error pattern observed in Fig. 7.7(d) is of high spatial frequency. We have now illustrated the difference between the long and short wavelength regimes when the maximum slope of the surface is below the critical threshold. How does the local error looks like beyond the threshold? From the study of the integrated consistency error as the cutoff index, N_c, is increased, we expect that the error probably grows without bound since the relative integrated consistency error goes towards unity. Let us observe this phenomenon locally as the

index cutoff increased. We consider a system characterized by $a/\lambda_1 = 5$ and $\sqrt{2}\pi H/a = 0.5$, i.e. just beyond Millar's threshold. Note that for $a/\lambda_1 = 5$, if the slope would be less than Millar's slope, the local error would converge to 0, i.e. without residual error, and this uniformly as illustrated in Fig. 7.7(d). Figures 7.8 (a)-(c) present the map within the unit cell of the local consistency error in a logarithmic scale respectively for $N_c = 10$, $N_c = 20$ and $N_c = 30$. For $N_c = 10$, we can see that the local error is concentrated around the surface extrema and reach a value of about 40 ($\log \varepsilon^{(p)}_{\mathrm{con,\,E}} \approx 1.6$, see Fig. 7.8 (a)). As N_c is increased to $N_c = 20$ and $N_c = 30$, the local error seems to increase exponentially with N_c, as $\log \varepsilon^{(p)}_{\mathrm{con,\,E}}(N_c = 20) \approx 3.7$ ($\varepsilon^{(p)}_{\mathrm{con,\,E}} \approx 4.5 \times 10^3$) and $\log \varepsilon^{(p)}_{\mathrm{con,\,E}}(N_c = 30) \approx 5.3$ ($\varepsilon^{(p)}_{\mathrm{con,\,E}} \approx 2.1 \times 10^5$). The error blow-up is then unambiguous. Furthermore, note that as N_c is increased the local error pattern is of higher and higher spatial frequency. This seems to indicate that the mismatch of the boundary conditions is due to components of the field on at least one side of the surface which are of high spatial frequency.

Remark 7.6. For the simulations achieved in Fig. 7.8, we have observed that the unitarity converges to 1 as N_c increases and this to machine precision, even though the boundary conditions are clearly violated.

Let us explain the behaviors observed in Fig. 7.8 from a more mathematical point of view. Consider an evanescent mode of the form $\mathbf{E}_1^{(\boldsymbol{\ell})}(\mathbf{p}_0)\,\exp(i\,\mathbf{k}_1^+(\mathbf{p}_{\boldsymbol{\ell}})\cdot\mathbf{r})$ which is defined above the interface. Without loss of generality, let us assume for simplicity that $\mathbf{p}_0 = \mathbf{0}$, and that we consider $\boldsymbol{\ell} = (\ell_1, 0)$ so that we have $\mathbf{p}_{\boldsymbol{\ell}} = \mathbf{G}^{(\ell_1,0)} = \frac{2\pi\ell_1}{a}\,\hat{\mathbf{e}}_1$. The mode then reduces to

$$\mathbf{E}_1^{(\boldsymbol{\ell})}(\mathbf{p}_0)\,\exp(i\,\mathbf{k}_1^+(\mathbf{p}_{\boldsymbol{\ell}})\cdot\mathbf{r}) = \mathbf{E}_1^{(\ell_1,0)}(\mathbf{p}_0)\,\exp(2\pi i\ell_1 x_1/a)\exp\left[-2\pi|\ell_1|\left(1-\frac{a^2}{\lambda_1^2\ell_1^2}\right)^{1/2} x_3/a\right] \tag{7.11}$$

$$\sim \mathbf{E}_1^{(\ell_1,0)}(\mathbf{p}_0)\,\exp(2\pi i\ell_1\,x_1/a)\exp(-2\pi|\ell_1|\,x_3/a) \tag{7.12}$$

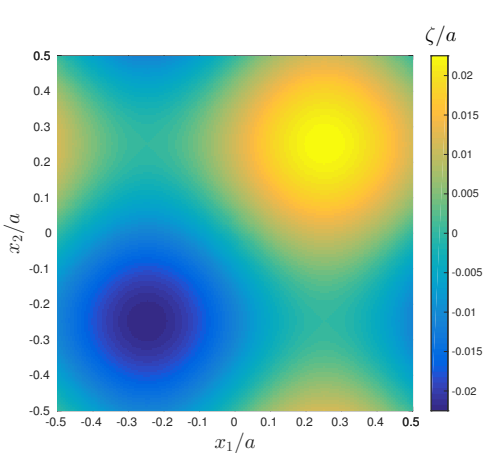
where we have used that for an evanescent mode

$$\alpha_1(\mathbf{G}^{(\ell_1,0)}) = i\sqrt{|\,\mathbf{G}^{(\ell_1,0)}\,|^2 - \epsilon_1\frac{\omega^2}{c^2}} = i\frac{2\pi|\ell_1|}{a}\sqrt{1-\frac{a^2}{\lambda_1^2\ell_1^2}} \sim i\frac{2\pi|\ell_1|}{a} \tag{7.13}$$

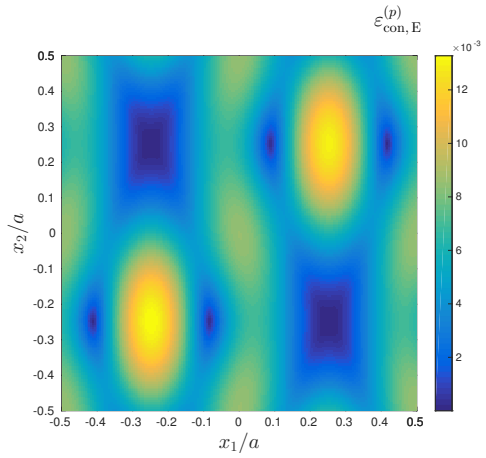
when $|\ell_1| \to \infty$. Similarly a similar evanescent mode below the interface reads

$$\mathbf{E}_2^{(\boldsymbol{\ell})}(\mathbf{p}_0)\,\exp(i\,\mathbf{k}_2^-(\mathbf{p}_{\boldsymbol{\ell}})\cdot\mathbf{r}) \sim \mathbf{E}_2^{(\ell_1,0)}(\mathbf{p}_0)\,\exp(2\pi i\ell_1\,x_1/a)\exp(2\pi|\ell_1|\,x_3/a)\,. \tag{7.14}$$

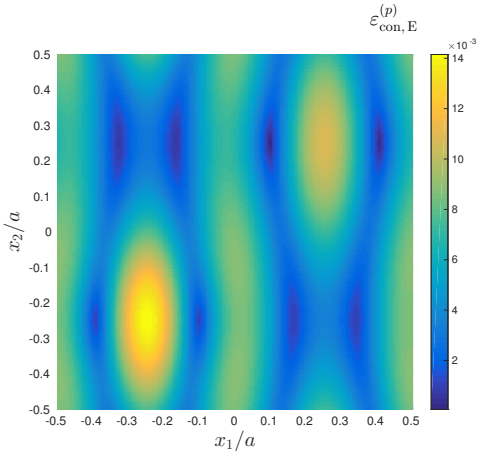
We verify that both modes, indeed decays exponentially respectively when $x_3 \to \pm\infty$. Let us analyze now what happens near an extremum of the surface profile, say at a minimum for which $\zeta = -H$ when $|\ell_1|$ increases. As $|\ell_1|$ increases, the x_3-dependent factor for the mode $\mathbf{E}_1^{(\boldsymbol{\ell})}(\mathbf{p}_0)$ *increases* exponentially inside the groove as $\exp(2\pi|\ell_1|\,H/a)$, whereas the x_3-dependent factor for the mode $\mathbf{E}_2^{(\boldsymbol{\ell})}(\mathbf{p}_0)$ *decreases* exponentially as $\exp(-2\pi|\ell_1|\,H/a)$. In other words, as more and more evanescent modes are added, the x_3-factor of each added mode becomes larger than the preceding in medium 1 near the surface minimum while, the x_3-factor of each added term becomes smaller than the preceding in medium 2. The opposite effect is of course valid near a maximum of the surface. How can then the field on both sides of the interface satisfy the boundary conditions? The only way to obtain finite fields near the surface as more modes are summed up, is by ensuring that the amplitude of the modes decay sufficiently fast. As a rule of thumb, they should decay with $|\ell_1|$ at least as fast as the x_3-dependent factor taken at an



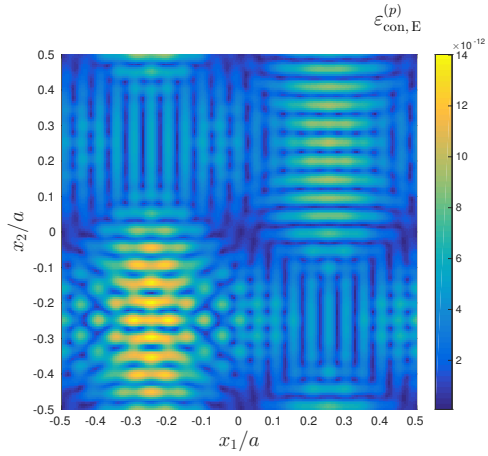
(a) Surface profile



(b) Local consistency error, $a/\lambda_1 = 0.1$



(c) Local consistency error, $a/\lambda_1 = 1$



(d) Local consistency error, $a/\lambda_1 = 10$

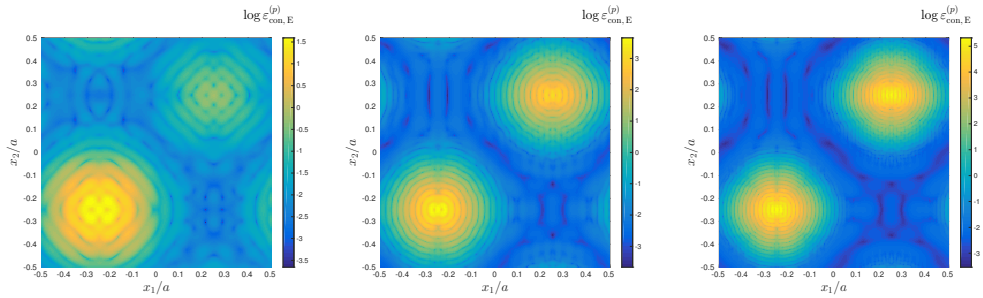
Figure 7.7: (a) A unit cell of the two-dimensional sinusoidal surface profile in units of a, characterized by a maximum slope of $\sqrt{2}\pi H/a = 0.1$. Surface maps of local consistency error $\varepsilon^{(p)}_{\mathrm{con,E}}$ for (b) $a/\lambda_1 = 0.1$, (c) $a/\lambda_1 = 1$ and (d) $a/\lambda_1 = 10$. The results were obtained for a dielectric contrast fixed to $z^2 = 2.25$. Note the scale of error between (b,c) and (d).

extrema of the surface grows. We then must have a criterion for the decay of the amplitude which reads

$$\mathbf{E}_j^{(\boldsymbol{\ell})}(\mathbf{p}_0) = \mathcal{O}\Big(\exp(-\kappa\,|\ell_1|)\Big) \tag{7.15}$$

as $|\ell_1| \to \infty$, with $\kappa = 2\pi\|\zeta\|_\infty/a$ being a critical decay rate for the amplitude.

Asymptotic decay of the reflection and transmission amplitude – Based on the above local analysis, we now understand why in Fig. 7.8 (i) the local error was localized near the extrema



(a) $N_c = 10$ (b) $N_c = 20$ (c) $N_c = 30$

Figure 7.8: Surface maps of local consistency error $\log \varepsilon^{(p)}_{\text{con, E}}$ for $a/\lambda_1 = 5$, obtained for (a) $N_c = 10$, (b) $N_c = 20$ and (c) $N_c = 30$. The results were obtained for two-dimensional surface characterized by a maximum slope of $\sqrt{2}\pi H/a = 0.5$ and the dielectric contrast is fixed to $z^2 = 2.25$.

of the surface, (ii) why the local error grows exponentially with N_c since mode with larger $|\boldsymbol{\ell}|$ are taken into account, and (iii) why the pattern of error becomes contains more and more high spatial frequency in the (x_1, x_2)-plane since the modes which dominates are the one for $|\boldsymbol{\ell}| = N_c$ and then the x_1-dependence of the modes read $\exp(2\pi i \ell_1 x_1/a)$ which oscillates more and more rapidly as $|\ell_1|$ is increased and similarly for x_2. Consequently, the fundamental difference between convergence and divergence of the near field resides in the decay rate the reflection and transmission amplitudes for large mode index. If the amplitudes decay fast enough (i.e. according to Eq. (7.15)) the near-field is likely to converge, whereas if they do not decay fast enough, the near-field will necessarily blow up. Let us verify this claim by plotting, in Fig. 7.9, the modulus of the reflection amplitudes as a function of ℓ_1 in the cases where the maximum slope of the surface is below Millar's threshold (Fig. 7.9(a)), at Millar's threshold (Fig. 7.9(b)) and above it (Fig. 7.9(c)), and compare the decay scaling with Eq. (7.15). Figures 7.9 (a) to (c) indeed confirm what was intuited from the local analysis. For $\sqrt{2}\pi H/a < 0.448$ the decay of the reflection amplitude is faster than $\exp(-\kappa|\ell_1|)$ (Fig. 7.9(a)), equally fast to $\exp(-\kappa|\ell_1|)$ for $\sqrt{2}\pi H/a = 0.448$ (Fig. 7.9(b)) and slower than $\exp(-\kappa|\ell_1|)$ for $\sqrt{2}\pi H/a > 0.448$ (Fig. 7.9(c)). Figure 7.9 hence validate the local analysis which gives a rather physical or at least mathematical understanding of the issue of convergence of the near-field.

What decides then the decay rate of the amplitudes? The amplitudes are solutions of the reduced Rayleigh equations for a given surface profile. So the surface profile dictates the decay of the amplitudes obtained from the reduced Rayleigh equations, via the scalar kernel factor $\mathcal{J}^{ba}_{lm}$. Since we are interested in the scaling of amplitudes for large $|\boldsymbol{\ell}|$ it seems reasonable to think that this scaling is related to that of $\mathcal{J}^{ba}_{lm}(\mathbf{p}_{\boldsymbol{\ell}}\,|\,\mathbf{p}_{\mathbf{m}})$ for somewhat large $|\boldsymbol{\ell}|$ and/or $|\mathbf{m}|$. In other words, the questions are the following. How does the scaling of $\mathcal{J}^{ba}_{lm}(\mathbf{p}_{\boldsymbol{\ell}}\,|\,\mathbf{p}_{\mathbf{m}})$ at infinity influences that of $\mathbf{R}^{(\boldsymbol{\ell})}(\mathbf{p}_0)$ at infinity (i.e. for large $|\boldsymbol{\ell}|$)? Is there a simple link between the behavior of $\mathcal{J}^{ba}_{lm}(\mathbf{p}_{\boldsymbol{\ell}}\,|\,\mathbf{p}_{\mathbf{m}})$ at infinity and geometrical parameters of the surface profile? Can we deduce a simple criterion on the surface profile for ensuring the convergence of the near-field and hence the validity of the reduced Rayleigh equations?

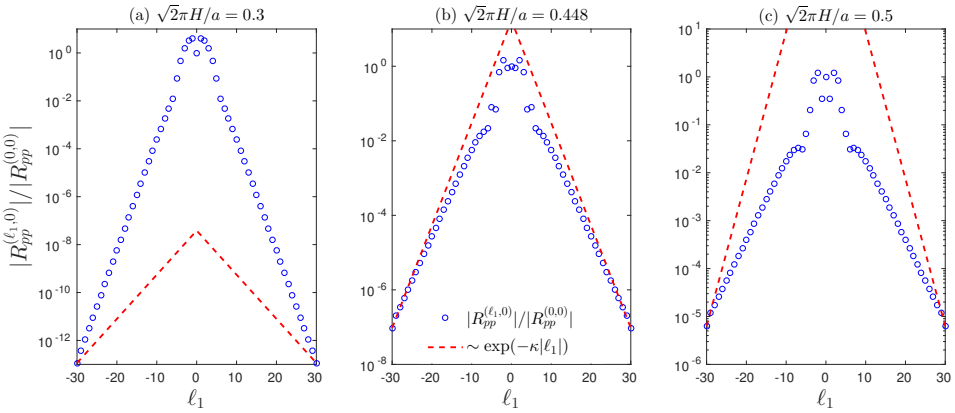


Figure 7.9: Modulus of the normalized reflection amplitude $|R_{pp}^{(\ell_1,0)}|/|R_{pp}^{(0,0)}|$ as a function of ℓ_1 for a maximum slope of (a) $\sqrt{2}\pi H/a = 0.3$, (b) $\sqrt{2}\pi H/a = 0.448$, and (c) $\sqrt{2}\pi H/a = 0.5$. The red dashed line is proportional to $\exp(-\kappa|\ell_1|)$, with $\kappa = 2\pi H/a$, and such that it coincides with the last data point at $\ell_1 = 30$ for a clear comparison of the decay rates. The dielectric contrast is fixed to $z^2 = 2.25$ and $a/\lambda_1 = 5$.

7.4.4 Towards a rigorous mathematical analysis?

Let us summarize the conclusions we have drawn from our numerical experiments and the local analysis for *sinusoidal surfaces*:

- There is a maximal surface slope threshold under which the solution of the reduced Rayleigh equations satisfies the boundary conditions.
- The value of the slope threshold is independent of whether a perfect electrical conductor or a dielectric system is considered and is equal to 0.448, value which is predicted by Millar's theorem for a one-dimensional sinusoidal perfectly conducting surface.
- Millar's threshold holds both for one- and two-dimensional surfaces.
- The local analysis relates the satisfaction of the boundary conditions with the decay rate of the amplitudes, namely for a one-dimensional surface, the amplitudes should decay faster than $\exp(-2\pi H|\ell_1|/a)$ as $|\ell_1| \to \infty$.
- The decay rate of the amplitude is necessarily linked to the scalar kernel (and right-hand side) factor $\mathcal{J}_{lm}^{ba}(\mathbf{p}_{\boldsymbol{\ell}} \,|\, \mathbf{p}_{\mathbf{m}})$, most probably its behavior at infinity. Hence by finding the link between the behavior of $\mathcal{J}_{lm}^{ba}(\mathbf{p}_{\boldsymbol{\ell}} \,|\, \mathbf{p}_{\mathbf{m}})$ at infinity and that of the amplitudes, one may hope to find a simple geometrical criterion for the validity of the reduced Rayleigh equations. This criterion should lead to $2\pi H/a < 0.448$ for a one-dimensional sinusoidal surface.

We will now give the ideas for a mathematical proof of these conjectures. We believe that a general rigorous mathematical proof may be rather technical in terms of analysis. Therefore we content ourselves with the case of a one-dimensional sinusoidal surface between two dielectrics and we will make some rough assumptions to keep what we believe are the key elements in

understanding Millar's threshold without requiring complex analysis but only very simple ideas.

Scalar wave approximation – Instead of working with electromagnetic waves, we simplify the analysis by considering scalar waves. This simplification is unlikely to change the end result for a simple reason. We know that to go from electromagnetic wave to scalar wave, it is sufficient to replace the polarization coupling matrices $\mathbf{M}_{lm}^{ba}$ by some scalar factor, which is a constant in the case of scalar wave subjected to the continuity of the field and its normal derivative at the boundary. Since the matrices $\mathbf{M}_{lm}^{ba}$ are essentially matrices of change of basis, the norm of this matrix can be bounded from above with some constant, so it does not play a role in the scaling. Alternatively, if one is uncomfortable with the scalar wave assumption, then one may regard our analysis as being that of an incident s-polarized wave. Thus we consider the following reduced Rayleigh equation for the scalar amplitude

$$\sum_{\mathbf{m}\in\mathbb{Z}^2} \tilde{\mathcal{J}}_{21}^{++}(\mathbf{p}_{\boldsymbol{\ell}}\,|\,\mathbf{p}_{\mathbf{m}})\, R^{(\mathbf{m})}(\mathbf{p}_0) = -\,\tilde{\mathcal{J}}_{21}^{+-}(\mathbf{p}_{\boldsymbol{\ell}}\,|\,\mathbf{p}_0)\,. \tag{7.16}$$

For the sake of simplicity, we will choose to work for normal incidence $\mathbf{p}_0 = \mathbf{0}$ and since we consider a one-dimensional sinusoidal surface, only the modes of the form $\boldsymbol{\ell} = (\ell_1, 0)$ remain. So the above equation reduces to

$$\sum_{m_1\in\mathbb{Z}} \tilde{\mathcal{J}}_{21}^{++}(\mathbf{p}_{\ell_1}\,|\,\mathbf{p}_{m_1})\, R^{(m_1)}(\mathbf{p}_0) = -\,\tilde{\mathcal{J}}_{21}^{+-}(\mathbf{p}_{\ell_1}\,|\,\mathbf{p}_0)\,, \tag{7.17}$$

where we have simplified the notations to keep only the first mode index.

Diagonal approximation – Now comes an approximation which may be hard to justify, or may be not be valid at all. We assume that for ℓ_1 large enough the diagonal element $\tilde{\mathcal{J}}_{21}^{++}(\mathbf{p}_{\ell_1}\,|\,\mathbf{p}_{\ell_1})$ dominates with respect to all the other row elements $\tilde{\mathcal{J}}_{21}^{++}(\mathbf{p}_{\ell_1}\,|\,\mathbf{p}_{m_1})$. In other word, we assume that the matrix $(\tilde{\mathcal{J}}_{21}^{++}(\mathbf{p}_{\ell_1}\,|\,\mathbf{p}_{m_1}))_{\ell_1,m_1\in\mathbb{N}}$ is roughly diagonal after a certain row index $|\ell_1|$. Within this assumption, for a large enough value of $|\ell_1|$ we have simply

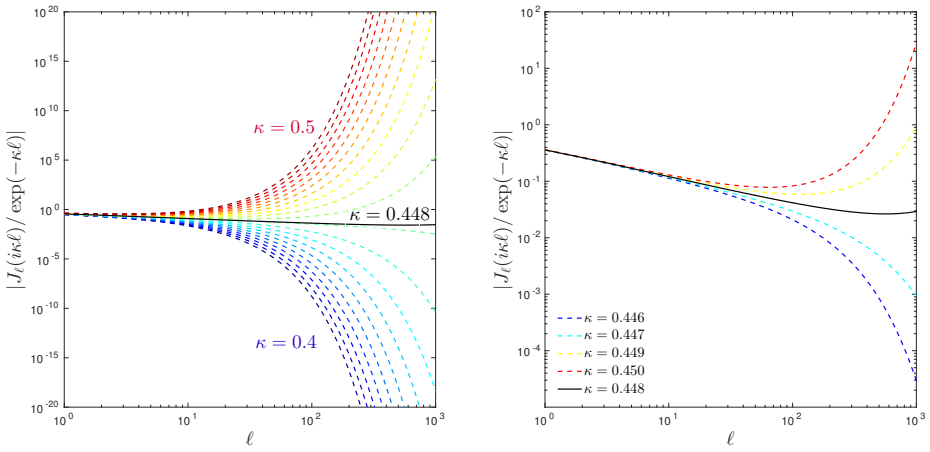
$$R^{(\ell_1)}(\mathbf{p}_0) \approx -\frac{\tilde{\mathcal{J}}_{21}^{+-}(\mathbf{p}_{\ell_1}\,|\,\mathbf{p}_0)}{\tilde{\mathcal{J}}_{21}^{++}(\mathbf{p}_{\ell_1}\,|\,\mathbf{p}_{\ell_1})}\,. \tag{7.18}$$

So under this assumption, we can now see the link between the scaling of $R^{(\ell_1)}(\mathbf{p}_0)$ and the scaling of $\tilde{\mathcal{J}}_{21}^{ba}$ as $|\ell_1| \to \infty$. So, according to the local analysis, a general criterion for the validity of the reduced Rayleigh equations for a scalar one-dimensional periodic surface under non-conical incidence may read

$$\left|\frac{\tilde{\mathcal{J}}_{21}^{+-}(\mathbf{p}_{\ell_1}\,|\,\mathbf{p}_0)}{\tilde{\mathcal{J}}_{21}^{++}(\mathbf{p}_{\ell_1}\,|\,\mathbf{p}_{\ell_1})}\right| = o\Big(\exp(-\kappa|\ell_1|)\Big)\,, \tag{7.19}$$

as $|\ell_1| \to \infty$ where we recall that $\kappa = 2\pi\|\zeta\|_\infty/a$. Let us further expand the criterion in the case of a sinusoidal surface under normal incidence. In this case we have (see Eq. (A.21))

$$\left|\frac{\tilde{\mathcal{J}}_{21}^{+-}(\mathbf{p}_{\ell_1}\,|\,\mathbf{0})}{\tilde{\mathcal{J}}_{21}^{++}(\mathbf{p}_{\ell_1}\,|\,\mathbf{p}_{\ell_1})}\right| = \left|\frac{\alpha_2(\mathbf{p}_{\ell_1}) - \alpha_1(\mathbf{p}_{\ell_1})}{\alpha_2(\mathbf{p}_{\ell_1}) + \alpha_1(\mathbf{0})}\right| \left|\frac{J_{\ell_1}\Big[\big(\alpha_2(\mathbf{p}_{\ell_1}) + \alpha_1(\mathbf{0})\big)H\Big]}{J_0\Big[\big(\alpha_2(\mathbf{p}_{\ell_1}) - \alpha_1(\mathbf{p}_{\ell_1})\big)H\Big]}\right|\,, \tag{7.20}$$



(a) (b)

Figure 7.10: Ratio $|J_\ell(i\kappa|\ell|)/\exp(-\kappa|\ell|)|$ as a function of ℓ for different values of κ. (a) The values of κ are taken evenly spaced between 0.4 (blue lower curve) to 0.5 (red upper curve) in steps of $\Delta\kappa \approx 0.005$. (b) Values of κ in the vicinity of $\kappa^\star = 0.448$.

where J_ν is the Bessel function of the first kind of order ν. When $|\ell_1| \to \infty$ we have the following expansion of $\alpha_1(\mathbf{p}_{\ell_1})$ and $\alpha_2(\mathbf{p}_{\ell_1})$,

$$\alpha_1(\mathbf{p}_{\ell_1}) = i\,\frac{2\pi|\ell_1|}{a}\left(1-\frac{a^2}{\lambda_1^2\ell_1^2}\right)^{1/2} = i\frac{2\pi|\ell_1|}{a}\left(1-\frac{a^2}{2\lambda_1^2\ell_1^2}+o(\ell_1^{-2})\right) \tag{7.21a}$$

$$\alpha_2(\mathbf{p}_{\ell_1}) = i\,\frac{2\pi|\ell_1|}{a}\left(1-\frac{z^2a^2}{\lambda_1^2\ell_1^2}\right)^{1/2} = i\frac{2\pi|\ell_1|}{a}\left(1-\frac{z^2a^2}{2\lambda_1^2\ell_1^2}+o(\ell_1^{-2})\right). \tag{7.21b}$$

Hence we have

$$\alpha_2(\mathbf{p}_{\ell_1}) - \alpha_1(\mathbf{p}_{\ell_1}) = i\frac{\pi a\,(1-z^2)}{\lambda_1^2|\ell_1|} + o(\ell_1^{-1}) \xrightarrow[|\ell_1|\to\infty]{} 0 \tag{7.22a}$$

$$\alpha_2(\mathbf{p}_{\ell_1}) + \alpha_1(\mathbf{0}) = \frac{2\pi}{\lambda_1} + i\frac{2\pi}{a}\,|\ell_1| + o(\ell_1^{-1})\,. \tag{7.22b}$$

Consequently the first factor in the right-hand side of Eq. (7.20) decays to zero as $|\ell_1|^{-2}$ and it is therefore not this factor which controls the expect exponential decay of the reflection amplitude. We have to analyze the ratio of Bessel functions. The denominator of the ratio of Bessel functions in the right-hand side of Eq. (7.20) reads

$$J_0\Big[\big(\alpha_2(\mathbf{p}_{\ell_1}) - \alpha_1(\mathbf{p}_{\ell_1})\big)H\Big] = J_0\Big[i\frac{\pi aH\,(1-z^2)}{\lambda_1^2|\ell_1|} + o(\ell_1^{-1})\Big] \xrightarrow[|\ell_1|\to\infty]{} 1\,. \tag{7.23}$$

The denominator hence tends to unity when $|\ell_1|$ tends to infinity. It then remains to analyze the numerator.

$$J_{\ell_1}\Big[\big(\alpha_2(\mathbf{p}_{\ell_1}) + \alpha_1(\mathbf{0})\big)H\Big] = J_{\ell_1}\left[\frac{2\pi H}{\lambda_1} + i\frac{2\pi H}{a}\,|\ell_1| + o(\ell_1^{-1})\right]. \tag{7.24}$$

How does the above term behave as $|\ell_1| \to \infty$? This is not a quite simple question to answer for two reasons. First the argument of the Bessel function is complex, and second the imaginary part of the argument is proportional to $|\ell_1|$, which up to a sign, is also the order of the Bessel function. For large enough $|\ell_1|$, the argument will essentially be $\sim i|\ell_1|2\pi H/a$. So it is sufficient for our purpose to study $J_{\ell_1}(i\kappa|\ell_1|)$ (note that $\kappa = 2\pi H/a$ for a sinusoidal surface with our definition of H) and in particular for which values of κ is $J_{\ell_1}(i\kappa|\ell_1|) = o\Big(\exp(-\kappa|\ell_1|)\Big)$. Instead of doing the analysis analytically, which would be tedious, we content ourselves by studying numerically the ratio $|J_{\ell_1}(i\kappa|\ell_1|)/\exp(-\kappa|\ell_1|)|$ for different values κ, and estimate the threshold value $\kappa^\star$ below which $|J_{\ell_1}(i\kappa|\ell_1|)/\exp(-\kappa|\ell_1|)| \to 0$ for large ℓ_1 and above which $|J_{\ell_1}(i\kappa|\ell_1|)/\exp(-\kappa|\ell_1|)| \to \infty$. Figures 7.10 illustrate this ratio as a function ℓ_1 for $\ell_1 > 0$ for different values of κ. We recover indeed that $\kappa^\star = 0.448$, i.e. Millar's threshold. We believe that this illustrate well the relevance of the presented analysis and we hope it will lead to a rigorous proof in a near future.

Remark 7.7. We would like to make a remark which we believe to be of particular interest. We have seen repeatedly in Chapters 5, 6, that there seemed to be two regimes of convergence rate when solving the reduced Rayleigh equations. The amplitude dominate regime and the slope dominate regime. Equation (7.24) gives us a remarkably clear expression of this fact. Indeed, the argument in the Bessel function essential contains the sum of two terms, $2\pi H/\lambda_1$ and $i|\ell_1|2\pi H/a$, i.e. the sum of the reduced amplitude and a term proportional to the characteristic slope of the surface. The two terms are exactly the parameters which we have identified numerically as being the one relevant for the study of the convergence rate. We believe that the fact that these two parameters appear so clearly in our analysis is no coincidence. Depending on the relative values of $2\pi H/\lambda_1$ and $2\pi H/a$, the behaviors of $J_{\ell_1}\Big[\big(\alpha_2(\mathbf{p}_{\ell_1}) + \alpha_1(\mathbf{0})\big)H\Big]$ for somewhat intermediate $|\ell_1|$ will be different. Indeed, if $2\pi H/\lambda_1 \gg 2\pi H/a$ i.e. $a/\lambda_1 \gg 1$ (which we recognized also as a threshold between two regimes of consistency errors previously), $J_{\ell_1}\Big[\big(\alpha_2(\mathbf{p}_{\ell_1}) + \alpha_1(\mathbf{0})\big)H\Big]$ behaves roughly as $J_{\ell_1}\left[\frac{2\pi H}{\lambda_1}\right]$. On the contrary if $a/\lambda_1 \ll 1$, $J_{\ell_1}\Big[\big(\alpha_2(\mathbf{p}_{\ell_1}) + \alpha_1(\mathbf{0})\big)H\Big]$ behaves roughly as $J_{\ell_1}\left[i|\ell_1|2\pi H/a\right]$. There are two fundamental difference here between the two cases. In the first case, the argument of the Bessel function is real and independent of ℓ_1, while in the second case it is imaginary and proportional to $|\ell_1|$. Whether the argument is real or imaginary can be dramatically different behavior for a Bessel function, changing from an oscillating, converging function to 0, to quickly converging or diverging one. We will not analyze this more in detail, but we believe such a careful analysis could lead to understanding unambiguously the different regimes of convergence identified previously.

7.4.5 Non-sinusoidal profile

So far we have based our numerical study and analysis on the case of sinusoidal profiles. The choice of a sinusoidal profile was motivated by the fact that theoretical results such as Millar's theorem are known and the condition for the validity of the Rayleigh hypothesis takes a simple form, namely a threshold for the maximum slope on the surface. In addition, the analytical expression of the $\mathcal{J}$-integral in terms of Bessel functions makes the analysis relatively simple. The next steps in this study would be to consider other types of surface profiles, in particular non-differentiable profiles. According to Millar's theorem, the Rayleigh hypothesis should not hold for non-differentiable profiles due to the fact that the profile function is not analytic. Nevertheless, the problem considered by Millar was that of a field satisfying the boundary conditions on every points on the surface. For weaker types of boundary conditions, it would be interesting to see whether the Rayleigh hypothesis yields satisfactory results also for non-differentiable profiles.

We have some preliminary numerical results partly answering this question but we choose not to expose them here as more analysis remains to be done. These results may become part of a future research article on the numerical analysis of the reduced Rayleigh equations. More than the purely theoretical aspect of the question of the use of the reduced Rayleigh equations, its application side is important. Indeed, one may be interested in simulated the response of a surface profile which as been triangulated, or of surfaces which physically exhibits pyramidal shapes for examples.

7.5 Summary

The present chapter may be considered as one of the most crucial in this thesis. It aims at answering the delicate question of the range of validity of the reduced Rayleigh equations, which is intrinsically related to the so-called Rayleigh hypothesis on which the method is built on. We have first given an historical review of the different results known concerning the range of validity of the Rayleigh hypothesis by stressing the different conclusions drawn by the different authors and trying to resolve the seemingly contradictory claims. Then we have demonstrated numerically, with an exhaustive parameter study, for the case of a sinusoidal perfectly conducting surface and for dielectric surfaces, that: (i) convergent results may be obtained with satisfaction of energy conservation for a characteristic slope larger than that given by Millar's theorem, (ii) within the convergent results only those under Millar's threshold indeed satisfy the boundary conditions and this to machine precision[3]. These results resolve the contradictions between the different claims in the literature. In fact, the results obtained by the different authors are all consistent; only their interpretation was subjected to the confusion between a convergent solution and a consistent or correct solution. As a theoretical backup of our numerical analysis, these results were supported by a simple asymptotic analysis of the near-field where we have recovered exactly Millar's criterion analytically by looking for the condition on the asymptotic scaling of the scattering amplitudes for large in-plane wave vectors which leads to a finite field everywhere inside the grooves. We believe our analysis clarifies most of the confusion one may get by reading the literature on the Rayleigh hypothesis and gives elements for assessing the validity of the hypothesis for any type of surface. The results presented in this chapter will form the basis of a devoted paper not included in this thesis.

[3]Apart from the region of parameter space where a residual error was observed which do not yet understand.

Chapter 8

Approximation methods for randomly rough surfaces

"Never tell me the odds!" – Han Solo
Star Wars, The Empire Strikes Back (1980).

In this chapter, the focus will be on developing a few approximation methods for obtaining solutions to the problem of scattering by randomly rough surfaces, or to be more accurate, to *ensemble averaged observables* of such scattering experiments. Indeed, for randomly rough surfaces, one is often primarily interested in the statistical properties of the light scattered by a set of randomly drawn surface realizations sharing common statistical properties before studying the speckle pattern from the scattering by a single surface realization. We want to precise though, that we do *not* claim that the study of speckle patterns is of little interest. In fact, the analysis of speckle patterns from a single surface can also be of great interest and reveal information about the surface, especially via the analysis of speckle correlation (see e.g. [79, 80]). The study of the probability density of speckle intensity for scattering by a disordered medium is also known to follow a Rayleigh probability distribution for unpolarized light and a modified Rayleigh distribution under some assumptions (see e.g. [81, 82]). Recently, a theoretical study supported by extensive numerical simulation has revealed the correlation or mutual information between the reflected and transmitted speckle intensities in the case of waves scattered in a disordered medium in a multiple scattering regimes [83]. Here, we simply choose to limit ourselves to some observables. To be more specific, we will be interested in the coherent and incoherent components of the mean differential reflection and transmission coefficients as defined in Eqs. (1.123a) and (1.123b). We recall that, the coherent and incoherent components of the mean differential reflection coefficient are given as a limit when the surface covered by the randomly rough surface goes to infinity of

$$\left\langle \frac{\partial R^{(S)}_{\alpha\beta}}{\partial \Omega_r}(\mathbf{p} \mid \mathbf{p}_0) \right\rangle_{\text{coh}} = \frac{\epsilon_1 \left(\frac{\omega}{2\pi c}\right)^2 \cos^2\theta_r}{S \cos\theta_0} \left| \left\langle R^{(S)}_{\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0) \right\rangle \right|^2$$

$$\left\langle \frac{\partial R^{(S)}_{\alpha\beta}}{\partial \Omega_r}(\mathbf{p} \mid \mathbf{p}_0) \right\rangle_{\text{incoh}} = \frac{\epsilon_1 \left(\frac{\omega}{2\pi c}\right)^2 \cos^2\theta_r}{S \cos\theta_0} \left[\left\langle |R^{(S)}_{\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0)|^2 \right\rangle - \left| \left\langle R^{(S)}_{\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0) \right\rangle \right|^2 \right],$$

and with similar definitions for the mean differential transmission coefficient. To fix the notations, we are interested in surface profiles that are realizations of a continuous, differentiable,

stationary stochastic process. If not otherwise specified, we assume in this chapter that the system under consideration is composed of two media separated by a single randomly rough interface. The stochastic process will then be assumed to have a zero mean, i.e. that the average surface is the plane $x_3 = 0$. Furthermore, to fully specify the process we need to define a height probability density and an auto-correlation function. For simplicity, all the examples treated hereafter will have a *Gaussian height probability density*. The important properties that we will then need are summarized by

$$\langle \zeta(\mathbf{x}) \rangle = 0 \,, \tag{8.2a}$$

$$\big\langle \zeta(\mathbf{x}) \zeta(\mathbf{x}') \big\rangle = \sigma^2 \, W(\mathbf{x}, \mathbf{x}') \,, \tag{8.2b}$$

where $\sigma > 0$ denotes the rms-roughness of the surface profile and W is the two-point auto-correlation function.

8.1 Direct Monte Carlo method

From the definition of the coherent and incoherent components of the mean differential scattering coefficients, it is clear that one needs to estimate the average of the scattering amplitudes and the average of its modulus squared. The most straightforward strategy that should come to mind would be (i) to generate realizations of the stochastic process with specified properties given by Eq. (8.2) over a surface area S, (ii) solve the reduced Rayleigh equations for each generated surface, and (iii) to gather statistics on the scattering amplitudes. If the characteristic length L of the surface is large enough compared to both the wavelength of the incident light and the characteristic correlation length of the surface profile, and that the number of generated surfaces N_{surf} is large enough, we can assume that Eqs. (1.123a) and (1.123b) give good estimates of the coherent and incoherent components of the mean differential scattering coefficients. To be more accurate, if we index the surface realizations $(\zeta_w)_{1 \leq w \leq N_{\text{surf}}}$, and denote by $R_{\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0, \zeta_w)$ the reflection amplitudes obtained by the direct method at the sample in-plane wave vector $\mathbf{p}$ for a given in-plane incident wave vector $\mathbf{p}_0$ and surface realization ζ_w, then the averages in Eqs. (1.123a) and (1.123b) are estimated by

$$\langle R_{\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0) \rangle \approx \frac{1}{N_{\text{surf}}} \sum_{w=1}^{N_{\text{surf}}} R_{\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0, \zeta_w) \tag{8.3a}$$

$$\big\langle |R_{\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0)|^2 \big\rangle \approx \frac{1}{N_{\text{surf}}} \sum_{w=1}^{N_{\text{surf}}} |R_{\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0, \zeta_w)|^2 \,. \tag{8.3b}$$

This approach that we will refer to as the direct Monte Carlo method, and that can be considered as a brute force method, has been used in numerous works to explore the scattering from randomly rough surfaces on dielectrics, metals, or perfect conductors, or even systems containing a film deposited on a substrate with either a single randomly rough interface or two [41, 43, 45, 1, 46]. This method is tractable with a desktop computer for one-dimensional surfaces but becomes rather demanding in terms of memory and speed for two-dimensional surfaces and super-computers may be preferred for the task. Generally, the characteristic length of the numerical surface is about twenty wavelength and the number of surface realizations is about a few thousands to ensure a rather small statistical noise to signal ratio. More details

about this method can be found in the aforementioned references.

Even though the results obtained by the direct Monte Carlo method are of fundamental importance for understanding and exploring the physics of light scattering by systems with randomly rough interfaces, the demanding computational power required by the method makes it unpractical for applications such as online characterization or for designing surfaces with specific optical properties. For these purposes, we are then after fast and accurate approximate methods, and we will present some in details hereafter. In the following, the direct Monte Carlo method will be beneficial for us in order to assess the accuracy of the approximate methods.

8.2 Iterative Monte Carlo methods

A first strategy that should come to mind to reduce the computational cost of the direct Monte Carlo method is to solve each scattering problem faster by using iterative methods. Indeed, for light scattering by a single randomly rough surface the computational cost associated with solving the linear system resulting from the discretization of the reduced Rayleigh equations, is significantly larger than that to set up the system (since N_c is large enough for the total complexity to be dominated by the solving part see Section 5.2.2) and we have seen in Chapter 6 that, in some cases, iterative methods may yield a significant speed up for solving the linear system. We have also seen that the perturbative method (SAPT) has a total complexity that may be advantageous if only a few orders are needed to get a good approximation of the solution, so typically for rather weakly rough surfaces. Therefore, one may simply replace the direct method by iterative methods and keep averaging the observables over surface realizations. This is expected to work well in practice if the surfaces are weakly rough, but the speed up will obviously depend on the tolerance set on the solution (in practice it is enough that the tolerance set on each solution is smaller than that set for the statistical noise to signal ratio) and the convergence rate of the iterative method, which is system dependent. Moreover, even though using an iterative method may yield a significant speed up, there is absolutely no gain in the statistical noise to signal ratio, and we still need to gather statistics over as many surface realizations as were necessary with the direct method. The second idea is then to look for approximate solutions in closed form that may be easy to average analytically.

8.3 Isserlis-Wick theorem

Before embarking on the discussion of approximation techniques for computing directly the mean differential scattering coefficients, for which we wish to avoid solving the reduced Rayleigh equation for a large number of surface realizations and then average the differential scattering coefficients, we make a mathematical detour in the realm of probability. The approximation techniques we will present in subsequent sections are based on a expansion of the kernel of the reduced Rayleigh equation in Fourier moments of the surface profile. This has already been encountered when we presented the small amplitude perturbation theory in Chapter 6. We will see that such an expansion, together with ensemble averaging of the differential scattering coefficients leads to the evaluation of covariances of the form $\langle \zeta^n(\mathbf{x}+\mathbf{u})\zeta^m(\mathbf{x})\rangle$, where n and m are two non-negative integers.

We restrict ourselves to the case where the surface profile is a realization of a stochastic process with *Gaussian height distribution*, zero average, and the auto-correlation is homogeneous. These

properties are summarized as

$$\langle \zeta(\mathbf{x}) \rangle = 0 \tag{8.4a}$$
$$\langle \zeta(\mathbf{x}) \zeta(\mathbf{x}') \rangle = \sigma^2 \, W(\mathbf{x} - \mathbf{x}') \, , \tag{8.4b}$$

where σ is the rms-roughness and W is the auto-correlation function normalized such that $W(\mathbf{0}) = 1$.

Isserlis-Wick theorem

The Isserlis-Wick theorem gives us exactly a way to compute the covariance $\langle \zeta^n(\mathbf{x}+\mathbf{u}) \zeta^m(\mathbf{x}) \rangle$ we are after. This theorem from probability theory was proved by Isserlis in 1916-18 and re-discovered by Wick in the field of particle physics in 1950 [84, 85, 86]. This theorem is valid for a set of n identically distributed, with zero average, *Gaussian random variables* that we may denote abstractly as $(X_1, \cdots, X_n)$. The theorem gives the average of the product of the $(X_i)_{1 \le i \le n}$, namely

$$\langle X_1 \cdots X_n \rangle = \begin{cases} 0 & \text{if } n \text{ is odd} \\ \sum \prod \langle X_i X_j \rangle & \text{if } n \text{ is even} \end{cases} . \tag{8.5}$$

Here the notation $\sum \prod$ denotes a sum over all possible *distinct* ways of partitioning the n variables in $n/2$ pairs of variables X_i and X_j. Each term in the sum is the product of the $n/2$ covariances of the pairs of the associated partitioning. As an example, for four random variables Eq. (8.5) reads

$$\langle X_1 X_2 X_3 X_4 \rangle = \langle X_1 X_2 \rangle \langle X_3 X_4 \rangle + \langle X_1 X_3 \rangle \langle X_2 X_4 \rangle + \langle X_1 X_4 \rangle \langle X_2 X_3 \rangle \, . \tag{8.6}$$

As a first example of application of the Isserlis-Wick theorem, let $X_1 = \cdots = X_n = \zeta(\mathbf{x})$, we obtain

$$\langle \zeta^n(\mathbf{x}) \rangle = \begin{cases} 0 & \text{if } n \text{ is odd} \\ \sum \prod \langle \zeta^2(\mathbf{x}) \rangle = \sum \prod \sigma^2 & \text{if } n \text{ is even} \end{cases} . \tag{8.7}$$

Following our example, for $n = 4$, this yields $\left\langle \zeta^4(\mathbf{x}) \right\rangle = 3 \, \sigma^4$. Higher order moments can thus be generated simply by counting the number of terms in the Isserlis-Wick sum which can be shown to be equal to

$$\sum \prod 1 = \frac{(2k)!}{2^k \, k!} \, , \tag{8.8}$$

for $n = 2k$ (see proof in Appendix B).

More generally, let us now consider covariances of the form $\langle \zeta^n(\mathbf{x}+\mathbf{u}) \zeta^m(\mathbf{x}) \rangle$. By considering the following $n+m$ random variables: $X_1 = \cdots = X_n = \zeta(\mathbf{u}+\mathbf{x})$ and $X_{n+1} = \cdots = X_{n+m} = \zeta(\mathbf{x})$, we obtain, according to the Isserlis-Wick theorem, that if $n+m$ is odd the covariance $\langle \zeta^n(\mathbf{u}+\mathbf{x}) \zeta^m(\mathbf{x}) \rangle$ vanishes and if $n+m$ is even it equals the sum of products of all possible

pairing of the random variables. The covariance of each pair can only take the following two possible values

$$\langle X_i X_j \rangle = \begin{cases} \sigma^2 & \text{if } i,j < n+1/2, \text{ or } i,j > n+1/2 \\ \sigma^2\, W(\mathbf{u}) & \text{if } i < n+1/2,\ j > n+1/2, \text{ or } i > n+1/2,\ j < n+1/2 \end{cases} . \tag{8.9a}$$

Indeed, on the one hand, if both i and j are on the same side from the index threshold $n+1/2$, then X_i and X_j are either both equal to $\zeta(\mathbf{u}+\mathbf{x})$ or are both equal to $\zeta(\mathbf{x})$ and hence $\langle X_i X_j \rangle = \sigma^2$. On the other hand, if i and j are on different sides of index threshold $n+1/2$, then X_i and X_j are distinct and are equal to either $\zeta(\mathbf{u}+\mathbf{x})$ or $\zeta(\mathbf{x})$ and hence $\langle X_i X_j \rangle = \langle \zeta(\mathbf{u}+\mathbf{x})\zeta(\mathbf{x}) \rangle = \sigma^2\, W(\mathbf{u})$ by definition of the auto-correlation function. Consequently, the sum of products of covariances in the Isserlis-Wick formula together with the aforementioned fact imply that $\langle \zeta^n(\mathbf{u}+\mathbf{x})\zeta^m(\mathbf{x}) \rangle$ is a *polynomial* in $W(\mathbf{u})$ of degree at most $(n+m)/2$, i.e.

$$\langle \zeta^n(\mathbf{u}+\mathbf{x})\zeta^m(\mathbf{x}) \rangle = \sigma^{n+m} \sum_{k=0}^{(n+m)/2} p_{n,m}^{(k)}\, W^k(\mathbf{u}) = \sigma^{n+m}\, P_{n,m}(W(\mathbf{u})) , \tag{8.10}$$

where $P_{n,m}$ is a polynomial with integer coefficients $(p_{n,m}^{(k)})_{0 \le k \le (n+m)/2}$ that depend on n and m.

Example: as an illustrative example, let us compute $\langle \zeta^2(\mathbf{u}+\mathbf{x})\zeta^2(\mathbf{x}) \rangle$ and $\langle \zeta^3(\mathbf{u}+\mathbf{x})\zeta^1(\mathbf{x}) \rangle$ that are both of order $n+m=4$. For $n=m=2$, according to the Isserlis-Wick theorem, Eq. (8.6) with $X_1 = X_2 = \zeta(\mathbf{u}+\mathbf{x})$ and $X_3 = X_4 = \zeta(\mathbf{x})$ gives

$$\begin{aligned} \langle \zeta^2(\mathbf{u}+\mathbf{x})\zeta^2(\mathbf{x}) \rangle &= \langle \zeta(\mathbf{u}+\mathbf{x})\zeta(\mathbf{u}+\mathbf{x}) \rangle \langle \zeta(\mathbf{x})\zeta(\mathbf{x}) \rangle \\ &\quad + \langle \zeta(\mathbf{u}+\mathbf{x})\zeta(\mathbf{x}) \rangle \langle \zeta(\mathbf{u}+\mathbf{x})\zeta(\mathbf{x}) \rangle \\ &\quad + \langle \zeta(\mathbf{u}+\mathbf{x})\zeta(\mathbf{x}) \rangle \langle \zeta(\mathbf{u}+\mathbf{x})\zeta(\mathbf{x}) \rangle \\ &= \sigma^4 \left[1 + 2\, W^2(\mathbf{u}) \right] . \end{aligned} \tag{8.11}$$

In addition, for $n=3$ and $m=1$, according to Isserlis-Wick theorem, Eq. (8.6) with $X_1 = X_2 = X_3 = \zeta(\mathbf{u}+\mathbf{x})$ and $X_4 = \zeta(\mathbf{x})$ gives

$$\begin{aligned} \langle \zeta^3(\mathbf{u}+\mathbf{x})\zeta(\mathbf{x}) \rangle &= \langle \zeta(\mathbf{u}+\mathbf{x})\zeta(\mathbf{u}+\mathbf{x}) \rangle \langle \zeta(\mathbf{u}+\mathbf{x})\zeta(\mathbf{x}) \rangle \\ &\quad + \langle \zeta(\mathbf{u}+\mathbf{x})\zeta(\mathbf{u}+\mathbf{x}) \rangle \langle \zeta(\mathbf{u}+\mathbf{x})\zeta(\mathbf{x}) \rangle \\ &\quad + \langle \zeta(\mathbf{u}+\mathbf{x})\zeta(\mathbf{x}) \rangle \langle \zeta(\mathbf{u}+\mathbf{x})\zeta(\mathbf{u}+\mathbf{x}) \rangle \\ &= 3\, \sigma^4\, W(\mathbf{u}) . \end{aligned} \tag{8.12}$$

Note the parity of the polynomials $P_{2,2}$ and $P_{3,1}$. We leave to the reader to show that $P_{1,3} = P_{3,1}$ and that the extreme (and trivial) cases $P_{4,0} = P_{0,4} = 3$. We can summarize our finding in the following table for the coefficients of the polynomial $P_{n,m}$ for $n+m=4$.

Remark 8.1. One can show that the polynomial $P_{n,m}$ is even if n and m are both even, and $P_{n,m}$ is odd if n and m are both odd. In addition, since each term in the Isserlis-Wick sum contributes to one unit for a single coefficient $p_{n,m}^{(k)}$, it is clear that the coefficients are all natural integers and that the sum of all coefficients equals the number of terms in the Isserlis-Wick sum. Hence we have

$$\sum_{k=0}^{(n+m)/2} p_{n,m}^{(k)} = \sum \prod 1 = \frac{(n+m)!}{2^{(n+m)/2}\, [(n+m)/2]!} . \tag{8.13}$$

(n,m)	$p^{(0)}_{n,m}$	$p^{(1)}_{n,m}$	$p^{(2)}_{n,m}$
$(4,0),(0,4)$	3	0	0
$(3,1),(1,3)$	0	3	0
$(2,2)$	1	0	2

Table 8.1: Coefficients of the polynomials $P_{n,m}$ for $n+m=4$.

We have now shown that the problem of computing $\langle \zeta^n(\mathbf{u}+\mathbf{x})\zeta^m(\mathbf{x})\rangle$ reduces to counting terms in the Isserlis-Wick sum that contribute to each monome in $W(\mathbf{u})$. It is clear from Eq. (8.13) that this combinatorial problem will grow in complexity rather quickly with $n+m$. The number of terms in the Isserlis-Wick sum is 3 for $n+m=4$, 15 for $n+m=6$, and 105 for $n+m=8$. Although it may be instructive to determine by hand the coefficients $p^{(k)}_{n,m}$ for $n+m=6$ as an exercise, it seems tedious to do it for larger orders, and this is where an algorithm must take over. We present such an algorithm in Appendix B capable of computing the coefficients $p^{(k)}_{n,m}$.

8.4 Small amplitude perturbation theory

We will now derive a closed form approximation of the mean differential scattering coefficient based on the small amplitude perturbative expansion of the reflection and transmission amplitudes to first order in the surface profile function. If we now use the expression found in Section 6.2 for the reflection amplitudes to first order in the product of surface profile function,

$$\mathbf{R}\left(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}\right) \approx \mathbf{R}^{(0)}\left(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}\right) - i\,\mathbf{R}^{(1)}\left(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}\right), \tag{8.14}$$

where $\mathbf{R}^{(0)}\left(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}\right)$ is the response from the corresponding system with a planar interface (i.e. the Fresnel amplitudes), Eq. (6.8), and $\mathbf{R}^{(1)}\left(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}\right)$ is given in Eq. (6.12), we obtain that the average amplitude (approximated to first order) is given by

$$\begin{aligned}\langle \mathbf{R}\left(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}\right)\rangle &= (2\pi)^2\delta(\mathbf{p}-\mathbf{p}_0)\,\boldsymbol{\rho}^{(0)}(\mathbf{p}_0) - i\left\langle \hat{\zeta}(\mathbf{p}-\mathbf{p}_0)\right\rangle \boldsymbol{\rho}^{(1)}\left(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}\right)\\ &= (2\pi)^2\delta(\mathbf{p}-\mathbf{p}_0)\,\boldsymbol{\rho}^{(0)}(\mathbf{p}_0).\end{aligned} \tag{8.15}$$

The above equation simply states that to first order in the surface profile function, the average field corresponds to that reflected by the flat interface. Since the average amplitude is proportional to a Dirac mass, the average field is a plane wave. Hence the coherent component of the mean differential reflection coefficient cannot be defined by Eq. (1.123a) but we must use the definition of the efficiency instead, i.e. Eq. (1.62), in order to evaluate the energy carried away by the averaged field. The incoherent component of the mean differential reflection coefficient is of greater interest. By using the small amplitude perturbation expansion to first order in the surface profile function, the factor in the square bracket in Eq. (1.123b) reads

$$\begin{aligned}\left\langle |R_{\alpha\beta}(\mathbf{p}\mid\mathbf{p}_0)|^2\right\rangle - |\langle R_{\alpha\beta}(\mathbf{p}\mid\mathbf{p}_0)\rangle|^2 &= \left\langle |R^{(1)}_{\alpha\beta}(\mathbf{p}\mid\mathbf{p}_0)|^2\right\rangle\\ &= \left\langle |\hat{\zeta}_S(\mathbf{p}-\mathbf{p}_0)|^2\right\rangle |\rho^{(1)}_{\alpha\beta}\left(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}\right)|^2.\end{aligned} \tag{8.16}$$

We remind the reader that here we are dealing with a scattering system whose surface profiles are flat outside the disk of radius r, hence the subscript S (see Section 1.2.4). For the statistical properties attributed to the surface profiles in Eq. (8.2), we have

$$\begin{aligned}\left\langle \hat{\zeta}_S(\mathbf{q})\hat{\zeta}_S^*(\mathbf{q})\right\rangle &= \left\langle \int_S\int_S \zeta(\mathbf{x})\zeta(\mathbf{x}')\exp(i\,\mathbf{q}\cdot(\mathbf{x}-\mathbf{x}'))\,\mathrm{d}^2x\,\mathrm{d}^2x'\right\rangle \\ &= \int_S\int_S \left\langle \zeta(\mathbf{x})\zeta(\mathbf{x}')\right\rangle \exp(i\,\mathbf{q}\cdot(\mathbf{x}-\mathbf{x}'))\,\mathrm{d}^2x\,\mathrm{d}^2x' \\ &= \int_S\int_S \sigma^2\,W(\mathbf{x}-\mathbf{x}')\,\exp(i\,\mathbf{q}\cdot(\mathbf{x}-\mathbf{x}'))\,\mathrm{d}^2x\,\mathrm{d}^2x'\,. \end{aligned} \tag{8.17}$$

Here we have used the definition of the Fourier transform, and the fact that the ensemble average commutes with the integration of the surfaces and the definition of the auto-correlation function. Via the change of variable $\mathbf{u}=\mathbf{x}-\mathbf{x}'$ we obtain

$$\left\langle \hat{\zeta}_S(\mathbf{q})\hat{\zeta}_S^*(\mathbf{q})\right\rangle = S\,\sigma^2\int_S W(\mathbf{u})\,\exp(i\,\mathbf{q}\cdot\mathbf{u})\,\mathrm{d}^2u = S\,\sigma^2\,g_S(\mathbf{q})\,. \tag{8.18}$$

Thus we have

$$\left\langle |R_{\alpha\beta}(\mathbf{p}\mid\mathbf{p}_0)|^2\right\rangle - |\langle R_{\alpha\beta}(\mathbf{p}\mid\mathbf{p}_0)\rangle|^2 = S\,g_S(\mathbf{p}-\mathbf{p}_0)\,\sigma^2\,|\rho^{(1)}_{\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o})\,|^2\,. \tag{8.19}$$

Finally, by plugging the above equation into Eq. (1.123b), the surface area S cancels and letting $r\to\infty$, $g_S\to g$ (where we remind the reader that g is the power spectrum of the surface profiles) and we finally obtain the expression for the incoherent component of the mean differential reflection coefficient for the entire (infinite) system under the first order approximation of the reflected amplitudes in the surface profile

$$\left\langle \frac{\partial R^{(\mathrm{SAPT1})}_{\alpha\beta}}{\partial\Omega_r}(\mathbf{p}\,|\,\mathbf{p}_0)\right\rangle_{\mathrm{incoh}} = \epsilon_1\left(\frac{\omega}{2\pi c}\right)^2\frac{\cos^2\theta_r}{\cos\theta_0}\,g(\mathbf{p}-\mathbf{p}_0)\,\sigma^2\,|\rho^{(1)}_{\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o})\,|^2\,. \tag{8.20}$$

Applying the same reasoning to the transmitted light we obtain that the averaged transmitted field corresponds to that refracted by the planar interface and the incoherent component of the mean differential transmission coefficient is given by

$$\left\langle \frac{\partial T^{(\mathrm{SAPT1})}_{\alpha\beta}}{\partial\Omega_t}(\mathbf{p}\,|\,\mathbf{p}_0)\right\rangle_{\mathrm{incoh}} = \left(\frac{\omega}{2\pi c}\right)^2\frac{\epsilon_2^{3/2}\cos^2\theta_t}{\epsilon_1^{1/2}\cos\theta_0}\,g(\mathbf{p}-\mathbf{p}_0)\,\sigma^2\,|\tau^{(1)}_{\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o})\,|^2\,. \tag{8.21}$$

Let us make a few comments on the obtained results. We have found that, to second order in the surface profile[1] the diffusely scattered intensity is proportional to the square of the rms-roughness, proportional to the power spectrum shifted by $\mathbf{p}_0$ and inversely proportional to the square of the wavelength. The presence of the power spectrum can be interpreted physically as a generalization of the grating formula. Taking the picture of single scattering, the probability that an elementary plane wave is scattered with a change of in-plane wave vector $\mathbf{p}-\mathbf{p}_0$ is proportional to $g(\mathbf{p}-\mathbf{p}_0)$. Hence if the power spectrum has its weight concentrate around the origin, most of the light will be scattered around the specular direction, while the broader the power spectrum the broader is the intensity scattered diffusely. How does it relates to the actual

[1] second order for the intensity but first order for the amplitude

geometry of the surface? A narrow power spectrum corresponds to a broad, slowly varying auto-correlation function. In other words, the correlation length is large compared to the wavelength. This indicates that the surface is rather smooth and can be well approximated by a planar interface, hence the observed weakly diffusely scattered light. On the contrary, for a correlation function concentrated within a fraction of the wavelength, the surface would look rougher and hence we would observe a more broadly scattered diffused light. For example, one can consider a Gaussian auto-correlation function defined by

$$W(\mathbf{x}) = \exp\left(-\frac{|\,\mathbf{x}\,|^2}{a^2}\right), \tag{8.22}$$

then the power spectrum is given by

$$g(\mathbf{p}) = \pi a^2 \exp\left(-\frac{|\,\mathbf{p}\,|^2 a^2}{4}\right). \tag{8.23}$$

It is then clear, and this is a general property of the Fourier transform, that the width of the auto-correlation function and the power spectrum vary inversely proportionally with respect to one anther. Note that if we denote by a the correlation length of the surface profile, the power spectrum is proportional to a^2, but its width is of the order a, so that the integral of the power spectrum is unity. Thus, according to Eqs. (8.20) and (8.21), we can give the following simple scaling law of the integrated intensity that is diffusely scattered by a weakly rough surface

$$\int \left\langle \frac{\partial X_{\alpha\beta}}{\partial \Omega}(\mathbf{p}\,|\,\mathbf{p}_0) \right\rangle_{\text{incoh}} \mathrm{d}\Omega = C\,\frac{\sigma^2}{\lambda^2}, \tag{8.24}$$

where C is some constant that depends of the materials, the angle of incidence and the precise shape of the power spectrum, and whether the reflected or transmitted light is considered.

Remark 8.2. One should be careful in interpreting the formulas for the coherent and incoherent components of the mean differential scattering coefficients from an energetic point of view. Indeed, if one sums the efficiencies associated with the average reflected and transmitted wave, we obtain unity (for loss-less material) since these two waves correspond to that reflected and refracted by the corresponding system with a planar interface. Hence if we were to sum on top of that the incoherent component of the mean differential reflection and transmission coefficients integrated over all scattering angles, we will necessarily obtain a total value that is larger than unity. One would expect instead that even on average, the sum of the effeciencies of the coherent components of the reflected and transmitted light be less than unity as power is on average also scattered away. The way to resolve this apparent paradox is to remember that the above results are approximations to first order in the surface profile for the amplitudes and that higher orders have been neglected. In particular, one can show that the second order will contain a correction to the coherent component.

8.4.1 Higher orders

The derivation of the mean differential scattering coefficients in closed form when higher order terms are taken into account becomes quickly cumbersome. To our knowledge, such an expression obtained with small amplitude perturbation theory have been obtained up to order three in the scattering amplitude, resulting in expression of the incoherent component of the mean differential scattering coefficients derived consistently to order four. This was achieved by Soubret *et al.* in Ref. [53] and used to study the phenomenon of *enhanced backscattering*, which is a multiple scattering phenomenon.

8.5 Fixed point iterates of the RRE of the second kind

We have seen in Section 6.3 that the reduced Rayleigh equations of the second kind can be solved by successive iterations yielding what we called the method of fixed point iterates of the reduced Rayleigh equations of the second kind (FIRES). When studying randomly rough surfaces, one is often interested in the incoherent component of the mean differential reflection coefficient averaged over realizations of the surface profile. Therefore, we focus our attention now on the approximation given by the first iterate of the algorithm of FIRES which read

$$\begin{aligned}\Delta\,\mathbf{R}^{(1)}(\mathbf{p}\mid\mathbf{p}_0) &= (\alpha_1(\mathbf{p})-\alpha_2(\mathbf{p}))\,\left[\mathbf{M}_{21}^{++}(\mathbf{p}\mid\mathbf{p})\right]^{-1}\,\Big(\mathcal{K}_{21}^{++}(\mathbf{p}\mid\mathbf{p}_0)\,\mathbf{M}_{21}^{++}(\mathbf{p}\mid\mathbf{p}_0)\,\boldsymbol{\rho}^{(0)}(\mathbf{p}_0)\\ &\qquad + \mathcal{K}_{21}^{+-}(\mathbf{p}\mid\mathbf{p}_0)\,\mathbf{M}_{21}^{+-}(\mathbf{p}\mid\mathbf{p}_0)\Big)\\ &\stackrel{\text{def}}{=} \mathcal{K}_{21}^{++}(\mathbf{p}\mid\mathbf{p}_0)\,\boldsymbol{\rho}^{+}(\mathbf{p}\,|\,\mathbf{p}_\text{o}) + \mathcal{K}_{21}^{+-}(\mathbf{p}\mid\mathbf{p}_0)\,\boldsymbol{\rho}^{-}(\mathbf{p}\,|\,\mathbf{p}_\text{o})\,, \qquad (8.25a)\\ \Delta\,\mathbf{T}^{(1)}(\mathbf{p}\mid\mathbf{p}_0) &= (\alpha_1(\mathbf{p})-\alpha_2(\mathbf{p}))\,\left[\mathbf{M}_{12}^{-,-}(\mathbf{p}\mid\mathbf{p})\right]^{-1}\,\mathcal{K}_{12}^{-,-}(\mathbf{p}\mid\mathbf{p}_0)\,\mathbf{M}_{12}^{-,-}(\mathbf{p}\mid\mathbf{p}_0)\,\boldsymbol{\tau}^{(0)}(\mathbf{p}_0)\\ &\stackrel{\text{def}}{=} \mathcal{K}_{12}^{-,-}(\mathbf{p}\mid\mathbf{p}_0)\,\boldsymbol{\tau}\,(\mathbf{p}\,|\,\mathbf{p}_\text{o})\,. \qquad (8.25b)\end{aligned}$$

Remark 8.3. Note that by approximating $\mathcal{K}_{21}^{ba}$ to the first term in its series expansion, one recovers exactly the first order correction $\mathbf{R}^{(1)}$ given by small amplitude perturbation theory (and similarly for the transmission amplitudes). Hence the first iterate of the method of FIRES gives a distinct approximation from small amplitude perturbation theory to first order and can possibly be more accurate as it takes into account higher order Fourier moments of the surface profile. The physical interpretation is that the first iterate of the method of FIRES gives a single scattering approximation similar to the first order of SAPT, but taking into account the detail of the surface more accurately than SAPT does. It should *not* be interpreted as taking into account multiple scattering since this would involve integrals over intermediary in-plane wave vectors, which is not the case for the first iterate.

Before we proceed with the computation of the incoherent component of the mean differential reflection and transmission coefficients we analyze the following ensemble averages

$$\left\langle\mathcal{K}_{lm}^{ba}(\mathbf{p}\mid\mathbf{q})\right\rangle = \sum_{n=1}^{\infty}\frac{(-i)^n}{n!}(b\alpha_l(\mathbf{p})-a\alpha_m(\mathbf{q}))^{n-1}\,\left\langle\hat{\zeta}^{(n)}(\mathbf{p}-\mathbf{q})\right\rangle \qquad (8.26)$$

$$\begin{aligned}\left\langle\mathcal{K}_{lm}^{ba}(\mathbf{p}\mid\mathbf{q})\,\mathcal{K}_{lm}^{b'a'}(\mathbf{p}\mid\mathbf{q})^*\right\rangle &= \sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\frac{(-i)^n i^m}{n!\,m!}(b\alpha_l(\mathbf{p})-a\alpha_m(\mathbf{q}))^{n-1}\,(b'\alpha_l(\mathbf{p})^*-a'\alpha_m(\mathbf{q})^*)^{m-1}\\ &\qquad\left\langle\hat{\zeta}^{(n)}(\mathbf{p}-\mathbf{q})\hat{\zeta}^{(m)}(\mathbf{p}-\mathbf{q})^*\right\rangle\,, \qquad (8.27)\end{aligned}$$

where the * denotes the complex conjugate. It is clear that the difficulty in evaluating these expressions resides in the averages of the Fourier moments of the surface profile (or their products). In practice, two strategies may be considered: the first consists in evaluating numerically these averages by taking averages over realization of the stochastic process (i.e. the surface), the second consists in computing them analytically. In general, these averages of Fourier moments do not let themselves express easily in closed form for general stochastic processes. There is, however, a non-trivial specific case for which this can be done, namely the case where the height distribution is Gaussian thanks to the Isserlis-Wick theorem.

Let us consider the average of the n^{th} Fourier moment

$$\left\langle\hat{\zeta}^{(n)}(\mathbf{p}-\mathbf{q})\right\rangle = \int e^{-i(\mathbf{p}-\mathbf{q})\cdot\mathbf{x}}\,\left\langle\zeta^n(\mathbf{x})\right\rangle\,\mathrm{d}^2x\,. \qquad (8.28)$$

For $n=0$, we easily obtain $\left\langle \hat{\zeta}^{(0)}(\mathbf{p}-\mathbf{q}) \right\rangle = (2\pi)^2\delta(\mathbf{p}-\mathbf{q})$, for $n=1$, we obtain $\left\langle \hat{\zeta}^{(1)}(\mathbf{p}-\mathbf{q}) \right\rangle = 0$ due to Eq. (8.4a), and for $n=2$, we get $\left\langle \hat{\zeta}^{(2)}(\mathbf{p}-\mathbf{q}) \right\rangle = \sigma^2\,(2\pi)^2\delta(\mathbf{p}-\mathbf{q})$ due to Eq. (8.4b) with $\mathbf{x}=\mathbf{x}'$. Then higher order moments are given by Isserlis-Wick theorem (see Eq. (8.8)):

$$\left\langle \hat{\zeta}^{(n)}(\mathbf{p}-\mathbf{q}) \right\rangle = \begin{cases} 0 & \text{if } n \text{ is odd} \\ \frac{n!}{2^{n/2}\,(n/2)!}\,\sigma^n\,(2\pi)^2\delta(\mathbf{p}-\mathbf{q}) & \text{if } n \text{ is even} \end{cases} . \tag{8.29}$$

Remark 8.4. Equation (8.29) shows that only even terms in Eq. (8.26) remain and that all terms are proportional to the Dirac mass $\delta(\mathbf{p}-\mathbf{q})$. We will see that these constitute correction to the coherent component of the mean differential coefficient.

Let us now consider covariances of Fourier moments of the form:

$$\left\langle \hat{\zeta}_S^{(n)}(\mathbf{p}-\mathbf{q})\hat{\zeta}_S^{(m)}(\mathbf{p}-\mathbf{q})^* \right\rangle = \int_S\int_S e^{-i(\mathbf{p}-\mathbf{q})\cdot(\mathbf{x}-\mathbf{x}')}\ \langle \zeta^n(\mathbf{x})\zeta^m(\mathbf{x}')\rangle\ \mathrm{d}^2x\,\mathrm{d}^2x' . \tag{8.30}$$

Here S denote the part of the surface of the disk of radius $r>0$ where the random profile is found for the *truncated* copy, $\zeta_S = \mathbb{1}_S\,\zeta$, of the full randomly rough surface. This will be needed in order to take the limit when $r\to\infty$ in the definition of the mean differential scattering coefficients (see Section 1.2.4). By doing the change of variable $\mathbf{u}=\mathbf{x}-\mathbf{x}'$ we obtain

$$\left\langle \hat{\zeta}_S^{(n)}(\mathbf{p}-\mathbf{q})\hat{\zeta}_S^{(m)}(\mathbf{p}-\mathbf{q})^* \right\rangle = \int_S\int_S e^{-i(\mathbf{p}-\mathbf{q})\cdot\mathbf{u}}\ \langle \zeta^n(\mathbf{u}+\mathbf{x})\zeta^m(\mathbf{x})\rangle\ \mathrm{d}^2u\,\mathrm{d}^2x . \tag{8.31}$$

According to the discussion from Section 8.3, and more specifically Eq. (8.10), we know that $\langle \zeta^n(\mathbf{u}+\mathbf{x})\zeta^m(\mathbf{x})\rangle$ either vanishes if $n+m$ is odd or can be expressed as a polynomial in the auto-correlation function. Hence

$$\begin{aligned} \left\langle \hat{\zeta}_S^{(n)}(\mathbf{p}-\mathbf{q})\hat{\zeta}_S^{(m)}(\mathbf{p}-\mathbf{q})^* \right\rangle &= S\,\sigma^{n+m}\sum_{k=0}^{(n+m)/2} p_{n,m}^{(k)}\int_S e^{-i(\mathbf{p}-\mathbf{q})\cdot\mathbf{u}}\,W^k(\mathbf{u})\,\mathrm{d}^2u \\ &\overset{\text{def}}{=} S\,\sigma^{n+m}\sum_{k=0}^{(n+m)/2} p_{n,m}^{(k)}\quad g_S^{(k)}(\mathbf{p}-\mathbf{q}) . \end{aligned} \tag{8.32}$$

Here we have defined $g_S^{(k)}$ as the Fourier transform of $\mathbb{1}_S\,W^k$.

Based on Eqs.(8.25a, 8.25b, 8.29) and (8.32) we can deduce that the average amplitudes of the first iterate of the method of FIRES are given by

$$\Big\langle \mathbf{R}^{(\mathrm{FIRES1})}(\mathbf{p} \mid \mathbf{p}_0) \Big\rangle = \mathbf{R}^{(0)}(\mathbf{p} \mid \mathbf{p}_0) + \Big\langle \Delta\, \mathbf{R}^{(1)}(\mathbf{p} \mid \mathbf{p}_0) \Big\rangle$$
$$= \Big[\boldsymbol{\rho}^{(0)}(\mathbf{p}_0) + \sum_{k=1}^{\infty} \frac{(-1)^k \sigma^{2k}}{2^k k!} \Big(\boldsymbol{\rho}^{+,(2k)}(\mathbf{p} \mid \mathbf{p}_0) + \boldsymbol{\rho}^{-,(2k)}(\mathbf{p} \mid \mathbf{p}_0) \Big) \Big] (2\pi)^2 \delta(\mathbf{p} - \mathbf{p}_0) \tag{8.33a}$$
$$\Big\langle \mathbf{T}^{(\mathrm{FIRES1})}(\mathbf{p} \mid \mathbf{p}_0) \Big\rangle = \mathbf{T}^{(0)}(\mathbf{p} \mid \mathbf{p}_0) + \Big\langle \Delta\, \mathbf{T}^{(1)}(\mathbf{p} \mid \mathbf{p}_0) \Big\rangle$$
$$= \left[\boldsymbol{\tau}^{(0)}(\mathbf{p}_0) + \sum_{k=1}^{\infty} \frac{(-1)^k \sigma^{2k}}{2^k k!} \boldsymbol{\tau}^{(2k)}(\mathbf{p} \mid \mathbf{p}_0) \right] (2\pi)^2 \delta(\mathbf{p} - \mathbf{p}_0) \tag{8.33b}$$
$$\boldsymbol{\rho}^{+,(k)}(\mathbf{p} \mid \mathbf{p}_0) \stackrel{\mathrm{def}}{=} (\alpha_2(\mathbf{p}) - \alpha_1(\mathbf{p}_0))^{k-1} \boldsymbol{\rho}^{+}(\mathbf{p} \mid \mathbf{p}_0) \tag{8.33c}$$
$$\boldsymbol{\rho}^{-,(k)}(\mathbf{p} \mid \mathbf{p}_0) \stackrel{\mathrm{def}}{=} (\alpha_2(\mathbf{p}) + \alpha_1(\mathbf{p}_0))^{k-1} \boldsymbol{\rho}^{-}(\mathbf{p} \mid \mathbf{p}_0) \tag{8.33d}$$
$$\boldsymbol{\tau}^{(k)}(\mathbf{p} \mid \mathbf{p}_0) \stackrel{\mathrm{def}}{=} (\alpha_1(\mathbf{p}_0) - \alpha_2(\mathbf{p}))^{k-1} \boldsymbol{\tau}(\mathbf{p} \mid \mathbf{p}_0) \tag{8.33e}$$

for which the efficiencies are easy to compute, and the incoherent component of the mean differential scattering coefficients are given by

$$\left\langle \frac{\partial R_{\alpha\beta}^{(\mathrm{FIRES1})}}{\partial \Omega_r}(\mathbf{p} \mid \mathbf{p}_0) \right\rangle_{\mathrm{incoh}} = \epsilon_1 \left(\frac{\omega}{2\pi c} \right)^2 \frac{\cos^2 \theta_r}{\cos \theta_0} \sum_{\substack{n=1 \\ m=1 \\ n+m \text{ even}}}^{\infty} \psi_{n,m}(\mathbf{p} \mid \mathbf{p}_0)$$
$$\times \Big(\rho_{\alpha\beta}^{+,(n)}(\mathbf{p} \mid \mathbf{p}_0) + \rho_{\alpha\beta}^{-,(n)}(\mathbf{p} \mid \mathbf{p}_0) \Big) \Big(\rho_{\alpha\beta}^{+,(m)}(\mathbf{p} \mid \mathbf{p}_0) + \rho_{\alpha\beta}^{-,(m)}(\mathbf{p} \mid \mathbf{p}_0) \Big)^* \tag{8.34a}$$
$$\left\langle \frac{\partial T_{\alpha\beta}^{(\mathrm{FIRES1})}}{\partial \Omega_t}(\mathbf{p} \mid \mathbf{p}_0) \right\rangle_{\mathrm{incoh}} = \left(\frac{\omega}{2\pi c} \right)^2 \frac{\epsilon_2^{3/2} \cos^2 \theta_t}{\epsilon_1^{1/2} \cos \theta_0} \sum_{\substack{n=1 \\ m=1 \\ n+m \text{ even}}}^{\infty} \psi_{n,m}(\mathbf{p} \mid \mathbf{p}_0)$$
$$\times \tau_{\alpha\beta}^{(n)}(\mathbf{p} \mid \mathbf{p}_0) \tau_{\alpha\beta}^{(m)*}(\mathbf{p} \mid \mathbf{p}_0) \tag{8.34b}$$
$$\psi_{n,m}(\mathbf{p} \mid \mathbf{p}_0) \stackrel{\mathrm{def}}{=} \frac{(-i)^n i^m \sigma^{n+m}}{n!\, m!} \sum_{k=0}^{(n+m)/2} p_{n,m}^{(k)}\, g^{(k)}(\mathbf{p} - \mathbf{p}_0)\,. \tag{8.34c}$$

Here we have defined $g^{(k)}$ as the Fourier transform for the k^{th} power of the auto-correlation function W.

Remark 8.5. We decided here to show the expressions of the mean differential scattering coefficients in an expanded form to illustrate what goes into them. However, it is not necessary, nor advisable, to implement them in a computer program in such expanded forms. In practice, we first use a fast Fourier transform algorithm to compute the $g^{(k)}$ on a grid (or we can also compute the $g^{(k)}$ analytically in the case of a Gaussian correlation function) and compute the Isserlis coefficients $p_{n,m}^{(k)}$ thanks to a dedicated algorithm detailed in Appendix B. Then we compute the $\Big\langle \mathcal{K}_{lm}^{ba}(\mathbf{p} \mid \mathbf{p}_0)\, \mathcal{K}_{lm}^{b'a'}(\mathbf{p} \mid \mathbf{p}_0)^* \Big\rangle$ according to Eqs.(8.27) and (8.32) that we use in evaluating $\Big\langle |\Delta R_{\alpha\beta}^{(1)}|^2 \Big\rangle$. Such an algorithm is rather simple to implement and computationally efficient (assuming fast convergence with the order $(n+m)/2$ for truncating the sum over (n, m)).

8.5.1 Higher order iterates

As in the case of small amplitude perturbation theory, more accurate expressions for the mean differential scattering coefficients could in principle be obtained by considering the amplitudes obtained with higher order iterates. Such an approach becomes quickly cumbersome.

8.6 Method comparison

The comparison of the performance of small amplitude perturbation theory and the method of FIRES have been extensively documented in Chapter 6 for sinusoidal surfaces. In particular, we have seen that the method of FIRES is less complex and converges faster than SAPT. These characteristics holds for randomly rough surfaces. As an additional illustration of the performance of the different methods, now in the case of randomly rough surfaces, with respect to convergence with the number of iterations or perturbative order, we will fix a system made of two semi-infinite dielectric media of respective dielectric constant ϵ_1 and ϵ_2 separated by a randomly rough surface with Gaussian height distribution and Gaussian auto-correlation function characterized by a rms-roughness of $\sigma = \lambda/20$ and a correlation length of $a = \lambda/3$. The dielectric constant are chosen to take values in $\{1, 2.64\}$, i.e. that we will consider the cases where the medium of incidence is either the optical denser or less dense medium. This set-up corresponds exactly to that studied by Hetland *et al.* in Ref. [45] whose data have generously been provided to us by the authors. This set of data was obtained by using what we described as the direct method where the ensemble average was performed over a set of 4500 surface realizations. The data from Ref. [45] will serve as a reference for the presented approximation methods.

Analytically averaged first order SAPT and FIRES approximations

We start by investigating how accurate are the closed form expressions for the incoherent component of the mean differential reflection coefficient as given by Eq. (8.20) in the case of small amplitude perturbation theory to lowest non-trivial order and Eq. (8.34a) for the first iterate of the method of FIRES. Figure 8.1 shows the incoherent component of the mean differential reflection coefficient as a function of the polar angle of scattering for in-plane co-polarized scattering, obtained by the direct Monte Carlo method, Eq. (8.20) and Eq. (8.34a).

Monte Carlo based on iterative methods

We now turn to Monte Carlo simulations based on the iterative methods given by the recursive equation Eq. (6.11) from small amplitude perturbation theory, and Eq. (6.39) from the method of FIRES, that we have already presented in Chapter 6. Instead of making a full study of convergence for all the iterative schemes, as we did in Chapter 6, we restrict ourselves here to SAPT and the method of FIRES, and only inspect visually how the successive iterates approach the reference solution in the case where the rms-roughness is equal to $\sigma = \lambda/20$, and the dielectric constants are $\epsilon_1 = 2.64$ and $\epsilon_2 = 1$. The specific choice of $\sigma = \lambda/20$ is simply guided by the fact that we have seen that the first order analytically averaged SAPT and FIRES approximation yield a significant deviation from the reference solution for such a roughness parameter. Moreover, we chose the case of reflection in the denser medium simply because it exhibits more features, in particular that of the Yoneda phenomenon which will be described in details in Chapter 10.

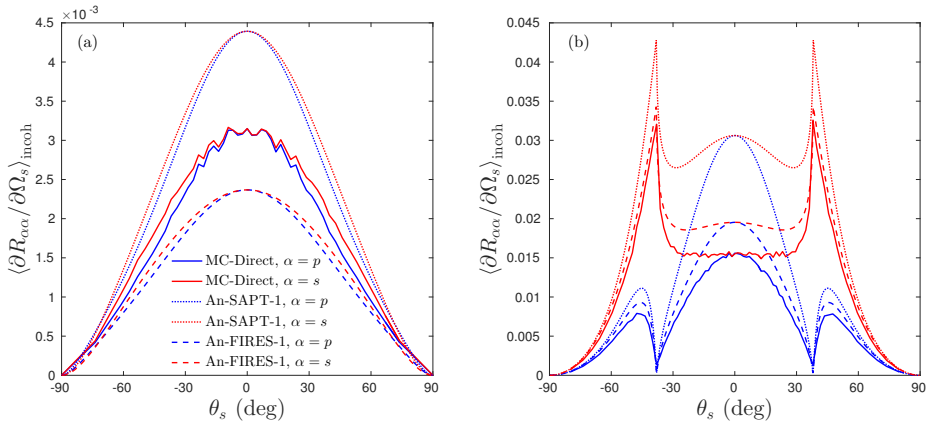


Figure 8.1: In-plane co-polarized incoherent mean differential reflection coefficient as a function of the polar scattering angle θ_s, for normal incidence. System composed of: (a) vacuum in medium 1 and a dielectric ($\epsilon_2 = 2.64$) in medium 2, (b) vacuum in medium 2 and a dielectric ($\epsilon_1 = 2.64$) in medium 1. Wavelength $\lambda = 632.8$ nm, rms-roughness $\sigma = \lambda/20$, correlation length $a = \lambda/4$. Solid lines: results obtained from the direct method and an ensemble average over 4500 realizations of the surface profile (courtesy of Øyvind Hetland [45]); dotted lines: results obtained by using small amplitude perturbation theory to order one in the surface profile of the reflection amplitude; dashed lines: results obtained by the first iterate of the method of FIRES. Blue: p-polarized light; red: s-polarized light. Figure taken from **Paper** [4].

We start by studying the convergence behaviour of the small amplitude perturbation expansion. Figure 8.2 shows the solutions obtained for the incoherent component of the mean differential reflection coefficient, averaged over 4500 realizations and for $N_c = 50$, when the perturbative expansion for the reflection amplitude is truncated at the k^{th} order (included) for $k \in [\![1, 6]\!]$. We can see in Fig. 8.2(a) that for $k = 1$ we recover rather closely the analytically averaged result, which deviates significantly from the reference solution. For $k = 2$ (Fig. 8.2(b)) the perturbative solution deviates even more from the reference solution than did that for $k = 1$. This is a good illustration that one should not necessarily always trust a higher order approximation without having an idea of the error bounds, or hints of convergence. The perturbation method may often start by deviating more and more from the exact solution before starting to converge. Here the convergence seems to be re-established already for $k = 3$ (Fig. 8.2(c)), and continues to approach the exact solution by oscillating around it as k increases (Fig. 8.2(d-f)).

Concerning the successive iterates of the method of FIRES, the first iterate is closer to the reference solution as already illustrated by the analytically averaged solution (Fig. 8.3(a)). As for SAPT, the second iterates seems to deviate more than the first iterate but not significantly (Fig. 8.3(b)). Then the third and following iterates, shown in Figs. 8.3(c-f), are almost not distinguishable from the reference solution and the deviation lies within the statistical noise. Based on these illustrations of the convergence of SAPT and the method of FIRES, the latter one would be recommended in practice. We can also mention that the simple iterative methods such as the Jacobi method or the successive over-relaxation also perform similarly well. An advantage of SAPT or the method of FIRES over the standard iterative methods for solving linear systems, is that they yield closed form approximations which may be interpreted on physical ground as we will see in Chapter 10.

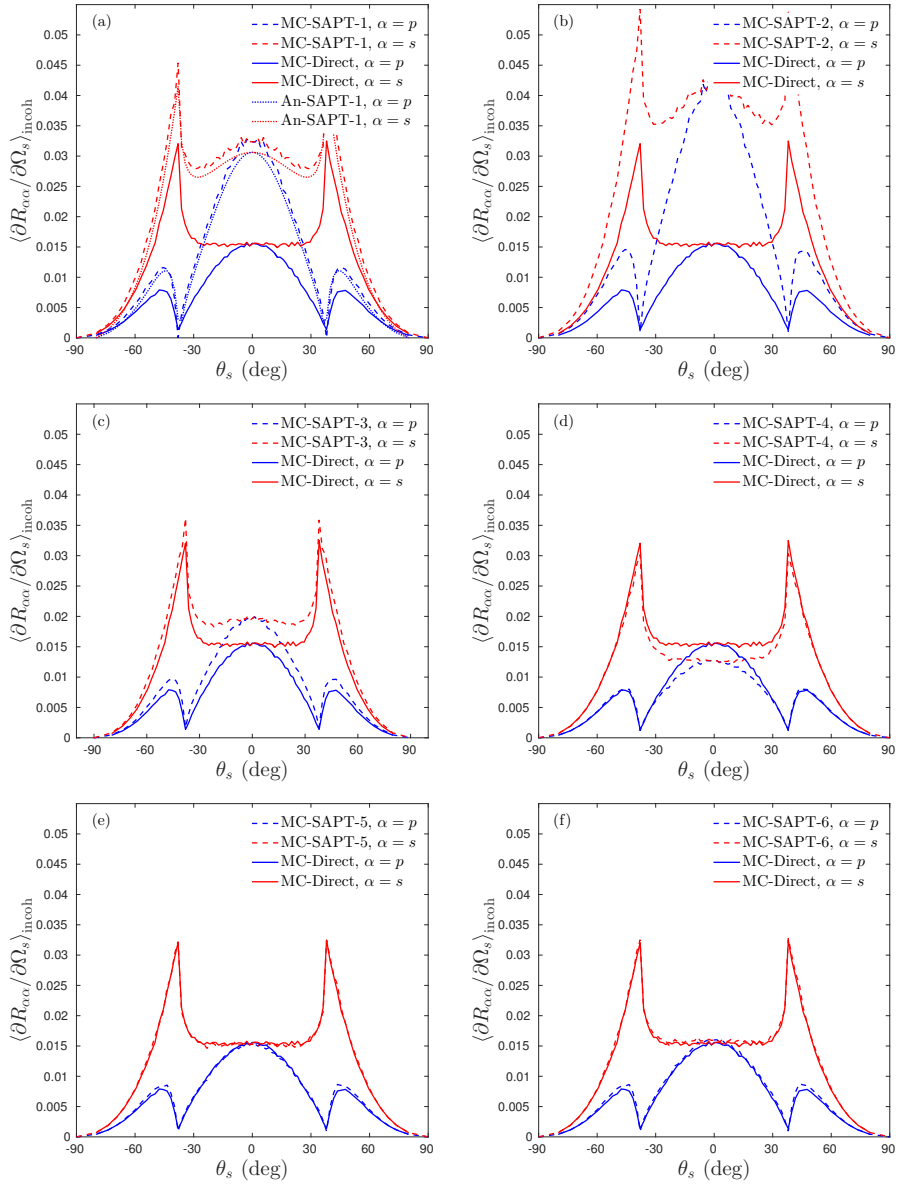


Figure 8.2: Incoherent component of the mean differential reflection coefficient for in-plane co-polarized scattering as a function of the polar scattering angle θ_s, for normal incidence. In all panels, the solid lines show results obtained from the direct method and an ensemble average over 4500 realizations of the surface profile (courtesy of Øyvind Hetland [45]); (a) dotted lines: results obtained by using SAPT to order one in the surface profile of the reflection amplitude analytically averaged (An-SAPT-1); (a-f) dashed lines: results obtained by the Monte Carlo SAPT method (MC-SAPT-k) to order $k = 1$ to 6 (resp. for panels (a) to (f)). Blue: p-polarized light; red: s-polarized light. The physical parameters are the same as the one from Fig. 8.1(b). Figure taken from **Paper** [4].

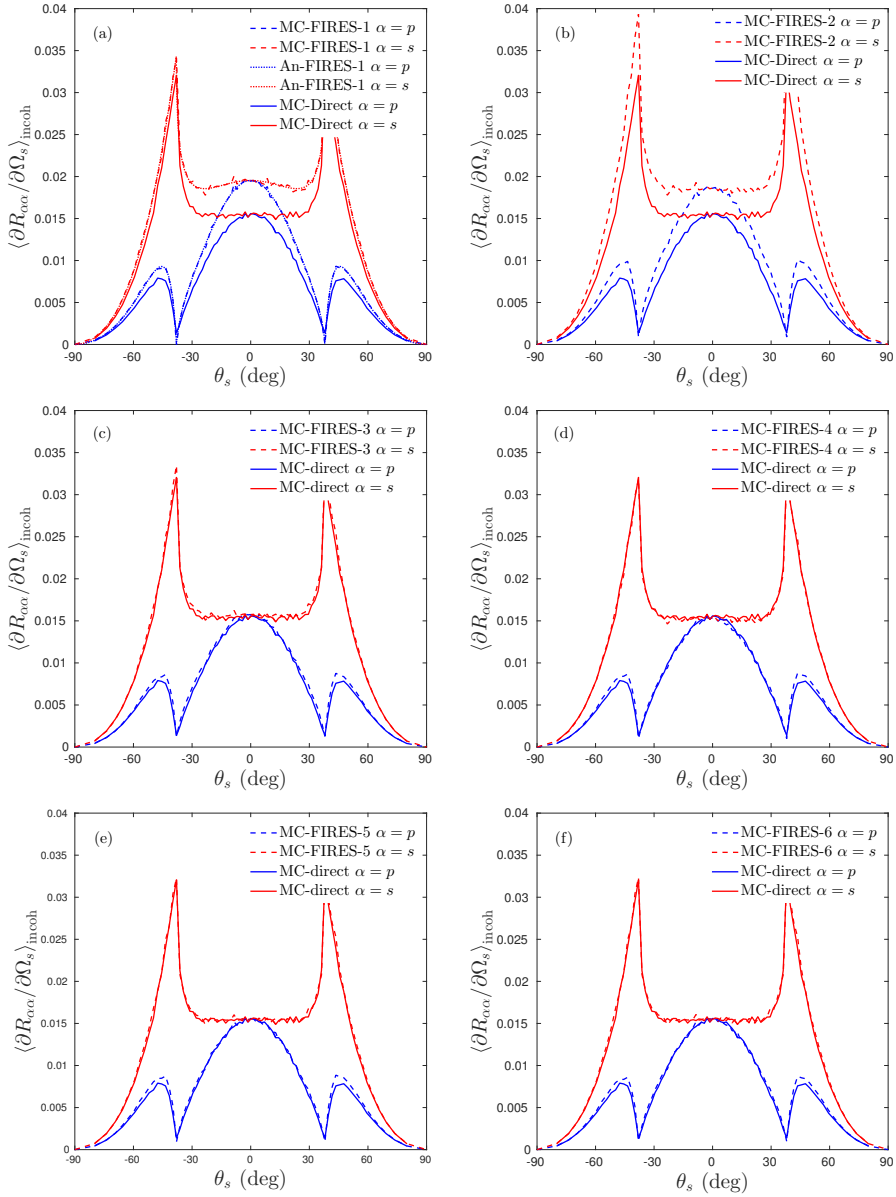


Figure 8.3: Incoherent component of the mean differential reflection coefficient for in-plane co-polarized scattering as a function of the polar scattering angle θ_s, for normal incidence. In all panels, the solid lines show results obtained from the direct method and an ensemble average over 4500 realizations of the surface profile (courtesy of Øyvind Hetland [45]); (a) dotted lines: results obtained by using the first iterate of the method of FIRES analytically averaged (An-FIRES-1); (a-f) dashed lines: results obtained by the Monte Carlo FIRES method (MC-FIRES-k) for iterates from 1 to 6 (resp. for panels (a) to (f)). Blue: p-polarized light; red: s-polarized light. The physical parameters are the same as the one from Fig. 8.1(b). Figure taken from **Paper** [4].

8.7 Reduction of variance

We have seen in the previous sections essentially three strategies to estimate the coherent and incoherent components of the mean differential scattering coefficients. The first one, the brute force Monte Carlo method consisted in solving the scattering problem for each surface realization by the use of the direct method and then in averaging observables over realizations of the surface profile function. The advantage of this method is that no approximation is made, apart from the truncation in wave-space and discretization of the surface and in wave-space. The two main inconvenients are that each scattering problem takes significant computational resources since we need to set up and solve a large linear system for each surface realization and that many realizations are needed to obtain a satisfactory small variance for the observables compared to their estimated averages. We have also seen that solving the linear system with iterative methods can considerably reduce the computational cost associated with solving each scattering problem, which is of great practical interest. Nevertheless, for each surface realization, the solution will be the same (within some set tolerance) whether the linear system is solved iteratively or not. The statistical noise hence remains identical, and we still need to average the observables over many realizations. Alternatively, we have seen that some approximation methods like small amplitude perturbation theory and the method of FIRES not only can be used as iterative methods but allow for simple closed form analytically averaged observables, but such expressions are simple only for the lowest order. They can provide important physical insights, and have an almost free computational cost, but when the roughness parameters of the surface become too large, the quantitative deviation to the exact solution increases.

In this section, we present an original method that makes use of the analytically approximated observables to *reduce the variance* of the observables estimated by the naive Monte Carlo method. The technique is known as reduction of variance by the use of a *control variable* [87, 88].

Control variable based on an analytically averaged approximation

The idea of reduction of variance by the use of a control variable is the following. Let X be a random variable of expected value $\langle X \rangle$ and variance $\mathbb{V}\mathrm{ar}[X] = \langle X^2 \rangle - \langle X \rangle^2$. Let $\mathbf{x} = (x_w)_{1 \leq w \leq N}$ be $N > 2$ realizations of the random variable X. We can estimate the expected value and the variance of X by

$$\langle X \rangle \approx \frac{1}{N} \sum_{w=1}^{N} x_w \stackrel{\text{def}}{=} \mathrm{E}(\mathbf{x}) \tag{8.35a}$$

$$\mathbb{V}\mathrm{ar}[X] \approx \frac{1}{N-1} \sum_{w=1}^{N} (x_w - \mathrm{E}(\mathbf{x}))^2 \stackrel{\text{def}}{=} \mathrm{V}(\mathbf{x}) \,. \tag{8.35b}$$

The uncertainty associated with the estimated expected value can be measured by

$$\mathbb{V}\mathrm{ar}\left[\mathrm{E}(\mathbf{x}) - \langle X \rangle\right] = \frac{\mathbb{V}\mathrm{ar}[\mathbf{x}]}{N} \approx \frac{\mathrm{V}(\mathbf{x})}{N} \,. \tag{8.36}$$

This last estimate tells something simple, that we have already experienced earlier, namely that the more realizations of the random variable we have, the more certain we are about its average estimated by Eq. (8.35a). In addition, for a set tolerance on our estimate of the average, a random variable with small variance will require fewer realizations than one with a large

variance to reach the tolerance. The goal of reduction of variance techniques, is to construct a random variable Z that has the same expected value as X, which is easy to compute, and with smaller variance than that of X. Hence instead of simulating X, it is more advantageous to simulate Z instead since fewer realizations will be necessary to reach an accurate estimation of the expected value. The method of the control variable is one method for doing exactly this. Consider any random variable Y whose expected value is zero. Then the random variable $Z = X - Y$ has the property that $\langle Z \rangle = \langle X \rangle$. However, the variance of Z is *a priori* different from that of X. Indeed, we have

$$\mathbb{V}\mathrm{ar}[Z] = \mathbb{V}\mathrm{ar}[X] + \mathbb{V}\mathrm{ar}[Y] - 2\mathbb{C}\mathrm{ov}[X, Y] \,, \tag{8.37}$$

with $\mathbb{C}\mathrm{ov}[X, Y] \stackrel{\text{def}}{=} \langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle$. The question now is following: How do we construct a random variable Y, called a *control variable*, such that the variance of Z is minimized, and is hopefully smaller than that of X? A first observation shows that any choice of control variable will not reduce the variance. Indeed, if X and Y are not correlated then the variance of Z will clearly be larger than that of X. Then it is expected that X and Y should be strongly correlated. Let us now ask the following question. If we knew exactly $\langle X \rangle$, what would be the best, ideal, choice for Y? Well, it does not require much effort to see that if we knew $\langle X \rangle$, then the ideal candidate for Y is $Y = X - \langle X \rangle$. Indeed, in that case we have

$$\mathbb{V}\mathrm{ar}[Z] = \mathbb{V}\mathrm{ar}[X] + \mathbb{V}\mathrm{ar}[X] - 2\mathbb{V}\mathrm{ar}[X] = 0 \,.$$

What this means is simply that if we knew $\langle X \rangle$, then $Y = X - \langle X \rangle$ is the fluctuation of X around its average and Z is X to which the fluctuation is removed, that is precisely $\langle X \rangle$. Of course, this is not of any practical interest because if we could remove exactly the fluctuation of X to X, that means that we already know $\langle X \rangle$. However, this trivial example illustrates very well what is the essence of the method. *We want to construct a control variable Y that behaves similarly to the fluctuations of X.* Thus, finding a good control variable Y will depend in general on the problem, and what the random variable X is actually modeling, and requires us to understand or to be able to approximate the fluctuation of X.

It is now time to quit the beautiful abstract world of mathematics and come back to our concrete scattering problem. In our case, the role of the random variable X would typically be $X = R_{\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0)$ or $X = |R_{\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0)|^2$ (remember what needs to be estimated to compute the incoherent component of the mean differential reflection coefficient for example). Each realization of X, say $x_w = R_{\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0, \zeta_w)$, is simulated by either the direct or an iterative method up to a set tolerance for each surface realization ζ_w. Now what should we choose for the control variable Y? We want to construct of a control variable that resembles the fluctuation of X and is easy (cheap) to compute. What if we used the deviation from a realization of the solution given by an approximate method such as, small amplitude perturbation theory or the method of FIRES to lowest order, from its analytical average? Indeed, if we had to bet on a good control variable that resembles the fluctuation of the rigorous method, why not bet on the fluctuation of an approximate method, fluctuation which, can be extracted exactly since we already know the average in closed form? Physically, this means that we bet that for a given surface realization, the speckle pattern obtained by the approximation method would be somewhat close to that obtain with the rigorous method, precisely because the two methods try to solve the same scattering problem for the same surface. Of course, there is no guarantee that the two speckle patterns will fully match, in fact they most probably will not. However, it seems intuitive to think that the two speckle patterns, produced by the two methods, will be correlated, and correlation is all we ask for. So this is our bet! If we now denote by $x_w^{(\mathrm{rig})}$ $(= R_{\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0, \zeta_w)$ or

$|R_{\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0, \zeta_w)|^2$, etc) the solution obtained by the rigorous method for the realization of the surface profile function ζ_w and by $x_w^{(\mathrm{app})}$ the corresponding solution for the *same* realization of the surface profile function computed with an approximation method (for which the average is known in closed form), and finally by $\left\langle X^{(\mathrm{app})} \right\rangle$ the exact average, analytically evaluated, then the corresponding realization of the control variable reads

$$y_w = x_w^{(\mathrm{app})} - \left\langle X^{(\mathrm{app})} \right\rangle , \tag{8.38}$$

and the true average of X is estimated by

$$\langle X \rangle \approx \frac{1}{N_{\mathrm{surf}}} \sum_{w=1}^{N_{\mathrm{surf}}} z_w = \frac{1}{N_{\mathrm{surf}}} \sum_{w=1}^{N_{\mathrm{surf}}} x_w^{(\mathrm{rig})} - y_w \, . \tag{8.39}$$

Concretely, if $X = |R_{\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0)|^2$ we have

$$\left\langle |R_{\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0)|^2 \right\rangle \approx \frac{1}{N_{\mathrm{surf}}} \sum_{w=1}^{N_{\mathrm{surf}}} \left[|R_{\alpha\beta}^{(\mathrm{rig})}(\mathbf{p} \mid \mathbf{p}_0, \zeta_w)|^2 - |R_{\alpha\beta}^{(\mathrm{app})}(\mathbf{p} \mid \mathbf{p}_0, \zeta_w)|^2 + \left\langle |R_{\alpha\beta}^{(\mathrm{app})}(\mathbf{p} \mid \mathbf{p}_0)|^2 \right\rangle \right] . \tag{8.40}$$

Let us once again stress what this means. Each term of the sum is composed of a solution realization computed with the rigorous simulation $|R_{\alpha\beta}^{(\mathrm{rig})}(\mathbf{p} \mid \mathbf{p}_0, \zeta_w)|^2$ to which we subtract the deviation of the approximate solution from its exactly known average $|R_{\alpha\beta}^{(\mathrm{app})}(\mathbf{p} \mid \mathbf{p}_0, \zeta_w)|^2 - \left\langle |R_{\alpha\beta}^{(\mathrm{app})}(\mathbf{p} \mid \mathbf{p}_0)|^2 \right\rangle$. The hope is that if the fluctuations of the approximate solution are similar to that of the exact one, then we effectively partly remove the fluctuation in the rigorous method and the variance is reduced.

Remark 8.6. Note that even if we make a poor choice of control variable, say it ends up to be uncorrelated from the fluctuations of X, then the expected value remains unchanged since the expected value of the control variable is zero. So in the worst case scenario, we simply have not managed to reduce the variance and still need many realizations to reach an accurate estimate of the expected value. Therefore there is no risk of getting a convergence to a wrong expected value.

Remark 8.7. The chosen control variable here was arbitrary but motivated by the idea that the approximate solution should be somewhat close to the rigorous one. However, we know that there will be deviation between the averages of the two methods (see for example the comparisons in the previous section). So we may expect that the fluctuations produced by the approximation method are indeed correlated to those produced by the rigorous method but may have a somewhat stronger or weaker scaling. This means that the control variable can, in principle, be optimized further. Indeed, we can multiply the control variable Y by a scaling factor η and optimize η to minimize the resulting variance of $Z(\eta) = X - \eta Y$,

$$\mathbb{V}\mathrm{ar}[Z(\eta)] = \mathbb{V}\mathrm{ar}[X] - 2\eta \mathbb{C}\mathrm{ov}[X, Y] + \eta^2 \mathbb{V}\mathrm{ar}[Y] \, . \tag{8.41}$$

$\mathbb{V}\mathrm{ar}[Z(\eta)]$ is a convex quadratic function of η and hence has a unique minimum at

$$\eta_{\mathrm{opt}} = \frac{\mathbb{C}\mathrm{ov}[X, Y]}{\mathbb{V}\mathrm{ar}[Y]} \, . \tag{8.42}$$

At the optimum, the variance of Z thus reads

$$\mathbb{V}\mathrm{ar}[Z(\eta_{\mathrm{opt}})] = \mathbb{V}\mathrm{ar}[X] \left(1 - \frac{\mathbb{C}\mathrm{ov}[X, Y]^2}{\mathbb{V}\mathrm{ar}[X]\mathbb{V}\mathrm{ar}[Y]} \right) = \mathbb{V}\mathrm{ar}[X] \left(1 - \rho_{X,Y}^2 \right) , \tag{8.43}$$

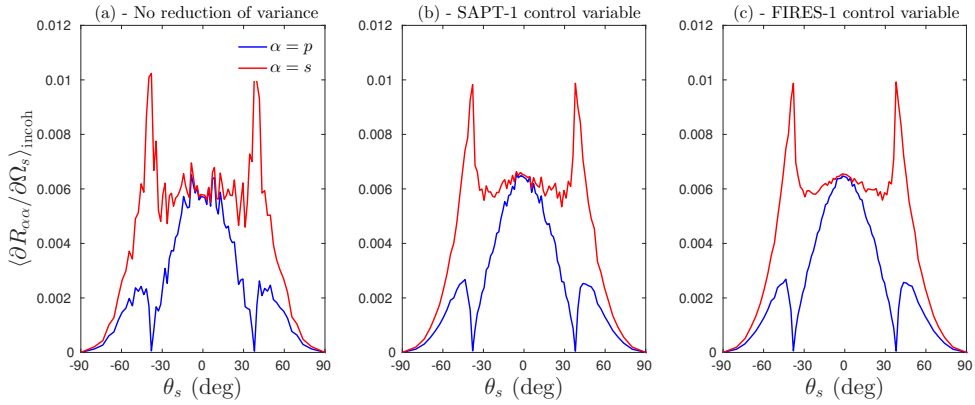


Figure 8.4: Incoherent component of the mean differential reflection coefficient for in-plane co-polarized scattering as a function of the polar scattering angle θ_s, for normal incidence. (a) Simulation without reduction of variance, (b) simulation with reduction of variance based on a SAPT control variable and $\eta = \eta_{\text{opt}}$, (c) simulation with reduction of variance based on a FIRES control variable and $\eta = \eta_{\text{opt}}$. All simulation results were obtained by averaging over 100 realizations of the surface profile function. Blue: p-polarized light; red: s-polarized light. Physical parameters: $\epsilon_1 = 2.64$, $\epsilon_2 = 1$, $a = \lambda/4$, $\sigma = \lambda/40$. Figure taken from **Paper** [4].

where it is now explicitly clear from the second term in the parenthesis that the closer to one the correlation $\rho_{X,Y}$, between X and Y is, the more significant the variance reduction. It is then convenient to define the *variance reduction factor* vrf defined as

$$\text{vrf} \overset{\text{def}}{=} \frac{\mathbb{V}\text{ar}[Z]}{\mathbb{V}\text{ar}[X]} , \tag{8.44}$$

which in the case of the optimum control variable $Z(\eta_{\text{opt}})$ reads

$$\text{vrf} = 1 - \rho^2_{X,Y} . \tag{8.45}$$

Note that in practice, we only have access to estimates of the variances of X and Y and the covariance of X and Y. The optimum scaling parameter η_{opt} must be replaced by its estimated value given by replacing $\mathbb{V}\text{ar}[Y]$ and $\mathbb{C}\text{ov}[X, Y]$ by their respective estimates in Eq. (8.42). One way of doing this is to use the n first realizations to estimate η_{opt} and the $N_{\text{surf}} - n$ remaining realizations to apply the control variable. Alternatively, we can update the estimate of η_{opt} for each realization as we gather statistics.

Example

It is now time to verify whether our bet was a good one. In the following, we present a set of three simulation results obtained on the basis of the iterative Monte Carlo method using the Jacobi method (see Chapter 6) since it is faster than the direct method for each realization and give accurate results. The first set of results is obtained without applying the method of reduction of variance and will serve as a reference. The second and third sets of results are obtained by applying the method of reduction of variance presented above, based on a control

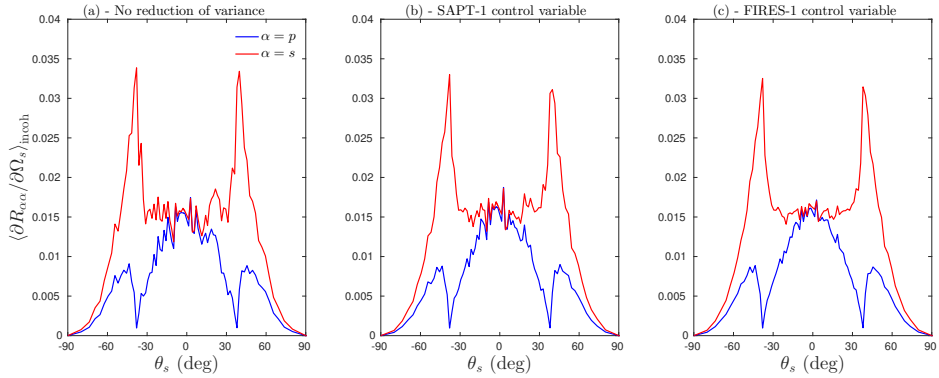


Figure 8.5: Same as Fig. 8.4, but for $\sigma = \lambda/20$. Figure taken from **Paper** [4].

variable constructed either from the first iterate of the method of FIRES or small amplitude perturbation theory to first order in the surface profile (for the reflection amplitude). The control variables are optimized, i.e. $\eta = \eta_{\mathrm{opt}}$ as previously explained. All the simulations are run with the following parameters, similar to those used previously: $\epsilon_1 = 2.64$, $\epsilon_2 = 1$, $\sigma \in \{\lambda/40, \lambda/20\}$, $a = \lambda/4$, $N_c = 50$, $N_{\mathrm{surf}} = 100$. For the computation of the optimal value of η, we simply store the results obtained for all the simulations, and we estimate η_{opt} according to Eq. (8.42). We focus here on the method applied to the incoherent component of the mean differential coefficient. Note that we obtain a value η_{opt} for each sampled $\mathbf{p}$ given an incident in-plane wave vector $\mathbf{p}_0$, or equivalently for each scattering angles (θ_s, ϕ_s), and for each polarization coupling, hence $\eta_{\mathrm{opt}} = \eta_{\mathrm{opt},\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0)$. We will restrict ourselves to co-polarized scattering in the plane of incidence for simplicity.

Figure 8.4 shows the incoherent component of the mean differential reflection coefficient for co-polarized scattering in the plane of incidence as a function of the polar scattering angle for the aforementioned three sets of simulations and assuming $\sigma = \lambda/40$. By visual comparison of the sets of data, it seems clear that the method of reduction of variance by the use of a control variable works well. Indeed, the statistical noise seems to be lower in the plots presented in Figs. 8.4(b) and 8.4(c) than in the plot presented in Fig. 8.4(a). The aim of reducing the variance is then achieved. Still by visual inspection of Figs. 8.4(b) and 8.4(c), it seems clear that the statistical noise is smaller when the FIRES control variable is used as compared with the SAPT control variable. The control variable based on FIRES is then more efficient than that based on SAPT for reducing the variance. This is understandable since we have seen previously that the approximation based on the method of FIRES is generally more accurate than SAPT. Thus we expect higher correlation between the approximate speckle pattern and the exact speckle pattern for FIRES than for SAPT. Repeating the numerical experiment for a rougher surface, $\sigma = \lambda/20$, we can see in Fig. 8.5 that the reduction of variance is still working relatively well but the statistical noise seems not to be reduced as much as in the case of the weakly rough surface ($\sigma = \lambda/40$, compare Fig. 8.5 with Fig. 8.4). This can be understood by the fact that for higher rms-roughness the approximate solutions deviate significantly from the exact solution and the speckle correlations between the approximate solutions and the exact one decrease, and consequently, the reduction of variance becomes less efficient.

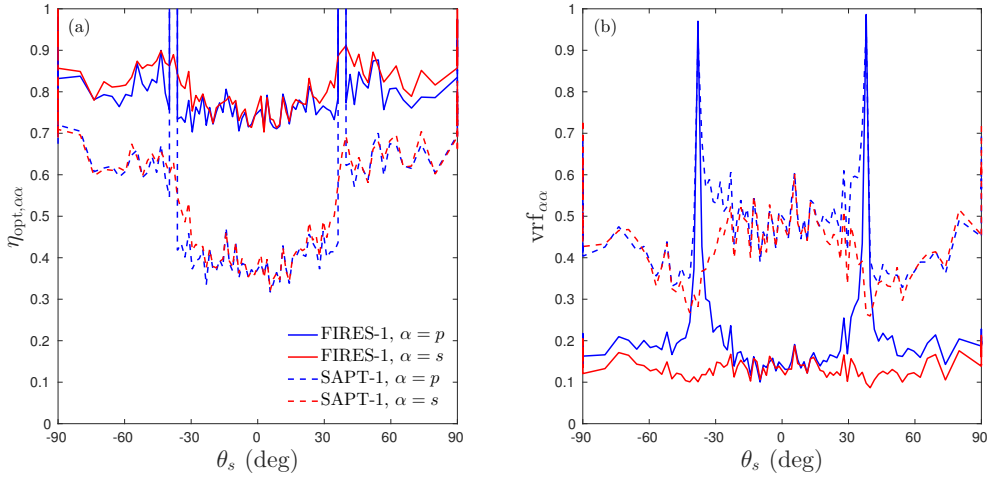


Figure 8.6: (a) Optimal scaling parameter $\eta_{\mathrm{opt},\alpha\alpha}$ and (b) corresponding variance reduction factor $\mathrm{vrf}_{\alpha\alpha}$ as a function of the polar scattering angle. The solid lines correspond to the case where the control variable is based on the first iterate of the method of FIRES and the dashed lines correspond to the case where the control variable is based on small amplitude perturbation theory to first order (for the reflection amplitude). All simulation results were obtained by averaging over 500 realizations of the surface profile function in order to obtain satisfactory averages. Blue: p-polarized light; red: s-polarized light. The physical parameters are the same as the one from Fig. 8.1(b). Figure taken from **Paper** [4].

Figure 8.6(a) shows the value of the optimal scaling parameter η_{opt} as a function of the polar scattering angle in the case of the rougher surface $\sigma = \lambda/20$. It can be seen that the optimal value roughly oscillates around 0.8 for FIRES and 0.5 for SAPT in average over scattering angle and explodes for p-polarized light at the critical angle (due to the zero of the scattering intensity at the critical angle). The optimal scaling factor is smaller for SAPT than for FIRES since it has to compensate for a large deviation of the approximation which, in both cases, overestimates the exact solution (see Fig. 8.1(b)). Figure 8.6(b) shows the variance reduction factor in the case where the scaling parameter η is optimized. The variance reduction factor reaches an average value (over scattering angles) around 0.2 fro FIRES. Therefore one would need about five times fewer realizations to reach a desired variance on the estimated average as compared to what would one need without applying the method of reduction of variance. Concerning the control variable based on small amplitude perturbation theory, the control variable does not reduce the variance as much as that based on the method of FIRES, even when optimized, as can be seen from Fig. 8.6(b). The variance is reduced only by about 50% in average over the scattering angle. This is due to the fact that the perturbation approximation is not as good as the FIRES approximation, thus the correlation between the fluctuations of the exact solution and that of the approximation is smaller.

The method of reduction of variance is then expected to work best with decreasing level of

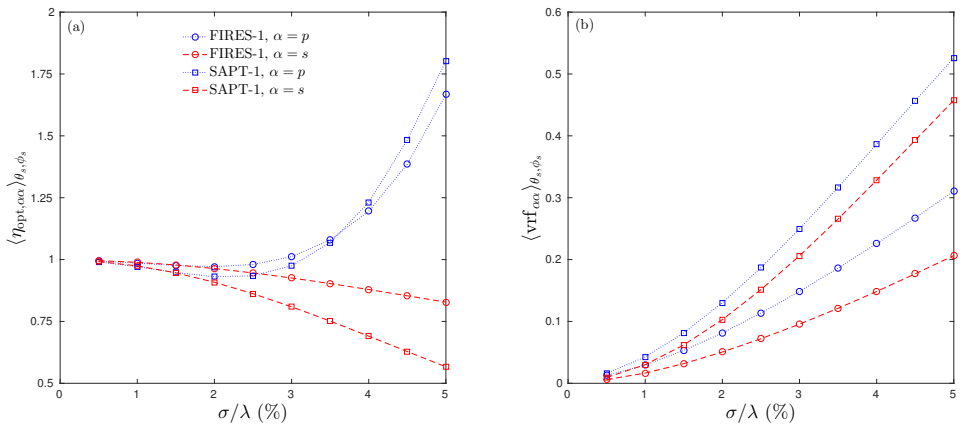


Figure 8.7: (a) Optimal scaling parameter η_{opt} and (b) variance reduction factor vrf as a function of the polar scattering angle. The solid lines correspond to the case where the control variable is based on the first iterate of the method of FIRES and the dashed lines correspond to the case where the control variable is based on small amplitude perturbation theory to first order (for the reflection amplitude). All simulation results were obtained by averaging over 500 realizations of the surface profile function. Blue: p-polarized light; red: s-polarized light. The physical parameters are the same as the one from Fig. 8.1(b). Figure taken from **Paper** [4].

roughness since the approximations then become more accurate. To illustrate this fact, let us now quantify the variance reduction factor averaged over scattering angles as a function of the rms-roughness. Figure 8.7 shows how the optimal scaling factor (Fig. 8.7(a)) and the corresponding variance reduction factor (Fig. 8.7(b)) averaged over scattering angles evolve as a function of the rms-roughness (other parameters such as the correlation length and the dielectric constants are kept fixed). The average optimal scaling factor $\langle \eta_{\mathrm{opt},ss} \rangle_{\theta_s,\phi_s}$ decays monotonically from unity for $\sigma \to 0$ to values around 0.8 and 0.6 for the FIRES and SAPT control variables respectively for $\sigma = 0.05\lambda$, as shown in Fig. 8.7(a). Such a behavior is consistent with our previous observations in Figs. 8.1 and 8.6. On the one hand, as the rms-roughness is decreased, the agreement between the approximations and the exact solution becomes better and no scaling is need, hence $\eta_{\mathrm{opt}} \to 1$ when $\sigma \to 0$. On the other hand, the approximations deviate more and more from the exact solution for increasing σ, over-estimating it for the considered system (see e.g. Fig. 8.1(b)), hence the decay of the scaling factor. In particular, since SAPT deviates faster than FIRES from the exact solution with increasing σ, the corresponding scaling factor decays faster. In the case of p-polarized light, the angle-averaged scaling factor starts to decrease for small values of the rms-roughness before increasing for larger rms-roughness (for roughly $\sigma/\lambda > 2\%$). The increasing behavior is an artifact due to the divergence of the scaling factor at the Brewster scattering angle. By excluding the Brewster scattering directions from the angle-average, the scaling factor can be shown to decay in a similar way as for s-polarized light. Now turning to the variance reduction factor, we observe, as expected, that the angle-averaged vrf increases with the rms-roughness as illustrated in Fig. 8.7(b). In other words, the reduction of variance becomes less and less efficient with increasing roughness as the approximations become poorer and poorer. Not surprisingly, the variance reduction factor

for the control variable based on the method of FIRES is always smaller than that based on SAPT. We note that as the rms-roughness goes to zero the vrf goes to zero, which is consistent with the fact that both approximations converge towards the exact solution in the limit of the planar surface. Figure 8.7(b) shows that $\langle \mathrm{vrf}_{\alpha\alpha} \rangle_{\theta_s,\phi_s}$ is consistently greater for p-polarized light than for s-polarized light which can partly be explained due to the Brewster scattering artifact.

The presented method of reduction of variance does not require any significant additional computational burden and can be considered as free compared to solving the reduced Rayleigh equation for a given surface realization, even when using an iterative method. The presented method hence has a rather interesting computational cost to variance reduction ratio, and we would recommend it in practice, in particular with the use of a control variable based on the FIRES approximation. As illustrated in Fig. 8.7(b), the reduction of variance is significantly reduced for very weakly rough surfaces, and only a few realizations would be needed to obtain a trustful average. However, for such very weakly rough surfaces, it may be argued that the analytically averaged approximations already represent excellent solutions and there is no real need for the Monte Carlo simulation. The real interest of the application of reduction of variance techniques lies in the intermediate roughness regime for which analytically averaged approximations deviate significantly from the exact solution, but not too much as to have the potential of yielding a significant reduction of variance. In the case of rougher surfaces, more accurate approximations should be developed in order to increase the correlation between the speckles obtained by the approximate and the exact solution. Such approximations could be made by attempting to average observables based on higher orders of SAPT (see e.g. Soubret *et al.* [53]) or higher iterates of the method of FIRES although such a process becomes cumbersome. Finally, when the surface becomes too rough, it should be kept in mind that the reduced Rayleigh equations are expected to fail to model the problem of scattering as was illustrated in Chapter 7. Other approaches must then be considered.

8.8 Summary

The main results obtained in the present chapter are the following. In complement to the study of the convergence of iterative methods for sinusoidal surfaces in Chapter 6, some illustrative examples of the use of iterative methods have been given in the case of randomly rough surfaces. The use of iterative methods was shown to yield a significant speed up for solving each scattering problem involved in the Monte Carlo simulation of the average intensity diffusely scattered by random surfaces. Moreover, we have derived closed form approximations for the mean differential scattering coefficients obtained from the scattering amplitudes given by SAPT and the method of FIRES to order one. These expressions are accurate for weakly rough surfaces and have an insignificant computational cost. Finally, we have exposed a technique of reduction of variance based on control variables built from the aforementioned approximations, and we have shown that the variance of the averaged observables can indeed be reduced on a few examples. The variance reduction factor was also quantified as a function of the rms-roughness. It was shown to increase (i.e. become worse) as the rms-roughness is increased which intrinsically is linked to the fact that the approximations become poorer with increasing rms-roughness hence degrading the correlations between the approximate speckle pattern and the exact one.

Chapter 9

More on perturbation theory

"There are two great rules of life: 1) never tell everything at once."
Ken Venturi.

In the present chapter, we develop the framework of small amplitude perturbation theory on two fronts. First, we extend the derivation of SAPT which was presented in Section 6.2 to the case of a multi-layer system for which the reduced Rayleigh equations were derived in Chapter 3. As for the case of a single interface, we will obtain recursive expressions to determine iteratively the successive orders of the expansion for the reflection and transmission amplitudes. In particular, special emphasis will be given to the amplitudes expanded to first order and the corresponding incoherent component of mean differential reflection and transmission coefficients. Within the first order approximation for the amplitudes, we will show that: (i) In the case of non-cross correlated interfaces, the overall incoherent component of mean differential reflection (resp. transmission) coefficient is the sum of the mean differential reflection (resp. transmission) coefficient for the corresponding systems for which only one of the interface is rough at a time and the remaining are planar; (ii) In the case of cross-correlated interfaces, cross terms induced by the interface-cross-correlations must be taken into account in addition to the uncorrelated case. This results will be at the basis of the study of Selényi rings and their interface-cross-correlation induced enhancement and attenuation which will be studied in Chapter 10. The second development of SAPT which is suggested in this chapter is to consider perturbation of a structured surface, like a periodic grating for example. So far we have considered a surface profile to be a perturbation of a planar surface. Instead of considering that the reference interface is planar, we will assume that the reference system is a structured surface, for which we assume that the reduced Rayleigh equations can be solved, and then we study small perturbations of the reference structure. A typical example of interest would be the study of the influence of small defects, or roughness, on top of a clean periodic grating. The hope of the presented method is that it may yield within a first order of the perturbation more accurate results than SAPT applied directly to the whole surface seen as a perturbation of the planar interface for somewhat higher order. Indeed, proceeding by perturbation of the structured surface, the zero order corresponds to the exact solution of the structured unperturbed surface, which would take *a priori* infinitely many orders to represent if it were considered itself as a perturbation of a planar surface. In other words, the reference unperturbed system would contain all the multiple scattering events and the perturbed system, whose amplitude is expanded, say, to order one, would yield a single scattering correction to the unperturbed solution for the structured system.

9.1 Generalization of SAPT to layered systems

We have seen in Chapter 6 how to compute the perturbative solution to any given order for a system composed of two media separated by a single randomly rough or periodic interface. What about generalizing this calculus to a layered system such as the one presented in Chapter 3? To this end, let us consider the system composed of a stack of $n+1$ media separated by n interfaces from Chapter 3. By using the notations introduced there, we know that the reduced Rayleigh equations for the reflection amplitude is given by Eq. (3.20)

$$\int \boldsymbol{\Theta}_{n+1,1}^{+,+}(\mathbf{p}\,|\,\mathbf{q})\;\mathbf{R}(\mathbf{q}\,|\,\mathbf{p}_{\mathrm{o}})\,\frac{\mathrm{d}^2q}{(2\pi)^2} = -\,\boldsymbol{\Theta}_{n+1,1}^{+,-}(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}})\;, \tag{9.0}$$

where we recall that the *forward n-interface transfer kernel* is defined as

$$\boldsymbol{\Theta}_{n+1,1}^{a_{n+1},a_1}(\mathbf{p}_{n+1}\,|\,\mathbf{p}_1) = \sum_{a_n=\pm} a_n \int \cdots \sum_{a_2=\pm} a_2 \int \boldsymbol{\Theta}_{n+1,n}^{a_{n+1},a_n}(\mathbf{p}_{n+1}\,|\,\mathbf{p}_n)\cdots\boldsymbol{\Theta}_{2,1}^{a_2,a_1}(\mathbf{p}_2\,|\,\mathbf{p}_1)\,\frac{\mathrm{d}^2p_2}{(2\pi)^2}\cdots\frac{\mathrm{d}^2p_n}{(2\pi)^2}\;, \tag{9.1}$$

with the *single-interface kernels* $\boldsymbol{\Theta}_{l,m}^{b,a}$ defined for successive media, i.e. $l,m\in[\![1,n+1]\!]$ such that $|l-m|=1$, $a,b\in\{\pm\}$, by Eq. (3.4) as

$$\boldsymbol{\Theta}_{l,m}^{b,a}(\mathbf{p}\,|\,\mathbf{q}) = \alpha_l^{-1}(\mathbf{p})\,\mathcal{J}_{l,m}^{b,a}(\mathbf{p}\,|\,\mathbf{q})\;\mathbf{M}_{l,m}^{b,a}(\mathbf{p}\,|\,\mathbf{q})\;. \tag{9.2}$$

The perturbative method consists in expanding each single-interface kernel in a series of Fourier moments as we have done in the previous section. However, in order to avoid unnecessary lengthy expansions, we first introduce some notations that will allow us to keep a compact derivation and that can further be useful for numerical implementation. We define

$$\begin{aligned}
&\tilde{\boldsymbol{\Theta}}_{n+1,1}^{a_{n+1},a_1,(\mathbf{m})}(\mathbf{p}_{n+1}\,|\,\mathbf{p}_n\,|\,\cdots\,|\,\mathbf{p}_2\,|\,\mathbf{p}_1) \overset{\text{def}}{=}\\
&\sum_{a_n=\pm} a_n\,\alpha_{n+1}^{-1}(\mathbf{p}_{n+1})\big[a_{n+1}\alpha_{n+1}(\mathbf{p}_{n+1}) - a_n\alpha_n(\mathbf{p}_n)\big]^{m_n-1}\,e^{-i[a_{n+1}\alpha_{n+1}(\mathbf{p}_{n+1})-a_n\alpha_n(\mathbf{p}_n)]d_n}\;\mathbf{M}_{n+1,n}^{a_{n+1},a_n}(\mathbf{p}_{n+1}\,|\,\mathbf{p}_n)\cdots\\
&\sum_{a_2=\pm} a_2\,\alpha_2^{-1}(\mathbf{p}_2)\big[a_2\alpha_2(\mathbf{p}_2) - a_1\alpha_1(\mathbf{p}_1)\big]^{m_1-1}\,e^{-i[a_2\alpha_2(\mathbf{p}_2)-a_1\alpha_1(\mathbf{p}_1)]d_1}\;\mathbf{M}_{2,1}^{a_2,a_1}(\mathbf{p}_2\,|\,\mathbf{p}_1)\,,
\end{aligned} \tag{9.3}$$

where $\mathbf{m}=(m_1,\cdots,m_n)\in\mathbb{N}^n$ is a multi-index. Here we have made the choice of factorizing the phase factor $e^{-i(a_{j+1}\alpha_{j+1}(\mathbf{p}_{j+1})-a_j\alpha_j(\mathbf{p}_j))d_j}$, with $d_j=\langle\zeta_j\rangle$ being the offset height of the j^{th} interface, from each factor $\mathcal{J}_{j+1,j}^{a_{j+1},a_j}(\mathbf{p}_{j+1}\,|\,\mathbf{p}_j)$ for later convenience. Given this definition, an expansion of the n-interface kernel in Fourier moments is given by

$$\begin{aligned}
\boldsymbol{\Theta}_{n+1,1}^{a_{n+1},a_1}(\mathbf{p}_{n+1}\,|\,\mathbf{p}_1) &= \sum_{\mathbf{m}=0}^{\infty}\frac{(-i)^{|\mathbf{m}|}}{\mathbf{m}!}\int\cdots\int \hat{h}_n^{(m_n)}(\mathbf{p}_{n+1}-\mathbf{p}_n)\cdots\hat{h}_1^{(m_1)}(\mathbf{p}_2-\mathbf{p}_1)\\
&\qquad \tilde{\boldsymbol{\Theta}}_{n+1,1}^{a_{n+1},a_1,(\mathbf{m})}(\mathbf{p}_{n+1}\,|\,\mathbf{p}_n\,|\,\cdots\,|\,\mathbf{p}_2\,|\,\mathbf{p}_1)\,\frac{\mathrm{d}^2p_2}{(2\pi)^2}\cdots\frac{\mathrm{d}^2p_n}{(2\pi)^2}\\
&\overset{\text{def}}{=} \sum_{\mathbf{m}=0}^{\infty}\frac{(-i)^{|\mathbf{m}|}}{\mathbf{m}!}\;\mathbf{Z}_{n+1,1}^{a_{n+1},a_1,(\mathbf{m})}(\mathbf{p}_{n+1}\,|\,\mathbf{p}_1)\,,
\end{aligned} \tag{9.4}$$

where $\sum_{\mathbf{m}=0}^{\infty}\overset{\text{def}}{=}\sum_{m_1=0}^{\infty}\cdots\sum_{m_n=0}^{\infty}$, $|\mathbf{m}|=\sum_{j=1}^n m_j$ is the length of the multi-index, and $\mathbf{m}!=m_1!\cdots m_n!$, and for all $j\in[\![1,n]\!]$,

$$\hat{h}_j^{(m_j)}(\mathbf{q}) = \int e^{-i\,\mathbf{q}\cdot\mathbf{x}}\,(\zeta_j-d_j)^{m_j}(\mathbf{x})\,\mathrm{d}^2x\,, \tag{9.5}$$

is the Fourier moment of $h_j \overset{\text{def}}{=} \zeta_j - d_j$ of order m_j. It is then clear that $\mathbf{Z}_{n+1,1}^{a_{n+1},a_1,(\mathbf{m})}(\mathbf{p}_{n+1}\,|\,\mathbf{p}_1)$, defines by Eq. (9.4), is a term of order $|\mathbf{m}|$ in products of surface profiles. As before, the reflection amplitude can be expanded as in Eq. (6.3b). We are now ready to start the derivation of perturbative expansion. By plugging Eqs. (9.4) and (6.3b) into Eq. (3.20) we obtain

$$\sum_{\substack{\mathbf{m}'=0\\ j=0}}^{\infty} \frac{(-i)^{|\mathbf{m}'|+j}}{\mathbf{m}'!\,j!} \int \mathbf{Z}_{n+1,1}^{+,+,(\mathbf{m}')}(\mathbf{p}\,|\,\mathbf{q})\ \mathbf{R}^{(j)}(\mathbf{q}\,|\,\mathbf{p}_\mathrm{o})\,\frac{\mathrm{d}^2q}{(2\pi)^2} = -\sum_{\mathbf{m}=0}^{\infty} \frac{(-i)^{|\mathbf{m}|}}{\mathbf{m}!}\,\mathbf{Z}_{n+1,1}^{+,-,(\mathbf{m})}(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o})\ . \tag{9.6}$$

Summing over all multi-index $\mathbf{m}$ is equivalent to summing over subsets $\mathcal{S}_m = \{\mathbf{m} \in \mathbb{N}^n \mid |\mathbf{m}| = m\}$ of multi-index of constant length m, i.e. that we have $\sum_{\mathbf{m}=0}^{\infty} \equiv \sum_{m=0}^{\infty}\sum_{\mathbf{m}\in\mathcal{S}_m}$, therefore the previous equation can be re-written as

$$\sum_{\substack{m'=0\\ j=0}}^{\infty}\sum_{\mathbf{m}'\in\mathcal{S}_{m'}} \frac{(-i)^{m'+j}}{\mathbf{m}'!\,j!} \int \mathbf{Z}_{n+1,1}^{+,+,(\mathbf{m}')}(\mathbf{p}\,|\,\mathbf{q})\ \mathbf{R}^{(j)}(\mathbf{q}\,|\,\mathbf{p}_\mathrm{o})\,\frac{\mathrm{d}^2q}{(2\pi)^2} = -\sum_{m=0}^{\infty}\sum_{\mathbf{m}\in\mathcal{S}_m} \frac{(-i)^{m}}{\mathbf{m}!}\,\mathbf{Z}_{n+1,1}^{+,-,(\mathbf{m})}(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o})\ . \tag{9.7}$$

We then use the definition of the multinomial coefficient in multi-index form as $|\mathbf{m}|!/\mathbf{m}! = \binom{|\mathbf{m}|}{\mathbf{m}}$ to obtain

$$\begin{aligned}
&\sum_{\substack{m'=0\\ j=0}}^{\infty} \frac{(-i)^{m'+j}}{m'!\,j!} \sum_{\mathbf{m}'\in\mathcal{S}_{m'}} \binom{m'}{\mathbf{m}'} \int \mathbf{Z}_{n+1,1}^{+,+,(\mathbf{m}')}(\mathbf{p}\,|\,\mathbf{q})\ \mathbf{R}^{(j)}(\mathbf{q}\,|\,\mathbf{p}_\mathrm{o})\,\frac{\mathrm{d}^2q}{(2\pi)^2}\\
&= -\sum_{m=0}^{\infty} \frac{(-i)^{m}}{m!} \sum_{\mathbf{m}\in\mathcal{S}_m} \binom{m}{\mathbf{m}}\,\mathbf{Z}_{n+1,1}^{+,-,(\mathbf{m})}(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o})\ .
\end{aligned}$$

We now make a change of summation index $j \leftrightarrow m - m'$ on the left hand side of the above equation. This makes clearly appear terms of order m in product of surface profiles. We obtain

$$\begin{aligned}
&\sum_{m=0}^{\infty}\sum_{m'=0}^{m} \frac{(-i)^{m}}{m'!\,(m-m')!} \sum_{\mathbf{m}'\in\mathcal{S}_{m'}} \binom{m'}{\mathbf{m}'} \int \mathbf{Z}_{n+1,1}^{+,+,(\mathbf{m}')}(\mathbf{p}\,|\,\mathbf{q})\ \mathbf{R}^{(m-m')}(\mathbf{q}\,|\,\mathbf{p}_\mathrm{o})\,\frac{\mathrm{d}^2q}{(2\pi)^2}\\
&= -\sum_{m=0}^{\infty} \frac{(-i)^{m}}{m!} \sum_{\mathbf{m}\in\mathcal{S}_m} \binom{m}{\mathbf{m}}\,\mathbf{Z}_{n+1,1}^{+,-,(\mathbf{m})}(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o})\,,
\end{aligned}$$

which can be re-written with the use of the definition of the binomial coefficient $\binom{m}{m'}$ as

$$\begin{aligned}
&\sum_{m=0}^{\infty} \frac{(-i)^{m}}{m!} \sum_{m'=0}^{m} \binom{m}{m'} \sum_{\mathbf{m}'\in\mathcal{S}_{m'}} \binom{m'}{\mathbf{m}'} \int \mathbf{Z}_{n+1,1}^{+,+,(\mathbf{m}')}(\mathbf{p}\,|\,\mathbf{q})\ \mathbf{R}^{(m-m')}(\mathbf{q}\,|\,\mathbf{p}_\mathrm{o})\,\frac{\mathrm{d}^2q}{(2\pi)^2}\\
&= -\sum_{m=0}^{\infty} \frac{(-i)^{m}}{m!} \sum_{\mathbf{m}\in\mathcal{S}_m} \binom{m}{\mathbf{m}}\,\mathbf{Z}_{n+1,1}^{+,-,(\mathbf{m})}(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o})\ .
\end{aligned}$$

It is now time to identify terms of same orders in the left and right hand sides. For $m = 0$, only the term for $\mathbf{m}' = (0,\cdots,0)$ remains in the left hand side, only the term $\mathbf{m} = (0,\cdots,0)$ remains in the right hand side and we have

$$\int \mathbf{Z}_{n+1,1}^{+,+,(\mathbf{0})}(\mathbf{p}\,|\,\mathbf{q})\ \mathbf{R}^{(0)}(\mathbf{q}\,|\,\mathbf{p}_\mathrm{o})\,\frac{\mathrm{d}^2q}{(2\pi)^2} = -\,\mathbf{Z}_{n+1,1}^{+,-,(\mathbf{0})}(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o})\,, \tag{9.5}$$

which, when expanded, reads

$$\int\!\!\int\cdots\int \hat{h}_n^{(0)}(\mathbf{p}-\mathbf{p}_n)\cdots\hat{h}_1^{(0)}(\mathbf{p}_2-\mathbf{q})\,\tilde{\mathbf{\Theta}}_{n+1,1}^{+,+,(\mathbf{0})}(\mathbf{p}\,|\,\mathbf{p}_n\,|\cdots|\,\mathbf{p}_2\,|\,\mathbf{q})\;\mathbf{R}^{(0)}(\mathbf{q}\,|\,\mathbf{p}_o)\,\frac{\mathrm{d}^2p_2}{(2\pi)^2}\cdots\frac{\mathrm{d}^2p_n}{(2\pi)^2}\,\frac{\mathrm{d}^2q}{(2\pi)^2}$$
$$=-\int\cdots\int \hat{h}_n^{(0)}(\mathbf{p}-\mathbf{p}_n)\cdots\hat{h}_1^{(0)}(\mathbf{p}_2-\mathbf{p}_0)\,\tilde{\mathbf{\Theta}}_{n+1,1}^{+,-,(\mathbf{0})}(\mathbf{p}\,|\,\mathbf{p}_n\,|\cdots|\,\mathbf{p}_2\,|\,\mathbf{p}_o)\,\frac{\mathrm{d}^2p_2}{(2\pi)^2}\cdots\frac{\mathrm{d}^2p_n}{(2\pi)^2}\;.$$

From the definition of the zero order Fourier moment, we have $\hat{h}_j^{(0)}(\mathbf{q})=(2\pi)^2\,\delta(\mathbf{q})$, which yields

$$\tilde{\mathbf{\Theta}}_{n+1,1}^{+,+,(\mathbf{0})}(\mathbf{p}\,|\,\mathbf{p}\,|\cdots|\,\mathbf{p}\,|\,\mathbf{p})\;\mathbf{R}^{(0)}(\mathbf{p}\,|\,\mathbf{p}_o)=-(2\pi)^2\,\tilde{\mathbf{\Theta}}_{n+1,1}^{+,-,(\mathbf{0})}(\mathbf{p}_o\,|\,\mathbf{p}_o\,|\cdots|\,\mathbf{p}_o\,|\,\mathbf{p}_o)\;\delta(\mathbf{p}-\mathbf{p}_0)\,. \tag{9.5}$$

Here we recognize the result obtained in Chapter 3 for a stack of flat interfaces

$$\mathbf{R}^{(0)}(\mathbf{p}\,|\,\mathbf{p}_0)=-\left[\bar{\mathbf{\Theta}}_{n+1,1}^{+,+}(\mathbf{p}_0)\right]^{-1}\bar{\mathbf{\Theta}}_{n+1,1}^{+,-}(\mathbf{p}_0)\,(2\pi)^2\delta(\mathbf{p}-\mathbf{p}_0)\overset{\text{def}}{=}\boldsymbol{\rho}_{n+1,1}^{(0)}(\mathbf{p}_0)\,(2\pi)^2\delta(\mathbf{p}-\mathbf{p}_0)\,, \tag{9.6}$$

where $\tilde{\mathbf{\Theta}}_{n+1,1}^{+,+,(\mathbf{0})}(\mathbf{p}_o\,|\,\mathbf{p}_o\,|\cdots|\,\mathbf{p}_o\,|\,\mathbf{p}_o)=\bar{\mathbf{\Theta}}_{n+1,1}^{+,+}(\mathbf{p}_0)$ and $\tilde{\mathbf{\Theta}}_{n+1,1}^{+,-,(\mathbf{0})}(\mathbf{p}_o\,|\,\mathbf{p}_o\,|\cdots|\,\mathbf{p}_o\,|\,\mathbf{p}_o)=\bar{\mathbf{\Theta}}_{n+1,1}^{+,-}(\mathbf{p}_0)$ as can been seen from the definitions in Eqs. (3.25) and (9.3). This result must be understood as a generalization of the similar result obtained for a single interface. Indeed, for a single interface, we have shown that the order zero of the perturbative expansion corresponds to the Fresnel amplitudes for the corresponding planar interface.

For orders $m\geq 1$, we have

$$\sum_{m'=0}^{m}\binom{m}{m'}\sum_{\mathbf{m}'\in\mathcal{S}_{m'}}\binom{m'}{\mathbf{m}'}\int\mathbf{Z}_{n+1,1}^{+,+,(\mathbf{m}')}(\mathbf{p}\,|\,\mathbf{q})\;\mathbf{R}^{(m-m')}(\mathbf{q}\,|\,\mathbf{p}_o)\,\frac{\mathrm{d}^2q}{(2\pi)^2}=-\sum_{\mathbf{m}\in\mathcal{S}_m}\binom{m}{\mathbf{m}}\mathbf{Z}_{n+1,1}^{+,-,(\mathbf{m})}(\mathbf{p}\,|\,\mathbf{p}_o)\;. \tag{9.7}$$

By isolating the term corresponding to $m'=0$, hence $\mathbf{m}'=(0,\cdots,0)$ and using that for all $j\in[\![1,n]\!]$ we have $\hat{h}_j^{(0)}(\mathbf{q})=(2\pi)^2\,\delta(\mathbf{q})$, we obtain

$$\tilde{\mathbf{\Theta}}_{n+1,1}^{+,+,(\mathbf{0})}(\mathbf{p}\,|\,\mathbf{p}\,|\cdots|\,\mathbf{p}\,|\,\mathbf{p})\;\mathbf{R}^{(m)}(\mathbf{p}\,|\,\mathbf{p}_o)=-\sum_{\mathbf{m}\in\mathcal{S}_m}\binom{m}{\mathbf{m}}\mathbf{Z}_{n+1,1}^{+,-,(\mathbf{m})}(\mathbf{p}\,|\,\mathbf{p}_o)$$
$$-\sum_{m'=1}^{m}\binom{m}{m'}\sum_{\mathbf{m}'\in\mathcal{S}_{m'}}\binom{m'}{\mathbf{m}'}\int\mathbf{Z}_{n+1,1}^{+,+,(\mathbf{m}')}(\mathbf{p}\,|\,\mathbf{q})\;\mathbf{R}^{(m-m')}(\mathbf{q}\,|\,\mathbf{p}_o)\,\frac{\mathrm{d}^2q}{(2\pi)^2}\,,$$

which yields

$$\mathbf{R}^{(m)}(\mathbf{p}\,|\,\mathbf{p}_o)=-\left[\bar{\mathbf{\Theta}}_{n+1,1}^{+,+}(\mathbf{p})\right]^{-1}\left[\sum_{\mathbf{m}\in\mathcal{S}_m}\binom{m}{\mathbf{m}}\mathbf{Z}_{n+1,1}^{+,-,(\mathbf{m})}(\mathbf{p}\,|\,\mathbf{p}_o)\right.$$
$$\left.+\sum_{m'=1}^{m}\binom{m}{m'}\sum_{\mathbf{m}'\in\mathcal{S}_{m'}}\binom{m'}{\mathbf{m}'}\int\mathbf{Z}_{n+1,1}^{+,+,(\mathbf{m}')}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{R}^{(m-m')}(\mathbf{q}\,|\,\mathbf{p}_o)\,\frac{\mathrm{d}^2q}{(2\pi)^2}\right]\;. \tag{9.7}$$

We have finally obtained a recursive expression giving the m^{th} order term in the reflection amplitude expansion as a function of terms of lower order terms. Note that Eq. (9.7) has the

same structure as Eq. (6.11), with the major difference lying in the complexity of the kernel containing product of Fourier moments of the surface profiles. One can verify that one recovers Eq. (6.11) by applying Eq. (9.7) in the case $n = 1$.

Remark 9.1. We have treated the case of reflection for fixing the ideas, but the same method applies for transmission. We will avoid going through yet another detailed similar derivation, and simply give key steps here. We recall that, with the notations from Chapter 3, the reduced Rayleigh equation for transmission through a stack of $n+1$ media is given by Eq. (3.21)

$$\int \boldsymbol{\Theta}^{-,-}_{1,n+1}\left(\mathbf{p}\,|\,\mathbf{q}\right)\,\mathbf{T}\left(\mathbf{q}\,|\,\mathbf{p}_{\mathrm{o}}\right)\frac{\mathrm{d}^2q}{(2\pi)^2} = -\eta_{1,n+1}\,(2\pi)^2\,\delta(\mathbf{p}-\mathbf{p}_0)\,\mathbf{I}_2\ , \tag{9.7}$$

where we recall that the *backward n-interface transfer kernel* is defined as

$$\boldsymbol{\Theta}^{a_1,a_{n+1}}_{1,n+1}\left(\mathbf{p}_1\,|\,\mathbf{p}_{n+1}\right) = \sum_{a_2=\pm} a_2 \int\cdots\sum_{a_n=\pm} a_n \int \boldsymbol{\Theta}^{a_1,a_2}_{1,2}\left(\mathbf{p}_1\,|\,\mathbf{p}_2\right)\cdots\boldsymbol{\Theta}^{a_n,a_{n+1}}_{n,n+1}\left(\mathbf{p}_n\,|\,\mathbf{p}_{n+1}\right)\frac{\mathrm{d}^2p_n}{(2\pi)^2}\,\frac{\mathrm{d}^2p_2}{(2\pi)^2}\ . \tag{9.8}$$

We define

$$\begin{aligned}
&\tilde{\boldsymbol{\Theta}}^{a_1,a_{n+1},(\mathbf{m})}_{1,n+1}\left(\mathbf{p}_1\,|\,\mathbf{p}_2\,|\cdots|\,\mathbf{p}_n\,|\,\mathbf{p}_{n+1}\right) = \\
&\sum_{a_2=\pm} a_2\,\alpha_1^{-1}(\mathbf{p}_1)(a_1\alpha_1(\mathbf{p}_1)-a_2\alpha_2(\mathbf{p}_2))^{m_1-1}\,e^{-i(a_1\alpha_1(\mathbf{p}_1)-a_2\alpha_2(\mathbf{p}_2))d_1}\ \mathbf{M}^{a_1,a_2}_{1,2}\left(\mathbf{p}_1\,|\,\mathbf{p}_2\right)\cdots \\
&\sum_{a_n=\pm} a_n\,\alpha_n^{-1}(\mathbf{p}_n)(a_n\alpha_n(\mathbf{p}_n)-a_{n+1}\alpha_{n+1}(\mathbf{p}_{n+1}))^{m_n-1}\,e^{-i(a_n\alpha_n(\mathbf{p}_n)-a_{n+1}\alpha_{n+1}(\mathbf{p}_{n+1}))d_n}\ \mathbf{M}^{a_n,a_{n+1}}_{n,n+1}\left(\mathbf{p}_n\,|\,\mathbf{p}_{n+1}\right),
\end{aligned} \tag{9.9}$$

and the expansion of the kernel in series of Fourier moments is given by

$$\begin{aligned}
\boldsymbol{\Theta}^{a_1,a_{n+1}}_{1,n+1}\left(\mathbf{p}_1\,|\,\mathbf{p}_{n+1}\right) &= \sum_{\mathbf{m}=0}^{\infty}\frac{(-i)^{|\mathbf{m}|}}{\mathbf{m}!}\int\cdots\int \hat{h}_1^{(m_1)}(\mathbf{p}_1-\mathbf{p}_2)\cdots\hat{h}_n^{(m_n)}(\mathbf{p}_n-\mathbf{p}_{n+1}) \\
&\qquad \tilde{\boldsymbol{\Theta}}^{a_1,a_{n+1},(\mathbf{m})}_{1,n+1}\left(\mathbf{p}_1\,|\,\mathbf{p}_2\,|\cdots|\,\mathbf{p}_n\,|\,\mathbf{p}_{n+1}\right)\frac{\mathrm{d}^2p_n}{(2\pi)^2}\,\frac{\mathrm{d}^2p_2}{(2\pi)^2} \\
&= \sum_{\mathbf{m}=0}^{\infty}\frac{(-i)^{|\mathbf{m}|}}{\mathbf{m}!}\ \mathbf{Z}^{a_1,a_{n+1},(\mathbf{m})}_{1,n+1}\left(\mathbf{p}_1\,|\,\mathbf{p}_{n+1}\right)\ .
\end{aligned} \tag{9.10}$$

After expanding the reduced Rayleigh equation and identifying orders, we obtain that the zero order satisfies

$$\int \mathbf{Z}^{-,-,(\mathbf{0})}_{1,n+1}\left(\mathbf{p}\,|\,\mathbf{q}\right)\,\mathbf{T}^{(0)}\left(\mathbf{q}\,|\,\mathbf{p}_{\mathrm{o}}\right)\frac{\mathrm{d}^2q}{(2\pi)^2} = -\eta_{1,n+1}\,(2\pi)^2\,\delta(\mathbf{p}-\mathbf{p}_0)\,\mathbf{I}_2\ , \tag{9.11}$$

which yields

$$\mathbf{T}^{(0)}(\mathbf{p}\,|\,\mathbf{p}_0) = -\eta_{1,n+1}\,\left[\bar{\boldsymbol{\Theta}}^{-,-}_{1,n+1}(\mathbf{p}_0)\right]^{-1}\,(2\pi)^2\delta(\mathbf{p}-\mathbf{p}_0) \overset{\text{def}}{=} \boldsymbol{\tau}^{(0)}_{n+1,1}(\mathbf{p}_0)\,(2\pi)^2\delta(\mathbf{p}-\mathbf{p}_0)\,, \tag{9.12}$$

where $\tilde{\boldsymbol{\Theta}}^{-,-,(\mathbf{0})}_{1,n+1}\left(\mathbf{p}_{\mathrm{o}}\,|\,\mathbf{p}_{\mathrm{o}}\,|\cdots|\,\mathbf{p}_{\mathrm{o}}\,|\,\mathbf{p}_{\mathrm{o}}\right) = \bar{\boldsymbol{\Theta}}^{-,-}_{1,n+1}(\mathbf{p}_0)$. For orders $m \geq 1$, we have

$$\sum_{m'=0}^{m}\binom{m}{m'}\sum_{\mathbf{m}'\in\mathcal{S}_{m'}}\binom{m'}{\mathbf{m}'}\int \mathbf{Z}^{-,-,(\mathbf{m}')}_{1,n+1}\left(\mathbf{p}\,|\,\mathbf{q}\right)\,\mathbf{T}^{(m-m')}\left(\mathbf{q}\,|\,\mathbf{p}_{\mathrm{o}}\right)\frac{\mathrm{d}^2q}{(2\pi)^2} = \mathbf{0}\,, \tag{9.13}$$

which yields

$$\tilde{\mathbf{\Theta}}^{-,-,(\mathbf{0})}_{1,n+1}(\mathbf{p}\,|\,\mathbf{p}\,|\,\cdots\,|\,\mathbf{p}\,|\,\mathbf{p})\;\mathbf{T}^{(m)}(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}}) = -\sum_{m'=1}^{m}\binom{m}{m'}\sum_{\mathbf{m}'\in\mathcal{S}_{m'}}\binom{m'}{\mathbf{m}'}\int \mathbf{Z}^{-,-,(\mathbf{m}')}_{1,n+1}(\mathbf{p}\,|\,\mathbf{q})\;\mathbf{T}^{(m-m')}(\mathbf{q}\,|\,\mathbf{p}_{\mathrm{o}})\,\frac{\mathrm{d}^2q}{(2\pi)^2}\,,$$

and finally, the recursive formula for successive orders reads

$$\mathbf{T}^{(m)}(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}}) = -\left[\bar{\mathbf{\Theta}}^{-,-}_{1,n+1}(\mathbf{p})\right]^{-1}\sum_{m'=1}^{m}\binom{m}{m'}\sum_{\mathbf{m}'\in\mathcal{S}_{m'}}\binom{m'}{\mathbf{m}'}\int \mathbf{Z}^{-,-,(\mathbf{m}')}_{1,n+1}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{T}^{(m-m')}(\mathbf{q}\,|\,\mathbf{p}_{\mathrm{o}})\,\frac{\mathrm{d}^2q}{(2\pi)^2}\,. \tag{9.13}$$

First order and mean differential scattering coefficients

In the same way we obtained simple closed form expressions for the mean differential scattering coefficients when restricting ourselves to the first order amplitudes in the surface profile function in Chapter 8 in the case of a single interface, we can obtain similar expressions when considering the first order amplitudes in product of the surface profile functions in the case of a multilayer system. Applying Eq.(9.7) for $m = 1$ we obtain

$$\begin{aligned}\mathbf{R}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}}) &= -\left[\bar{\mathbf{\Theta}}^{+,+}_{n+1,1}(\mathbf{p})\right]^{-1}\sum_{\mathbf{m}\in\mathcal{S}_1}\left[\mathbf{Z}^{+,-,(\mathbf{m})}_{n+1,1}(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}}) + \int \mathbf{Z}^{+,+,(\mathbf{m})}_{n+1,1}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{R}^{(0)}(\mathbf{q}\,|\,\mathbf{p}_{\mathrm{o}})\,\frac{\mathrm{d}^2q}{(2\pi)^2}\right]\\ &= -\left[\bar{\mathbf{\Theta}}^{+,+}_{n+1,1}(\mathbf{p})\right]^{-1}\sum_{\mathbf{m}\in\mathcal{S}_1}\left[\mathbf{Z}^{+,-,(\mathbf{m})}_{n+1,1}(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}}) + \mathbf{Z}^{+,+,(\mathbf{m})}_{n+1,1}(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}})\;\boldsymbol{\rho}^{(0)}_{n+1,1}(\mathbf{p}_0)\right].\end{aligned} \tag{9.14}$$

Here we have used the previously obtained expression for $\mathbf{R}^{(0)}(\mathbf{q}\,|\,\mathbf{p}_{\mathrm{o}})$ in Eq. (9.6), and in particular the fundamental property of the Dirac delta. From the definition of $\mathbf{Z}^{a_{n+1},a_1,(\mathbf{m})}_{n+1,1}$ [Eq. (9.4)] it is clear that for $\mathbf{m} = (1,0,\cdots,0)$, $(0,\cdots,0,1,0,\cdots,0)$ (with a 1 in position j) or $(0,\cdots,0,1)$ the integration reduces to

$$\mathbf{Z}^{a_{n+1},a_1,(1,0,\cdots,0)}_{n+1,1}(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}}) = \hat{h}^{(1)}_1(\mathbf{p}-\mathbf{p}_0)\,\tilde{\mathbf{\Theta}}^{a_{n+1},a_1,(1,0\cdots,0)}_{n+1,1}(\mathbf{p}\;|\;\cdots\;|\;\mathbf{p}\;|\;\mathbf{p}_0) \tag{9.15a}$$

$$\mathbf{Z}^{a_{n+1},a_1,(0,\cdots,0,1,0,\cdots,0)}_{n+1,1}(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}}) = \hat{h}^{(1)}_j(\mathbf{p}-\mathbf{p}_0)\,\tilde{\mathbf{\Theta}}^{a_{n+1},a_1,(0,\cdots,0,1,0,\cdots,0)}_{n+1,1}(\mathbf{p}\;|\;\cdots\;|\;\mathbf{p}\;|\;\underbrace{\mathbf{p}_0\;|\;\cdots\;|\;\mathbf{p}_0}_{j\text{ times}}) \tag{9.15b}$$

$$\mathbf{Z}^{a_{n+1},a_1,(0,\cdots,0,1)}_{n+1,1}(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}}) = \hat{h}^{(1)}_n(\mathbf{p}-\mathbf{p}_0)\,\tilde{\mathbf{\Theta}}^{a_{n+1},a_1,(0,\cdots,0,1)}_{n+1,1}(\mathbf{p}\;|\;\mathbf{p}_0\;|\;\cdots\;|\;\mathbf{p}_0). \tag{9.15c}$$

It is convenient to group terms with common factor $\hat{h}_j \equiv \hat{h}^{(1)}_j$ in Eq. (9.14), which leads to

$$\mathbf{R}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}}) = \sum_{j=1}^{n}\hat{h}_j(\mathbf{p}-\mathbf{p}_0)\,\boldsymbol{\rho}^{(1)}_j(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}})\,, \tag{9.16}$$

with

$$\boldsymbol{\rho}_j^{(1)}(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}) = -\left[\bar{\boldsymbol{\Theta}}_{n+1,1}^{+,+}(\mathbf{p})\right]^{-1}\left[\tilde{\boldsymbol{\Theta}}_{n+1,1}^{+,-,(0,\cdots,0,1,0,\cdots,0)}(\mathbf{p}\mid\cdots\mid\mathbf{p}\mid\underbrace{\mathbf{p}_0\mid\cdots\mid\mathbf{p}_0}_{j\text{ times}}) + \tilde{\boldsymbol{\Theta}}_{n+1,1}^{+,+,(0,\cdots,0,1,0,\cdots,0)}(\mathbf{p}\mid\cdots\mid\mathbf{p}\mid\underbrace{\mathbf{p}_0\mid\cdots\mid\mathbf{p}_0}_{j\text{ times}})\,\boldsymbol{\rho}_{n+1,1}^{(0)}(\mathbf{p}_0)\right]. \tag{9.17}$$

Similarly, the transmission amplitude to first order in products of surface profiles reads

$$\mathbf{T}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}) = \sum_{j=1}^{n}\hat{h}_j(\mathbf{p}-\mathbf{p}_0)\,\boldsymbol{\tau}_j^{(1)}(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o})\,, \tag{9.18}$$

with

$$\boldsymbol{\tau}_j^{(1)}(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}) = -\left[\bar{\boldsymbol{\Theta}}_{1,n+1}^{-,-}(\mathbf{p})\right]^{-1}\tilde{\boldsymbol{\Theta}}_{1,n+1}^{-,-,(0,\cdots,0,1,0,\cdots,0)}(\underbrace{\mathbf{p}\mid\cdots\mid\mathbf{p}}_{j\text{ times}}\mid\mathbf{p}_0\mid\cdots\mid\mathbf{p}_0)\,\boldsymbol{\tau}_{n+1,1}^{(0)}(\mathbf{p}_0)\,. \tag{9.19}$$

The corresponding incoherent component of the mean differential scattering coefficient are then easily expressed as

$$\left\langle\frac{\partial R_{\alpha\beta}}{\partial\Omega_r}(\mathbf{p}\,|\,\mathbf{p}_0)\right\rangle_\mathrm{incoh} = \left(\frac{\omega}{2\pi c}\right)^2\frac{\epsilon_1\cos^2\theta_r}{\cos\theta_0}\sum_{j=1}^{n}\sum_{k=1}^{n}G_{jk}(\mathbf{p}-\mathbf{p}_0)\rho_{j,\alpha\beta}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o})\,\rho_{k,\alpha\beta}^{(1)*}(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}) \tag{9.20a}$$

$$\left\langle\frac{\partial T_{\alpha\beta}}{\partial\Omega_t}(\mathbf{p}\,|\,\mathbf{p}_0)\right\rangle_\mathrm{incoh} = \left(\frac{\omega}{2\pi c}\right)^2\frac{\epsilon_{n+1}^{3/2}\cos^2\theta_t}{\epsilon_1^{1/2}\cos\theta_0}\sum_{j=1}^{n}\sum_{k=1}^{n}G_{jk}(\mathbf{p}-\mathbf{p}_0)\tau_{j,\alpha\beta}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o})\,\tau_{k,\alpha\beta}^{(1)*}(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}) \tag{9.20b}$$

$$G_{jk}(\mathbf{p}-\mathbf{p}_0) \stackrel{\mathrm{def}}{=} \left\langle\hat{h}_j(\mathbf{p}-\mathbf{p}_0)\hat{h}_k^*(\mathbf{p}-\mathbf{p}_0)\right\rangle. \tag{9.20c}$$

We can give an intuitive physical interpretation of the above expressions. The term G_{jk} is the Fourier transform of the covariance of the fluctuations h_j and h_k of the surface profiles ζ_j and ζ_k. It is then expected to be proportional to $\sigma_j\sigma_k$ where the $(\sigma_j)_{1\le j\le n}$ are the rms-roughness parameters the respective interfaces. Consequently, if assume all the interfaces but the ℓth one to be planar, all the covariance terms but that for which $j=\ell$ and $k=\ell$ vanish which leads to

$$\left\langle\frac{\partial R_{\alpha\beta}^{(\ell)}}{\partial\Omega_r}(\mathbf{p}\,|\,\mathbf{p}_0)\right\rangle_\mathrm{incoh} = \left(\frac{\omega}{2\pi c}\right)^2\frac{\epsilon_1\cos^2\theta_r}{\cos\theta_0}\,\sigma_\ell^2\,g_\ell(\mathbf{p}-\mathbf{p}_0)|\rho_{\ell,\alpha\beta}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o})\,|^2 \tag{9.21a}$$

$$\left\langle\frac{\partial T_{\alpha\beta}^{(\ell)}}{\partial\Omega_t}(\mathbf{p}\,|\,\mathbf{p}_0)\right\rangle_\mathrm{incoh} = \left(\frac{\omega}{2\pi c}\right)^2\frac{\epsilon_{n+1}^{3/2}\cos^2\theta_t}{\epsilon_1^{1/2}\cos\theta_0}\,\sigma_\ell^2\,g_\ell(\mathbf{p}-\mathbf{p}_0)|\tau_{\ell,\alpha\beta}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o})\,|^2\,. \tag{9.21b}$$

Here we have denoted by g_k the power spectrum of the k^{th} interface. The resulting expressions correspond to the diffusely scattered intensity scattered by a system for which all interfaces but the ℓ^{th} are planar under the *single scattering approximation*. This means that, for this system, we can picture the overall scattering amplitude to be a sum of probability amplitudes over all possible optical paths for which one single scattering event is allowed on the rough surface, but allowing any arbitrary number of specular reflections and refractions within the layers prior

to and posterior to the scattering event. In particular, the above formula takes into account interference effects due to the phase differences underwent by different paths within the layers, and is encoded in the amplitudes $\rho^{(1)}_{\ell,\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o})$ and $\tau^{(1)}_{\ell,\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o})$. Now let us imagine that the system is made of layers whose all interfaces are rough, but for which *no* pairs of interfaces are cross-correlated, i.e. that $G_{jk} = \delta_{jk}\,\sigma_j^2\,g_j$ for all $j,k \in [\![1,n]\!]$. Then the incoherent component of the mean differential scattering coefficients reads

$$\left\langle \frac{\partial R^{(\mathrm{uncorr})}_{\alpha\beta}}{\partial\Omega_r}(\mathbf{p}\,|\,\mathbf{p}_0)\right\rangle_{\mathrm{incoh}} = \sum_{j=1}^{n}\left\langle \frac{\partial R^{(j)}_{\alpha\beta}}{\partial\Omega_r}(\mathbf{p}\,|\,\mathbf{p}_0)\right\rangle_{\mathrm{incoh}} \tag{9.22a}$$

$$\left\langle \frac{\partial T^{(\mathrm{uncorr})}_{\alpha\beta}}{\partial\Omega_t}(\mathbf{p}\,|\,\mathbf{p}_0)\right\rangle_{\mathrm{incoh}} = \sum_{j=1}^{n}\left\langle \frac{\partial T^{(j)}_{\alpha\beta}}{\partial\Omega_t}(\mathbf{p}\,|\,\mathbf{p}_0)\right\rangle_{\mathrm{incoh}} . \tag{9.22b}$$

This means that the intensity of the diffusely scattered light for a system with *un-cross-correlated interfaces* is, within the single scattering approximation, the sum of the intensity one would obtain for the corresponding system where only one of the interfaces is rough at a time. This indicates that, on average, we would not observe interference effects resulting from a path which experienced a scattering event on a surface j with a path which experienced a scattering event on a surface k. This is due to the fact that the interfaces are not cross-correlated, therefore from one realization of the stochastic process to the next the phase difference between two such paths would be an uncorrelated random variable hence resulting in optical incoherence. Note however, that paths experiencing a scattering event on the same interface still have correlated phase differences and may lead to interference effects (as one would observe if one only of the interface were rough). In the most general case where the interfaces may be uncorrelated, two paths experiencing a scattering event on different interface may interfere constructively or destructively, hence modifying the interference pattern one would obtain by summing the interference patterns of the corresponding systems where only one of the interfaces is rough. Indeed, we can re-write Eq.(9.20) as

$$\begin{aligned}\left\langle \frac{\partial R_{\alpha\beta}}{\partial\Omega_r}(\mathbf{p}\,|\,\mathbf{p}_0)\right\rangle_{\mathrm{incoh}} &= \left\langle \frac{\partial R^{(\mathrm{uncorr})}_{\alpha\beta}}{\partial\Omega_r}(\mathbf{p}\,|\,\mathbf{p}_0)\right\rangle_{\mathrm{incoh}} \\ &\quad + 2\left(\frac{\omega}{2\pi c}\right)^2 \frac{\epsilon_1\cos^2\theta_r}{\cos\theta_0}\,\mathrm{Re}\left[\sum_{j<k} G_{jk}(\mathbf{p}-\mathbf{p}_0)\rho^{(1)}_{j,\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o})\,\rho^{(1)*}_{k,\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o})\right] \end{aligned} \tag{9.23a}$$

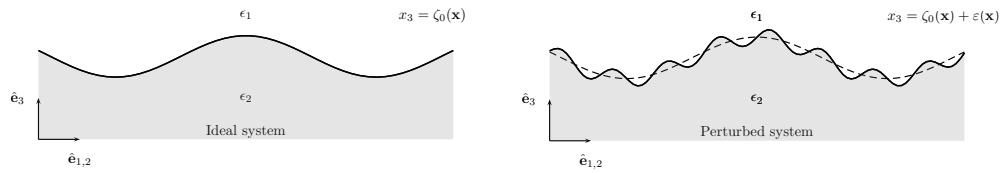
$$\begin{aligned}\left\langle \frac{\partial T_{\alpha\beta}}{\partial\Omega_t}(\mathbf{p}\,|\,\mathbf{p}_0)\right\rangle_{\mathrm{incoh}} &= \left\langle \frac{\partial T^{(\mathrm{uncorr})}_{\alpha\beta}}{\partial\Omega_t}(\mathbf{p}\,|\,\mathbf{p}_0)\right\rangle_{\mathrm{incoh}} \\ &\quad + 2\left(\frac{\omega}{2\pi c}\right)^2 \frac{\epsilon_{n+1}^{3/2}\cos^2\theta_t}{\epsilon_1^{1/2}\cos\theta_0}\,\mathrm{Re}\left[\sum_{j<k} G_{jk}(\mathbf{p}-\mathbf{p}_0)\tau^{(1)}_{j,\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o})\,\tau^{(1)*}_{k,\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o})\right] , \end{aligned} \tag{9.23b}$$

where it should now be clear that the cross terms which can take both positive and negative values may yield interference enhancement or attenuation. Note that the non-negativity of the intensity is preserved since we always have $2\mathrm{Re}(zz') \leq |z|^2 + |z'|^2$ for any two complex number z and z'. Such an interference phenomena induced by the cross-correlation between the interfaces was studied in **Paper** [3] demonstrating the mechanism of selective enhancement and attenuation of the Selényi rings in the case of a system with two cross-correlated interfaces. In

Chapter 10, we will explain in more details the analysis based on optical paths experiencing a single scattering event by constructing a dedicated deductive model which we will show to be numerically indistinguishable from the small amplitude perturbation results exposed here, hence justifying the described physical picture, and how this analysis is used to interpret the phenomenon of selective enhancement and attenuation of interference rings in the intensity of the diffusely scattered light.

9.2 Perturbation of a structured surface

We have seen previously how to construct approximate solutions of the reduced Rayleigh equations by considering the surface profile as a perturbed planar interface. Indeed, we have shown that the zero order term in the power series expansion corresponds to Fresnel scattering amplitudes. This approach is found to be widely used in the literature for studying the scattering of light by randomly weakly rough surfaces, where the average surface is planar. It could also be of interest to study the effect of random deviation from a structured surface. For example, a manufacturer of gratings may wish to asses how random small[1] scale defects affect the optical diffraction response. In particular, one may expect that if the random deviations of the grating profile are two large in some sense, the diffraction peaks would smear out and thus lower its angular resolution. Another case of interest, in radar detection, could be the study of electromagnetic wave scattering by the ocean. Water waves at the surface of the ocean can be described by the superposition of waves of different spatial scales. Typically, this could be a long range tide wave and smaller wavelets induced by the wind. Therefore, the surface encountered from the electromagnetic wave can be seen as a perturbation of the long range waves. In these cases, it seems more appropriate to consider perturbation of the grating profile, or the long range water wave, rather than perturbation of the planar interface. To illustrate the method we will consider the case of single interface system.



(a) (b)

Figure 9.1: Illustration of an ideal system defined by a surface profile ζ_0 (a) to which a perturbation ε is added (b).

Let the surface profile ζ be defined as the sum of two terms, a large scale component ζ_0 that may be thought of as an ideal grating profile, and a small scale component ε that corresponds

[1] small compared with the grating parameters

to a perturbation. The kernel scalar factor $\mathcal{J}^{b,a}_{l,m}$ is now expanded as follow

$$\begin{aligned}\mathcal{J}^{b,a}_{l,m}(\mathbf{p}\,|\,\mathbf{q}) &= (b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q}))^{-1} \int e^{-i(\mathbf{p}-\mathbf{q})\cdot\mathbf{x}}\, e^{-i(b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q}))\,(\zeta_0(\mathbf{x}) + \varepsilon(\mathbf{x}))}\,\mathrm{d}^2x \\ &= \sum_{n=0}^{\infty} \frac{(-i)^n}{n!}\,(b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q}))^{n-1} \int e^{-i(\mathbf{p}-\mathbf{q})\cdot\mathbf{x}}\, e^{-i(b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q}))\,\zeta_0(\mathbf{x})}\,\varepsilon^n(\mathbf{x})\,\mathrm{d}^2x \\ &= \sum_{n=0}^{\infty} \frac{(-i)^n}{n!}\,(b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q}))^{n-1}\,\varepsilon^{b,a,(n)}_{l,m}(\mathbf{p}\,|\,\mathbf{q}) \end{aligned} \tag{9.24}$$

Here, we have defined the following integral transform of the n^{th} power of ε as

$$\varepsilon^{b,a,(n)}_{l,m}(\mathbf{p}\,|\,\mathbf{q}) = \int e^{-i(\mathbf{p}-\mathbf{q})\cdot\mathbf{x}}\, e^{-i(b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q}))\,\zeta_0(\mathbf{x})}\,\varepsilon^n(\mathbf{x})\,\mathrm{d}^2x\ . \tag{9.25}$$

Remark 9.2. Note that the zero order term in Eq. (9.24), or equivalently Eq. (9.24) in the limit $\varepsilon \to 0$, corresponds to $\mathcal{J}^{b,a}_{l,m}$ for the *unperturbed* system that we will specifically denote $\mathring{\mathcal{J}}^{b,a}_{l,m}$ in this case. We will denote similarly all quantities relative to the unperturbed system.

In a similar fashion as what we have seen in previous sections, we obtain that the reflection expansion amplitudes satisfy

$$\begin{aligned}&\sum_{m=0}^{\infty} \frac{(-i)^m}{m!} \sum_{m'=0}^{m} \binom{m}{m'} \int \varepsilon^{+,+,(m')}_{2,1}(\mathbf{p}\,|\,\mathbf{q})\ \tilde{\mathbf{\Theta}}^{+,+,(m')}_{2,1}(\mathbf{p}\,|\,\mathbf{q})\ \mathbf{r}^{(m-m')}(\mathbf{q}\,|\,\mathbf{p}_{\mathrm{o}})\,\frac{\mathrm{d}^2q}{(2\pi)^2} \\ &= -\sum_{m=0}^{\infty} \frac{(-i)^m}{m!}\,\varepsilon^{+,-,(m)}_{2,1}(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}})\ \tilde{\mathbf{\Theta}}^{+,-,(m)}_{2,1}(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}})\ .\end{aligned}$$

Identifying orders yields for $m = 0$,

$$\int \mathring{\mathbf{\Theta}}^{+,+}_{2,1}(\mathbf{p}\,|\,\mathbf{q})\ \mathbf{r}^{(0)}(\mathbf{q}\,|\,\mathbf{p}_{\mathrm{o}})\,\frac{\mathrm{d}^2q}{(2\pi)^2} = -\ \mathring{\mathbf{\Theta}}^{+,-}_{2,1}(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}})\,, \tag{9.25}$$

and for $m \geq 1$,

$$\begin{aligned}&\int \mathring{\mathbf{\Theta}}^{+,+}_{2,1}(\mathbf{p}\,|\,\mathbf{q})\ \mathbf{r}^{(m)}(\mathbf{q}\,|\,\mathbf{p}_{\mathrm{o}})\,\frac{\mathrm{d}^2q}{(2\pi)^2} = -\,\varepsilon^{+,-,(m)}_{2,1}(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}})\ \tilde{\mathbf{\Theta}}^{+,-,(m)}_{2,1}(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}}) \\ &- \sum_{m'=1}^{m} \binom{m}{m'} \int \varepsilon^{+,+,(m')}_{2,1}(\mathbf{p}\,|\,\mathbf{q})\ \tilde{\mathbf{\Theta}}^{+,+,(m')}_{2,1}(\mathbf{p}\,|\,\mathbf{q})\ \mathbf{r}^{(m-m')}(\mathbf{q}\,|\,\mathbf{p}_{\mathrm{o}})\,\frac{\mathrm{d}^2q}{(2\pi)^2}\ .\end{aligned} \tag{9.26}$$

We observe that Eq. (9.25) is nothing but the reduced Rayleigh equations for the unperturbed system and does not depend on the perturbation ε. Hence the zero order term in the reflection amplitudes expansion is solution of the unperturbed problem. The higher order terms can be computed recursively as can be seen in Eq. (9.26). However, a significant difference compared to the method of perturbation of a flat surface is that one must solve an integral equation for each order. Nevertheless, this should not a problem. Indeed, as we saw in the case of the direct method, Chapter 5, solving a reduced Rayleigh equation translates numerically to solving a linear system. In fact, if one is able to solve for the zero order term, solving for higher order terms is almost free if the right hand side is given. Indeed, the kernel stays the same at all orders and is identical to that of the unperturbed system. A good numerical strategy then consists in making a LU factorization of the unperturbed kernel for the zero order and then re-use this factorization for each order. In fact, the challenging part of this method is to compute efficiently the factors $\varepsilon^{b,a,(n)}_{l,m}(\mathbf{p}\,|\,\mathbf{q})$.

9.2.1 Perturbed grating

Let us know analyze the case where the unperturbed profile ζ_0 is periodic. In this case, we know from Chapter 4 that the unperturbed scalar factor $\mathring{\mathcal{J}}_{2,1}^{+,+}$ in the kernel and the unperturbed solution $\mathbf{r}^{(0)}$ read

$$\mathring{\mathcal{J}}_{2,1}^{+,+}(\mathbf{p}\,|\,\mathbf{q}) = (2\pi)^2 \sum_{\boldsymbol{\ell}\in\mathbb{Z}^2} \delta(\mathbf{p}^{(\boldsymbol{\ell})} - \mathbf{q})\ \mathring{\tilde{\mathcal{J}}}_{2,1}^{+,+}(\mathbf{p}\,|\,\mathbf{q})\,, \tag{9.27}$$

$$\mathbf{r}^{(0)}(\mathbf{q}\,|\,\mathbf{p}_{\mathrm{o}}) \equiv \mathring{\mathbf{r}}(\mathbf{q}\,|\,\mathbf{p}_{\mathrm{o}}) = (2\pi)^2 \sum_{\mathbf{m}\in\mathbb{Z}^2} \delta(\mathbf{q} - \mathbf{p}_0^{(\mathbf{m})})\ \mathring{\mathbf{r}}^{(\mathbf{m})}(\mathbf{p}_0)\,. \tag{9.28}$$

with the short hand notation $\mathbf{p}^{(\boldsymbol{\ell})} = \mathbf{p} - \mathbf{G}^{(\boldsymbol{\ell})}$ and $\mathbf{p}_0^{(\mathbf{m})} = \mathbf{p}_0 + \mathbf{G}^{(\mathbf{m})}$ and $\mathbf{G}^{(\cdot)}$ are reciprocal lattice vectors. Plugging Eqs. (9.27) and (9.28) into Eq. (9.26) for $m = 1$ we obtain

$$\begin{aligned}\sum_{\boldsymbol{\ell}\in\mathbb{Z}^2} \mathring{\tilde{\boldsymbol{\Theta}}}_{2,1}^{+,+}(\mathbf{p}\ |\ \mathbf{p}^{(\boldsymbol{\ell})})\ \mathbf{r}^{(1)}(\mathbf{p}^{(\boldsymbol{\ell})}\ |\ \mathbf{p}_0) &= -\,\varepsilon_{2,1}^{+,-,(1)}(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}})\ \tilde{\boldsymbol{\Theta}}_{2,1}^{+,-,(1)}(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}}) \\ -\sum_{\mathbf{m}\in\mathbb{Z}^2} \varepsilon_{2,1}^{+,+,(1)}(\mathbf{p}\ |\ \mathbf{p}_0^{(\mathbf{m})})\ \tilde{\boldsymbol{\Theta}}_{2,1}^{+,+,(1)}(\mathbf{p}\ |\ \mathbf{p}_0^{(\mathbf{m})})\ \mathring{\mathbf{r}}^{(\mathbf{m})}(\mathbf{p}_0)\,.\end{aligned} \tag{9.29}$$

Remark 9.3. Compared with the ideal periodic case, Eq. (9.29) requires to sample $\mathbf{p}$. This is to be expected since the unperturbed grating is *a priori* not periodic. If the perturbation ε is periodic with the same period as the ideal profile, then factors $\varepsilon_{2,1}^{b,a,(1)}$ can be expressed as a sum of Dirac deltas, similarly as $\mathring{\mathcal{J}}_{2,1}^{+,+}$. In that case, $\mathbf{r}^{(1)}(\mathbf{p}\ |\ \mathbf{p}_0)$ will also necessarily be expressed as a sum of Dirac deltas as is $\mathbf{r}^{(0)}(\mathbf{p}\,|\,\mathbf{p}_{\mathrm{o}})$, and the linear system will not require to sample $\mathbf{p}$ similarly to what we have seen in Chapter 4. To be more specific, in that case, it is straightforward to show, following the method from Chapter 4, that the first order correction satisfies

$$\mathbf{r}^{(1)}(\mathbf{q}\,|\,\mathbf{p}_{\mathrm{o}}) = (2\pi)^2 \sum_{\mathbf{m}\in\mathbb{Z}^2} \delta(\mathbf{q} - \mathbf{p}_0^{(\mathbf{m})})\ \mathbf{r}^{(1,\mathbf{m})}(\mathbf{p}_0)\,, \tag{9.30}$$

and for all $\boldsymbol{\ell} \in \mathbb{Z}^2$,

$$\begin{aligned}\sum_{\mathbf{m}\in\mathbb{Z}^2} \mathring{\tilde{\boldsymbol{\Theta}}}_{2,1}^{+,+}(\mathbf{p}_0^{(\boldsymbol{\ell})}\ |\ \mathbf{p}_0^{(\mathbf{m})})\ \mathbf{r}^{(1,\mathbf{m})}(\mathbf{p}_0) &= -\,\tilde{\varepsilon}_{2,1}^{+,-,(1)}(\mathbf{p}_0^{(\boldsymbol{\ell})}\ |\ \mathbf{p}_0)\ \tilde{\boldsymbol{\Theta}}_{2,1}^{+,-,(1)}(\mathbf{p}_0^{(\boldsymbol{\ell})}\ |\ \mathbf{p}_0) \\ -\sum_{\mathbf{m}\in\mathbb{Z}^2} \tilde{\varepsilon}_{2,1}^{+,+,(1)}(\mathbf{p}_0^{(\boldsymbol{\ell})}\ |\ \mathbf{p}_0^{(\mathbf{m})})\ \tilde{\boldsymbol{\Theta}}_{2,1}^{+,+,(1)}(\mathbf{p}_0^{(\boldsymbol{\ell})}\ |\ \mathbf{p}_0^{(\mathbf{m})})\ \mathring{\mathbf{r}}^{(\mathbf{m})}(\mathbf{p}_0)\,,\end{aligned} \tag{9.31}$$

with

$$\tilde{\varepsilon}_{l,m}^{b,a,(n)}(\mathbf{p}\,|\,\mathbf{q}) = \frac{1}{a_c}\int_{a_c} e^{-i(\mathbf{p}-\mathbf{q})\cdot\mathbf{x}}\, e^{-i(b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q}))\,\zeta_0(\mathbf{x})}\,\varepsilon^n(\mathbf{x})\,\mathrm{d}^2x\ . \tag{9.32}$$

It is clear that by induction, the same can be said for Eq. (9.26) for $m > 1$.

But what if the perturbation is not periodic? In this case, $\mathbf{r}^{(1)}(\mathbf{p}\ |\ \mathbf{p}_0)$ has no reason to have weight on a discrete set of points anymore and we would in fact rather expect it to be a density from a physical point of view. Indeed, it is known from experiment that random defects on grating have a tendency to smear out the discrete diffraction peaks. Therefore, we understand

that it is quite natural to obtain an equation that requires us to sample the in-plane wavevector space, or $\mathbf{p}$-plane. However, we notice that it is not necessary to sample the whole $\mathbf{p}$-plane. Indeed, for a given $\mathbf{p}$, solving Eq. (9.29) gives $\mathbf{r}^{(1)}(\mathbf{p} \mid \mathbf{p}_0)$ evaluated a the set of points $\{\mathbf{p} - \mathbf{G}^{(\boldsymbol{\ell})} \mid \boldsymbol{\ell} \in \mathbb{Z}^2\}$. Therefore, solving Eq. (9.29) for a set of $\mathbf{p}$ spanning the primitive reciprocal lattice cell will yield knowledge of the solution (to first order) in the whole plane.

9.2.2 First order coherence matrix for a randomly perturbed grating

We are know assuming that the perturbation is a stochastic process with the following properties

$$\langle \varepsilon(\mathbf{x}) \rangle = 0 \,, \tag{9.33a}$$

$$\langle \varepsilon(\mathbf{x})\, \varepsilon(\mathbf{x}') \rangle = \sigma^2\, W(\mathbf{x}, \mathbf{x}') \,. \tag{9.33b}$$

Where W known as the auto-correlation function. Here the angle brackets denote an ensemble average over realizations of the stochastic process.

For a given realization of ε, we consider Eq. (9.29) for incident polarization β, i.e. we consider incoming p- and s-polarization separately. Equation (9.29) multiplied to the right by its Hermitian conjugate yields

$$\begin{aligned}
&\sum_{\boldsymbol{\ell}\in\mathbb{Z}^2} \sum_{\boldsymbol{\ell}'\in\mathbb{Z}^2} \mathring{\tilde{\boldsymbol{\Theta}}}{}^{+,+}_{2,1}(\mathbf{p} \mid \mathbf{p}^{(\boldsymbol{\ell})})\; \mathbf{r}^{(1)}_{\cdot\beta}(\mathbf{p}^{(\boldsymbol{\ell})} \mid \mathbf{p}_0)\; \mathbf{r}^{(1)}_{\cdot\beta}(\mathbf{p}^{(\boldsymbol{\ell}')} \mid \mathbf{p}_0)^\dagger\, \mathring{\tilde{\boldsymbol{\Theta}}}{}^{+,+}_{2,1}(\mathbf{p} \mid \mathbf{p}^{(\boldsymbol{\ell}')})^\dagger \\
&\quad = |\varepsilon^{+,-,(1)}_{2,1}(\mathbf{p} \,|\, \mathbf{p}_\mathrm{o})\,|^2\; \tilde{\boldsymbol{\Theta}}^{+,-,(1)}_{2,1,\cdot\beta}(\mathbf{p} \,|\, \mathbf{p}_\mathrm{o})\; \tilde{\boldsymbol{\Theta}}^{+,-,(1)}_{2,1,\cdot\beta}(\mathbf{p} \,|\, \mathbf{p}_\mathrm{o})^\dagger \\
&\quad + \sum_{\mathbf{m}\in\mathbb{Z}^2} \varepsilon^{+,-,(1)}_{2,1}(\mathbf{p} \,|\, \mathbf{p}_\mathrm{o})\; \varepsilon^{+,+,(1)}_{2,1}(\mathbf{p} \mid \mathbf{p}_0^{(\mathbf{m})})^*\; \tilde{\boldsymbol{\Theta}}^{+,-,(1)}_{2,1,\cdot\beta}(\mathbf{p} \,|\, \mathbf{p}_\mathrm{o})\; \mathring{\mathbf{r}}^{(\mathbf{m})}_{\cdot\beta}(\mathbf{p}_0)^\dagger\; \tilde{\boldsymbol{\Theta}}^{+,+,(1)}_{2,1}(\mathbf{p} \mid \mathbf{p}_0^{(\mathbf{m})})^\dagger \\
&\quad + \sum_{\mathbf{m}\in\mathbb{Z}^2} \varepsilon^{+,+,(1)}_{2,1}(\mathbf{p} \mid \mathbf{p}_0^{(\mathbf{m})})\; \varepsilon^{+,-,(1)}_{2,1}(\mathbf{p} \,|\, \mathbf{p}_\mathrm{o})^*\; \tilde{\boldsymbol{\Theta}}^{+,+,(1)}_{2,1}(\mathbf{p} \mid \mathbf{p}_0^{(\mathbf{m})})\; \mathring{\mathbf{r}}^{(\mathbf{m})}_{\cdot\beta}(\mathbf{p}_0)\; \tilde{\boldsymbol{\Theta}}^{+,-,(1)}_{2,1,\cdot\beta}(\mathbf{p} \,|\, \mathbf{p}_\mathrm{o})^\dagger \\
&\quad + \sum_{\mathbf{m}\in\mathbb{Z}^2} \sum_{\mathbf{m}'\in\mathbb{Z}^2} \varepsilon^{+,+,(1)}_{2,1}(\mathbf{p} \mid \mathbf{p}_0^{(\mathbf{m})})\; \varepsilon^{+,+,(1)}_{2,1}(\mathbf{p} \mid \mathbf{p}_0^{(\mathbf{m}')})^* \\
&\qquad \times \tilde{\boldsymbol{\Theta}}^{+,+,(1)}_{2,1}(\mathbf{p} \mid \mathbf{p}_0^{(\mathbf{m})})\; \mathring{\mathbf{r}}^{(\mathbf{m})}_{\cdot\beta}(\mathbf{p}_0)\; \mathring{\mathbf{r}}^{(\mathbf{m}')}_{\cdot\beta}(\mathbf{p}_0)^\dagger\; \tilde{\boldsymbol{\Theta}}^{+,+,(1)}_{2,1}(\mathbf{p} \mid \mathbf{p}_0^{(\mathbf{m}')})^\dagger \,.
\end{aligned} \tag{9.34}$$

Now taking the ensemble average of the above equation yields

$$\begin{aligned}
&\sum_{\boldsymbol{\ell}\in\mathbb{Z}^2} \sum_{\boldsymbol{\ell}'\in\mathbb{Z}^2} \mathring{\tilde{\boldsymbol{\Theta}}}{}^{+,+}_{2,1}(\mathbf{p} \mid \mathbf{p}^{(\boldsymbol{\ell})}) \left\langle \mathbf{r}^{(1)}_{\cdot\beta}(\mathbf{p}^{(\boldsymbol{\ell})} \mid \mathbf{p}_0)\; \mathbf{r}^{(1)}_{\cdot\beta}(\mathbf{p}^{(\boldsymbol{\ell}')} \mid \mathbf{p}_0)^\dagger \right\rangle \mathring{\tilde{\boldsymbol{\Theta}}}{}^{+,+}_{2,1}(\mathbf{p} \mid \mathbf{p}^{(\boldsymbol{\ell}')})^\dagger \\
&\quad = \left\langle |\varepsilon^{+,-,(1)}_{2,1}(\mathbf{p} \,|\, \mathbf{p}_\mathrm{o})\,|^2 \right\rangle \tilde{\boldsymbol{\Theta}}^{+,-,(1)}_{2,1,\cdot\beta}(\mathbf{p} \,|\, \mathbf{p}_\mathrm{o})\; \tilde{\boldsymbol{\Theta}}^{+,-,(1)}_{2,1,\cdot\beta}(\mathbf{p} \,|\, \mathbf{p}_\mathrm{o})^\dagger \\
&\quad + \sum_{\mathbf{m}\in\mathbb{Z}^2} \left\langle \varepsilon^{+,-,(1)}_{2,1}(\mathbf{p} \,|\, \mathbf{p}_\mathrm{o})\; \varepsilon^{+,+,(1)}_{2,1}(\mathbf{p} \mid \mathbf{p}_0^{(\mathbf{m})})^* \right\rangle \tilde{\boldsymbol{\Theta}}^{+,-,(1)}_{2,1,\cdot\beta}(\mathbf{p} \,|\, \mathbf{p}_\mathrm{o})\; \mathring{\mathbf{r}}^{(\mathbf{m})}_{\cdot\beta}(\mathbf{p}_0)^\dagger\; \tilde{\boldsymbol{\Theta}}^{+,+,(1)}_{2,1}(\mathbf{p} \mid \mathbf{p}_0^{(\mathbf{m})})^\dagger \\
&\quad + \sum_{\mathbf{m}\in\mathbb{Z}^2} \left\langle \varepsilon^{+,+,(1)}_{2,1}(\mathbf{p} \mid \mathbf{p}_0^{(\mathbf{m})})\; \varepsilon^{+,-,(1)}_{2,1}(\mathbf{p} \,|\, \mathbf{p}_\mathrm{o})^* \right\rangle \tilde{\boldsymbol{\Theta}}^{+,+,(1)}_{2,1}(\mathbf{p} \mid \mathbf{p}_0^{(\mathbf{m})})\; \mathring{\mathbf{r}}^{(\mathbf{m})}_{\cdot\beta}(\mathbf{p}_0)\; \tilde{\boldsymbol{\Theta}}^{+,-,(1)}_{2,1,\cdot\beta}(\mathbf{p} \,|\, \mathbf{p}_\mathrm{o})^\dagger \\
&\quad + \sum_{\mathbf{m}\in\mathbb{Z}^2} \sum_{\mathbf{m}'\in\mathbb{Z}^2} \left\langle \varepsilon^{+,+,(1)}_{2,1}(\mathbf{p} \mid \mathbf{p}_0^{(\mathbf{m})})\; \varepsilon^{+,+,(1)}_{2,1}(\mathbf{p} \mid \mathbf{p}_0^{(\mathbf{m}')})^* \right\rangle \\
&\qquad \times \tilde{\boldsymbol{\Theta}}^{+,+,(1)}_{2,1}(\mathbf{p} \mid \mathbf{p}_0^{(\mathbf{m})})\; \mathring{\mathbf{r}}^{(\mathbf{m})}_{\cdot\beta}(\mathbf{p}_0)\; \mathring{\mathbf{r}}^{(\mathbf{m}')}_{\cdot\beta}(\mathbf{p}_0)^\dagger\; \tilde{\boldsymbol{\Theta}}^{+,+,(1)}_{2,1}(\mathbf{p} \mid \mathbf{p}_0^{(\mathbf{m}')})^\dagger \,.
\end{aligned} \tag{9.35}$$

Equation (9.35) is an equation for the so-called *first order coherence matrix*

$$\left(\left\langle \mathbf{r}^{(1)}_{\cdot,\beta}(\mathbf{p}^{(\boldsymbol{\ell})} \mid \mathbf{p}_0)\ \mathbf{r}^{(1)}_{\cdot,\beta}(\mathbf{p}^{(\boldsymbol{\ell}')} \mid \mathbf{p}_0)^{\dagger} \right\rangle\right)_{\boldsymbol{\ell},\boldsymbol{\ell}'\in\mathbb{Z}^2} . \tag{9.35}$$

The elements of the coherence matrix contain covariances of the random first order reflection amplitude perturbations in the directions $\mathbf{p}^{(\boldsymbol{\ell})}$ and $\mathbf{p}^{(\boldsymbol{\ell}')}$ for incoming β polarized light. In other words, it contains information about speckle correlations.

Remark 9.4. The diagonal of the coherence matrix are proportional to the first order incoherent mean differential reflection coefficient.

In order to solve Eq. (9.35), we need to first set up its right hand side, which contains factors of the form $\left\langle \varepsilon^{b,a,(1)}_{2,1}(\mathbf{p} \mid \mathbf{q})\, \varepsilon^{b',a',(1)}_{2,1}(\mathbf{p} \mid \mathbf{q}')^{*} \right\rangle$. Let us analyze such factors in more details. By definition of $\varepsilon^{b,a,(1)}_{2,1}$, Eq. (9.25), and interchanging integrals and ensemble average, we obtain

$$\begin{aligned}
&\left\langle \varepsilon^{b,a,(1)}_{2,1}(\mathbf{p} \mid \mathbf{q})\, \varepsilon^{b',a',(1)}_{2,1}(\mathbf{p} \mid \mathbf{q}')^{*} \right\rangle \\
&= \int\!\!\int e^{-i(\mathbf{p}-\mathbf{q})\cdot\mathbf{x}+i(\mathbf{p}-\mathbf{q}')\cdot\mathbf{x}'}\, e^{-i(b\alpha_l(\mathbf{p})-a\alpha_m(\mathbf{q}))\,\zeta_0(\mathbf{x})+i(b'\alpha_l(\mathbf{p})-a'\alpha_m(\mathbf{q}'))^{*}\,\zeta_0(\mathbf{x}')}\ \langle \varepsilon(\mathbf{x})\varepsilon(\mathbf{x}')\rangle\, \mathrm{d}^2x'\ \mathrm{d}^2x\,, \\
&= \int\!\!\int e^{-i(\mathbf{p}-\mathbf{q})\cdot\mathbf{x}+i(\mathbf{p}-\mathbf{q}')\cdot\mathbf{x}'}\, e^{-i(b\alpha_l(\mathbf{p})-a\alpha_m(\mathbf{q}))\,\zeta_0(\mathbf{x})+i(b'\alpha_l(\mathbf{p})-a'\alpha_m(\mathbf{q}'))^{*}\,\zeta_0(\mathbf{x}')}\ \sigma^2 W(\mathbf{x},\mathbf{x}')\, \mathrm{d}^2x'\ \mathrm{d}^2x\ .
\end{aligned} \tag{9.36}$$

Homogeneous pertubation

In the case where the perturbations have an isotropic homogeneous auto-correlation function we can write that the auto-correlation function is a function only of the difference between two points, i.e. $\langle \varepsilon(\mathbf{x})\, \varepsilon(\mathbf{x}')\rangle = \sigma^2\, W(\mathbf{x}-\mathbf{x}')$. In this case, we can write

$$\begin{aligned}
&\left\langle \varepsilon^{b,a,(1)}_{2,1}(\mathbf{p} \mid \mathbf{q})\, \varepsilon^{b',a',(1)}_{2,1}(\mathbf{p} \mid \mathbf{q}')^{*} \right\rangle \\
&= \int\!\!\int e^{-i(\mathbf{p}-\mathbf{q})\cdot\mathbf{x}+i(\mathbf{p}-\mathbf{q}')\cdot\mathbf{x}'}\, e^{-i(b\alpha_l(\mathbf{p})-a\alpha_m(\mathbf{q}))\,\zeta_0(\mathbf{x})+i(b'\alpha_l(\mathbf{p})-a'\alpha_m(\mathbf{q}'))^{*}\,\zeta_0(\mathbf{x}')}\ \sigma^2 W(\mathbf{x}-\mathbf{x}')\, \mathrm{d}^2x'\ \mathrm{d}^2x \\
&= \int\!\!\int e^{-i\,\mathbf{p}\cdot(\mathbf{x}-\mathbf{x}')} e^{-i(\mathbf{q}'\cdot\mathbf{x}'-\mathbf{q}\cdot\mathbf{x})}\, e^{-i\gamma^{b,a}_{l,m}(\mathbf{p}\,|\,\mathbf{q})\,\zeta_0(\mathbf{x})+i\gamma^{b',a'}_{l,m}(\mathbf{p}\,|\,\mathbf{q}')^{*}\,\zeta_0(\mathbf{x}')}\ \sigma^2 W(\mathbf{x}-\mathbf{x}')\, \mathrm{d}^2x'\ \mathrm{d}^2x\ .
\end{aligned} \tag{9.37}$$

Here we have introduced the shorthand notation $\gamma^{b,a}_{l,m}(\mathbf{p}\,|\,\mathbf{q}) = b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q})$. Via a change of variable $\mathbf{u} = \mathbf{x}-\mathbf{x}'$, we obtain

$$\begin{aligned}
&\left\langle \varepsilon^{b,a,(1)}_{2,1}(\mathbf{p} \mid \mathbf{q})\, \varepsilon^{b',a',(1)}_{2,1}(\mathbf{p} \mid \mathbf{q}')^{*} \right\rangle \\
&= \sigma^2 \int e^{-i\,\mathbf{p}\cdot\mathbf{u}}\, W(\mathbf{u}) \int e^{-i(\mathbf{q}'\cdot(\mathbf{x}-\mathbf{u})-\mathbf{q}\cdot\mathbf{x})}\, e^{-i\gamma^{b,a}_{l,m}(\mathbf{p}\,|\,\mathbf{q})\,\zeta_0(\mathbf{x})+i\gamma^{b',a'}_{l,m}(\mathbf{p}\,|\,\mathbf{q}')^{*}\,\zeta_0(\mathbf{x}-\mathbf{u})}\, \mathrm{d}^2x\ \mathrm{d}^2u \\
&= \sigma^2 \int e^{-i\,\mathbf{p}\cdot\mathbf{u}}\, W(\mathbf{u}) \left(f^{b,a}_{l,m}[\mathbf{p}\,|\,\mathbf{q}] * f^{b',a'}_{l,m}[\mathbf{p}\,|\,\mathbf{q}']^{*}\right)(\mathbf{u})\, \mathrm{d}^2u\,.
\end{aligned} \tag{9.38}$$

In the last step, the inner-most integral has been written as a convolution product of the functions $f^{b,a}_{l,m}[\mathbf{p}\,|\,\mathbf{q}]$ and $f^{b',a'}_{l,m}[\mathbf{p}\,|\,\mathbf{q}']^{*}$ defined as

$$f^{b,a}_{l,m}[\mathbf{p}\,|\,\mathbf{q}](\mathbf{x}) = e^{i\,\mathbf{q}\cdot\mathbf{x}} e^{-i\gamma^{b,a}_{l,m}(\mathbf{p}\,|\,\mathbf{q})\,\zeta_0(\mathbf{x})}\,. \tag{9.39}$$

The square bracket $[\mathbf{p}\,|\,\mathbf{q}]$ simply serves to stress a dependence on $\mathbf{p}$ and $\mathbf{q}$ which play the role of parameters here. The right hand side in Eq. (9.38) reads as the Fourier transform of the product of the functions W and $f^{b,a}_{l,m}[\mathbf{p}\,|\,\mathbf{q}] * f^{b',a'}_{l,m}[\mathbf{p}\,|\,\mathbf{q}']^{*}$, which, by the use of the convolution theorem, can be written as the convolution product of the Fourier transform of

How do we solve it?

Equation (9.35) may seem rather complicated to solve at first sight. It looks complicated for several reasons. First, we need several lines to write it down, but the brave reader who made it here should not have any difficulty to identify each term and convince oneself that each term can be computed separately using techniques we have seen previously. Second, Eq. (9.35) looks like the linear systems we have been used to so far in this work, but not quite. On one hand, the "rows" of the linear system are indexed by a continuous variable $\mathbf{p}$, and on the other the unknown is a matrix sandwiched between two other matrices. The fact that the rows are indexed by a continuous variable $\mathbf{p}$ is rather natural from a physical point of view as already mentioned in a previous remark. It suffices indeed numerically to sample $\mathbf{p}$ on a discrete set of points within the unit reciprocal cell of the grating. Concerning the matrix-sandwiched form of the linear system, we will expose now how some appropriate re-indexing can be used to recover a good old matrix-vector form. To this end, it suffices to consider the term $\overset{\circ}{\tilde{\mathbf{\Theta}}}{}^{+,+}_{2,1}(\mathbf{p} \mid \mathbf{p}^{(\boldsymbol{\ell})}) \left\langle \mathbf{r}^{(1)}_{\cdot\beta}(\mathbf{p}^{(\boldsymbol{\ell})} \mid \mathbf{p}_0)\, \mathbf{r}^{(1)}_{\cdot\beta}(\mathbf{p}^{(\boldsymbol{\ell}')^\dagger} \mid \mathbf{p}_0) \right\rangle \overset{\circ}{\tilde{\mathbf{\Theta}}}{}^{+,+}_{2,1}(\mathbf{p} \mid \mathbf{p}^{(\boldsymbol{\ell}')})^\dagger$ for a given $\mathbf{p}$, $\boldsymbol{\ell}$ and $\boldsymbol{\ell}'$, and abstract it to a product of three 2×2 matrices $\mathbf{A}\,\mathbf{X}\,\mathbf{B}$ for clarity. Now let us compute explicitly the product

$$\begin{aligned}\mathbf{C} = \mathbf{A}\,\mathbf{X}\,\mathbf{B} &= \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \\ &= \begin{pmatrix} a_{11}b_{11}x_{11} + a_{11}b_{12}x_{12} + a_{12}b_{11}x_{21} + a_{12}b_{12}x_{22} & a_{11}b_{21}x_{11} + a_{11}b_{22}x_{12} + a_{12}b_{21}x_{21} + a_{12}b_{22}x_{22} \\ a_{21}b_{11}x_{11} + a_{21}b_{12}x_{12} + a_{22}b_{11}x_{21} + a_{22}b_{12}x_{22} & a_{21}b_{21}x_{11} + a_{21}b_{22}x_{12} + a_{22}b_{21}x_{21} + a_{22}b_{22}x_{22} \end{pmatrix}.\end{aligned}$$

Now, instead of seeing $\mathbf{C} = (c_{ij})_{1\le i,j\le 2}$ as a 2×2 matrix, let us rather see it as a four component vector, and similarly for $\mathbf{X}$. To be more precise, we construct the isomorphism ψ between $\mathcal{M}_{2,2}(\mathbb{C})$ and $\mathbb{C}^4$ defined as

$$\mathbf{C}' = \psi(\mathbf{C}) = \begin{pmatrix} c_{11} \\ c_{12} \\ c_{21} \\ c_{22} \end{pmatrix}, \qquad \mathbf{X}' = \psi(\mathbf{X}) = \begin{pmatrix} x_{11} \\ x_{12} \\ x_{21} \\ x_{22} \end{pmatrix}. \tag{9.40}$$

Then the equality $\mathbf{C} = \mathbf{A}\,\mathbf{X}\,\mathbf{B}$ in the space of 2×2 matrices is equivalent in the space for four component vectors to the equality

$$\begin{pmatrix} c_{11} \\ c_{12} \\ c_{21} \\ c_{22} \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} & a_{11}b_{12} & a_{12}b_{11} & a_{12}b_{12} \\ a_{11}b_{21} & a_{11}b_{22} & a_{12}b_{21} & a_{12}b_{22} \\ a_{21}b_{11} & a_{21}b_{12} & a_{22}b_{11} & a_{22}b_{12} \\ a_{21}b_{21} & a_{21}b_{22} & a_{22}b_{21} & a_{22}b_{22} \end{pmatrix} \begin{pmatrix} x_{11} \\ x_{12} \\ x_{21} \\ x_{22} \end{pmatrix}, \tag{9.41}$$

which we can write as

$$\mathbf{C}' = \mathbf{A} \otimes \mathbf{B}\,\mathbf{X}', \tag{9.42}$$

where $\mathbf{A} \otimes \mathbf{B}$ denotes the Kronecker product of the matrices $\mathbf{A}$ and $\mathbf{B}$ (and is defined by Eq. (9.41)).

9.3 Summary

In this last chapter of Part II, we have presented the detailed derivation of small amplitude perturbation theory applied in the case of multilayer systems and a derivation of small perturbation theory for a structured surface. We have shown that in the case of the multilayer system, for which the reduced Rayleigh equations have been derived in Chapter 3, a perturbative expansion of the scattering amplitudes is obtained in a very similar fashion as in the case of a single interface system with the difference that the expansion must be taken in order of products of surface profiles. In addition, we have derived closed form expressions for the mean differential scattering coefficients obtained when the amplitudes were approximated to order one in product of surface profiles. We have shown that the diffuse intensity can be expressed as the sum of the diffuse intensity of the associated systems where only one of the interface is rough at a time and the remaining are planar in the case where all the interfaces are uncorrelated, and that additional cross terms encoding the interference induced by the correlations between the surface profiles must be taken into account when the surface profiles are correlated. We have then explored how the perturbative method may be adapted to structured surface. The idea is that the surface profile may have two scale components; a large scale component representing an ideal structured profile and a small scale perturbation added to it. Instead of treating the whole profile as a perturbation of a planar surface, which is expected to yield poor results if the large scale component as a significant amplitude, we have chosen to adapt the technique to perturb from the structured surface with the hope of obtaining more accurate approximations. Indeed, the response of the ideal structured can be computed exactly once and then the correction due to the small scale component remains to be computed approximately. Focusing on the first order, we have obtained a linear system for the average coherency matrix and we have briefly explained how the problem could be solved numerically. The latter method could be of particular interest for applications ranging from the design and quality control of optical grating to the scattering of electromagnetic wave from the ocean.

Part III

Applications

Chapter 10

Single randomly rough interface

"Oh, la belle obscurité. [...] On ne sait plus ce que c'est que l'obscurité.
A force de vouloir faire la lumière sur tout, on ne distingue plus rien ![...]
Regardez, regardez ce coin sombre ! Tout à l'heure à la lumière il passait inaperçu."
Les ombres d'antan, Raymond Devos (1973).

In this chapter, we present numerical calculations and a theoretical analysis of the scattering response of systems made of two media separated by a randomly rough interface. We will limit ourselves to a regime of weakly rough surfaces, for which approximations based on first order small amplitude perturbation theory are sufficient to accurately describe the scattering response. In fact, we will start by considering the case of a system made of two dielectric media separated by a randomly weakly rough surface for which one can observe phenomena associated with physical mechanism which have been identified only recently **Paper** [2]: the optical Yoneda effect and the Brewster scattering effect. The solution given by small amplitude perturbation theory, which we have treated simply as a particular iterative method up to this point, will prove to be a powerful tool which can lead to the understanding of the underlying physics of the observed phenomena. This is the reason why, we will first revisit the solution given by the first order perturbation theory and re-write it in a form which intimately relates to physical mechanisms. The discussion is adapted from **Paper** [2]. Finally, effects due to higher order scattering events for metallic surfaces will be mentioned briefly.

10.1 Scattering by a dielectric interface

The present section is devoted to the study of the scattering of light by a randomly rough surface separating two dielectric media. Numerical simulations based on Monte Carlo simulations (see Chapter 8 for details about the methods) of the reduced Rayleigh equations and the solution given by small amplitude perturbation theory (SAPT) to first order (for the scattering amplitude) in the surface profile will be used for our exploration. The Monte Carlo simulation will essentially be used to justify the use of the pertubative solution or stress deviations which then may be explored further by analyzing higher order terms. In the following, the dielectric constants of the media are denoted ϵ_1 and ϵ_2, the *smallest* and *largest* of the two dielectric constants are denoted respectively $\epsilon_{\min}$ and $\epsilon_{\max}$, σ and a denote the rms-roughness and the correlation length of the surface with Gaussian height distribution and Gaussian auto-correlation function respectively (unless otherwise specified).

10.1.1 Physical interpretation of SAPT to first order

Let us start by revisiting the solution of the reduced Rayleigh equations given by SAPT to first order in the surface profile function. The aim here is to use the approximate solution to shed some light on the physical interpretation.

Order zero - Revisiting the Fresnel amplitudes

Consider first the zero order reflection amplitudes given by Eq. (6.8), and which are nothing else but the Fresnel amplitudes, and hence can be written as (see Section 2.3 for derivation)

$$\rho_{ss}^{(0)}(\mathbf{p}_0) = \frac{\alpha_1(\mathbf{p}_0) - \alpha_2(\mathbf{p}_0)}{\alpha_1(\mathbf{p}_0) + \alpha_2(\mathbf{p}_0)} \tag{10.1a}$$

$$\rho_{pp}^{(0)}(\mathbf{p}_0) = \frac{\epsilon_2\alpha_1(\mathbf{p}_0) - \epsilon_1\alpha_2(\mathbf{p}_0)}{\epsilon_2\alpha_1(\mathbf{p}_0) + \epsilon_1\alpha_2(\mathbf{p}_0)} \,. \tag{10.1b}$$

The Fresnel amplitudes as written above, in terms of components of the wave vectors along $\hat{\mathbf{e}}_3$, are common expressions. They can be recast to make appear the angle of incidence and the angle of reflection if we so wished. The form of the Fresnel amplitudes given in Eq. (10.1) is rather compact but hides the microscopic physical mechanisms underlying the macroscopic response. The reflection amplitudes can be re-written in a different manner, which are obtained directly from Eq. (6.8) together with the definition of the $\mathbf{M}_{l,m}^{b,a}$ matrix in terms of the change of local basis between polarization vectors, Eq. (2.45), and read

$$\rho_{ss}^{(0)}(\mathbf{p}_0) = \rho^{(0)}(\mathbf{p}_0)\, \frac{\hat{\mathbf{e}}_s(\mathbf{p}_0) \cdot \hat{\mathbf{e}}_s(\mathbf{p}_0)}{\hat{\mathbf{e}}_s(\mathbf{p}_0) \cdot \hat{\mathbf{e}}_s(\mathbf{p}_0)} \tag{10.2a}$$

$$\rho_{pp}^{(0)}(\mathbf{p}_0) = \rho^{(0)}(\mathbf{p}_0)\, \frac{\hat{\mathbf{e}}_{p,2}^{+}(\mathbf{p}_0) \cdot \hat{\mathbf{e}}_{p,1}^{-}(\mathbf{p}_0)}{\hat{\mathbf{e}}_{p,2}^{+}(\mathbf{p}_0) \cdot \hat{\mathbf{e}}_{p,1}^{+}(\mathbf{p}_0)} \tag{10.2b}$$

$$\rho^{(0)}(\mathbf{p}_0) = \frac{\alpha_1(\mathbf{p}_0) - \alpha_2(\mathbf{p}_0)}{\alpha_1(\mathbf{p}_0) + \alpha_2(\mathbf{p}_0)} \,. \tag{10.2c}$$

Here we have defined the *scalar reflection amplitude* $\rho^{(0)}(\mathbf{p}_0)$, which corresponds to the reflection amplitude for a scalar wave reflected by a planar interface subjected to the continuity of the field and its normal derivative with respect to the surface (see Section 2.6). The factorization in Eqs. (10.2a) and (10.2b) is elementary, but nevertheless interesting when it comes to the physical interpretation of the Fresnel amplitudes. Indeed, the rather unusual factorized form of the Fresnel amplitudes expressed in Eqs. (10.2a) and (10.2b) reveals that the reflection amplitude of the zero order field is the product of the reflection amplitude of the corresponding scalar problem $\rho^{(0)}(\mathbf{p}_0)$ (which is independent of polarization) and a factor which is dependent on the polarization state. We have chosen on purpose not to simplify the dot products (all equal to unity) for s-polarized light in Eq. (10.2a) to show the similarity with the second factor in Eq. (10.2b), and that both are related to the dipolar nature of the radiation emitted by the elementary scatterers in the media (atoms, molecules). The factorization can be interpreted microscopically in the following way. The amplitude of the reflected wave is given by the contribution of arrays of scatterers emitting coherently, or in phase, after being excited by the incident wave as viewed as propagating in vacuum between the scatterers. The superposition of the elementary wavelets emitted by each scatterer forms the reflected and transmitted waves. The resultant waves can be described by the superposition of spherical waves and their amplitudes are encoded by the term $\rho^{(0)}(\mathbf{p}_0)$ for the reflected wave and $\tau^{(0)}(\mathbf{p}_0)$ for the transmitted wave (see Eq. (10.3)). This

result is what one would get for the reflection of a scalar wave based on the Huygens principle. However, the polarization of the incident field excites the elementary dipole associated with each scatterer along different directions depending on the polarization state of the incident wave. For a s-polarized wave incident in the $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_3)$-plane each dipole is excited along $\hat{\mathbf{e}}_2$. Knowing that the radiation of a dipole is isotropic in the plane orthogonal to the direction of oscillation (see Section 1.3), the dipolar polarization projection reduces to unity independently of the angle of incidence in Eq. (10.2a). This is the reason why the reflection amplitude for a s-polarized wave reduces to that of the scalar wave in Eq. (10.2a). For a p-polarized wave incident in the $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_3)$-plane each dipole is excited along a direction in the $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_3)$-plane. This direction is a priori different in the different media and depends of all the waves viewed by the scatterers in the stationary regime, and is a priori not straightforward to determine based on a microscopic theory. Nevertheless, for obvious symmetry reason, the oscillations are always in the plane of incidence and result in a modulation of the scalar reflection amplitude due to the dipole radiation pattern in the plane of incidence. This is precisely the nature of the modulation obtained in the second factor of Eq. (10.2b). This factorization and its microscopic interpretation was suggested by W. T. Doyle in Ref. [27] with the aim of providing a solid physical interpretation of Brewster's law. The detailed analysis and physical interpretation given by Doyle was based on the Ewald-Oseen extinction theorem, in particular, the original derivation given by Ewald based on microscopic optics [28]. Similarly the transmission amplitudes are given by

$$\tau_{ss}^{(0)}(\mathbf{p}_0) = \frac{\tau^{(0)}(\mathbf{p}_0)}{\hat{\mathbf{e}}_s(\mathbf{p}_0) \cdot \hat{\mathbf{e}}_s(\mathbf{p}_0)} \tag{10.3a}$$

$$\tau_{pp}^{(0)}(\mathbf{p}_0) = \frac{\tau^{(0)}(\mathbf{p}_0)}{\hat{\mathbf{e}}_{p,1}^{-}(\mathbf{p}_0) \cdot \hat{\mathbf{e}}_{p,2}^{-}(\mathbf{p}_0)} \tag{10.3b}$$

$$\tau^{(0)}(\mathbf{p}_0) = \frac{c^2}{\omega^2} \frac{2\alpha_1(\mathbf{p}_0)}{(\epsilon_1 - \epsilon_2)} (\alpha_1(\mathbf{p}_0) - \alpha_2(\mathbf{p}_0)) \,, \tag{10.3c}$$

where we also make explicit the scalar-polarization factorization.

Total internal reflection – We have seen in Chapter 1 that the phenomenon of total internal reflection occurs under the condition that light is reflected in the optically denser medium ($\epsilon_1 > \epsilon_2$). We have seen that the critical angle of incidence $\theta_0 = \theta_c = \arcsin(n_2/n_1)$ for total internal reflection corresponds to the transition at which the transmitted wave goes from being progressive for $\theta_0 < \theta_c$ to being evanescent for $\theta_0 > \theta_c$. Equivalently, this condition may be written for a critical norm of the in-plane wave vector $p_c = n_2\,\omega/c$, such that for all in-plane wave vectors where $|\,\mathbf{p}_0\,| > p_c$, the incident power is entirely reflected. The phenomenon of total internal reflection is controlled by the factor $\rho^{(0)}(\mathbf{p}_0)$ which is present for both polarizations. This observation indicates that the phenomenon can be analyzed from a scalar wave picture decoupled from polarization effects. In a few words, total internal reflection occurs whenever the refracted wave is evanescent in the medium of transmission, and therefore it cannot transport energy away from the surface. Let us analyze in some details the behavior of the reflection amplitude $\rho^{(0)}(\mathbf{p}_0)$ as one varies the incident in-plane wave vector $\mathbf{p}_0$ and in particular across the threshold $|\,\mathbf{p}_0\,| = p_c$. At normal incidence, $\mathbf{p}_0 = \mathbf{0}$, and $\rho^{(0)}(\mathbf{0}) = (n_1 - n_2)/(n_1 + n_2) > 0$ which lies on the real line. For $|\,\mathbf{p}_0\,| < p_c$, both $\alpha_1(\mathbf{p}_0)$ and $\alpha_2(\mathbf{p}_0)$ are real and hence so is $\rho^{(0)}(\mathbf{p}_0)$. As $|\,\mathbf{p}_0\,| \to p_c$, $\rho^{(0)}(\mathbf{p}_0)$ increases on the real line towards the value 1 when α_2 vanishes, and $\alpha_1(\mathbf{p}_0) = \sqrt{\epsilon_1 - \epsilon_2}\,\omega/c$ for $|\,\mathbf{p}_0\,| = p_c$. When $|\,\mathbf{p}_0\,| > p_c$, $\alpha_2(\mathbf{p}_0) = i\sqrt{\mathbf{p}_0^2 - \epsilon_2\,\omega^2/c^2}$ turns pure imaginary and $\rho^{(0)}(\mathbf{p}_0)$ draws a circular arc in the lower half of the complex plane (negative imaginary part) with unit modulus. It is immediate to see that $|\rho^{(0)}(\mathbf{p}_0)| = 1$ for $|\,\mathbf{p}_0\,| > p_c$ since

in this case, $\rho^{(0)}(\mathbf{p}_0)$ is of the form $z^*/z = \exp(-2i\varphi)$ where $z = r\exp(i\varphi)$ is a non-zero complex number. As $|\,\mathbf{p}_0\,| \to n_1\omega/c$, the reflected wave (as well as the incident wave) reaches the limit of propagation in the first medium and $\alpha_1(\mathbf{p}_0) \to 0$ which makes the reflection amplitude real and negative, equal to -1. To sum up, as $|\,\mathbf{p}_0\,|$ goes from the critical point p_c to grazing incidence $n_1\omega/c$, the reflection amplitude traces a half circle in the complex plane with unit modulus. The argument of the complex reflection amplitude, hence varies from 0 to $-\pi$ rad. This gradual change of phase with the angle of incidence in the regime of total internal reflection is known as the Goos-Hänchen phase shift [17, 59]. It can be interpreted as follows: If we regard the reflected and refracted waves as two pieces of a single mode, then as the wave enters the second medium, in which it is evanescent, it oscillates while propagating along the surface before it eventually goes back into the first medium where it can continue to propagate to infinity. As the wave propagates along the surface, it acquires a temporal delay which depends on its wave vector. This delay is translated into a phase shift with respect to the incident wave.

Remark 10.1. Note the presence of the factor $\alpha_1(\mathbf{p}_0) - \alpha_2(\mathbf{p}_0)$ in the transmission amplitude of the scalar wave in Eq. (10.3c) which is identical to the numerator of the reflection amplitude in Eq. (10.2c). The analysis of this term on total internal reflection hence leads to a similar behavior for the transmission amplitude, in the sense that $\tau^{(0)}(\mathbf{p}_0)$ leaves the real line and traces a path in the complex plane when total internal reflection occurs. This fact illustrates the coupling between the reflected and the transmitted waves, which may be interpreted as two components of the same mode.

Brewster's law, Ewald triad and dipole radiation – For non-magnetic media, the Brewster phenomenon for reflection by a planar surface can only be observed for p-polarized waves. We have seen in Chapter 1 that the so-called Brewster angle of incidence for which the reflection amplitude for a p-polarized wave vanishes was found to be equal to $\theta_0 = \theta_{\mathrm{B}} = \arctan(n_2/n_1)$. Moreover it was observed that the Brewster phenomenon occurs when the geometrical condition of orthogonality between the wave vector of the (non)reflected wave and the wave vector of the transmitted wave is satisfied. It is often argued in textbooks that the physical mechanism at the origin of the Brewster phenomenon is the fact that the dipoles in the medium of transmission oscillate along the electric field of the transmitted wave and consequently, since dipoles do not radiate energy along the axis of oscillation, no energy can be radiated in the direction given by the wave vector of the reflected wave under the aforementioned condition. However, this explanation, although attractive, cannot be complete. Indeed, as was pointed out by P. Lorrain and D. R. Corson in Ref. [89], the Brewster phenomenon can also be observed when the incident wave is incident in a dielectric medium and impinges on the surface between this dielectric and vacuum. In such a case, there is no dipole oscillating in the vacuum. Doyle resolved this issue in Ref. [27] by careful examination of the so-called *Ewald wave triads* in each medium based on the original microscopic derivation of the extinction theorem by Ewald [28]. Let us analyze the Brewster's law in the light of Eq. (10.2b) following Doyle's reasoning.

From Eq. (10.2b), it is clear that the Fresnel amplitude for p-polarized light is proportional to $\hat{\mathbf{e}}^+_{p,2}(\mathbf{p}_0)\cdot\hat{\mathbf{e}}^-_{p,1}(\mathbf{p}_0)$. In other words, the Fresnel amplitude is proportional to the component of the unit incident electric field (given by $\hat{\mathbf{e}}^-_{p,1}(\mathbf{p}_0)$) along the direction given by $\hat{\mathbf{e}}^+_{p,2}(\mathbf{p}_0)$. What does $\hat{\mathbf{e}}^+_{p,2}(\mathbf{p}_0)$ correspond to? The direction given by $\hat{\mathbf{e}}^+_{p,2}(\mathbf{p}_0)$ corresponds to the local p-polarization direction for a wave whose wave vector is given by $\mathbf{k}^+_2(\mathbf{p}_0)$, i.e. a wave which propagates upwards as if it were propagating in medium 2. Therefore, it seems that the reflected amplitude for p-polarized light depends on a projection of the incident field along the polarization vector of a seemingly *nonexisting* or *virtual* wave, propagating along the wave vector $\mathbf{k}^+_2(\mathbf{p}_0)$. The virtual

wave does, in fact, have a physical interpretation which was provided by Doyle [27] based on the concept of *wave triads* introduced by Ewald [28]. The physical system considered by Ewald is made of a dense array of dipole scatterers, constituting the dielectric medium, occupying a half space. This array of dipole scatterers is excited by an incident plane wave incident from the vacuum half-space and which is present in the whole space, including the space in between the scatterers. Ewald argued that, in a stationary regime, the dipole scatterers would respond to the excitation and to their mutual interactions, via the waves scattered by each elementary dipole, in such a way that there exist planes of scatterers of coherent response. This means that all dipoles within such a plane oscillate in phase. Consequently and due to the fact that the array of scatterers is bounded within a half space by a planar interface, Ewald showed that the superposition of the elementary waves emitted by each individual scatterer results in the propagation of three plane waves: two waves named *vacuum waves* propagating with a phase velocity equal to c and one wave propagating with phase velocity c/n named the *polarization wave*, where n corresponds to the refractive medium made of scatterers within the macroscopic picture. The polarization wave, which propagates with phase velocity c/n, corresponds to the transmitted wave in the macroscopic picture, while one of the waves propagating with phase velocity c serves to exactly cancel the incident wave within the dielectric medium. The remaining wave propagating with phase velocity c exits the medium and corresponds to the reflected wave. The wave vectors of the different waves are, as expected, given by Snell's law, and Ewald's derivation can be regarded as a microscopic validation of Snell's law.

Ewald's result can be generalized to the case of two half-spaces filled with dipoles of different dipole moments and separated by a planar interface. The main difference being that the superposition of the waves emitted by the scatterers on both sides of the interface must be taken into account. According to Doyle [27], this yields three wave triads: one triad for the incident wave, one for the reflected wave and one for the refracted wave. To be more accurate, it can be shown that to the incident polarization wave in medium 1, propagating with phase velocity c/n_1 and wave vector $\mathbf{k}_1^-(\mathbf{p}_0)$ is associated two waves propagating with phase velocity c/n_2 with wave vectors $\mathbf{k}_2^\pm(\mathbf{p}_0)$. Similarly, to the reflected (resp. refracted) polarization wave, whose wave vector is given by $\mathbf{k}_1^+(\mathbf{p}_0)$ (resp. $\mathbf{k}_2^-(\mathbf{p}_0)$), is associated two waves propagating with phase velocity c/n_2 (resp. c/n_1) and wave vectors $\mathbf{k}_2^\pm(\mathbf{p}_0)$ (resp. $\mathbf{k}_1^\pm(\mathbf{p}_0)$). The amplitudes of the waves in the triads are linked by the so-called dynamical conditions [27]. The amplitudes of the different waves are such that (i) the wave associated with the refracted polarization wave and propagating along $\mathbf{k}_1^-(\mathbf{p}_0)$ in medium 2 cancels the incident wave and (ii) that the superposition of waves associated with the incident and reflected polarization waves and propagating along $\mathbf{k}_2^+(\mathbf{p}_0)$ vanishes (more details can be found in Refs. [27, 28]).

In our analysis of Brewster's law and, more generally, in the analysis of the Brewster scattering effect in the case of the diffusely scattered light, it is convenient to introduce the concept of *Snell-conjugate* wave vectors, or in a more abusive language Snell-conjugate waves. The wave vectors $\mathbf{k}_1^a(\mathbf{p})$ and $\mathbf{k}_2^a(\mathbf{p})$, i.e. wave vectors satisfying the dispersion relation respectively in medium 1 and 2, sharing the *same* projection in the $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2)$-plane and pointing either both upward or downward, will be referred to as *Snell-conjugate* wave vectors.

Let us come back now to the polarization dependence of the reflection amplitudes for p-polarized light in Eq. (10.2b). We readily observe that $\rho_{pp}^{(0)}(\mathbf{p}_0)$ is proportional to $\hat{\mathbf{e}}_{p,2}^+(\mathbf{p}_0) \cdot \hat{\mathbf{e}}_{p,1}^-(\mathbf{p}_0)$ which indicates that the amplitude of the reflected wave is controlled by the component of the incident field along the p-polarization vector associated with the Snell-conjugate wave vector

of the wave vector of the reflected wave. This is a clear indication that the direction of the dipole oscillation is intimately linked to waves in the Ewald triad. How can Brewster's law be rephrased in view of Eq. (10.2b) and with the concept of Snell-conjugate waves in mind? In the case of a planar interface between two dielectrics, the Brewster phenomenon is commonly defined as the extinction of the p-polarized reflected wave, that is when the right-hand side in Eq. (10.2b) vanishes. It is clear that, assuming $\epsilon_1 \neq \epsilon_2$, the Fresnel amplitude vanishes if and only if $\hat{\mathbf{e}}^+_{p,2}(\mathbf{p}_0) \cdot \hat{\mathbf{e}}^-_{p,1}(\mathbf{p}_0) = 0$. In other words, the condition for Brewster's law can be rephrased as the orthogonality between the polarization vector of the incident wave and the polarization vector of the *Snell-conjugate* wave associated with the reflected wave. This slightly different way of stating Brewster's law is of course equivalent to the common result from textbooks. Indeed, if $\hat{\mathbf{e}}^-_{p,1}(\mathbf{p}_0) \cdot \hat{\mathbf{e}}^+_{p,2}(\mathbf{p}_0) = 0$ it is immediate that $\mathbf{k}^+_1(\mathbf{p}_0) \cdot \mathbf{k}^-_2(\mathbf{p}_0) = 0$, i.e. that the angle between the (non)reflected wave and the transmitted wave reaches 90°. However, we will see below that the new geometrical criterion proposed above holds when applied to a Snell-conjugate wave vector associated with a non-specularly scattered wave.

Concerning the transmitted wave now, Eq. (10.3) shows that neither the s- nor p-polarized zero order transmitted wave vanishes in general. No Brewster angle is known for transmission through a *planar* interface. This fact does not, however, prevent the existence of Brewster *scattering* angles in the diffusely transmitted light as we will see when analyzing the first order term of SAPT.

First order

Scalar waves – We analyze now the first order amplitude given by SAPT. Let us consider first the case of a scalar wave subjected to the continuity of the field and its normal derivative with respect to the surface (see Section 2.6). According to Section 2.6, it suffices to replace the $\mathbf{M}^{ba}_{lm}$ matrices by $k_1 k_2 \, \omega^2/c^2$ and the identity matrix by 1 in Eqs. (6.12) and (6.19) to obtain the first order reflection and transmission amplitudes for the scalar problem. This yields

$$R^{(1)}(\mathbf{p} \,|\, \mathbf{p}_0) = (\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p}))\, \hat{\zeta}(\mathbf{p} - \mathbf{p}_0)\, [1 + \rho^{(0)}(\mathbf{p}_0)] \tag{10.4a}$$

$$T^{(1)}(\mathbf{p} \,|\, \mathbf{p}_0) = (\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p}))\, \hat{\zeta}(\mathbf{p} - \mathbf{p}_0)\, \tau^{(0)}(\mathbf{p}_0)\,. \tag{10.4b}$$

The first point to notice in Eqs. (10.4a) and (10.4b) is that *both* the reflection *and* transmission amplitudes are proportional to $(\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p}))\, \hat{\zeta}(\mathbf{p} - \mathbf{p}_0)$. Second, this common factor encodes to whole $\mathbf{p}$-dependence of the amplitudes. Indeed, the remaining factors $1 + \rho^{(0)}(\mathbf{p}_0)$ and $\tau^{(0)}(\mathbf{p}_0)$ respectively, only depend on the incident in-plane wave vector $\mathbf{p}_0$. How should we interpret such a factorization? Since the reflection and transmission amplitudes vary in the same way with $\mathbf{p}$, we can say that both the elementary reflected wave with in-plane wave vector $\mathbf{p}$ and the elementary transmitted wave with the *same* wave vector $\mathbf{p}$ are coupled. They can be viewed as two pieces of a single scattered mode. In fact, for scalar waves, we have $1 + \rho^{(0)}(\mathbf{p}_0) = \tau^{(0)}(\mathbf{p}_0)$, which gives $R^{(1)}(\mathbf{p} \,|\, \mathbf{p}_0) = T^{(1)}(\mathbf{p} \,|\, \mathbf{p}_0)$; the first order reflection and transmission amplitudes are equal! This strengthen the coupled mode interpretation. Furthermore, since there is no summation over intermediate wave vectors $\mathbf{q}$, we can consider that a scattered mode characterized by a wave vector $\mathbf{p}$ is decoupled from a scattered mode characterized by a wave vector $\mathbf{q} \neq \mathbf{p}$. The scattered modes are, however, coupled to the incident wave which acts as a source. It is also fruitful to go back to the derivation of SAPT to get a intuitive understanding of the meaning of the factor $(\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p}))\, \hat{\zeta}(\mathbf{p} - \mathbf{p}_0)$. This factor originates from the Taylor expansion of the exponential factor $\exp[-i(b\alpha_2(\cdot) - \alpha_1(\cdot))\zeta(\cdot)]$ in the definition of the $\mathcal{J}^{ba}_{lm}$-integral. What the

$\mathcal{J}^{ba}_{lm}$-integral encodes is intuitively a sum of complex amplitudes along the surface where each scattered path will experience different phase shift depending on their scattering event along the surface. The factor $\exp[-i(b\alpha_2(\cdot) - \alpha_1(\cdot))\zeta(\cdot)]$ can be thought as a phase factor varying along the surface. The factor $(\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p}))\,\hat{\zeta}(\mathbf{p} - \mathbf{p}_0)$ originates from the linear approximation of this phase factor. Thus it corresponds to the approximation of the *interference* pattern resulting from single scattering events with phases linearly approximated. In other words, it gives the speckle pattern. Taking one step further in our interpretation, $(\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p}))\,\hat{\zeta}(\mathbf{p} - \mathbf{p}_0)$ can then be roughly said to be a *probability amplitude* for a change of in-plane wave vector (or momentum) from $\mathbf{p}_0$ to $\mathbf{p}$. This is not the entire reflection nor transmission amplitude yet. We still have to give an interpretation to the remaining factors in Eq. (10.4). The factor $1 + \rho^{(0)}(\mathbf{p}_0)$ is the sum of the incident unit field and the corresponding reflected zero order field, hence it describes the *total zero order field* in medium 1. Similarly, $\tau^{(0)}(\mathbf{p}_0)$ is the zero order transmitted amplitude and it describes the total zero order field in medium 2. Thus the first order reflected and transmitted amplitudes are proportional, respectively, to the total zero order field in the medium of reflection and transmission. In other words, we can say that it is not only the incident wave that acts as a source for the first order waves, but the sum of the incident and scattered zero order field. We can summarize our overall interpretation as follows: *the first order reflection and transmission amplitudes are the product of the total zero order field which characterizes a zero order state and a common probability amplitude for a single change of in-plane momentum from* $\mathbf{p}_0$ *to* $\mathbf{p}$ *given a zero order state characterized only by* $\mathbf{p}_0$.

Electromagnetic waves – Now that we have given a physical interpretation to the first order amplitudes for scalar waves, let us consider the case of electromagnetic waves. The first order amplitudes for electromagnetic waves are given by Eqs. (6.12) and (6.19). We have seen when discussing the Fresnel amplitudes that, in view of a physical interpretation, it was beneficial to express the $\mathbf{M}^{ba}_{lm}$ matrices in terms of the polarization vectors according to Eq. (2.45). Following the same idea, we express the first order amplitudes as a function of the polarization vectors as

$$\mathbf{R}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_0) = (\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p}))\,\hat{\zeta}(\mathbf{p} - \mathbf{p}_0)\,\hat{\boldsymbol{\rho}}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_0) \tag{10.5a}$$

$$\mathbf{T}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_0) = (\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p}))\,\hat{\zeta}(\mathbf{p} - \mathbf{p}_0)\,\hat{\boldsymbol{\tau}}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_0) \tag{10.5b}$$

with

$$\hat{\rho}^{(1)}_{ss}(\mathbf{p}\,|\,\mathbf{p}_0) = \frac{\hat{\mathbf{e}}_s(\mathbf{p})\cdot\hat{\mathbf{e}}_s(\mathbf{p}_0) + \rho^{(0)}_{ss}(\mathbf{p}_0)\;\hat{\mathbf{e}}_s(\mathbf{p})\cdot\hat{\mathbf{e}}_s(\mathbf{p}_0)}{\hat{\mathbf{e}}_s(\mathbf{p})\cdot\hat{\mathbf{e}}_s(\mathbf{p})} = \frac{\hat{\mathbf{e}}_s(\mathbf{p})\cdot\mathbf{E}^{(0)}_{1,s}(\mathbf{p}_0)}{\hat{\mathbf{e}}_s(\mathbf{p})\cdot\hat{\mathbf{e}}_s(\mathbf{p})} \tag{10.6a}$$

$$\hat{\rho}^{(1)}_{ps}(\mathbf{p}\,|\,\mathbf{p}_0) = \frac{\hat{\mathbf{e}}^{+}_{p,2}(\mathbf{p})\cdot\hat{\mathbf{e}}_s(\mathbf{p}_0) + \rho^{(0)}_{ss}(\mathbf{p}_0)\,\hat{\mathbf{e}}^{+}_{p,2}(\mathbf{p})\cdot\hat{\mathbf{e}}_s(\mathbf{p}_0)}{\hat{\mathbf{e}}^{+}_{p,2}(\mathbf{p})\cdot\hat{\mathbf{e}}^{+}_{p,1}(\mathbf{p})} = \frac{\hat{\mathbf{e}}^{+}_{p,2}(\mathbf{p})\cdot\mathbf{E}^{(0)}_{1,s}(\mathbf{p}_0)}{\hat{\mathbf{e}}^{+}_{p,2}(\mathbf{p})\cdot\hat{\mathbf{e}}^{+}_{p,1}(\mathbf{p})} \tag{10.6b}$$

$$\hat{\rho}^{(1)}_{sp}(\mathbf{p}\,|\,\mathbf{p}_0) = \frac{\hat{\mathbf{e}}_s(\mathbf{p})\cdot\hat{\mathbf{e}}^{-}_{p,1}(\mathbf{p}_0) + \hat{\mathbf{e}}_s(\mathbf{p})\cdot\hat{\mathbf{e}}^{+}_{p,1}(\mathbf{p}_0)\rho^{(0)}_{pp}(\mathbf{p}_0)}{\hat{\mathbf{e}}_s(\mathbf{p})\cdot\hat{\mathbf{e}}_s(\mathbf{p})} = \frac{\hat{\mathbf{e}}_s(\mathbf{p})\cdot\mathbf{E}^{(0)}_{1,p}(\mathbf{p}_0)}{\hat{\mathbf{e}}_s(\mathbf{p})\cdot\hat{\mathbf{e}}_s(\mathbf{p})} \tag{10.6c}$$

$$\hat{\rho}^{(1)}_{pp}(\mathbf{p}\,|\,\mathbf{p}_0) = \frac{\hat{\mathbf{e}}^{+}_{p,2}(\mathbf{p})\cdot\hat{\mathbf{e}}^{-}_{p,1}(\mathbf{p}_0) + \hat{\mathbf{e}}^{+}_{p,2}(\mathbf{p})\cdot\hat{\mathbf{e}}^{+}_{p,1}(\mathbf{p}_0)\rho^{(0)}_{pp}(\mathbf{p}_0)}{\hat{\mathbf{e}}^{+}_{p,2}(\mathbf{p})\cdot\hat{\mathbf{e}}^{+}_{p,1}(\mathbf{p})} = \frac{\hat{\mathbf{e}}^{+}_{p,2}(\mathbf{p})\cdot\mathbf{E}^{(0)}_{1,p}(\mathbf{p}_0)}{\hat{\mathbf{e}}^{+}_{p,2}(\mathbf{p})\cdot\hat{\mathbf{e}}^{+}_{p,1}(\mathbf{p})} \tag{10.6d}$$

for the reflection amplitudes and

$$\hat{\tau}^{(1)}_{ss}(\mathbf{p}\,|\,\mathbf{p}_0) = \tau^{(0)}_{ss}(\mathbf{p}_0)\,\frac{\hat{\mathbf{e}}_s(\mathbf{p})\cdot\hat{\mathbf{e}}_s(\mathbf{p}_0)}{\hat{\mathbf{e}}_s(\mathbf{p})\cdot\hat{\mathbf{e}}_s(\mathbf{p})} = \frac{\hat{\mathbf{e}}_s(\mathbf{p})\cdot\mathbf{E}^{(0)}_{2,s}(\mathbf{p}_0)}{\hat{\mathbf{e}}_s(\mathbf{p})\cdot\hat{\mathbf{e}}_s(\mathbf{p})} \tag{10.7a}$$

$$\hat{\tau}^{(1)}_{ps}(\mathbf{p}\,|\,\mathbf{p}_0) = \tau^{(0)}_{ss}(\mathbf{p}_0)\,\frac{\hat{\mathbf{e}}^{-}_{p,1}(\mathbf{p})\cdot\hat{\mathbf{e}}_s(\mathbf{p}_0)}{\hat{\mathbf{e}}^{-}_{p,1}(\mathbf{p})\cdot\hat{\mathbf{e}}^{-}_{p,2}(\mathbf{p})} = \frac{\hat{\mathbf{e}}^{-}_{p,1}(\mathbf{p})\cdot\mathbf{E}^{(0)}_{2,s}(\mathbf{p}_0)}{\hat{\mathbf{e}}^{-}_{p,1}(\mathbf{p})\cdot\hat{\mathbf{e}}^{-}_{p,2}(\mathbf{p})} \tag{10.7b}$$

$$\hat{\tau}^{(1)}_{sp}(\mathbf{p}\,|\,\mathbf{p}_0) = \tau^{(0)}_{pp}(\mathbf{p}_0)\,\frac{\hat{\mathbf{e}}_s(\mathbf{p})\cdot\hat{\mathbf{e}}^-_{p,2}(\mathbf{p}_0)}{\hat{\mathbf{e}}_s(\mathbf{p})\cdot\hat{\mathbf{e}}_s(\mathbf{p})} = \frac{\hat{\mathbf{e}}_s(\mathbf{p})\cdot\mathbf{E}^{(0)}_{2,p}(\mathbf{p}_0)}{\hat{\mathbf{e}}_s(\mathbf{p})\cdot\hat{\mathbf{e}}_s(\mathbf{p})} \tag{10.7c}$$

$$\hat{\tau}^{(1)}_{pp}(\mathbf{p}\,|\,\mathbf{p}_0) = \tau^{(0)}_{pp}(\mathbf{p}_0)\,\frac{\hat{\mathbf{e}}^-_{p,1}(\mathbf{p})\cdot\hat{\mathbf{e}}^-_{p,2}(\mathbf{p}_0)}{\hat{\mathbf{e}}^-_{p,1}(\mathbf{p})\cdot\hat{\mathbf{e}}^-_{p,2}(\mathbf{p})} = \frac{\hat{\mathbf{e}}^-_{p,1}(\mathbf{p})\cdot\mathbf{E}^{(0)}_{2,p}(\mathbf{p}_0)}{\hat{\mathbf{e}}^-_{p,1}(\mathbf{p})\cdot\hat{\mathbf{e}}^-_{p,2}(\mathbf{p})} \tag{10.7d}$$

for the transmission amplitudes. Here we have defined the total zero order field amplitudes in media 1 and 2, for s- and p-polarized incident light, as

$$\mathbf{E}^{(0)}_{1,s}(\mathbf{p}_0) = \left[1+\rho^{(0)}_{ss}(\mathbf{p}_0)\right]\,\hat{\mathbf{e}}_s(\mathbf{p}_0) \tag{10.8a}$$

$$\mathbf{E}^{(0)}_{1,p}(\mathbf{p}_0) = \hat{\mathbf{e}}^-_{p,1}(\mathbf{p}_0) + \rho^{(0)}_{pp}(\mathbf{p}_0)\,\hat{\mathbf{e}}^+_{p,1}(\mathbf{p}_0) \tag{10.8b}$$

$$\mathbf{E}^{(0)}_{2,s}(\mathbf{p}_0) = \tau^{(0)}_{ss}(\mathbf{p}_0)\,\hat{\mathbf{e}}_s(\mathbf{p}_0) \tag{10.8c}$$

$$\mathbf{E}^{(0)}_{2,p}(\mathbf{p}_0) = \tau^{(0)}_{pp}(\mathbf{p}_0)\,\hat{\mathbf{e}}^-_{p,2}(\mathbf{p}_0)\,. \tag{10.8d}$$

The factorization suggested by Eq. (10.5) is very similar to that given by Eq. (10.4) for scalar waves. We have already discussed the factor $(\alpha_1(\mathbf{p})-\alpha_2(\mathbf{p}))\,\hat{\zeta}(\mathbf{p}-\mathbf{p}_0)$ for scalar waves and it keeps its interpretation for electromagnetic waves. The difference between electromagnetic and scalar waves reside in the last factor, $\hat{\boldsymbol{\rho}}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_0)$ and $\hat{\boldsymbol{\tau}}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_0)$ respectively for the reflection and transmission amplitude. In the case of scalar waves, the last factor encoded the state of the scalar zero order field. The factors $\hat{\boldsymbol{\rho}}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_0)$ and $\hat{\boldsymbol{\tau}}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_0)$ also encodes the state of the electromagnetic zero order field. Indeed this is readily seen from Eq. (10.8) where $\mathbf{E}^{(0)}_{j,\beta}(\mathbf{p}_0)$ is the sum of the β-polarized unit incident (if $j=1$) field amplitude and the corresponding zero order scattered field. We will hence call $\mathbf{E}^{(0)}_{j,\beta}(\mathbf{p}_0)$ *the total zero order field amplitude*. What is even more interesting now, compared to the case of scalar waves, is that the amplitudes $\hat{\boldsymbol{\rho}}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_0)$ and $\hat{\boldsymbol{\tau}}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_0)$ also depend on $\mathbf{p}$ and not only on $\mathbf{p}_0$. By a close inspection of the expressions given in Eqs. (10.6) and (10.7), we note that all these amplitudes can be written as

$$\hat{x}^{(1)}_{\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_0) = \frac{\hat{\mathbf{e}}^{a_x}_{\alpha,\bar{j}_x}(\mathbf{p})\cdot\mathbf{E}^{(0)}_{j_x,\beta}(\mathbf{p}_0)}{\hat{\mathbf{e}}^{a_x}_{\alpha,\bar{j}_x}(\mathbf{p})\cdot\hat{\mathbf{e}}^{a_x}_{\alpha,j_x}(\mathbf{p})} \tag{10.9}$$

with $x=\rho$ or τ, $j_x=1$ or 2, and $a_x=\pm$ respectively for $x=\rho$ and $x=\tau$, and $\bar{j}_x=1$ if $j_x=2$ and $\bar{j}_x=2$ if $j_x=1$. We also adopt the convention that $\hat{\mathbf{e}}^a_{s,j}(\mathbf{p})=\hat{\mathbf{e}}_s(\mathbf{p})$ independently of j and a. Stated in a sentence, the amplitude is given by the *the total zero order field amplitude projected on the Snell-conjugate polarization vector of the measured polarization normalized by the similar projection as if the total zero order field were replaced by the measured polarization vector*. The normalization can be puzzling at first, so let us start by interpreting the numerator in some particular cases. In the case of $s\to s$ reflection we have that $\hat{\rho}^{(1)}_{ss}(\mathbf{p}\,|\,\mathbf{p}_0)$ is proportional to $\hat{\mathbf{e}}_s(\mathbf{p})\cdot\mathbf{E}^{(0)}_{1,s}(\mathbf{p}_0)$. This seems quite intuitive, since it means that the probability of scattering to a s-polarized wave with wave vector $\mathbf{p}$ is proportional to the projection of the zero order field along the outgoing polarization state vector. Consider now the first order reflection amplitude $\hat{\rho}^{(1)}_{pp}(\mathbf{p}\,|\,\mathbf{p}_0)$. It is proportional to $\hat{\mathbf{e}}^+_{p,2}(\mathbf{p})\cdot\mathbf{E}^{(0)}_{1,p}(\mathbf{p}_0)$, which means that the probability for scattering to a p-polarized wave with wave vector $\mathbf{p}$ is proportional to the projection of the zero order field along the polarization vector of the *Snell-conjugate wave* ($\hat{\mathbf{e}}^+_{p,2}(\mathbf{p})$) associated with the outgoing polarization vector ($\hat{\mathbf{e}}^+_{p,1}(\mathbf{p})$). This is slightly less intuitive than the $s\to s$ case. What this means is that, in virtue of Ewald's interpretation discussed for the Fresnel amplitudes, it is the Snell-conjugate wave that controls the polarization coupling. This fact is the analogue of what we found for the reflection of a p-polarized wave at a planar interface,

Eq. (10.2b). For s-polarized waves, this interpretation still holds since Snell-conjugate pairs always have the same s polarization vector.

Let us attempt to summarize the overall interpretation of the first order amplitudes for electromagnetic waves as we did for scalar waves in a few sentences. *The first order reflection and transmission amplitudes for polarized waves are the product of a common probability amplitude for a single change of in-plane momentum from* $\mathbf{p}_0$ *to* $\mathbf{p}$*, which is of a scalar nature, with a probability amplitude for a polarization coupling between the outgoing wave and the zero order state. The latter probability amplitude for polarization coupling, i.e. to scatter to a* α*-polarized wave with in-plane wave vector* $\mathbf{p}$ *from an incident* β*-polarized wave with in-plane wave vector* $\mathbf{p}_0$*, is proportional to the projection coefficient of the total zero order field* $\mathbf{E}^{(0)}_{j_x,\beta}(\mathbf{p}_0)$ *along the polarization vector of the Snell-conjugate wave associated with the measured wave.*

The following subsections aim at analyzing the consequences of the aforementioned amplitudes. We explain the physics of the Yoneda ring phenomenon which is of a scalar nature and is easily explained in terms of coupled modes. We will also analyze in detail the Brewster scattering phenomenon which is a generalization of the Brewster phenomenon for a planar interface, by giving a simple geometrical interpretation and a clear intuitive picture in terms of oscillating dipole radiation. Finally, we will predict a new effect, that of a circularly polarized Brewster scattering which occurs when the zero order wave undergoes total internal reflection. The physical mechanism will be clearly identified as linked to the radiation of rotating dipoles. The discussion on the radiation of an oscillating and of a rotating dipole in free space given in Section 1.3 will be particularly helpful.

10.1.2 Yoneda ring

The first phenomenon to be studied in terms of the first order reflection and transmission amplitudes is the so-called Yoneda ring phenomenon. A Yoneda ring is a ring of enhanced intensity for the diffusely scattered light which is observed when the light is observed in the denser medium. One also find the terms Yoneda peaks when one consider a cut of the intensity of the diffusely scattered light in the plane of incidence. The peaks or ring, are named after Y. Yoneda who first observed an "anomalous surface reflection" (abbreviated A.S.R) in the scattering of X-rays on different types of samples with refractive index less than unity[1] [90]. The experimental investigation carried out by Yoneda was remarkably complete on many aspects. Despite the beliefs of the time that only a specular reflection of X-rays could be observed, Yoneda managed, with an improved setup compared to those used at the time, to demonstrate that an off-specular peak is also present at large scattering angles (i.e. nearly grazing). Yoneda then studied thoroughly, (i) the wavelength of the anomalous peak signal, (ii) the angular position of the peak and its relative intensity as the angle of incidence is varied, (iii) the position of the anomalous peak as the wavelength is varied, and (iv) this for different materials. In addition, (v) the influence of surface roughness on the intensity of the specular and anomalous peak was studied. This extensive study, which is summarized in a paper of barely five pages, shows the will of the author to understand the origin of the phenomenon as well summarized by the sentence "*the ultimate solution of this problem involves the intrinsic mechanisms of A.S.R., which are not known at present.*", as written by Yoneda [90]. The experimental study, in fact, gave all the characteristic features of the anomalous peak and all the keys to understand the underlying mechanism. Let us summarize some of the key points observed by Yoneda:

[1] This is of importance here since it means that the medium of reflection is the denser one.

i The wavelength of the anomalous peak is the same as the wavelength of the incident beam.

ii As the angle of incidence is varied, the angle of reflection of the specular peak varies accordingly (Snell's law) but the angular position of the anomalous peak is unchanged. Only the relative intensity is varied. The closer the specular peak is to the anomalous peak the larger the intensity of the anomalous peak.

iii The angular position of the anomalous peak depends on the wavelength.

iv The angular position of the anomalous peak depends on the material.

v As the surface is deteriorated by etching, the intensity of the specular intensity decreases while the intensity of the anomalous peak increases.

Yoneda also identified that there may be a link between the position of the anomalous peak and the type of atoms constituting the material, and made a remark on the potential role of valence electrons. Let us comment each of the result obtained by Yoneda in a somewhat detective fashion and in light of the theory developed in this thesis. The first observation made by Yoneda is that the wavelength of the anomalous peak is the same as the wavelength of the incident beam (i). This strongly suggests that the effect is linear, and can be understood from a linear theory of the response of the material (which is what we are using in this thesis). On the angular position of the anomalous peak, we know that it is independent of the angle of incidence and characterized by the wavelength and material (ii-iv). Here we would like to note that it may not be obvious from Yoneda's experiment that the wavelength dependence is decoupled from the material dependence. Indeed, the bulk response of a material in principle depends on the wavelength. Hence, we can assume that the position depends on the dielectric function (which is frequency dependent) but not necessarily on the wavelength itself. Finally, and of crucial importance, the intensity of the anomalous peak is increased as the surface becomes rougher (v) which indicates that the underlying mechanism can be looked for in terms of scattering by the surface.

The Yoneda anomalous surface reflection have been investigated in subsequent works for X-rays scattering from metallic [90, 91, 92, 93, 94, 95] and non-metallic [96, 97, 98, 99] surfaces. The scattering of neutrons from a rough surface was also investigated theoretically by Sinha *et al* [92]. The possibility of the existence of the Yoneda phenomenon at optical frequency was first described in a paper by Nieto-Vesperinas and Sánchez-Gil [100], who observed numerically the presence of "sidelobes" in the angular intensity distributions of light scattered by a one-dimensional dielectric surface. Such side lobes were also shown to be present in the numerical investigation of light scattering from two-dimensional dielectric surfaces carried out by Kawanishi *et al.* [101] and was qualified as "quasi-anomalous scattering peaks ". In particular, Kawanishi *et al.* showed that the Yoneda peaks could be observed both in reflection and transmission depending on whether the light was incoming from the optically denser or less dense medium. They also observed what they called, the Brewster scattering phenomenon, which is a zero, or near zero of intensity in the diffusely scattered light for p-polarization. The explanation suggested by the authors was that the effects could be related to the excitation of resonant surface waves, hence implicitly requiring a multiple scattering picture. **Paper** [1] gives, to the best of our knowledge, the first experimental demonstration of the existence of the Yoneda ring at optical frequency, corroborated with numerical simulations based on the reduced Rayleigh equations and small amplitude perturbation theory expanded to first order in the surface profile [1, 45, 46]. The fact that a single scattering theory, such as SAPT to first order,

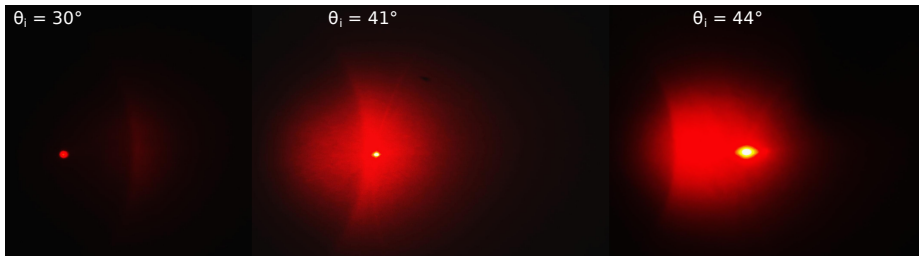


Figure 10.1: Snapshots of a red laser beam ($\lambda = 632.8$ nm) reflected diffusely in glass from a rough interface and projected onto a screen for different angles of incidence $\theta_0 \in \{30°, 41°, 44°\}$. The specular direction is given by the bright spot and the Yoneda ring is well visible at an angle of reflection around 41°. The experimental set-up is similar to the one used in **Paper** [1] but for a sample with estimated rms-roughness $\sigma \approx 20$ nm and correlation length $a \approx 7$ μm. Courtesy of A. K. González-Alcalde and E. R. Méndez.

catches the Yoneda phenomenon consequently questioned the speculated explanation based on the excitation of resonant surface waves. **Paper** [2] aims at giving a clear physical explanation of the Yoneda phenomenon within a single scattering picture, and shows the intrinsically scalar wave essence of the mechanism (i.e. that polarization is not required to explain it). We expose this theory in the following.

First, as an illustration of the Yoneda phenomenon in optics, Fig. 10.1 shows snapshots of a scattering experiment generously provided by A. K. González-Alcalde and E. R. Méndez. The set-up is similar to the one used in **Paper** [1]. A thin rough dielectric film is deposited on a glass sample to which is attached a hemispherical glass cap. The laser beam is incident on the cap, normally to the surface of the hemisphere, but making an angle of incidence with the rough interface. The cap is necessary to keep the angular distribution of the reflected light from the rough surface as it is inside the glass when it is finally exiting the cap. Without the cap, the light scattered above the critical angle for total internal reflection inside the glass cannot escape the sample and the Yoneda phenomena is not observable. We can clearly observe in Fig. 10.1 that the scattered light projected onto a screen exhibits a fixed ring, defined by a constant polar angle of scattering independent of the angle of incidence and above which the intensity of the diffusely scattered light is enhanced. This is the Yoneda phenomenon. Let us now turn to some theoretical curves. When the diffusely scattered intensity is measured as a function of the angle of scattering in the plane of incidence the Yoneda effect is identified as two peaks symmetrically positioned from the normal to the average plane (i.e. it is a cut of the two-dimensional ring in the full angular distribution of the MDRC or MDTC). A characteristic property of the Yoneda ring, is that the polar angle of scattering at which it is observed is independent of the angle of incidence, but only depends on the dielectric constants of the media. The top panels of the subfigures in Figs. 10.2–10.4 show the diffuse component of the MDRC and MDTC as a function of the in-plane wave vector and as a function of the angle of scattering (double axis) for different angles of incidence and in both cases where the incident wave is impinging from the vacuum side or the glass side. Focusing our attention on s-co-polarized scattering, we can observe that when the incident wave impinges from the vacuum side, sharp peaks are observed in the MDTC at $p_1 = \pm\omega/c$ for normal incidence in Fig. 10.2(d), and that these peaks remain at the same positions as the angle of incidence is increased, although the overall weight of the distribution

is shifted towards higher angles of scattering (see Figs. 10.2(e) and 10.2(f)). A similar behavior is observed for the reflected light in the case of reflection in the denser medium as can be readily seen in Figs. 10.3(a-c) and 10.4(a-c). A point worth noticing from Figs. 10.2–10.4 is that the positions of the peaks for the light scattered in the dense medium always corresponds to $|p_1| = n_{\min}\,\omega/c$ ($n_{\min} = 1$ in our examples), i.e. the value of the in-plane wave vector above which a wave scattered with $|\,\mathbf{p}\,| > n_{\min}\,\omega/c$ becomes evanescent in the less dense medium. This can be readily seen from the fact that the MDRC or MDTC becomes identically zero beyond this threshold. Turning now to p-polarized light, the same observation can be made, up to one small difference. At normal incidence, a zero of intensity is observed instead of a peak at $p_1 = \pm n_{\min}\omega/c$. As the angle of incidence is progressively increased, the zero of intensity shifts position and the Yoneda peaks can also be observed for p-polarized light. This zero of intensity for p-polarized light is known as the Brewster scattering effect and will be studied in details in Section 10.1.3. This effect is characteristic of polarized waves while the Yoneda effect is fundamentally of a scalar nature, and the effects are decoupled from each other. The fact that the Brewster scattering angle coincides with the critical angle of the Yoneda phenomenon for normal incidence can be taken as a simple coincidence for the time being and will be clarified later.

Let us analyze in more detail the behavior of the first order reflection and transmission amplitudes as the in-plane wave vector of scattering $\mathbf{p}$ is progressively increased. As an illustration of the discussion, the reader may find helpful to follow at the same time the different subfigures in Fig. 10.5 which illustrate the different configurations for the scattered mode as $\mathbf{p}$ is increased (Figs. 10.5(a-c)) together with the behavior of the factor $(\alpha_1 - \alpha_2)c/\omega$ in the complex plane (Fig. 10.5(d)). The scalar reflection and transmission amplitudes are given by Eq. (10.4) and we have already mentioned that the whole $\mathbf{p}$-dependence is encoded in the linear approximation of the phase factor $(\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p}))\,\hat{\zeta}(\mathbf{p} - \mathbf{p}_0)$. Since the intensity ring is observed on average over realizations of the surface profile, it is more precisely the factor $(\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p}))$ which is of interest. Hence we can assume for the sake of simplicity that $|\hat{\zeta}|$ varies slowly. First for $|\,\mathbf{p}\,| < p_c$, both the elementary reflected and transmitted waves propagate away from the surface (see Fig. 10.5(a)), and we have that both $\alpha_1(\mathbf{p})$ and $\alpha_2(\mathbf{p})$ are real. As $|\,\mathbf{p}\,| \to p_c$ from below, $\alpha_{\min}(\mathbf{p}) \to 0$ and $\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p}) \to \pm\alpha_c$, with $\alpha_c = \sqrt{\epsilon_{\max} - \epsilon_{\min}}\,\omega/c$ being a critical out-of-plane component of the wave vector of scattering in the medium of propagation for a critical mode characterized by $|\,\mathbf{p}\,| = p_c$ (see the segments on the real line in Fig. 10.5(d) corresponding to $0 < |\,\mathbf{p}\,| < p_c$). By writing $p = |\,\mathbf{p}\,| = p_c - \Delta p$, with $\Delta p > 0$ being a small variation of in-plane wave vector, we can make an asymptotic analysis of $|\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p})|$ as $p \to p_c$ from below ($\Delta p \to 0$). In this way, we obtain

$$
\begin{aligned}
&|\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p})|\frac{c}{\omega} = [\alpha_{\max}(\mathbf{p}) - \alpha_{\min}(\mathbf{p})]\frac{c}{\omega} \\
&= [\epsilon_{\max} - (\tilde{p}_c - \Delta\tilde{p})^2]^{1/2} - [\epsilon_{\min} - (\tilde{p}_c - \Delta\tilde{p})^2]^{1/2} \\
&= [\epsilon_{\max} - \epsilon_{\min} + 2\epsilon_{\min}^{1/2}\Delta\tilde{p} - \Delta\tilde{p}^2]^{1/2} - [2\epsilon_{\min}^{1/2}\Delta\tilde{p} - \Delta\tilde{p}^2]^{1/2} \\
&= \alpha_c\,\frac{c}{\omega} - [2\epsilon_{\min}^{1/2}\Delta\tilde{p}]^{1/2} + o(\Delta\tilde{p}^{1/2})\,.
\end{aligned}
\tag{10.10}
$$

We have chosen here to work with unit-less quantities and denoted $\tilde{p} = p\,c/\omega$ for conciseness. It follows from Eq. (10.10) that, as $\Delta p \to 0$, $|\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p})|$ increases towards α_c in an inner-neighborhood of the circle $p = p_c$. Furthermore, the asymptotic expansion reveals that the critical point will be reached with a sharp edge (infinite slope) for $p < p_c$ as can be deduced from the square root behavior in Δp. This is indeed what can be observed in the

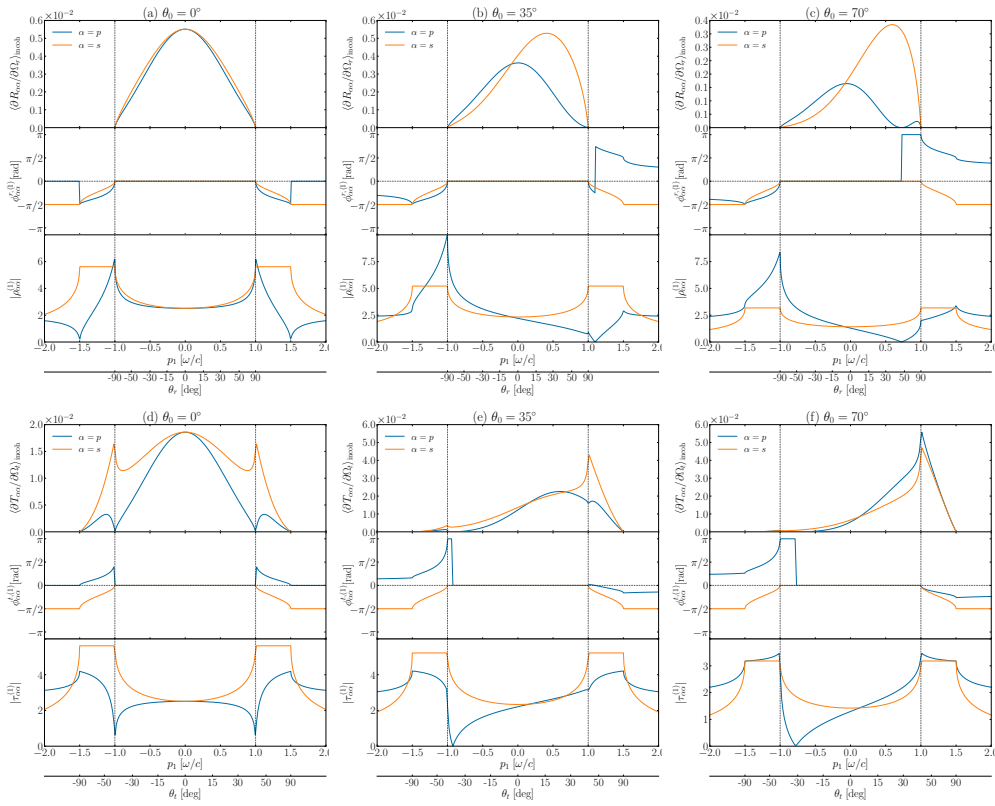


Figure 10.2: The diffuse component of the MDRC (top row) and MDTC (bottom row) for light incident from vacuum ($\epsilon_1 = 1.0$) onto a randomly rough interface with glass ($\epsilon = 2.25$) as a function of the in-plane wave vector and the angle of scattering for co-polarized scattering in the plane of incidence (top panel). The polar angle of incidence is indicated on top of each subfigure. The corresponding average phase of the reflection and transmission amplitudes is indicated in the middle panel, while the modulus of the amplitudes is given in the bottom panel. Note that evanescent modes are also represented. We adopt the convention $\theta_{r,t} > 0$ for an azimuthal angle of scattering $\phi = 0°$ (forward scattering) and $\theta_{r,t} < 0$ for an azimuthal angle of scattering $\phi = 180°$ (back scattering). Figure taken from **Paper** [2].

panels corresponding to $|\rho^{(1)}_{\alpha\alpha}|$ and $|\tau^{(1)}_{\alpha\alpha}|$ in Figs. 10.2-10.4. Note that this behavior is the same both for the reflection and transmission amplitudes independently of which medium is denser. This illustrates the fact that the two waves are part of the same mode. However, as the wave propagating in the less dense medium becomes a grazing wave, the corresponding differential scattering coefficient is forced to vanish due to the angular dependence in $\cos^2\theta_s$ ($\theta_s = \theta_r$ or θ_t depending on the context). The modulus of the complex amplitude is nevertheless enhanced for both the reflected and transmitted wave. This is illustrated for example in Figs. 10.2(a) and 10.2(d), which corresponds to a case for which the medium of incidence is vacuum. From the results presented in these figures, we can see that while the incoherent

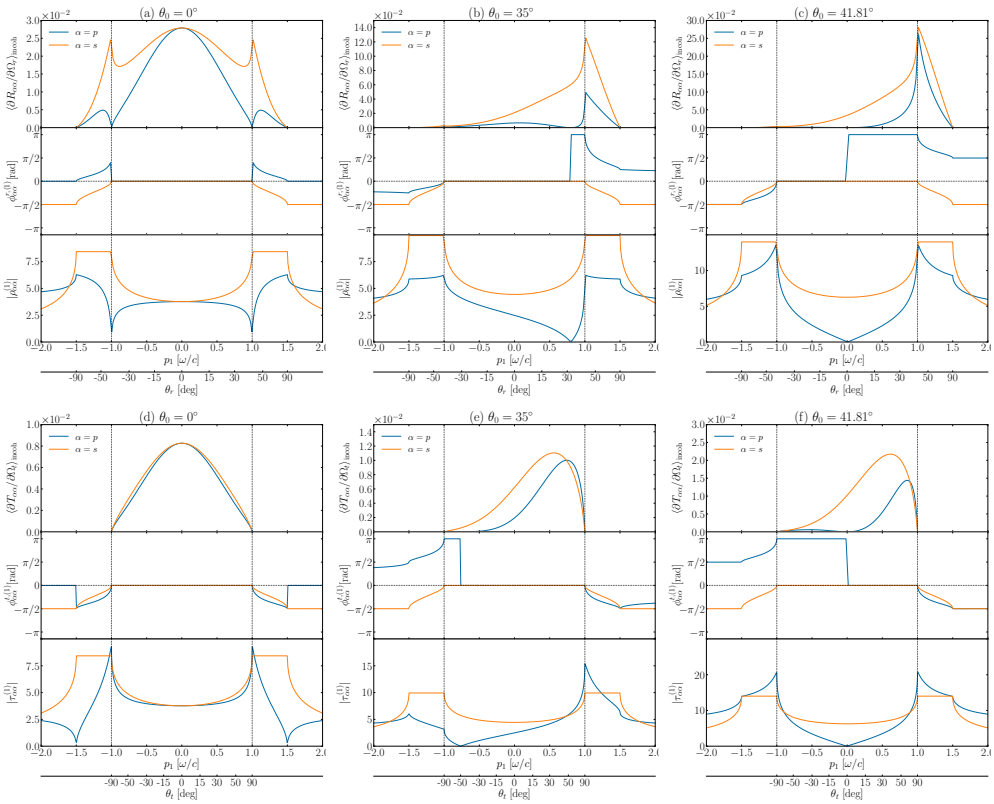


Figure 10.3: Same as Fig. 10.2, but for light incident from glass ($\epsilon_1 = 2.25$) onto a randomly rough interface with vacuum ($\epsilon = 1.0$). Figure taken from **Paper** [2].

component of the MDRC is forced to go to zero when $p_1 \to p_c = \omega/c$, the surface-independent factor of the reflection amplitude $\rho_{ss}^{(1)}$ exhibits a sharp increase in modulus. Simultaneously, the surface-independent factor of the transmission amplitude $\tau_{ss}^{(1)}$ also exhibits a similar sharp increase in modulus as p_1 approaches p_c. Consequently, since the wave can propagate away from the surface in the second medium (which consists of glass in this specific case), the corresponding incoherent component of the MDTC exhibits a similar increase. A physically intuitive picture for the enhancement of intensity in the denser medium is the following: We can imagine that there is a certain amount of power allocated to the mode characterized by the in-plane wave vector $\mathbf{p}$. When both the reflected and transmitted wave may propagate to infinity, they can share the total power allocated to the mode and transport their share of the power away from the surface. As the wave component of this mode in the less dense medium becomes evanescent, it is the role of the wave that can propagate in the denser medium to radiate away the whole power allocated to the mode. Note that both the phases associated with $\rho_{ss}^{(1)}$ and $\tau_{ss}^{(1)}$ remain constant and equal to 0 for $p_1 < p_c$ for all θ_0 in Fig. 10.2, since the complex amplitude stays on the real line in the case where $\epsilon_1 < \epsilon_2$ independent of the angle of incidence. Figures 10.3 and 10.4 support the same conclusion but by interchanging the role of the media. The only difference worth noting is that the phases $\phi_{ss}^{r,(1)}$ and $\phi_{ss}^{t,(1)}$

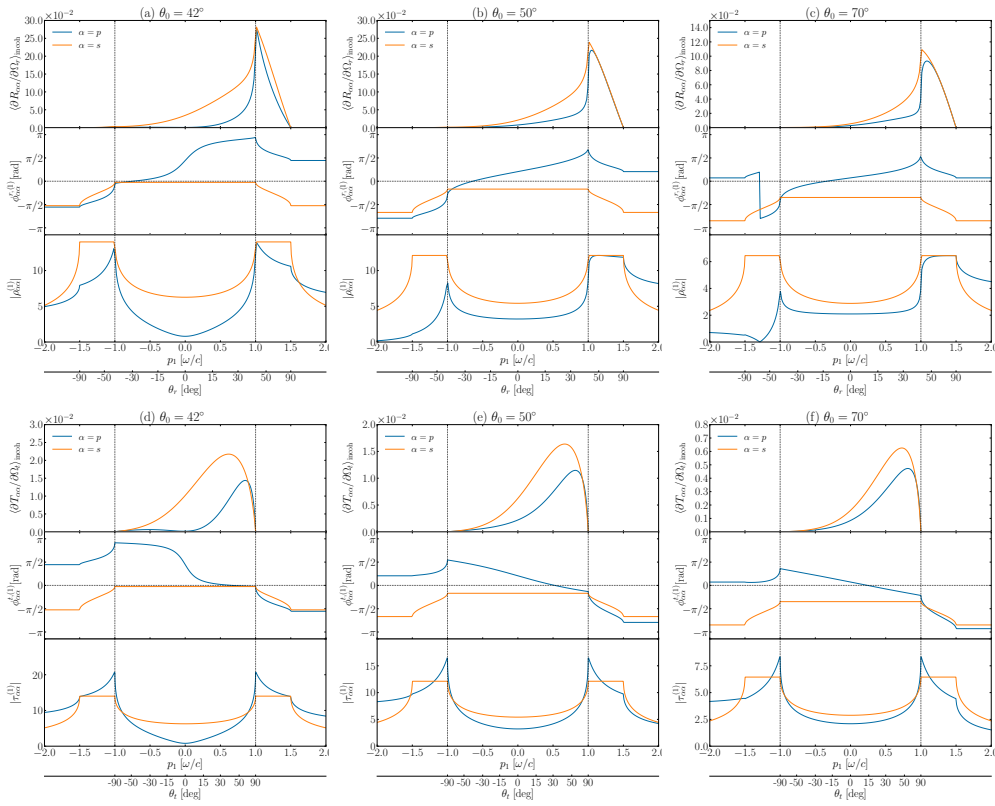
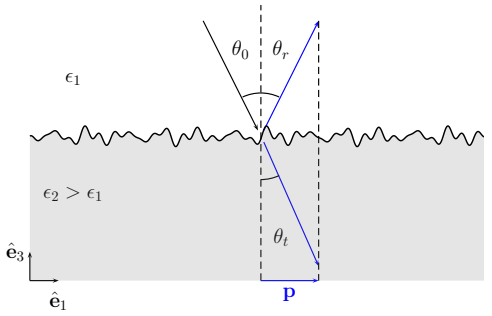


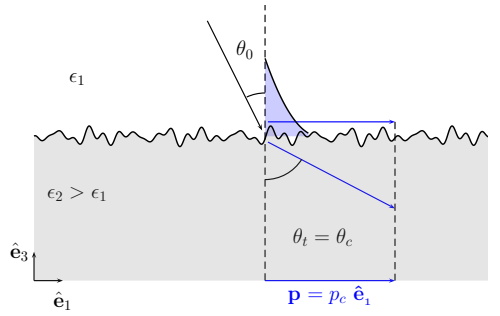
Figure 10.4: Same as Fig. 10.3, but for additional polar angles of incidence θ_0. Figure taken from **Paper** [2].

have a constant plateau for $p_1 < p_c$ which is equal to 0 only for $\theta_0 < \theta_c$. The plateau is offset for $\theta_0 > \theta_c$. This overall phase offset is due to the Goos-Hänchen phase shift associated with total internal reflection of the zero order wave (see the corresponding discussion in the previous section). Indeed, recall that the first order amplitudes are proportional to the total zero order field amplitudes (see Eq. 10.4). Therefore, if the zero order waves exhibit a phase shift, it will affect the first order amplitudes in the form of a constant phase offset for all $\mathbf{p}$.

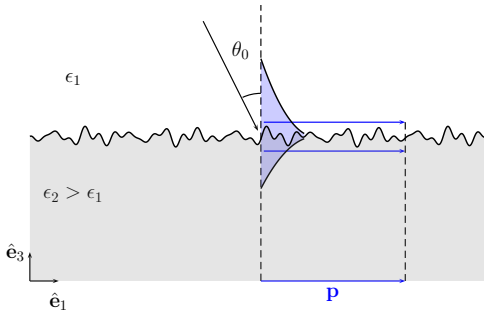
When $|\,\mathbf{p}\,| > p_c$, $\alpha_{\min}$ becomes pure imaginary and $\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p})$ thus moves off the real line. For $p_c < |\,\mathbf{p}\,| < n_{\max}\omega/c$, the wave in the less dense medium is evanescent and the wave in the denser medium is progressive (see Fig. 10.5(b)). In this regime, $\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p})$ keeps a constant



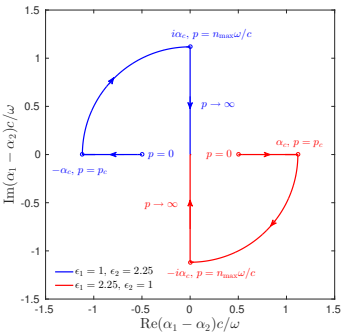
(a) $p < p_c$



(b) $p = p_c = n_{\min}\,\omega/c$



(c) $p > n_{\max}\,\omega/c$



(d)

Figure 10.5: Illustration of the nature of the scattered couple mode in blue as the in-plane wave vector of scattering $\mathbf{p} = p\,\hat{\mathbf{e}}_1$ is progressively increased and exhibit the Yoneda peak transition. In the figure, $\epsilon_1 < \epsilon_2$ is assumed. (a) In-plane wave vector of scattering below the Yoneda threshold, $p < p_c$, the reflected and transmitted components of the coupled mode propagate away from the surface. (b) In-plane wave vector of scattering equal to the Yoneda threshold, $p = p_c$, the reflected component of the coupled mode becomes evanescent and the transmitted component propagates away from the surface. (c) In-plane wave vector of scattering above the propagation threshold in both media $p > n_{\max}\,\omega/c$, both the reflected and transmitted components of the coupled mode are evanescent. (d) Path taken by $(\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p}))c/\omega$ as p goes from 0 to ∞.

modulus equal to α_c. Indeed, by writing $\alpha_{\min}(\mathbf{p}) = i\beta_{\min}(\mathbf{p})$ we have

$$\begin{aligned} |\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p})| &= |\alpha_{\max}(\mathbf{p}) - i\beta_{\min}(\mathbf{p})| \\ &= \left[\alpha_{\max}^2(\mathbf{p}) + \beta_{\min}^2(\mathbf{p})\right]^{1/2} \\ &= \left[\epsilon_{\max} - |\,\tilde{\mathbf{p}}\,|^2 + |\,\tilde{\mathbf{p}}\,|^2 - \epsilon_{\min}\right]^{1/2} \frac{\omega}{c} \\ &= \alpha_c\,. \end{aligned} \tag{10.11}$$

The complex number $\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p})$ thus traces a circular arc of radius α_c in the complex plane (see corresponding circular arcs in Fig. 10.5(d)). Finally, when $|\,\mathbf{p}\,| > n_{\max}\,\omega/c$, both the reflected and transmitted waves are evanescent, $\alpha_{\max}$ becomes pure imaginary and hence

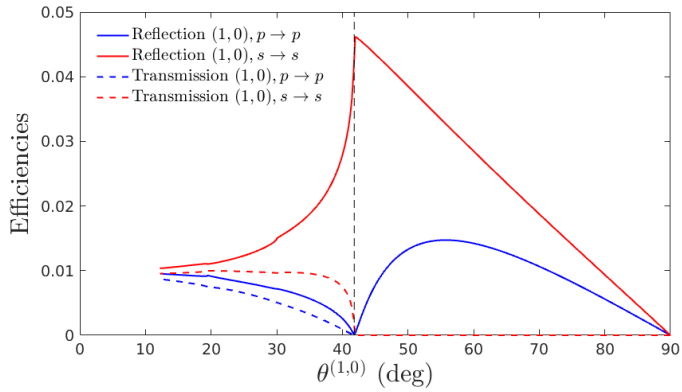


Figure 10.6: Diffraction efficiency of the reflected and transmitted $(1,0)$ order for p- and s-polarized waves for a set of one-dimensional sinusoidal gratings with small amplitude and with varying lattice constant ($\epsilon_1 = 2.25$, $\epsilon_2 = 1$). The efficiencies are plotted as a function of the angle of reflection of the $\theta^{(1,0)}$ of the first diffracted order which is parametrized by the lattice constant.

$\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p})$ moves along the imaginary axis (see corresponding segments on the imaginary line in Fig. 10.5(d)). The constant value of $|\rho_{ss}^{(1)}|$ and $|\tau_{ss}^{(1)}|$ in the regime $n_{\min}\,\omega/c < \mathbf{p} < n_{\max}\,\omega/c$ can be appreciated for all angles of incidence illustrated in Figs. 10.2 – 10.4, while the phases exhibit a smooth variation from their plateau value and decay by a total amount of $-\pi/2$ when reaching $p_1 = n_{\max}\,\omega/c$. Once the threshold of $n_{\max}\,\omega/c$ has been passed, the phases remain constant and the modulii decay towards zero which can easily be deduced from the fact that $\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p}) \sim i(\epsilon_{\max} - \epsilon_{\min})\omega^2/(2\,c^2 p)$ as $p \to \infty$ (see Fig. 10.5(d)). The phase change associated with the transition from the real line to the imaginary line in the complex plane is therefore $-\pi/2$. This gradual phase change is similar to that of the Goos-Hänchen phase shift discussed for the reflection by a planar surface. The difference of absolute total phase change, of π for the case of the Fresnel amplitude and $\pi/2$ in the case of the scattered waves, comes mathematically from the fact that in the former case the amplitude is written as the ratio of a complex number and its complex conjugate, while in the latter case there is no such ratio. The phase consequently turns twice as fast in the former case than in the latter. A physical interpretation of this difference is that for the Fresnel amplitude both the incident and outgoing wave vector must vary simultaneously (since they are the same), while in the case of a scattered wave the incident wave vector is fixed while only the outgoing wave vector is allowed to vary. In fact, we have only analyzed the phase associated with the factor $\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p})$ in Eq. (10.4). The phase of the overall complex amplitude will be the sum of the aforementioned phase, that given by the argument of $\hat{\zeta}(\mathbf{p} - \mathbf{p}_0)$, and the phase given by the argument of the total zero order amplitude $[1 + \rho^{(0)}(\mathbf{p}_0)]$ or $\tau^{(0)}(\mathbf{p}_0)$. In particular, if the angle of incidence is such that total internal reflection occurs for the zero order field, the overall phase of the scattered amplitude will contain a signature of the Goos-Hänchen phase shift associated with the total internal reflection of the zero order field in addition to the corresponding Goos-Hänchen phase shift associated with the Yoneda effect. Note that when averaged over surface realizations, the phase contribution coming from $\hat{\zeta}$ averages to zero. This supports our choice of limiting the detailed investigation to the surface-independent factors in Eq. (10.4).

Remark 10.2. The Yoneda phenomenon for weakly rough surfaces originates from the same physical mechanism as the Rayleigh anamolies for periodic dielectric gratings. The continuous set of scattered wave vectors in the case of a randomly rough surface can be viewed as probing a diffracted order scattered from a periodic surface with continuously changing lattice constant. It is easy to show numerically and with SAPT to first order, that the behavior of the efficiency of a given diffractive order as the lattice constant is changed exhibits the same characteristic peak as the Yoneda peak when its counter part in the less dense medium becomes evanescent. This is well illustrated in Fig. 10.6 in which the diffraction efficiencies of the first reflected and transmitted orders are plotted as functions of the angle of reflection as the lattice constant is varied. The perturbative analysis in the case of a periodic grating is exactly the same as in the case of a randomly rough surface with the only difference being that $\mathbf{p}$ must be replaced by the in-plane wave vector of the diffractive order of interest and make the lattice constant vary instead.

10.1.3 Brewster scattering

We can see in Figs. 10.2 – 10.4 that for co-polarized scattering in the plane of incidence, the diffusely scattered p-polarized light may exhibit a zero of intensity in reflection or transmission (see the panels corresponding to the MDRC and MDTC in Figs. 10.2 – 10.4) for a wide range of angles of incidence while s-polarized light has a strictly positive intensity for all angles of scattering. The phenomenon of vanishing of the intensity for p-polarized light is known as *Brewster scattering phenomenon* by analogy with the Brewster angle for a planar interface. There are three points already worth mentioning here based on Figs. 10.2 – 10.4. First, the Brewster scattering angle depends on the angle of incidence. Second, the Brewster scattering phenomenon may occur both for reflection and transmission. In addition, for a given angle of incidence, we generally observe that if a Brewster scattering occurs in reflection in the forward scattering direction, then a Brewster scattering occurs in transmission in the backward scattering direction (see e.g. Figs. 10.2(c) and 10.2(f)). Third, by considering the reflection and transmission amplitudes instead of solely considering the intensity, one finds that a zero of amplitude may be found for *evanescently* scattered waves (see e.g. $|\rho_{pp}^{(1)}|$ in Figs. 10.2(a-c)). This suggests to generalize the concept of Brewster scattering to evanescent waves. Furthermore, we will see later that, Brewster scattering may also be generalized to other polarizations when the full angular distribution is considered. The present section aims at explaining all these features based on the physical interpretation of SAPT to first order in the surface profile given in Section 10.1.1 and to build a physical intuition of the observations based on elementary concepts of oscillating and rotating dipole radiations presented in Section 1.3, which are the mechanisms at play here.

Scattering in the plane of incidence

We first analyze the case of co-polarized scattering in the plane of incidence.

s-polarized waves – Why does $s \to s$ scattering yield a non-vanishing intensity? We have seen in Section 10.1.1 that, to first order in the surface profile, the scattering amplitude is proportional to the projection of the total zero order field amplitude projected on the Snell-conjugate polarization vector of the measured polarization. In the case of $s \to s$ scattering in the plane of incidence, the reflection amplitude $\hat{\rho}_{ss}^{(1)}(\mathbf{p} \mid \mathbf{p}_0)$ is proportional to $\hat{\mathbf{e}}_s(\mathbf{p}) \cdot \mathbf{E}_{s,1}^{(0)}(\mathbf{p}_0)$ (see Eq. (10.6a)). Since the scattering is in the plane of incidence, we have $\hat{\mathbf{e}}_s(\mathbf{p}) = \hat{\mathbf{e}}_s(\mathbf{p}_0)$ and therefore the first order reflection amplitude reduces to that of the scalar wave Eq. (10.4a).

Consequently, there is no extinction for $s \to s$ scattering in the plane of incidence for any angle of incidence. The same analysis and conclusion hold for the transmitted s-polarized wave.

p-polarized waves – For p-polarized waves, the first order reflection amplitude is proportional to $\hat{\mathbf{e}}^+_{p,2}(\mathbf{p}) \cdot \mathbf{E}^{(0)}_{1,p}(\mathbf{p}_0)$, where we recall that $\mathbf{E}^{(0)}_{1,p}(\mathbf{p}_0)$ is the total zero order field amplitude for an incident p-polarized wave (see Eq. (10.6d)). This is very similar to what was found for the Fresnel amplitude for p-polarized light in Eq. (10.2b). Indeed, $\hat{\mathbf{e}}^+_{p,2}(\mathbf{p})$ plays now the role of $\hat{\mathbf{e}}^+_{p,2}(\mathbf{p}_0)$ as the Snell-conjugate polarization vector of the measured wave and $\mathbf{E}^{(0)}_{1,p}(\mathbf{p}_0)$ plays the role of $\hat{\mathbf{e}}^-_{p,1}(\mathbf{p}_0)$ as the source. Equation (10.6d) thus gives us a simple geometrical criterion for Brewster scattering within first order perturbation theory:

$$\hat{\mathbf{e}}^+_{p,2}(\mathbf{p}_\mathrm{B}) \cdot \mathbf{E}^{(0)}_{p,1}(\mathbf{p}_0) = 0\,, \tag{10.12}$$

where $\mathbf{p}_\mathrm{B}$ is an in-plane wave vector that satisfies Eq. (10.12) and hence defines a direction $\mathbf{k}^+_1(\mathbf{p}_\mathrm{B})$ along which the reflection amplitude vanishes. Stated in a sentence, the criterion reads as follows.

Criterion of $p \to p$ Brewster scattering in medium 1: *The set of lateral wave vectors $\mathbf{p}_\mathrm{B}$ of the elementary Brewster scattered waves, for which the reflection amplitude for a p-polarized reflected wave vanishes given a p-polarized incident wave with lateral wave vector $\mathbf{p}_0$ are given by the condition of orthogonality between the p-polarization vector of the Snell-conjugate scattered waves and the total zero order field in medium 1.*

Remark 10.3. Note that there may *a priori* be several solutions to Eq. (10.12) as we will soon see. Also, the criterion, Eq. (10.12), for $p \to p$ scattering is not restricted to scattering in the plane of incidence, but is valid for all scattering direction as well as evanescent waves.

A direct consequence of Eq. (10.12) is that for co-polarized scattering in the plane of incidence, the geometrical condition can be re-stated as a requirement on the colinearity between the Snell-conjugate wave vector and the total zero order field, which is exactly the same geometrical criterion found in the case of reflection from a planar interface. A second corollary for scattering in the plane of incidence is a fixed point property of the function that maps the angle of incidence to the Brewster scattering angle $\Theta_\mathrm{B} : \theta_0 \mapsto \Theta_\mathrm{B}(\theta_0)$. We find that the *Brewster scattering angle* is equal to the *Brewster angle for a planar interface* when the angle of incidence is equal to the Brewster angle for a planar interface, $\theta_0 = \theta_\mathrm{B}$, i.e. $\Theta_\mathrm{B}(\theta_\mathrm{B}) = \theta_\mathrm{B}$, where we recall that $\theta_\mathrm{B} = \arctan(n_2/n_1)$. This property is readily understood from the geometrical criterion expressed by Eq. (10.12). At Brewster incidence, $\theta_0 = \theta_\mathrm{B}$, the zero order reflected wave vanishes (by definition of Brewster incidence). Thus the total zero field amplitude is simply the incident field amplitude, $\mathbf{E}^{(0)}_{p,1}(\mathbf{p}_0) = \hat{\mathbf{e}}^-_{p,1}(\mathbf{p}_0)$, and consequently, the Brewster scattering angle is necessarily equal to θ_B since the criterion reduces then to that of the planar interface.

How does the Brewster scattering direction varies with the angle of incidence? We will answer this question by using the criterion Eq. (10.12) and pure geometrical constructions. Consider first the case of a wave incident from the less dense medium, for example in vacuum $\epsilon_1 = 1$ and is reflected after interaction with a glass substrate $\epsilon_2 = 2.25$. Figure 10.7 presents a set of sketches illustrating the geometrical construction leading to the Brewster direction for different angles of incidence θ_0. The dispersion relations ($|\,\mathbf{k}^\pm_j\,| = k_j = n_j\,\omega/c$) giving the set of wave vectors for progressive waves in the two media are represented as dashed circles. The wave vectors of the incident and reflected zero order wave, $\mathbf{k}^-_1(\mathbf{p}_0)$ and $\mathbf{k}^+_1(\mathbf{p}_0)$, are drawn in black respectively

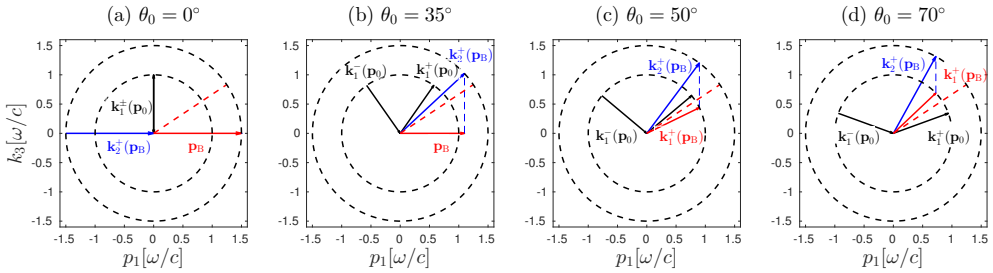


Figure 10.7: Illustration of the geometrical criterion for in-plane Brewster scattering for different polar angles of incidence (a-d). The dashed circles represent the norm of the full wave vectors, given by the dispersion relations, in vacuum ($\epsilon_1 = 1$ inner circle) and glass ($\epsilon_2 = 2.25$ outer circle). The black arrows represent respectively the incident wave vector $\mathbf{k}_1^-(\mathbf{p}_0)$, which is drawn as pointing towards the origin for clarity, and the wave vector of the reflected zero order wave, $\mathbf{k}_1^+(\mathbf{p}_0)$. The blue arrow represents the wave vector of the virtual wave, $\mathbf{k}_2^+(\mathbf{p}_B)$, from which the in-plane wave vector of the Brewster wave, $\mathbf{p}_B$, is deduced by projection along $\hat{\mathbf{e}}_1$. From $\mathbf{p}_B$, the full wave vector for the Brewster wave, $\mathbf{k}_1^+(\mathbf{p}_B)$, can be drawn (provided propagation in medium 1) as a red arrow. Note that if the Brewster wave is evanescent, only $\mathbf{p}_B$ is draw in red as the out-of-plane component of $\mathbf{k}_1^+(\mathbf{p}_B)$ is purely imaginary. The red dashed line indicate the Brewster angle for a planar surface approximately equal to 56.3° in this case. Figure taken from **Paper** [2].

pointing towards the origin and pointing outwards. Finally, with the intention of illustrating the fixed point property of the Brewster angle at $\theta_0 = \theta_B$, the direction corresponding to the wave vector of the zero order non-reflected wave under Brewster incidence is indicated as the red dashed line. The construction steps leading to the Brewster direction in Fig. 10.7 are the following:

1 The wave vectors of the incident and the reflected zero order waves are drawn in black.

2 The direction of the total zero order field given by Eq. (10.8b) is determined (not represented) and the wave vector of the virtual wave colinear to the total zero order field, is drawn as the blue wave vector $\mathbf{k}_2^+(\mathbf{p}_B)$. Note that the direction of the total zero order field may be constructed geometrically as explained in details in Ref. [102], but we skip these steps for simplicity.

3 The wave vector $\mathbf{k}_2^+(\mathbf{p}_B)$ lies on the circle of radius $n_2\omega/c$. Its projection along $\hat{\mathbf{e}}_1$ gives the Brewster lateral wave vector $\mathbf{p}_B$ from which we deduce $\mathbf{k}_1^+(\mathbf{p}_B)$ in red. Or equivalently, we project $\mathbf{k}_2^+(\mathbf{p}_B)$ on the circle of radius $n_1\omega/c$ along $\hat{\mathbf{e}}_3$. Note that the reflected wave associated with $\mathbf{k}_1^+(\mathbf{p}_B)$ may be evanescent, and in that case we simply represent its lateral component $\mathbf{p}_B$ as its component along $\hat{\mathbf{e}}_3$ is pure imaginary.

For normal incidence, the total zero order electric field, and consequently $\mathbf{k}_2^+(\mathbf{p}_B)$, is aligned with the $\hat{\mathbf{e}}_1$ direction as illustrated in Fig. 10.7(a). This particular case yields a degeneracy for the in-plane Brewster wave vector $\mathbf{p}_B$. Indeed, for $\mathbf{k}_2^+(\mathbf{p}_B)$ along $\hat{\mathbf{e}}_1$ there are two solutions to the Brewster criterion in the plane of incidence given by $p_1 = \pm n_2\,\omega/c$. Only the one pointing to the right is represented in Fig. 10.7(a) for clarity. Since $n_1 < n_2$, it follows that $|\,\mathbf{p}_B\,| > n_1\,\omega/c$ and the corresponding Brewster (non-) reflected wave is evanescent. This is of

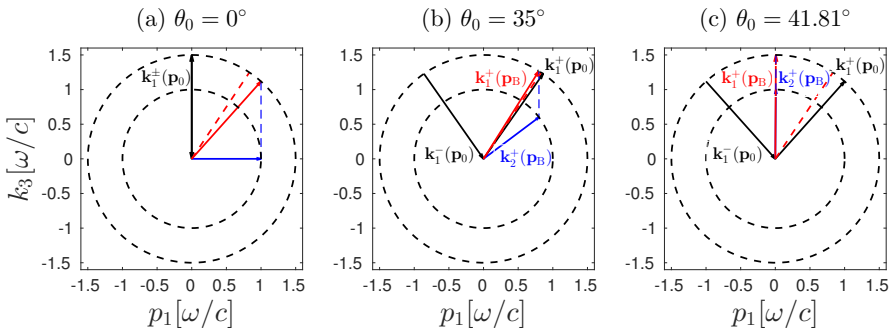


Figure 10.8: Illustration of the geometrical criterion for in-plane Brewster scattering for different polar angles of incidence (a-c). The dashed circles represent the norm of the full wave vectors, given by the dispersion relations, in glass ($\epsilon_1 = 2.25$ inner circle) and vacuum ($\epsilon_2 = 1$ outer circle). The black arrows represent respectively the incident wave vector $\mathbf{k}_1^-(\mathbf{p}_0)$, which is drawn as pointing towards the origin for clarity, and the wave vector of the reflected zero order wave, $\mathbf{k}_1^+(\mathbf{p}_0)$. The blue arrow represents the wave vector of the virtual wave, $\mathbf{k}_2^+(\mathbf{p}_B)$, from which the in-plane wave vector of the Brewster wave, $\mathbf{p}_B$, is deduced by projection along $\hat{\mathbf{e}}_1$. From $\mathbf{p}_B$, the full wave vector for the Brewster wave, $\mathbf{k}_1^+(\mathbf{p}_B)$, can be drawn (provided propagation in medium 1) as a red arrow. Note that if the Brewster wave is evanescent, only $\mathbf{p}_B$ is drawn in red as the out-of-plane component of $\mathbf{k}_1^+(\mathbf{p}_B)$ is purely imaginary. The red dashed line indicates the Brewster angle for a planar surface, approximately equal to 33.7° in this case. Figure taken from **Paper** [2].

course not visible when considering the incoherent component of mean differential reflection coefficient but it can be seen by inspection of the modulus of the amplitude, $\rho_{pp}^{(1)}$, in Fig. 10.2(a). There, we can readily observe that $\rho_{pp}^{(1)}$ vanishes at $p_1 = \pm n_2\,\omega/c$. Correspondingly, the phase $\phi_{pp}^{\mathrm{r},(1)}$ exhibits a jump which is characteristic of the Brewster effect. In this degenerate case, the phase jump is equal to $\pi/2$, while in general (non degenerate case) a π phase jump is observed. The $\pi/2$ phase jump seems to occur only in the degenerate case, which as far as we could see, only occur at normal incidence for this system. As the polar angle of incidence is progressively increased, the direction of the total zero order field varies, and so does the wave vector of the Brewster Snell-conjugate wave which is now found to be unique for scattering in the plane of incidence. Figure 10.7(b) illustrates the case where $\theta_0 = 35°$, where we notice that the projection of $\mathbf{k}_2^+(\mathbf{p}_B)$ along $\hat{\mathbf{e}}_1$ still leads to an evanescent Brewster wave. The lateral wave vector is nevertheless closer to the propagation limit in medium 1. The corresponding reflection amplitude is plotted in Fig. 10.2(b), where we can observe that $\rho_{pp}^{(1)}$ vanishes for p_1 just above $n_1\,\omega/c$, and that the phase exhibits a π jump. By further increasing the polar angle of incidence the Brewster wave is found in the propagating region as $|\,\mathbf{p}_B\,| < n_1\,\omega/c$. Its full wave vector, $\mathbf{k}_1^+(\mathbf{p}_B)$, can now be represented as following the inner dashed circle as illustrated in Fig. 10.7(c) for $\theta_0 = 50°$. When the polar angle of incidence approaches the Brewster angle of incidence, $\theta_0 \to \theta_B$, the wave vector associated with the reflected zero order wave, $\mathbf{k}_1^+(\mathbf{p}_0)$, and the wave vector of the Brewster scattered wave, $\mathbf{k}_1^+(\mathbf{p}_B)$, respectively drawn in black and red in Fig. 10.7(c), approach the red dashed line from either sides. At the Brewster incidence, $\theta_0 = \theta_B$, the two wave vectors are equal, $\mathbf{k}_1^+(\mathbf{p}_B) = \mathbf{k}_1^+(\mathbf{p}_0)$, and are aligned with the red dashed line as explained previously. Figure 10.7(c) shows the case where $\theta_0 = 50°$ at a slightly lower

angle than the Brewster angle of incidence (approximately equal to 56.3°), i.e. just before the cross-over. By increasing further θ_0, the lateral component of the Brewster wave vector continuously decreases. The case for $\theta_0 = 70°$ is illustrated in Fig. 10.7(d). The Brewster wave is well inside the propagating region as can be seen both from $\rho_{pp}^{(1)}$ and the extinction of the incoherent component of the MDRC in Fig. 10.2(c). Also observe the π jump in the phase. Eventually, as the polar angle of incidence approaches 90°, the wave vector $\mathbf{k}_2^+(\mathbf{p}_\mathrm{B})$, does *not* approach the vertical direction as one might naively expect. One must remember that the total zero order field does not become oriented along $\hat{\mathbf{e}}_3$ but along the direction given by the critical angle for total internal reflection.

Consider now the configuration in which the incident wave approaches the surface in the optically denser medium (e.g. $\epsilon_1 = 2.25$ and $\epsilon_2 = 1$). As for the previous configuration, the total zero order field is along $\hat{\mathbf{e}}_1$ for normal incidence. Hence the existence of two Brewster waves in the plane of incidence. However, the norm of the Snell-conjugate wave vectors are now such that $k_2 < k_1$ and the corresponding wave vectors are situated on the inner circle ($\epsilon_2 = 1$). Consequently, in virtue of the geometry imposed by the dispersion relations, the wave vectors $\mathbf{k}_1^+(\pm\,\mathbf{p}_\mathrm{B})$ correspond to propagating waves in glass, and coincide with the Yoneda threshold. This situation is shown in Fig. 10.8(a) and Fig. 10.3(a). The coincidence of the Yoneda critical angle and the Brewster scattering angle for reflection in the optically denser medium for normal incidence is now explained. Although the two effects are of different nature and decoupled, they occur simultaneously in this case as a simple consequence of the geometry imposed by the dispersion relations. Our analysis rules out the speculated explanation involving the excitation of a resonant surface wave found in the literature [101]. By increasing the angle of incidence, only one Brewster wave remains with a shrinking lateral wave vector (see Fig. 10.8(b) and Fig. 10.3(b)). As in the previous configuration, the wave vectors of the reflected zero order wave, $\mathbf{k}_1^+(\mathbf{p}_0)$, and of the Brewster wave, $\mathbf{k}_1^+(\mathbf{p}_\mathrm{B})$, cross each other at the Brewster angle of incidence ($\approx 33.7°$). We now discover an interesting effect which was not present when the wave was incident from the less dense medium. As the polar angle of incidence approaches the critical angle of total internal reflection of the zero order reflected wave, $\theta_0 = \theta_c$, both the Snell-conjugate wave vector and the Brewster wave vector approach the vertical direction and reach it for $\theta_0 = \theta_c$. This situation is illustrated in Fig. 10.3(c). When θ_0 is increased beyond θ_c a sudden transition occurs. Let us come back to Fig. 10.4(a), where the reflection amplitudes and incoherent component of the MDRC are shown for an angle of incidence just above the critical angle. It seems that the Brewster scattering angle is nowhere to be found, and comparing with Fig. 10.3(c), the curve of the modulus of the reflection amplitude seems to detach from the abscissa axis. The Brewster scattering angle, however, comes back from the backscattering side, visible in the evanescent region of Fig. 10.4(c) where the polar angle of incidence is 70°. Let us explain this curious behavior. The overall behavior of the phase in Figs. 10.4(a)–(c) gives us a good hint. We have mentioned earlier that for *s*-polarized light, when the zero order reflected wave undergoes total internal reflection, the phase must undergo a Goos-Hänchen shift with θ_0. This also holds for the *p*-polarized zero order reflected wave, and consequently the two terms in Eq. (10.8b) are *not* any longer in phase. In the configuration where $\epsilon_1 < \epsilon_2$, the arguments of the two complex amplitudes in Eq. (10.8b) are always either in phase or are separated by a phase difference of π. Therefore, as time progresses, the *real part* of the total zero order field keeps a fixed direction. In the configuration where $\epsilon_1 > \epsilon_2$, the Goos-Hänchen phase shift makes the real part of the total zero order field change direction and turn in the plane of incidence as time progresses (it describes an ellipse). This indicates that the corresponding dipole radiation associated with the microscopic scatterers is not expected to be that of an oscillating dipole

anymore but that of a *rotating* dipole. Equivalently, we can say that the reflection amplitude of the zero order wave, $\rho_{pp}^{(0)}$, draws a lower half circle in the complex plane from 1 to -1 as the angle of incidence is varied from the critical angle to 90° for the configuration $\epsilon_1 > \epsilon_2$, while it stays on the real line for $\epsilon_1 < \epsilon_2$. It follows that $\mathbf{E}_{p,1}^{(0)}(\mathbf{p}_0)$ has a complex amplitude with a non-zero imaginary part. Therefore, a *propagating* Snell-conjugate wave *cannot* satisfy the requirement of Eq. (10.12) since its p-polarization vector would be real. In order to satisfy Eq. (10.12) the polarization vector $\hat{\mathbf{e}}_{p,2}^{+}(\mathbf{p}_{\mathrm{B}})$ must itself be complex (with a non-zero imaginary part), and the Snell-conjugate wave is naturally to be expected in the *evanescent* region of medium 2. This is the reason why the lateral wave vector of the Brewster scattering wave seems to disappear at the transition $\theta_0 = \theta_c + \Delta\theta_0$ and then comes back from the negative p_1 side as the angle of incidence is increased, revealing that the Snell-conjugate wave is evanescent.

Remark 10.4. Note that according to the definition of a polarization vector for a p-polarized wave, $\hat{\mathbf{e}}_{p,j}^{\pm}(\mathbf{p})$, given in Eq. (1.32d), the corresponding wave takes an interesting structure when it is evanescent. For an evanescent wave, $\alpha_j(\mathbf{p})$ is pure imaginary and the polarization vector $\hat{\mathbf{e}}_{p,j}^{\pm}(\mathbf{p})$ hence has a real component along $\hat{\mathbf{e}}_3$ and a pure imaginary component along $\mathbf{p}$. The wave is then transverse in the sense of the complex vectors, but not in terms of their real parts. Indeed, the corresponding real electric field is then the sum of a wave polarized along $\hat{\mathbf{e}}_3$ and a longitudinal wave (*longitudinal* with respect to the lateral wave vector i.e. the real part of the wave vector) dephased by $\pi/2$ radians with respect to the first wave. The resulting field therefore describes an ellipse in the $(\hat{\mathbf{e}}_3, \hat{\mathbf{p}})$-plane.

In-plane transmission — The analysis for the Brewster scattering effect for the transmitted waves is similar to that for the reflected waves. The analysis is left to the reader. One difference worth mentioning, however, is that the Brewster scattering direction is generally found in the backscattering region when the corresponding Brewster scattering for reflection is found in the forward scattering direction. Intuitively, this effect can be related to the emission of an oscillating dipole which yields zero emitted power along the direction of oscillation, hence producing two antipodal zero intensity points when the intensity is mapped onto a sphere (for a free dipole see e.g. Fig. 1.6(a)). The dipolar nature of the radiation will be clearly apparent when considering the full angular scattering distribution in the next section.

Scattering in all directions

p-polarized Brewster scattering — The full angular distributions of the diffuse component of the MDRC and MDTC, for normal incidence and the parameters assumed in Figs. 10.2(a) and 10.2(d), are presented in Fig. 10.9. The lower left 2×2 panels in each collection of panels in Fig. 10.9 exhibit a pattern which is reminiscent of the polarization pattern of the dipole radiation in free space discussed in Section 1.3 in the case when the dipole oscillates in the $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2)$-plane (see Fig. 1.6(c)). Indeed, for normal incidence, the zero order waves and the incident wave have an oscillating electric field either along $\hat{\mathbf{e}}_1$ for p polarization or along $\hat{\mathbf{e}}_2$ for s polarization. Consequently, the dipoles in the media also oscillate along the direction of the incident field. In the case of an s-polarized wave (field along $\hat{\mathbf{e}}_2$) we have seen in Section 1.3 that the dipole radiation in free space yields zero s-polarized emission in the $(\hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3)$-plane and an overall $|\sin(\phi - \pi/2)|$ intensity, which is consistent with what is observed in Fig. 10.9(f). In fact, for a fixed polar angle of reflection θ_r, the variation along the azimuthal angle ϕ of the incoherent component of the MDRC to lowest non-zero order in the surface roughness for $s \to s$ polarized scattering is *exactly* proportional to $|\sin(\phi - \pi/2)|$ since $\rho_{ss}^{(1)}$ is proportional to $\hat{\mathbf{e}}_s(\mathbf{p}) \cdot \hat{\mathbf{e}}_s(\mathbf{p}_0)$, as can be seen from Eq. (10.6a), and this is the only ϕ dependence for normal

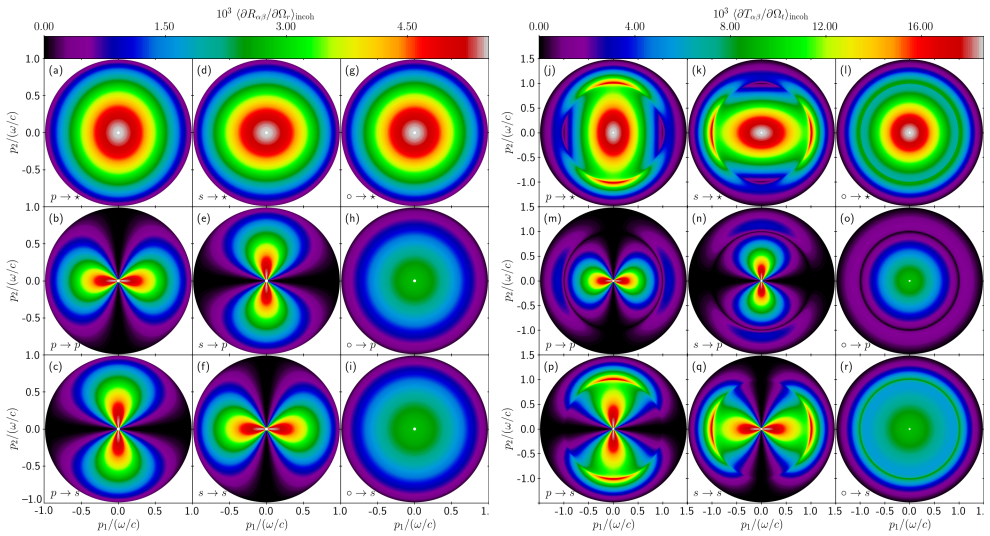


Figure 10.9: The full angular distribution of the incoherent component of the MDRC/MDTC, $\langle \partial X_{\alpha\beta}/\partial\Omega_x \rangle_{\rm incoh}$ for $X = R$ or T, in the $\mathbf{p}$-plane of the light that is scattered from a rough interface where the angle of incidence $\theta_0 = 0°$. The positions of the specular directions in reflection and transmission are indicated by white dots. The physical parameters were identical to those assumed in obtaining the results of Fig. 10.2. The sub-figures in Figs. 10.9(a)–(i) and 10.9(j)–(r) are both organized in the same manner and show how incident β-polarized light is scattered into α-polarized light ($\alpha, \beta \in \{p, s\}$) and denoted $\beta \to \alpha$. Moreover, the notation $\circ \to \star$ is taken to mean that the incident light was unpolarized while the polarization of the scattered light was not recorded. For instance, this means that the data shown in Fig. 10.9(a) are obtained by adding the data sets presented in Figs. 10.9(b)–(c); similarly, the data shown in Fig. 10.9(g) result from the addition and division by a factor two of the the data sets presented in Figs. 10.9(a) and 10.9(d); *etc.* Finally, the intensity variations in the plane of incidence ($\mathbf{p} = p_1\ \hat{\mathbf{e}}_1$) from Figs. 10.9(b, f) and 10.9(k, o) are the curves depicted in Figs. 10.2(a) and Figs. 10.2(d), respectively. Figure taken from **Paper** [2].

incidence. For normal incidence, this observation holds for all the polarization couplings up to an azimuthal rotation by $\pi/2$ for cross-polarization.

A similar dipole radiation pattern can be observed for the transmitted light (see Fig. 10.9(j-r)). Moreover, we also observe the Yoneda phenomenon. This is the enhancement of the diffuse contribution to the MDTC intensity above the critical lateral wave vector for the scattered waves with $|\,\mathbf{p}\,| > p_c$, as discussed extensively in Section 10.1.2. This is particularly visible for out-going s-polarized light, especially in Figs. 10.9(p)–(r). For out-going p-polarized light, we observe a dark ring of zero scattering intensity along the circle $|\,\mathbf{p}\,| = p_c$. This is the two-dimensional generalization of our discussion for scattering in the plane of incidence, where we found two Brewster waves with $\mathbf{p}_B = \pm p_c\ \hat{\mathbf{e}}_1$ for normal incidence. Now we see that in the two-dimensional $\mathbf{p}$-space the solutions to Eq. (10.12) are in fact given by $|\,\mathbf{p}\,| = p_c$. This set of solutions corresponds to the vanishing radiation of p-polarized light in the equatorial plane for the case $\vartheta = 90°$ as illustrated in Fig. 1.6(c) for dipole radiation in free space. To be more

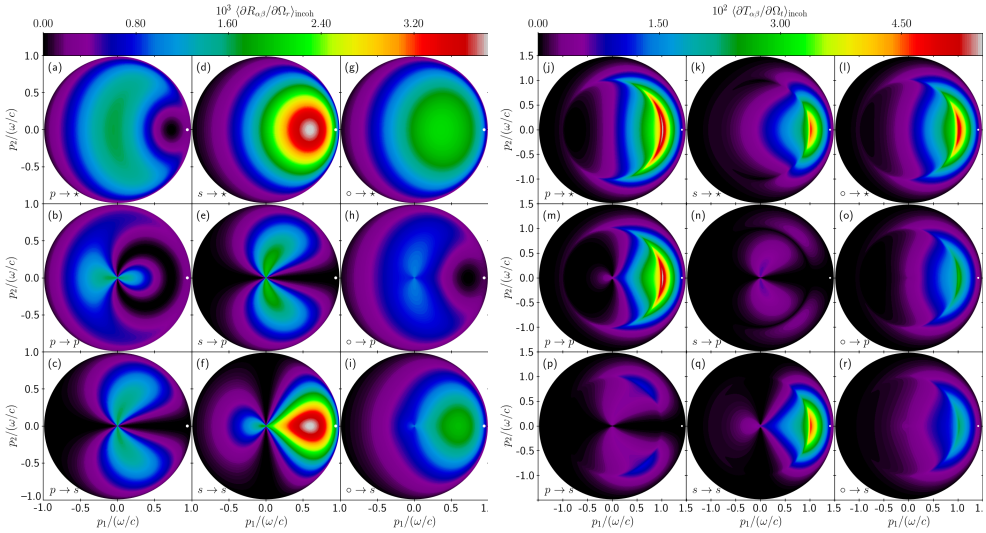


Figure 10.10: Same as Fig. 10.9, but now for the angle of incidence $\theta_0 = 70°$. Figure taken from **Paper** [2].

accurate we must add all the in-plane of wave vectors of the form $\mathbf{p} = p_2\,\hat{\mathbf{e}}_2$ to the previous set to obtain all the solutions of the criterion Eq. (10.12). We thus understand clearly the link between the criterion Eq. (10.12) and the radiation of the elementary dipoles in the media by comparison with what have been discussed in Section 1.3.

Remark 10.5. If we had plotted the modulus of the reflection amplitude in the **p**-plane instead of the incoherent component of the MDRC, a ring of zero amplitude for out-going p-polarized would also be seen, but of course in the evanescent regime.

Consider now a larger angle of incidence, $\theta_0 = 70°$, for which the diffuse contributions to the MDRC and MDTC for incidence in vacuum are shown in Fig. 10.10. First, we observe in Fig. 10.10(b) that for $p \to p$ reflection, there exists a closed curve of zero intensity in the forward scattering direction. Similarly, we observe in Fig. 10.10(m) a closed curve of zero intensity for $p \to p$ transmission but in the backscattering region. These features are analogous to those observed in the case of the p polarization component of the dipole radiation in free space studied in Section 1.3 in the case where the dipole tilting angle is such that $0° < \vartheta < 90°$ (see Fig. 1.6(b) for $\vartheta = 45°$). The curves of zero intensity for $p \to p$ scattering in Fig. 10.10 can then be interpreted as the signature of an overall dipole radiation whose dipole moment is tilted from the x_3-axis by some angle ϑ. The polarization of the reflected (resp. transmitted) light would then derive from the northern (resp. southern) hemisphere of the radiation polarization pattern.

Let us now make this statement more precise by interpreting geometrically the criterion for vanishing $p \to p$ scattering, Eq. (10.12), for the case of reflection and $\theta_0 = 70°$. The following construction is a generalization of the one presented previously for scattering in the plane of incidence. The main steps of the geometrical construction of the set of directions of zero $p \to p$ reflection are illustrated in Fig. 10.11. The steps go as follows:

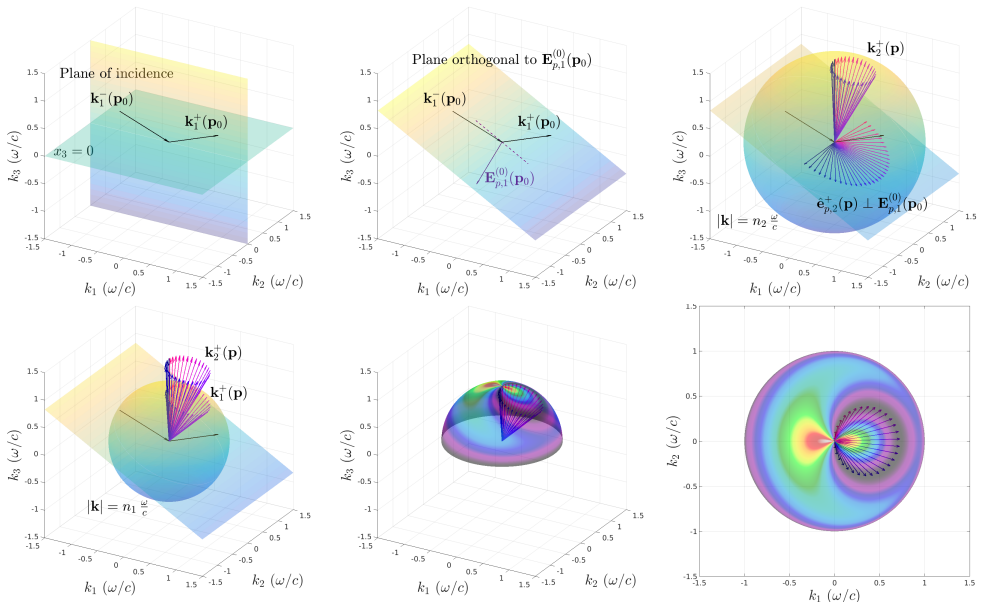


Figure 10.11: Illustration of the construction steps leading to the set of directions of zero $p \to p$ reflection for an angle of incidence of $\theta_0 = 70°$. (a) Sketch of the average surface, the plane of incidence and the considered wave vectors of the incident and reflected zero order waves ($\mathbf{k}_1^-(\mathbf{p}_0)$ and $\mathbf{k}_1^-(\mathbf{p}_0)$). (b) Construction of the total zero order field amplitude $\mathbf{E}_{p,1}^{(0)}(\mathbf{p}_0)$ and the plane orthogonal to it. Note that the incident wave vector does not in general belong in this plane as illustrated with the dashed indigo line indicating the intersection of the plane of incidence and the plane $\mathbf{E}_{p,1}^{(0)}(\mathbf{p}_0)^\perp$. (c) Unit vectors belonging to the lower half $\mathbf{E}_{p,1}^{(0)}(\mathbf{p}_0)^\perp$-plane. They correspond to the possible polarization vectors $\hat{\mathbf{e}}_{p,2}^+(\mathbf{p})$ of Eq. (10.12). The wave vectors $\mathbf{k}_2^+(\mathbf{p})$ associated with the polarization vectors $\hat{\mathbf{e}}_{p,2}^+(\mathbf{p})$ are then constructed according to Eqs. (1.32). Note that they lie on a sphere of radius $|\,\mathbf{k}\,| = n_2\,\omega/c$. The color associated with the vectors $\hat{\mathbf{e}}_{p,2}^+(\mathbf{p})$ and $\mathbf{k}_2^+(\mathbf{p})$ helps us to identify the $\mathbf{k}_2^+(\mathbf{p})$ associated with each $\hat{\mathbf{e}}_{p,2}^+(\mathbf{p})$ (they share the same color). (d) The wave vectors $\mathbf{k}_2^+(\mathbf{p})$ are projected on the sphere of radius $|\,\mathbf{k}\,| = n_1\,\omega/c$ following the x_3-direction, hence giving the wave vectors $\mathbf{k}_1^+(\mathbf{p})$ of zero $p \to p$ reflection. (e) The incoherent component of the mean differential reflection coefficient is shown on the scattering sphere together with the set of wave vectors $\mathbf{k}_1^+(\mathbf{p})$ obtained in (d). (f) Projection of (e) in the $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2)$-plane. We verify in (e) and (f) that the constructed wave vectors indeed follow the curve of zero scattering of the incoherent component of the MDRC. Figure taken from **Paper** [2].

1 The wave vectors of incidence $\mathbf{k}_1^-(\mathbf{p}_0)$ and of the reflected zero order wave $\mathbf{k}_1^+(\mathbf{p}_0)$ are drawn in black (see Fig. 10.11(a)).

2 The direction of the total zero order field amplitude $\mathbf{E}_{p,1}^{(0)}(\mathbf{p}_0)$ is determined[2]. Note that this direction lies in the plane of incidence (see Fig. 10.11(b)).

3 Based on $\mathbf{E}_{p,1}^{(0)}(\mathbf{p}_0)$, this determines the vector plane orthogonal to it, $\mathbf{E}_{p,1}^{(0)}(\mathbf{p}_0)^\perp$ (see

[2]We recall that the steps to geometrically construct the total zero order field have been treated in detail in Ref. [102], and we do not show these here.

Fig. 10.11(b)). Note that in general this plane does *not* contain the incident wave vector as made clear by the dashed line, showing the intersection of the plane of incidence with the plane $\mathbf{E}^{(0)}_{p,1}(\mathbf{p}_0)^\perp$ in Fig. 10.11(b).

4 Eq. (10.12) states that each polarization vector $\hat{\mathbf{e}}^+_{p,2}(\mathbf{p})$ contained in the plane $\mathbf{E}^{(0)}_{p,1}(\mathbf{p}_0)^\perp$ lead to a solution of the Brewster scattering criterion. Since the $\hat{\mathbf{e}}^+_{p,2}(\mathbf{p})$ vectors are normalized they are distributed on a circle of unit radius on the plane $\mathbf{E}^{(0)}_{p,1}(\mathbf{p}_0)^\perp$. The set of all $\hat{\mathbf{e}}^+_{p,2}(\mathbf{p})$ vectors satisfying Eq. (10.12) therefore spans a half circle in the plane $\mathbf{E}^{(0)}_{p,1}(\mathbf{p}_0)^\perp$ as shown in Fig. 10.11(c), where a sample of polarization vectors are represented. The fact that only the lower half circle is needed comes from the definition[3] of a polarization vector $\hat{\mathbf{e}}^+_{p,2}(\mathbf{p})$.

5 For each polarization vector satisfying Eq. (10.12), we can construct its corresponding wave vector $\mathbf{k}^+_2(\mathbf{p})$. This is done for example by using that its direction is given by $\hat{\mathbf{e}}^+_{p,2}(\mathbf{p}) \times [\hat{\mathbf{e}}^+_{p,2}(\mathbf{p}) \times \hat{\mathbf{e}}_3]$ and that $\mathbf{k}^+_2(\mathbf{p})$ lies on the northern hemisphere of the sphere of radius $k_2 = n_2\omega/c$. We thus obtain the cone of all wave vectors $\mathbf{k}^+_2(\mathbf{p})$ whose corresponding p polarization vector satisfies Eq. (10.12). A sample of such vectors are represented for $\hat{\mathbf{e}}^+_{p,2}(\mathbf{p})$ and $\mathbf{k}^+_2(\mathbf{p})$ in Fig. 10.11(c).

6 The last step consists in projecting the vectors $\mathbf{k}^+_2(\mathbf{p})$ along $\hat{\mathbf{e}}_3$ onto the sphere of radius $|\,\mathbf{k}\,| = n_1\,\omega/c$ to obtain the wave vectors $\mathbf{k}^+_1(\mathbf{p})$ of zero $p \to p$ reflection as illustrated in Fig. 10.11(d)).

The sampled wave vectors $\mathbf{k}^+_1(\mathbf{p})$ together with the diffuse contribution to the MDRC, mapped to the hemisphere and its projection in the $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2)$-plane constructed following the aforementioned steps are respectively shown in Fig. 10.11(e) and 10.11(f). We verify that the set of constructed wave vectors correspond to the observed curve of zero intensity for $p \to p$ reflection.

$s \to p$ Brewster scattering — Figure 10.10(n) shows that the $s \to p$ transmitted light exhibits a circle of zero intensity, for $|\,\mathbf{p}\,| = n_1\,\omega/c$ similar to what was observed for normal incidence in Fig. 10.9(n). This feature is also present in reflection but in the evanescent region, and is observed by considering the complex amplitude instead of the MDRC. The invariance of the circle of zero intensity with the angle of incidence for the $s \to p$ scattering is simple to understand in terms of the dipole radiation in free space. For s-polarized incident light the dipoles in the media are all oriented along $\hat{\mathbf{e}}_2$, independent of the angle of incidence. When measuring the p-polarized component of the radiated light, we thus expect to obtain an underlying pattern of zero intensity consistent with that obtained in the case of the oscillating dipole in free space as illustrated in Fig. 1.6(c).

S-black-out and circularly polarized Brewster scattering under TIR

In order to appreciate the behavior of the amplitudes of the waves scattered in the evanescent region as well as the ones scattered in the propagating region, it is instructive to study the modulus square of the amplitudes instead of the MDRC and MDTC. In this section, we analyze what is to our knowledge a new phenomenon and which can be considered as a generalization of

[3] One may extend the construction to all vectors on the circle defined as the intersection of the unit sphere and the plane $\mathbf{E}^{(0)}_{p,1}(\mathbf{p}_0)^\perp$, but it would result in constructing twice the same set of wave vectors of zero scattering.

the Brewster scattering effect for light scattered from linearly to circularly polarized light. To this end, we show in Figs. 10.12-10.15, $|\rho^{(1)}_{\alpha\beta}|^2$ in the $\mathbf{p}$-plane for different polar angles of incidence.

Let $\beta \in \{p, s, \circ\}$ and $\alpha \in \{p, s, \sigma^+, \sigma^-\}$ denote respectively the polarization of the incident wave and of the scattered wave. Here $\circ$ indicates unpolarized light. The states $\sigma^\pm$ denote respectively left and right circular polarization states and the corresponding reflection amplitudes are deduced from the p and s polarization amplitudes by

$$\rho^{(1)}_{\sigma^\pm\beta} = \frac{1}{\sqrt{2}} \left[\rho^{(1)}_{p\beta} \pm i\rho^{(1)}_{s\beta}\right] , \tag{10.13}$$

and similarly for the transmission amplitudes.

We consider here the case for which the medium of incidence is the optically denser one ($\epsilon_1 > \epsilon_2$), as an intersting transition occurs for the linearly to circularly polarized Brewster scattering effect only when the reflected zero order wave undergoes total internal reflection. Even though we treat only the reflected light as an illustration, the effect can be observed both for the reflected *and* the transmitted scattered light. When studying the Brewster scattering phenomenon in the configuration $\epsilon_1 > \epsilon_2$ restricted to scattering in the plane of incidence, we have seen that the Brewster scattering effect exhibits a sudden transition when the reflected zero order wave undergoes total internal reflection. First, the direction of zero p-polarized reflected intensity goes towards the x_3-direction as the polar angle of incidence approaches the critical angle for total internal reflection. As the polar angle of incidence goes beyond the critical angle for total internal reflection, the direction of zero scattering suddenly disappears from the propagating region. We have argued that this sudden transition is attributed to a transition of the dipolar response of the media, going from an oscillating behavior to a rotating behavior due to the phase shift between the incident excitation and the scattered zero order waves.

Let us now follow this transition in the full $\mathbf{p}$-plane with particular attention on the scattered *circularly polarized* light. Indeed, it was shown in Section 1.3 that the radiation emitted by a rotating dipole in free space exhibits characteristic signatures in the emitted circularly-polarized light out of the plane of incidence (see Figs. 1.6(e) and 1.6(f)). According to our understanding of the response in terms of dipole radiation, such signatures should then also be observed for the scattering by a rough interface under total internal reflection.

For polar angles of incidence smaller than the critical angle, $\theta_0 < \theta_c$, the amplitudes $\rho^{(1)}_{p\beta}$ and $\rho^{(1)}_{s\beta}$ are both real valued for scattering angles smaller than the Yoneda threshold. In that case, the circularly polarized amplitudes vanish if and only if both $\rho^{(1)}_{p\beta}$ and $\rho^{(1)}_{s\beta}$ vanish simultaneously according to Eq. (10.13). For an incident p-polarized wave (i.e. $\beta = p$) this occurs only where the curve of zero $p \to p$ scattering (cf. previous subsection) intersects with the plane of incidence in which $p \to s$ scattering is identically zero. This fact is illustrated for normal incidence in Figs. 10.12(a) and (d) showing $|\rho^{(1)}_{pp}|^2$ and $|\rho^{(1)}_{sp}|^2$ in the $\mathbf{p}$-plane. There we recognize the curves of zero scattering for the p- and s-polarized light discussed in previous sections. Accordingly, Figs. 10.12(g) and (j) show $|\rho^{(1)}_{\sigma^\pm p}|^2$ where two directions of zero $p \to \sigma^\pm$ scattering are present, namely at $\mathbf{p} = \pm p_c\, \hat{\mathbf{e}}_1$, although they are hardly visible on this figure. The effect is clearer for oblique incidence, as shown in Figs. 10.13(g) and (j) ($\theta_0 = 35°$), where a clear unique direction of zero intensity in $p \to \sigma^\pm$ scattering in the plane of incidence can be distinguished. This effect is also observed in the configuration $\epsilon_1 < \epsilon_2$ (not shown here). When the angle of incidence

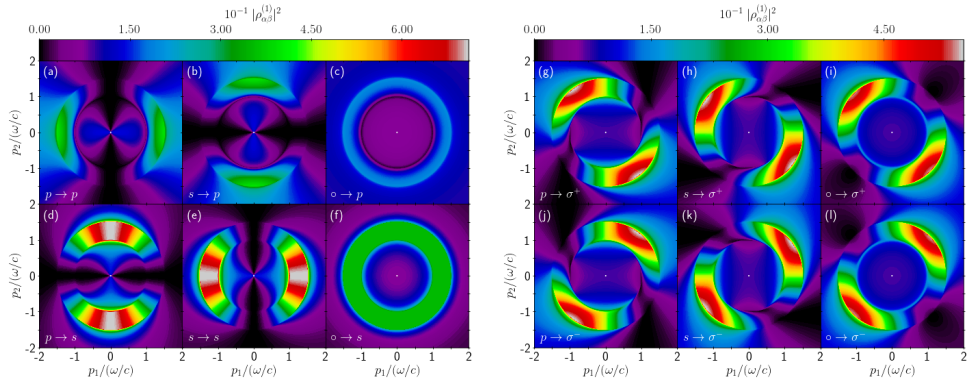


Figure 10.12: The full angular distribution of $|\rho^{(1)}_{\alpha\beta}|^2$ for normal incidence, $\theta_0 = 0°$, $\epsilon_1 = 2.25$, $\epsilon_2 = 1$, for incident polarization $\beta \in \{p, s\}$ or unpolarized ($\circ$) and outgoing polarization $\alpha \in \{p, s, \sigma^+, \sigma^-\}$. Figure taken from **Paper** [2].

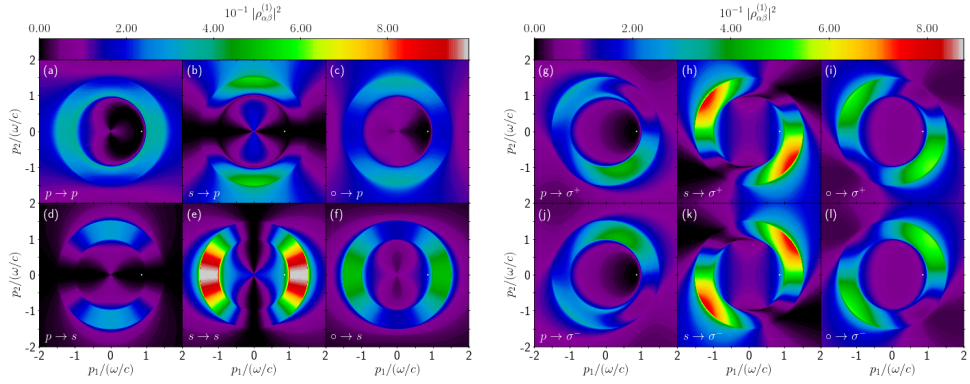


Figure 10.13: Same as Fig. 10.12 but for the angle of incidence $\theta_0 = 35°$. Figure taken from **Paper** [2].

reaches the critical angle of incidence, $\theta_0 = \theta_c$, the direction of zero intensity in $p \to \sigma^\pm$ scattering reaches the x_3-direction, as illustrated in Figs. 10.14(g) and (j). Note that $p \to p$ scattering also has a zero of intensity along the x_3-direction as explained earlier. Furthermore, we note that the distribution of $|\rho^{(1)}_{pp}|^2$ and $|\rho^{(1)}_{\sigma^\pm p}|^2$ are cylindrically symmetric as shown in Fig. 10.14(a), (g) and (j). The cylindrical symmetry can be understood based on the radiation of an oscillating dipole aligned with the x_3-axis. Indeed, we have seen in Section 1.3 that the p-polarized radiation from a dipole oscillating along the x_3-axis is cylindrically symmetric with vanishing emission at the poles of the unit sphere (see Fig. 1.6(a)). The radiation from such a dipole is also *purely* p-polarized, which has two consequences: (i) the s-polarized scattered light vanishes identically for *all* $\mathbf{p}$, a phenomena that we like to refer as *s-black-out* (see Fig. 10.14(d)); (ii) the radiation can be decomposed into σ^+ and σ^- components of *equal* intensity, as can be observed in Figs. 10.14(g) and (j).

Remark 10.6. Our interpretation of these observations is based on the radiation of an oscillat-

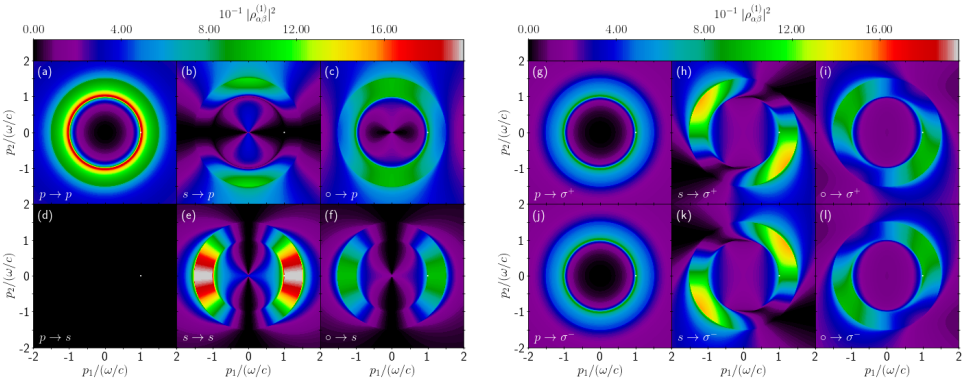


Figure 10.14: Same as Fig. 10.12 but for the angle of incidence equal to the critical angle for total internal reflection, $\theta_0 = \theta_c = 41.81°$. Figure taken from **Paper** [2].

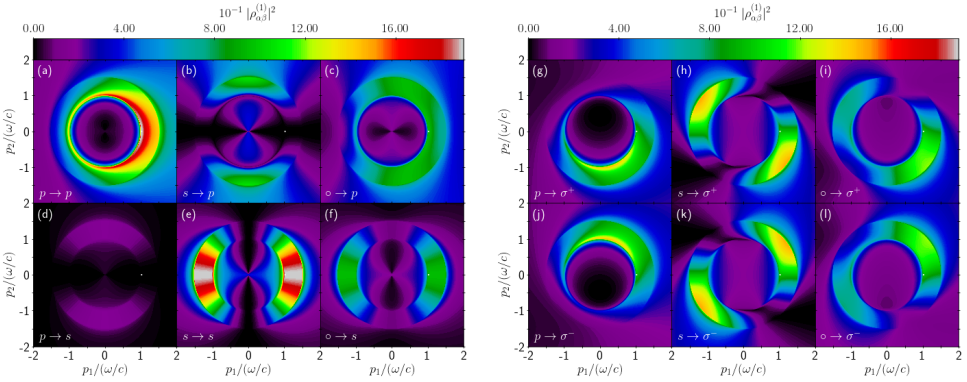


Figure 10.15: Same as Fig. 10.12 but for the angle of incidence $\theta_0 = 43°$. Figure taken from **Paper** [2].

ing dipole in free space for the sake of simplicity, and because we believe it illustrates well the fundamental mechanism behind the observed features. However, it is relatively simple to verify these assertions based on the expressions of the amplitudes given in Eq. (10.6). For example, it is clear that for $\theta_0 = \theta_c$, the total zero order field $\mathbf{E}^{(0)}_{1,p}(\mathbf{p}_0)$ is along $\hat{\mathbf{e}}_3$ and the dot product in Eq. (10.6c) vanishes for all $\mathbf{p}$.

For angles of incidence larger than the critical angle ($\theta_0 > \theta_c$), it is convenient to expand the right-hand side in Eq. (10.13) by inserting Eq. (10.6) for the reflections amplitudes. The reduced first order reflection amplitude, $\hat{\rho}^{(1)}_{\sigma^\pm p}$, for $\sigma^\pm$-polarized light scattered from incident p-polarized light is then given by

$$\hat{\rho}^{(1)}_{\sigma^\pm p}(\mathbf{p}\,|\,\mathbf{p}_0) = \frac{1}{\sqrt{2}}\left[\gamma(\mathbf{p})\,\hat{\mathbf{e}}^+_{p,2}(\mathbf{p}) \pm i\,\hat{\mathbf{e}}_s(\mathbf{p})\right] \cdot \mathbf{E}^{(0)}_{1,p}(\mathbf{p}_0)\,. \tag{10.14}$$

where we have used the short-hand notation $\gamma(\mathbf{p}) = (\hat{\mathbf{e}}^+_{p,2}(\mathbf{p}) \cdot \hat{\mathbf{e}}^+_{p,1}(\mathbf{p}))^{-1}$. For $\theta_0 > \theta_c$, the total zero order field amplitude $\mathbf{E}^{(0)}_{1,p}(\mathbf{p}_0)$ is complex with non-zero imaginary part. Thus neither $\rho^{(1)}_{pp}$

nor $\rho_{sp}^{(1)}$ can be zero for *propagating* waves. We have seen nevertheless that a vanishing $p \to p$ scattering amplitude point can be found in the evanescent region since $\hat{\mathbf{e}}_{p,2}^{+}(\mathbf{p})$ becomes complex (and non real). However, a zero point of intensity in $p \to \sigma^{\pm}$ scattering may be found in the *propagating* region. Indeed, the square bracket in Eq. (10.14) is complex even for purely real values of $\hat{\mathbf{e}}_{p,2}^{+}(\mathbf{p})$ and $\hat{\mathbf{e}}_{s}(\mathbf{p})$ may compensate for the fact that $\mathbf{E}_{1,p}^{(0)}(\mathbf{p}_0)$ is complex (and non real) and make the dot product in Eq. (10.14) vanish. Note the similarity with Eq. (1.85) for the case of the radiation emitted by a rotating dipole, with the important difference that the p polarization vector is that of the *Snell-conjugate wave*. Since $\mathbf{E}_{1,p}^{(0)}(\mathbf{p}_0)$ represents a state of polarization of the media in which the dipole rotates in the plane of incidence, we expect to find a zero in the $\sigma^{\pm}$ scattering intensity on each side ($\phi = \pm\pi/2$) of the plane of incidence. This is indeed what we observe in Figs. 10.15(a) and (d) in the $|\rho_{\sigma^{\pm}p}^{(1)}|^2$ distribution of $p \to \sigma^{\pm}$ scattering for $\theta_0 = 43°$. Finally, let us comment on $s \to \sigma^{\pm}$ scattering. From Figs. 10.12-10.15 it can be noticed that the distribution for $|\rho_{\sigma^{\pm}s}^{(1)}|^2$ stays identical, up to an overall scaling factor, as the angle of incidence varies. This can also be understood from the dipole picture. For a s-polarized incident wave, the incident and zero order waves are s-polarized, so the dipoles oscillate along the x_2-direction independently of the angle of incidence. For scattering in the plane of incidence the first order waves are purely s-polarized and the two $\sigma^{\pm}$ components have equal intensity. For scattering at azimuthal angles $\phi_{=} \pm\pi/2$, the first order waves are purely p-polarized and the two $\sigma^{\pm}$ components have again have equal intensity. We obtain the largest contrast between σ^{+} and σ^{-} for ϕ being a multiple of 45° since then the p- and s-polarized components have similar amplitudes.

10.2 Multiple scattering

We have discussed the physics of the scattering of light from a rough interface in terms of a single scattering picture, and we have interpreted the observed polarization features based on dipole radiation. What would happen if higher order scattering events were taken into account? When multiple scattering events are taken into account, one does not observe a strict zero for the Brewster scattering direction but a local minimum. In fact, the Brewster effect for the coherent component of the intensity, i.e. the intensity associated with the average field does not yield a zero anymore, like for the Fresnel amplitude, but also a local minimum. The angle of incidence associated with the minimum of reflection intensity is in general shown to be shifted compared with the Brewster angle of incidence for a planar interface due to multiple scattering. An intuitive picture for understanding the fact that one does not find a strict zero of intensity anymore is the following. The zero and first order fields acts as a source for the second order field and so on. Hence the light being scattered several times may in principle be redistributed into the directions of Brewster scattering. But why should not this redistribution keep the directions of zero scattering? The first order field had the zero order field as a source, and it was the fact that the zero order polarization induced by the incident and zero order reflected and refracted fields behaved as a planar interface response that yielded to an overall well defined dipole radiation pattern. The first order field may be seen as a probe, thanks to the scattering from the surface, for the zero order response in all the directions. Note that this is the reason why the directions of the zero scattering are *independent* of the surface profile parameters. When the second order is taken into account, the first order field gives a random contribution to the dipole polarization on top of the zero order polarization. These randomly perturbed dipoles hence do not all oscillate (and/or rotate) in the same direction anymore thus leading to a redistribution of power, overall in all the directions. The quantitative estimates of the resulting intensity at the

Brewster minimum and corresponding angular shift require some more detailed analysis of the higher order terms in the perturbative expansion of the field, and depends on the parameters of the surface profile such as the correlation length and rms-roughness. We refer here to the work of Saillard and Maystre [103], Nieto-Vesperinas *et al* [104] and Greffet [105] for the shift of the angular position of the Brewster angle for the coherent field.

10.3 Scattering by a metallic surface

The single scattering theory presented previously and illustrated for dielectric extends naturally to metallic surfaces. However, for metallic surfaces, one expects to find phenomena associated with the excitation of surface plasmon polaritons (see Section 1.4.2). As shown for periodic metallic gratings, where Wood anomalies can be observed, surface plasmon polaritons may be excited as the incoming wave can be scattered with, *a priori*, any change of in-plane momentum. The degree of coupling will hence depend on a combination of frequency, dielectric function, angle of incidence and on the power spectrum of the surface profile. The excited surface plasmon polaritons will, however, scatter or couple back to radiative, or progressive, modes due to multiple scattering events. When averaged over realizations of the surface profile, it was shown that certain multiple scattering paths exhibit a robust coherent effect upon averaging. They yield so-called enhanced forward [106] and backscattering [107, 108, 109, 110]. The enhanced backscattering peak can be observed in the light scattered in the retro-direction, i.e. back from where the incident wave came from. The common explanation is that this phenomenon is due to the constructive interference of two reciprocal paths: one scattering into a surface plasmon polariton at some point A on the surface and scattering into a radiative mode in the retro-direction at some other point B, and one scattering into a surface plasmon polariton at point B and scattering into a radiative mode in the retro-direction at point A. Such scattering paths and their reciprocal of course exists for all possible angle of scattering. But it can be shown that the backscattering direction gives paths that are phase coherent while when the angle of scattering moves away from this specific direction a loss of coherence is observed. This explanation holds for not too rough surfaces in a sense. For very rough surfaces and even for volume random media, the effect can also be observed also surface plasmon polariton may not be present as a mediator, but simply "regular" scattering events [36, 111, 112, 113].

10.4 Summary

This first chapter of Part III was focused on the physical understanding of a few phenomena observed in the scattering from a single dielectric or metallic interface.

For dielectric surfaces, the phenomena of the Yoneda ring, which is a ring of enhanced intensity in the diffuse light scattered in the optically denser medium, and of the Brewster scattering directions, which are directions of zero or local minima in the diffuse intensity, have been explained based on small amplitude perturbation theory to lowest non-trivial order. By a careful analysis of the reflection and transmission amplitudes, we were able to recast the expressions given by small amplitude perturbation theory to interpret the amplitudes in terms of the product of a scalar wave behavior and the signature of the polarization state of the media. Clear physical interpretations of the Yoneda phenomenon, which is intrinsically a scalar wave behavior and of the Brewster scattering phenomenon, which is microscopically linked to the dipolar nature of the radiation of the elementary constituents of the media, have then been

given. In order to quantitatively understand the Brewster phenomenon, the concept of Snell conjugate waves has been introduced as a generalization for scattering from a rough interface between two dielectric media of the Ewald wave triad for the reflection and refraction of a plane wave at the interface between two homogeneous media made of dipole arrays. We have also observed two new phenomena, the *s*-black-out and the Brewster scattering phenomena for linearly to circularly polarized light for a configuration of internal reflection under total internal reflection incidence for the zero order wave. The *s*-black-out was explained physically by the oscillation of dipoles aligned perpendicularly to the average plane of the interface, the scattered light hence being entirely *p*-polarized for all directions. The Brewster scattering angle for linear to circular polarization was explained in terms of rotating dipoles. The phenomena and related physical mechanisms are expected to hold for weakly rough surfaces and all require solely a single scattering picture.

For metallic surfaces, we have briefly discussed the phenomenon of enhanced backscattering and its multiple scattering nature and gave references for further readings.

Chapter 11

Multilayer with randomly rough interfaces

We are now familiar with the physics of scattering by a single rough interface between two dielectric media within a single scattering approximation. The aim of the present chapter is to take one step further and to consider a multilayer system whose interfaces are rough. The reduced Rayleigh equations for such systems were derived in Chapter 3 and the corresponding small amplitude perturbation theory developed in Chapter 9. For the sake of simplicity, we will study the case of a dielectric film deposited on a dielectric substrate, hence limiting the number of interfaces to two. The discussion can of course be generalized to several layers. Moreover, we will study the case where the two interfaces may be rough, but the rms-roughness will be assumed to be small enough so that small amplitude perturbation theory to first order in the product of the surface profiles can be applied within reasonable accuracy. Such a regime allows for interesting phenomena which can be explained in terms of a single scattering picture, which means that optical paths may be allowed to exhibit a single scattering event either on the top or bottom interface in a sense which will be precised in Section 11.1.

After recalling the expressions of, and giving a physical interpretation to, the scattering amplitudes given by SAPT to first order and the corresponding expressions for the mean differential scattering coefficients, we will explain the physics of Selényi rings. Selényi rings, named after the Hungarian physicist Pál Selényi, are rings of interference observed in the diffusely scattered light [114, 115, 116, 117]. We will explain the origin of the interference pattern (Section 11.2.1), and how the ring contrast is affected depending on which interface is rough (under the assumption that only one is rough at a time, Section 11.2.2). We will also explore the influence of the Brewster and Yoneda effects, studied in great detail in Chapter 10, on the Selényi rings and show that Selényi rings may be used to experimentally verify the phase jump and phase shift associated respectively with the Brewster scattering and Yoneda effects (Section 11.2.3). Furthermore, we will show that an additional degree of coherence can be present when considering interface-cross-correlation by demonstrating that cross-correlated interfaces may lead to significant enhancement and attenuation of some of the rings (Section 11.3). Finally, the chapter will close with a discussion about what may be observed for rougher interfaces, hence requiring a multiple scattering picture.

11.1 Interpretation of SAPT, single scattering picture

Definition of the scattering system – Consider a system made of three dielectric media of dielectric constants ϵ_j ($j \in \{1,2,3\}$) separated by two interfaces described by the equations $x_3 = \zeta_j(\mathbf{x}) = d_j + h_j(\mathbf{x})$ with $d_j = \langle \zeta_j \rangle$ being the offset of the average jth interface with respect to the origin and h_j is a realization of a continuous, differentiable, stationary, isotropic, Gaussian stochastic process. We will denote the average film thickness by $d = d_1 - d_2$. The statistical properties of the stochastic processes are given by

$$\langle h_j(\mathbf{x}) \rangle = 0 \tag{11.1}$$

$$\langle h_i(\mathbf{x})\, h_j(\mathbf{x}') \rangle = \left[\delta_{ij} + \gamma(1-\delta_{ij})\right] \sigma_i \sigma_j\, W(\mathbf{x}-\mathbf{x}') \tag{11.2}$$

for $i,j \in \{1,2\}$, where σ_j denotes the rms-roughness of the jth interface, W is an auto-correlation function which we choose to be identical for both surfaces, and $\gamma \in [-1,1]$ is an interface-cross-correlation coupling parameter. Hence we have $\langle h_j(\mathbf{x})\, h_j(\mathbf{x}') \rangle = \sigma_j^2\, W(\mathbf{x}-\mathbf{x}')$ and $\langle h_1(\mathbf{x})\, h_2(\mathbf{x}') \rangle = \gamma\, \sigma_1 \sigma_2\, W(\mathbf{x}-\mathbf{x}')$. For $\gamma > 0$ the interfaces are positively correlated (the extreme case $\gamma = 1$ corresponding to $h_2 = c\, h_1$ for each realization of the system and $c = \sigma_2/\sigma_1$), for $\gamma < 0$ the interfaces are negatively correlated (the extreme case $\gamma = -1$ corresponding to $h_2 = -c\, h_1$ for each realization of the system) and the interfaces are uncorrelated for $\gamma = 0$ (realizations of h_1 and h_2 are picked independently). The auto-correlation function will be taken to be Gaussian with auto-correlation length a,

$$W(\mathbf{x}) = \exp\left(-\frac{|\,\mathbf{x}\,|^2}{a^2}\right),$$

for which we recall that the power spectrum is given by

$$g(\mathbf{p}) = \pi a^2 \exp\left(-\frac{|\,\mathbf{p}\,|^2 a^2}{4}\right).$$

First order SAPT – With the aforementioned definitions and notations for the description of the scattering system, the first order terms in the reflection and transmission amplitudes expanded in products of surface profiles are given, according to Eqs. (9.16) and (9.18) for $n = 2$, by

$$\mathbf{R}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}) = \hat{h}_1(\mathbf{p}-\mathbf{p}_0)\, \boldsymbol{\rho}_1(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}) + \hat{h}_2(\mathbf{p}-\mathbf{p}_0)\, \boldsymbol{\rho}_2(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}), \tag{11.3a}$$

$$\mathbf{T}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}) = \hat{h}_1(\mathbf{p}-\mathbf{p}_0)\, \boldsymbol{\tau}_1(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}) + \hat{h}_2(\mathbf{p}-\mathbf{p}_0)\, \boldsymbol{\tau}_2(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}), \tag{11.3b}$$

where $\hat{h}_j$ is the Fourier transform of h_j and the amplitudes $\boldsymbol{\rho}_j$ and $\boldsymbol{\tau}_j$ are defined by (see Eqs. (9.17) and (9.19))

$$\boldsymbol{\rho}_1(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}) = \left[\bar{\boldsymbol{\Theta}}_{3,1}^{+,+}(\mathbf{p})\right]^{-1} \left[\tilde{\boldsymbol{\Theta}}_{3,1}^{+,+,(1,0)}(\mathbf{p}\,|\,\mathbf{p}\,|\,\mathbf{p}_0)\, \boldsymbol{\rho}_0(\mathbf{p}_0) - \tilde{\boldsymbol{\Theta}}_{3,1}^{+,-,(1,0)}(\mathbf{p}\,|\,\mathbf{p}\,|\,\mathbf{p}_0)\right], \tag{11.4a}$$

$$\boldsymbol{\rho}_2(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}) = \left[\bar{\boldsymbol{\Theta}}_{3,1}^{+,+}(\mathbf{p})\right]^{-1} \left[\tilde{\boldsymbol{\Theta}}_{3,1}^{+,+,(0,1)}(\mathbf{p}\,|\,\mathbf{p}_0\,|\,\mathbf{p}_0)\, \boldsymbol{\rho}_0(\mathbf{p}_0) - \tilde{\boldsymbol{\Theta}}_{3,1}^{+,-,(0,1)}(\mathbf{p}\,|\,\mathbf{p}_0\,|\,\mathbf{p}_0)\right], \tag{11.4b}$$

$$\boldsymbol{\tau}_1(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}) = -\left[\bar{\boldsymbol{\Theta}}_{1,3}^{-,-}(\mathbf{p})\right]^{-1} \tilde{\boldsymbol{\Theta}}_{1,3}^{-,-,(1,0)}(\mathbf{p}\,|\,\mathbf{p}_0\,|\,\mathbf{p}_0)\, \boldsymbol{\tau}_0(\mathbf{p}_0), \tag{11.4c}$$

$$\boldsymbol{\tau}_2(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}) = -\left[\bar{\boldsymbol{\Theta}}_{1,3}^{-,-}(\mathbf{p})\right]^{-1} \tilde{\boldsymbol{\Theta}}_{1,3}^{-,-,(0,1)}(\mathbf{p}\,|\,\mathbf{p}\,|\,\mathbf{p}_0)\, \boldsymbol{\tau}_0(\mathbf{p}_0), \tag{11.4d}$$

where the definitions of the various matrices involved here were defined in Section 9.1. Note that we omit the superscript (1) for clarity since there will be no ambiguity due to our restriction to work only with the first order amplitudes and that $\boldsymbol{\rho}_0(\mathbf{p}_0)$ and $\boldsymbol{\tau}_0(\mathbf{p}_0)$ correspond respectively to the reflection and transmission amplitudes for the corresponding system with planar interfaces (Fabry-Perot amplitudes). Remembering the details of the amplitudes $\boldsymbol{\rho}_j\left(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}\right)$ and $\boldsymbol{\tau}_j\left(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}\right)$ will not be so important in the following since we will soon give a simpler, physically intuitive, equivalent formulation. There are only a few important observations to note in Eq. (11.3), which were already mentioned in Section 9.1, and which we believe are worth recalling here.

First, in the special case where only one of the interfaces is rough, say for $j=\ell$, and the other interface is planar ($h_j=0$ for $j\neq\ell$), only the term corresponding to $j=\ell$ remains in the right-hand side of Eq. (11.3), i.e.

$$\begin{aligned}\mathbf{R}^{(1)}\left(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}\right)&=\hat{h}_\ell(\mathbf{p}-\mathbf{p}_0)\,\boldsymbol{\rho}_\ell\left(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}\right),\\ \mathbf{T}^{(1)}\left(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}\right)&=\hat{h}_\ell(\mathbf{p}-\mathbf{p}_0)\,\boldsymbol{\tau}_\ell\left(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}\right).\end{aligned}$$

In other words, *the first order amplitudes in the case of two rough interfaces can be interpreted as the sum of the first order amplitudes of the two associated subsystems for which only one of the interfaces is rough and the other planar.* This is only true for the first order amplitudes and it is a property of the single scattering picture. Turning now to the diffuse intensity, we have seen in Section 9.1 that the diffuse component of the mean differential scattering coefficients read (see Eq. (9.20) for $n=2$)

$$\begin{aligned}\left\langle\frac{\partial R_{\alpha\beta}}{\partial\Omega_r}(\mathbf{p}\,|\,\mathbf{p}_0)\right\rangle_{\mathrm{incoh}}&=\left(\frac{\omega}{2\pi c}\right)^2\frac{\epsilon_1\cos^2\theta_r}{\cos\theta_0}\Big[G_{11}(\mathbf{p}-\mathbf{p}_0)|\rho_{1,\alpha\beta}\left(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}\right)|^2+G_{22}(\mathbf{p}-\mathbf{p}_0)|\rho_{2,\alpha\beta}\left(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}\right)|^2\\&\quad+2\mathrm{Re}\left(G_{12}(\mathbf{p}-\mathbf{p}_0)\rho_{1,\alpha\beta}\left(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}\right)\rho^*_{2,\alpha\beta}\left(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}\right)\right)\Big],\end{aligned}\tag{11.6a}$$

$$\begin{aligned}\left\langle\frac{\partial T_{\alpha\beta}}{\partial\Omega_t}(\mathbf{p}\,|\,\mathbf{p}_0)\right\rangle_{\mathrm{incoh}}&=\left(\frac{\omega}{2\pi c}\right)^2\frac{\epsilon_3^{3/2}\cos^2\theta_t}{\epsilon_1^{1/2}\cos\theta_0}\Big[G_{11}(\mathbf{p}-\mathbf{p}_0)|\tau_{1,\alpha\beta}\left(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}\right)|^2+G_{22}(\mathbf{p}-\mathbf{p}_0)|\tau_{2,\alpha\beta}\left(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}\right)|^2\\&\quad+2\mathrm{Re}\left(G_{12}(\mathbf{p}-\mathbf{p}_0)\tau_{1,\alpha\beta}\left(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}\right)\tau^*_{2,\alpha\beta}\left(\mathbf{p}\,|\,\mathbf{p}_\mathrm{o}\right)\right)\Big],\end{aligned}\tag{11.6b}$$

$$G_{jk}(\mathbf{p}-\mathbf{p}_0)\overset{\mathrm{def}}{=}\left\langle\hat{h}_j(\mathbf{p}-\mathbf{p}_0)\hat{h}^*_k(\mathbf{p}-\mathbf{p}_0)\right\rangle.\tag{11.6c}$$

These expressions are typical intensity expressions one obtains when the field can be expressed as the sum of two fields. The resulting intensity is the sum of intensities one would obtain with either one field or the other plus an interference term, i.e. expressions of the form

$$I=|E_1|^2+|E_2|^2+2\mathrm{Re}(E_1E_2^*)=I_1+I_2+2\mathrm{Re}(E_1E_2^*).$$

Note, however, that each of the two first terms in the right-hand side of Eq. (11.6) already contains all the interference effects that may occur in each corresponding subsystem where only one of the interface is rough at a time. The last "interference" term in Eq. (11.6) corresponds to an additional degree of coherence induced by the correlation between the interfaces. In other words, it tells that two optical paths exhibiting a scattering event which happens on the different interfaces will, on average, keep a certain phase coherence due to the cross-correlation of the interfaces. In the case where the two interfaces are uncorrelated, this last term vanishes

and the resulting intensity is the sum of intensities of the two associated subsystem for which only one interface is rough.

In the particular case of correlations defined earlier in Eq. (11.2), Eq. (11.6) simplifies to

$$\left\langle \frac{\partial R_{\alpha\beta}}{\partial \Omega_r}(\mathbf{p}\,|\,\mathbf{p}_0) \right\rangle_{\text{incoh}} = \left(\frac{\omega}{2\pi c}\right)^2 \frac{\epsilon_1 \cos^2\theta_r}{\cos\theta_0}\, g(\mathbf{p}-\mathbf{p}_0) \left[\sigma_1^2\, |\rho_{1,\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_{\text{o}})|^2 + \sigma_2^2\, |\rho_{2,\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_{\text{o}})|^2 \right.$$
$$\left. + 2\gamma\sigma_1\sigma_2\, \text{Re}\left(\rho_{1,\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_{\text{o}})\, \rho^*_{2,\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_{\text{o}})\right) \right], \tag{11.7a}$$

$$\left\langle \frac{\partial T_{\alpha\beta}}{\partial \Omega_t}(\mathbf{p}\,|\,\mathbf{p}_0) \right\rangle_{\text{incoh}} = \left(\frac{\omega}{2\pi c}\right)^2 \frac{\epsilon_3^{3/2} \cos^2\theta_t}{\epsilon_1^{1/2} \cos\theta_0}\, g(\mathbf{p}-\mathbf{p}_0) \left[\sigma_1^2\, |\tau_{1,\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_{\text{o}})|^2 + \sigma_2^2\, |\tau_{2,\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_{\text{o}})|^2 \right.$$
$$\left. + 2\gamma\sigma_1\sigma_2\, \text{Re}\left(\tau_{1,\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_{\text{o}})\, \tau^*_{2,\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_{\text{o}})\right) \right], \tag{11.7b}$$

$$G_{jk}(\mathbf{p}-\mathbf{p}_0) = \left[\delta_{jk} + \gamma(1-\delta_{kj})\right] \sigma_j \sigma_k\, g(\mathbf{p}-\mathbf{p}_0)\,. \tag{11.7c}$$

Single scattering interpretation and optical paths model – We will now present a model for the scattering amplitudes where the notions of single scattering event and optical paths mentioned earlier takes a physically intuitive meaning. The model is based on the Fresnel amplitudes between two media, and the first order SAPT amplitudes between *two media* separated by a rough interface, i.e. those given by Eqs. (6.12) and (6.19). We believe that the presented model is equivalent to SAPT to first order for the whole system of three media separated by two rough interfaces as we have observed no numerical difference between the two models. The proof of the equivalence should be simple to derive but we have preferred a numerical illustration.

For the sake of simplicity, we focus the discussion for co-polarized reflection in the plane of incidence and we will therefore drop the polarization indices which could be either p or s. The model is built in the following way. Let $r_{ji}^{(\text{F})}(\mathbf{q})$ (resp. $t_{ji}^{(\text{F})}(\mathbf{q})$) denote the Fresnel reflection (resp. transmission) amplitude (for either p or s-polarized light depending on the considered polarization) for an incident plane wave with in-plane wave vector $\mathbf{q}$ reflected by a planar interface between media of dielectric constant ϵ_i and ϵ_j. The second index, i, denote the medium of incidence and the index j the substrate medium. Let us denote in a similar fashion the first order terms of the reflection (resp. transmission) amplitude given by SAPT, Eq. (6.12), (resp. Eq. (6.19)) for an elementary wave out-going with in-plane wave vector $\mathbf{p}$ given an incident in-plane wave vector $\mathbf{p}_0$, $r_{ji}(\mathbf{p}\,|\,\mathbf{p}_0)$ (resp. $t_{ji}(\mathbf{p}\,|\,\mathbf{p}_0)$), with the same convention for the indices as for the Fresnel amplitudes. An elementary scattering amplitude issued from a given elementary optical path is given by the product of the Fresnel and SAPT amplitudes according to the Fresnel reflection or refraction and the reflection or transmission scattering event which the path exhibits, and the product of the corresponding phase factor associated with the propagation of the path between two consecutive intersections. The overall reflection (resp. transmission) amplitude for the whole system is given by the sum of all the possible optical paths where we only allow for one single scattering event, i.e. a single SAPT reflection or transmission, and arbitrarily many Fresnel reflections and refractions (which are *not* considered as "scattering event"). Note that the intersections are assumed to all take place on the average surfaces. As an example, the path indicated in Fig. 11.1 leads to an elementary reflection amplitude

$$r_{\text{ex.path}} = t_{12}^{(\text{F})}(\mathbf{p})\, e^{i\varphi_r}\, r_{32}(\mathbf{p}\,|\,\mathbf{p}_0)\, e^{i\varphi_0}\, t_{21}^{(\text{F})}(\mathbf{p}_0)\,. \tag{11.8}$$

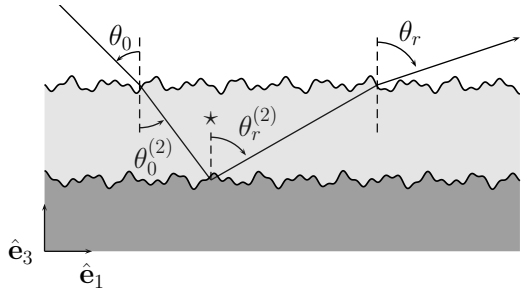


Figure 11.1: Example of optical path exhibiting a Fresnel refraction (from medium 1 to 2) through the first interface, a scattering event reflecting from the second interface and a Fresnel refraction (from medium 2 to 1) through the first interface. The angles θ_0 and θ_r are respectively the angle of incidence and of reflection. The angles $\theta_0^{(2)}$ and $\theta_r^{(2)}$ are respectively the angles associated to θ_0 and θ_r corresponding to Snell's law in medium 2. The star indicates the intersection at which the scattering event takes place.

We find it convenient to order the products from right to left following the sequence of events encountered during the path history. The phases φ_0 and φ_r are given by simple geometric consideration by

$$\varphi_0 = 2\pi\, d\, n_2 \cos\theta_0^{(2)}/\lambda \tag{11.9a}$$

$$\varphi_r = 2\pi\, d\, n_2 \cos\theta_r^{(2)}/\lambda\,, \tag{11.9b}$$

where λ denotes the wavelength in vacuum, n_2 is the refractive index of the film, and $\theta_0^{(2)}$ and $\theta_r^{(2)}$ are the angles conjugate to θ_0 and θ_r in Snell's law.

Remark 11.1. For a given surface profile, note that when considering a path scattered at the interface when the path is incoming with direction from $j+1$ to j, the x_3-axis hence becomes reversed in the convention adopted for Eqs. (6.12) and (6.19) compared with the case where the path is incoming from j to $j+1$. Thus the involved surface profile function must then changed to its opposite in Eqs. (6.12) and (6.19).

We are now ready to enumerate and categorize all the possible paths exhibiting a single scattering event and construct the overall reflection and transmission amplitudes. To follow easily the discussion, the reader may find helpful to refer oneself to Figs. 11.2(a-c) which illustrate the different types of optical paths. We have seen in Eq. (11.3) that the overall scattering amplitude can be written as the sum of the scattering amplitudes of the two associated subsystems where only one of the interfaces is planar. Hence for each of these subsystems, the possible paths exhibiting a single scattering event will be constraint to either exhibit a scattering event on the top surface (if the top surface is rough) or on the bottom surface (if the bottom surface is rough). We therefore analyze these two cases separately and we denote the reflection amplitude associated with the two subsystems respectively r_{RP} and r_{PR}, RP and PR standing for "rough-planar" (the top interface is rough) and "planar-rough" (the bottom interface is rough). Let us first analyze the possible single scattering optical paths in the configuration RP. Following Fig. 11.2(a), consider an optical path incoming with an angle θ_0 with respect to the x_3-direction. Such a path intersects the first interface, which is rough in the RP configuration. At the first

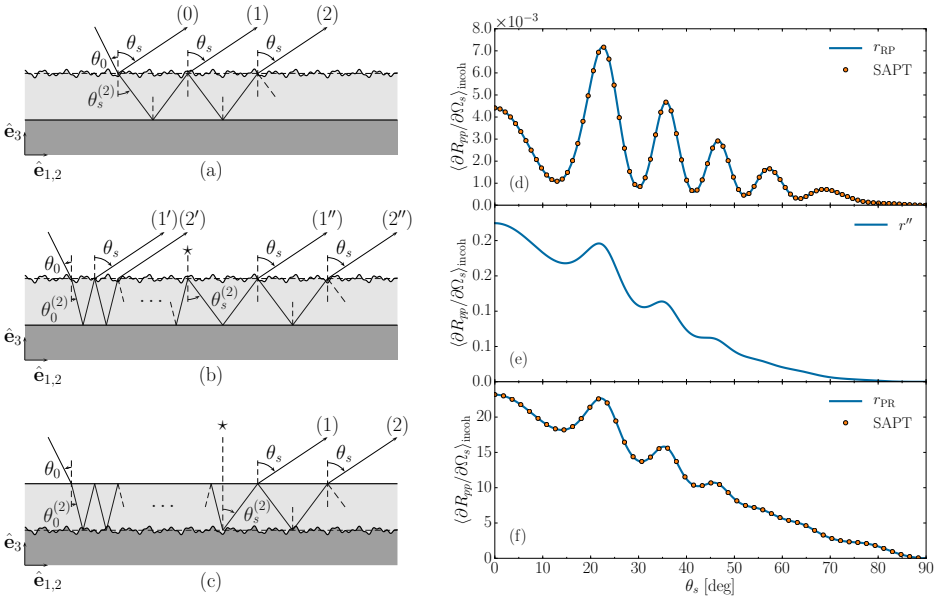


Figure 11.2: Sketch of the optical paths involved in the single scattering model in the case of scattering from the top surface (a) and (b), or from the bottom interface (c). Incoherent component of the MDRC for in-plane co-polarized scattering as a function of the polar angle of scattering for normal incidence for p-polarization (d) to (f). Physical parameters: $\epsilon_1 = 1$ (vacuum), $\epsilon_2 = 2.69$ (photo-resist), $\epsilon_3 = 15.08{+}0.15i$ (silicon substrate), $\lambda = 632.8$ nm, $\sigma = \lambda/30$ (rms-roughness of the rough interface), $a = \lambda/3$ (correlation length), $d = 8\lambda$ (average film thickness), $\theta_0 = 0°$. In panels (d) and (f), the results were obtained from SAPT (circles), and from the single scattering model Eqs. (11.13)(d) and (11.14)(e) (solid line) respectively for the cases illustrated in (a-b) and (c). In panel (e), only the contribution of r'' (Eq. (11.12)) to the incoherent component of the MDRC is shown. This figure is taken from **Paper** [3].

intersection, the path may exhibit a scattering event by reflecting with an angle of scattering θ_s[1] from the first interface. This is the path denoted (0) in Fig. 11.2(a) and the elementary reflection amplitude associated with this path is simply $r_{21}(\mathbf{p}\,|\,\mathbf{p}_0)$, where $\mathbf{p}$ is the in-plane wave vector associated with the scattering angle θ_s. Alternatively, the path may experience a scattering event and transmits inside the film. Then since the allowed single scattering event has taken place, all the remaining intersections with both interfaces will be of the Fresnel type, i.e. conserving the in-plane wave vector and contributing with the respective Fresnel amplitudes, hence bouncing with an angle $\theta_s^{(2)}$. The path hence may bounce an arbitrary number of times inside the film before exiting the film by a final Fresnel refraction[2]. These paths are denoted (1), (2), etc ... in Fig. 11.2(a). The path index hence counts the number of times the path as bounced on the bottom surface. Summing the elementary reflection amplitudes for the paths (0), (1), (2), etc ... we obtain the contribution of this first family of paths to the overall reflection amplitude $r_{\rm RP}$,

[1]The angle of scattering may be denoted θ_r, θ_t or simply θ_s in the following.

[2]Note, of course, that in principle the path may also be transmitted in medium 3, but then the path contributes to the transmission amplitude, and we are only analyzing the reflection amplitude here.

and it reads

$$r(\mathbf{p}\,|\,\mathbf{p}_0) = r_{21}(\mathbf{p}\,|\,\mathbf{p}_0) + t_{12}^{(\mathrm{F})}(\mathbf{p})\, r_{32}^{(\mathrm{F})}(\mathbf{p})\, t_{21}(\mathbf{p}\,|\,\mathbf{p}_0) \exp(2i\varphi_s) \sum_{n=0}^{\infty} \left[r_{12}^{(\mathrm{F})}(\mathbf{p})\, r_{32}^{(\mathrm{F})}(\mathbf{p})\, \exp(2i\varphi_s) \right]^n$$
$$= r_{21}(\mathbf{p}\,|\,\mathbf{p}_0) + \frac{t_{12}^{(\mathrm{F})}(\mathbf{p})\, r_{32}^{(\mathrm{F})}(\mathbf{p})\, t_{21}(\mathbf{p}\,|\,\mathbf{p}_0)\, \exp(2i\varphi_s)}{1 - r_{12}^{(\mathrm{F})}(\mathbf{p})\, r_{32}^{(\mathrm{F})}(\mathbf{p}) \exp(2i\varphi_s)}\,, \tag{11.10}$$

where we have recognized the sum of a geometric series. However, the attentive reader will object that these are not the only possible paths exhibiting a single scattering in the RP configuration. Indeed, the scattering event could occur at another intersection than the first one. Hence consider the family of path denoted (1'), (2'), etc ... in Fig. 11.2(b). For such a path, the path experiences first a Fresnel refraction at the first interface then may bounce an arbitrary number of Fresnel reflections inside the film, hence conserving its in-plane wave vector $\mathbf{p}_0$, i.e. angle of bounce inside the film $\theta_0^{(2)}$, before exiting the film with a final scattering event of transmission from medium 2 to medium 1. The index of the path for the primed paths hence counts the number of times the path experiences a Fresnel reflection on the bottom interface. The contribution to the reflection amplitude r_{RP} of the primed paths then reads

$$r'(\mathbf{p}\,|\,\mathbf{p}_0) = t_{12}(\mathbf{p}\,|\,\mathbf{p}_0)\, r_{32}^{(\mathrm{F})}(\mathbf{p}_0)\, t_{21}^{(\mathrm{F})}(\mathbf{p}_0) \exp(2i\varphi_0) \sum_{n=0}^{\infty} \left[r_{12}^{(\mathrm{F})}(\mathbf{p}_0)\, r_{32}^{(\mathrm{F})}(\mathbf{p}_0)\, \exp(2i\varphi_0) \right]^n$$
$$= \frac{t_{12}(\mathbf{p}\,|\,\mathbf{p}_0)\, r_{32}^{(\mathrm{F})}(\mathbf{p}_0)\, t_{21}^{(\mathrm{F})}(\mathbf{p}_0) \exp(2i\varphi_0)}{1 - r_{12}^{(\mathrm{F})}(\mathbf{p}_0)\, r_{32}^{(\mathrm{F})}(\mathbf{p}_0)\, \exp(2i\varphi_0)}\,. \tag{11.11}$$

There are still some paths to be enumerated. Indeed, consider now the paths denoted (1"), (2"), etc ... in Fig. 11.2(b). Let us consider a path following the initial history of a path of type prime, i.e. bouncing an arbitrary number of times inside the film with angle $\theta_0^{(2)}$. Instead of experiencing a scattering event on the top surface while transmitting, like for a primed path, a second (double primed) path experiences a scattering event while reflecting and continues to bounce an arbitrary number of times inside the film, with angle $\theta_s^{(2)}$, before exiting the film with a final Fresnel refraction in medium 1. In this case, the path index would correspond to the number of Fresnel reflections inside the film after the scattering event has occurred. However, this is of little use for the calculation since such a path correspond in fact to infinitely many paths having arbitrary number of Fresnel bounces *prior* to the scattering event. Nevertheless, the calculation of the contribution of the second paths to the reflection amplitude r_{RP} is relatively easy once we notice that it is the concatenation of paths of type (1'), (2'), ... before the scattering event and of paths of types (1), (2) ... after the scattering event. The corresponding reflection amplitude associated to the sum of all the paths (1"), (2"), etc ... thus reads

$$r''(\mathbf{p}\,|\,\mathbf{p}_0) = t_{21}^{(\mathrm{F})}(\mathbf{p}_0)\, r_{32}^{(\mathrm{F})}(\mathbf{p}_0)\, \exp(2i\varphi_0) \sum_{n=0}^{\infty} \left[r_{12}^{(\mathrm{F})}(\mathbf{p}_0)\, r_{32}^{(\mathrm{F})}(\mathbf{p}_0)\, \exp(2i\varphi_0) \right]^n$$
$$\times\, t_{12}^{(\mathrm{F})}(\mathbf{p})\, r_{32}^{(\mathrm{F})}(\mathbf{p})\, r_{12}(\mathbf{p}\,|\,\mathbf{p}_0)\, \exp(2i\varphi_s) \sum_{n'=0}^{\infty} \left[r_{12}^{(\mathrm{F})}(\mathbf{p})\, r_{32}^{(\mathrm{F})}(\mathbf{p})\, \exp(2i\varphi_s) \right]^{n'}$$
$$= \frac{t_{12}^{(\mathrm{F})}(\mathbf{p})\, r_{32}^{(\mathrm{F})}(\mathbf{p})\, r_{12}(\mathbf{p}\,|\,\mathbf{p}_0)\, r_{32}^{(\mathrm{F})}(\mathbf{p}_0)\, t_{21}^{(\mathrm{F})}(\mathbf{p}_0)\, \exp(2i(\varphi_0 + \varphi_s))}{\left[1 - r_{12}^{(\mathrm{F})}(\mathbf{p})\, r_{32}^{(\mathrm{F})}(\mathbf{p})\, \exp(2i\varphi_s)\right] \left[1 - r_{12}^{(\mathrm{F})}(\mathbf{p}_0)\, r_{32}^{(\mathrm{F})}(\mathbf{p}_0)\, \exp(2i\varphi_0)\right]}. \tag{11.12}$$

We have now enumerated all the possible single scattering paths whose scattering event occur on the top surface and their associated contribution to the overall reflection amplitude. We finally have

$$r_{\mathrm{RP}}(\mathbf{p}\,|\,\mathbf{p}_0) = r(\mathbf{p}\,|\,\mathbf{p}_0) + r'(\mathbf{p}\,|\,\mathbf{p}_0) + r''(\mathbf{p}\,|\,\mathbf{p}_0). \tag{11.13}$$

The PR configuration, i.e. the case where only the bottom interface is rough, can be analyzed in the same way. In fact, the PR configuration is relatively simpler in the sense that there is only one family of path to consider. Following Fig. 11.2(c) all the possible single scattering paths in the PR configuration are of a form that resembles that of the paths (1"), (2"), and so on from the RP configuration. To be more accurate, a single scattering path in the PR configuration must first exhibit a Fresnel refraction to enter the film (no scattering event allowed on the top interface) and may then undergo an arbitrary number of Fresnel bounces inside the film before a scattering event takes place in reflection on the bottom interface and then the path may undergo an arbitrary number of Fresnel bounces before eventually refracting into medium 1. The corresponding reflection amplitude is then given by

$$\begin{aligned} r_{\mathrm{PR}}(\mathbf{p}\,|\,\mathbf{p}_0) &= t_{21}^{(\mathrm{F})}(\mathbf{p}_0)\,\exp(i\varphi_0)\sum_{n=0}^{\infty}\Big[r_{12}^{(\mathrm{F})}(\mathbf{p}_0)\,r_{32}^{(\mathrm{F})}(\mathbf{p}_0)\,\exp(2i\varphi_0)\Big]^n \\ &\quad\times t_{12}^{(\mathrm{F})}(\mathbf{p})\,r_{32}(\mathbf{p}\,|\,\mathbf{p}_0)\,\exp(i\varphi_s)\sum_{n'=0}^{\infty}\Big[r_{12}^{(\mathrm{F})}(\mathbf{p})\,r_{32}^{(\mathrm{F})}(\mathbf{p})\,\exp(2i\varphi_s)\Big]^{n'} \\ &= \frac{t_{12}^{(\mathrm{F})}(\mathbf{p})\,r_{32}(\mathbf{p}\,|\,\mathbf{p}_0)\,t_{21}^{(\mathrm{F})}(\mathbf{p}_0)\,\exp(i(\varphi_0+\varphi_s))}{\Big[1-r_{12}^{(\mathrm{F})}(\mathbf{p})\,r_{32}^{(\mathrm{F})}(\mathbf{p})\,\exp(2i\varphi_s)\Big]\Big[1-r_{12}^{(\mathrm{F})}(\mathbf{p}_0)\,r_{32}^{(\mathrm{F})}(\mathbf{p}_0)\,\exp(2i\varphi_0)\Big]}. \end{aligned} \tag{11.14}$$

Now our claim is that $r_{RP}(\mathbf{p}\,|\,\mathbf{p}_0)$ and $r_{PR}(\mathbf{p}\,|\,\mathbf{p}_0)$ derived by considering all the possible single scattering diagrams are equal respectively to the first and second term in Eq. (11.3a) (assuming here co-polarized scattering in the plane of incidence, but this can easily be generalized to the whole cross-polarized and angular distribution). Let us verify this claim by plotting the different expressions. Figures 11.2(d) and (f) show the diffuse component of the mean differential reflection coefficient as a function of the angle of scattering respectively for the RP configuration and the PR configuration. The physical parameters considered are given in the figure caption. The solid curves are obtained based on the amplitude given by Eqs. (11.13) and (11.14) while the circles are obtained based on Eq. (11.3a) for the corresponding configurations. The agreement is perfect. In conclusion of this section, we can say that the scattering amplitudes given by SAPT to first order in the product of surface profiles correspond to the sum of the scattering amplitudes of all the possible elementary paths for which only a single scattering event is allowed, and the probability amplitude associated with this scattering event is given by SAPT to first order for the relevant system made of a single interface between two media. Although the expressions obtained with SAPT are compact, the equivalent expressions obtained by decomposing the different types of path will be useful for a careful analysis of the physical phenomena observed for the diffusely scattered light.

11.2 Physics of Selényi rings

We illustrate now the phenomenon of Selényi rings in the case of a photo-resist film deposited on a silicon substrate using the results presented in **Paper** [3]. The physical parameters used for the simulations were as indicated in the caption of Fig. 11.3. A point worth stressing is that

in order to observe Selényi rings, one needs a film thickness which is a few times the wavelength in the film medium.

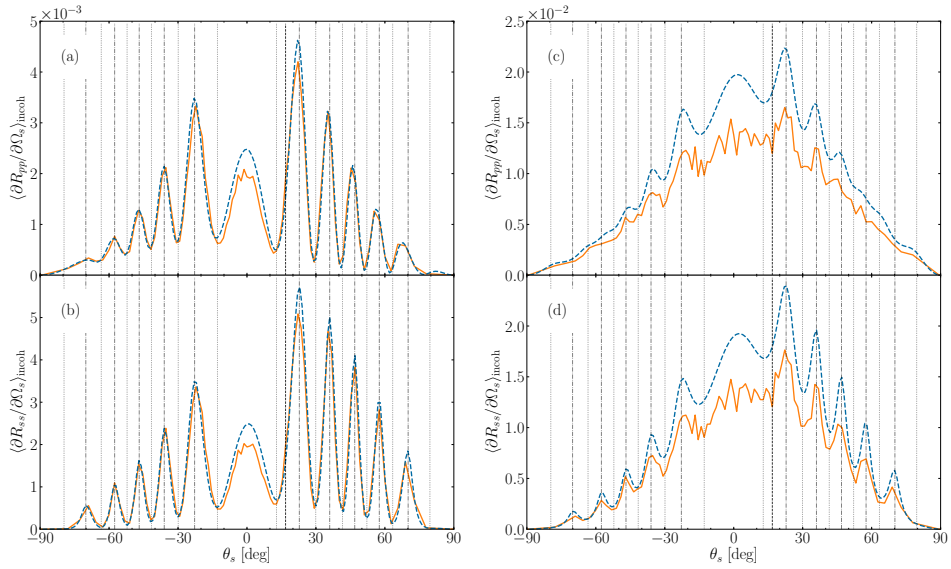


Figure 11.3: Incoherent components of the mean DRCs for in-plane co-polarized scattering as functions of the polar angle of scattering, θ_s (note the convention $\theta_s < 0$ for $\phi_s = \phi_0 + 180°$). The light of wavelength $\lambda = 632.8$ nm was incident from vacuum on the rough photoresist film supported by a silicon substrate ($\epsilon_1 = 1.0$, $\epsilon_2 = 2.69$, $\epsilon_3 = 15.08 + 0.15i$). The surface-height correlation length of the rough Gaussian correlated surface was $a = \lambda/3$, the mean film thickness was $d = 8\lambda$, and the angles of incidence were $(\theta_0, \phi_0) = (16.8°, 0°)$ in all cases. Panels (a) and (b) correspond to cases where only the top interface was rough, while panels (c) and (d) presents the results for a film where only the bottom interface of the film was rough. In both cases, the rms-roughness of the rough interface was set to $\sigma = \lambda/30$. The results obtained on the basis of the non-perturbative method are shown as solid lines while those obtained with the perturbative method, Eq. (11.7), are shown as dashed lines. The position of the specular direction in reflection is indicated by the vertical dashed lines. The vertical dash-dotted and dotted lines indicate the angular positions of the maxima and minima predicted by Eq. (11.16), respectively. This figure is taken from **Paper** [3].

Figure 11.3 shows the diffuse component of the mean differential reflection coefficient as a function of the angle of scattering for co-polarized scattering in the plane of incidence (the parameters are indicated in the figure caption). Figures 11.3(a-b) correspond to the RP configuration while Figs. 11.3(c-d) correspond to the PR configuration. The solid lines were obtained by the direct Monte Carlo method for the reduced Rayleigh equations derived for a film system where only one of the interfaces is rough. The numerically stable formulation in each case, RP or PR, as presented in Section 3.5.1 was used. The dashed lines are the corresponding results obtained with SAPT, Eq. (11.7a). The comparison between the Monte Carlo simulations

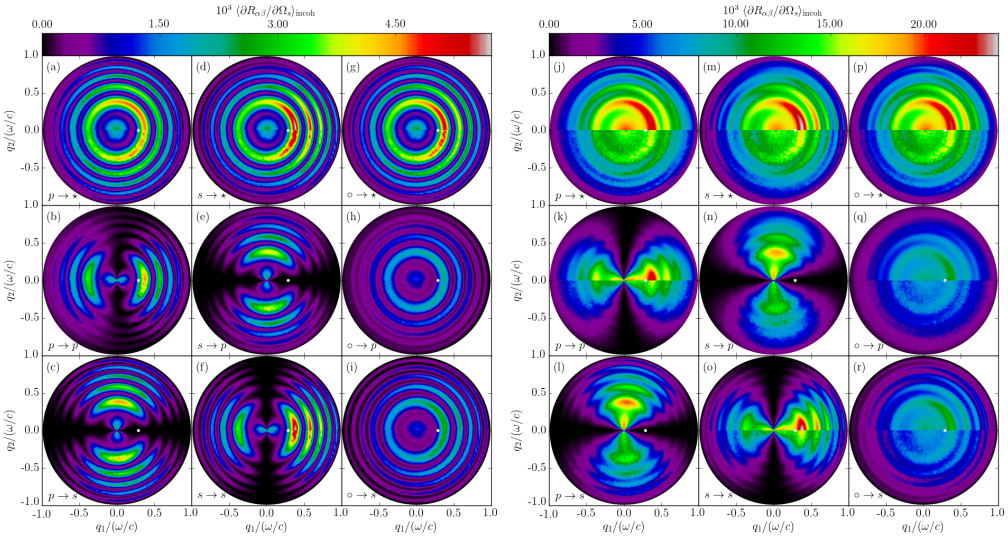


Figure 11.4: The full angular distribution of the incoherent component of the mean DRC, $\langle \partial R_{\alpha\beta}/\partial\Omega_s \rangle_{\rm incoh}$, as function of the lateral wave vector $\mathbf{q}$ of the light that is scattered from a rough film where either the top interface is rough (Figs. 11.4(a)–(i)) or the bottom interface is rough (Figs. 11.4(j)–(r)) and the other interface of the film is planar. The physical parameters are the same as those indicated in the caption of Fig. 11.3. The positions of the specular directions in reflection are indicated by white dots. The upper halves of all panels are results from the small amplitude perturbation method to leading order, while the lower halves show results obtained through the non-perturbative solutions of the RRE. The sub-figures in Figs. 11.4(a)–(i) and 11.4(j)–(r) are both organized in the same manner and show how incident β-polarized light is scattered by the one-rough-interface film geometry into α-polarized light (with $\alpha, \beta \in \{p, s\}$) and denoted $\beta \to \alpha$. Moreover, the notation $\circ \to \star$ is taken to mean that the incident light was unpolarized while the polarization of the scattered light was not recorded. For instance, this means that the data shown in Fig. 11.4(a) are obtained by adding the data sets presented in Figs. 11.4(b)–(c); similarly, the data shown in Fig. 11.4(g) result from the addition and division by a factor two of the data sets presented in Figs. 11.4(a) and 11.4(d); *etc.* Finally, the in-plane intensity variations from Figs. 11.4(b, f) and 11.4(k, o) are the curves depicted in Figs. 11.3(a)–(b) and Figs. 11.3(c)–(d), respectively. This figure is taken from **Paper** [3].

and first order SAPT shows a good quantitative agreement and the interference fringes are overall well resolved. This justifies the use of SAPT for the considered roughness level, and hence the use of a single scattering picture for interpreting the results. Let us now make some comments about the similarities and differences one can draw from a comparison between the two configurations, RP and PR. We will then explain in the following sections each of these observations. First, concerning the overall scale of the incoherent component of the MDRC, we can see that a larger power is scattered in the case where the rough interface is the bottom one than where it is the top one. This can easily be understood by the fact that the bottom interface is an interface between the photo-resist and the silicon substrate which corresponds to the largest jump in refractive index in the system. Consequently, the scattering is stronger

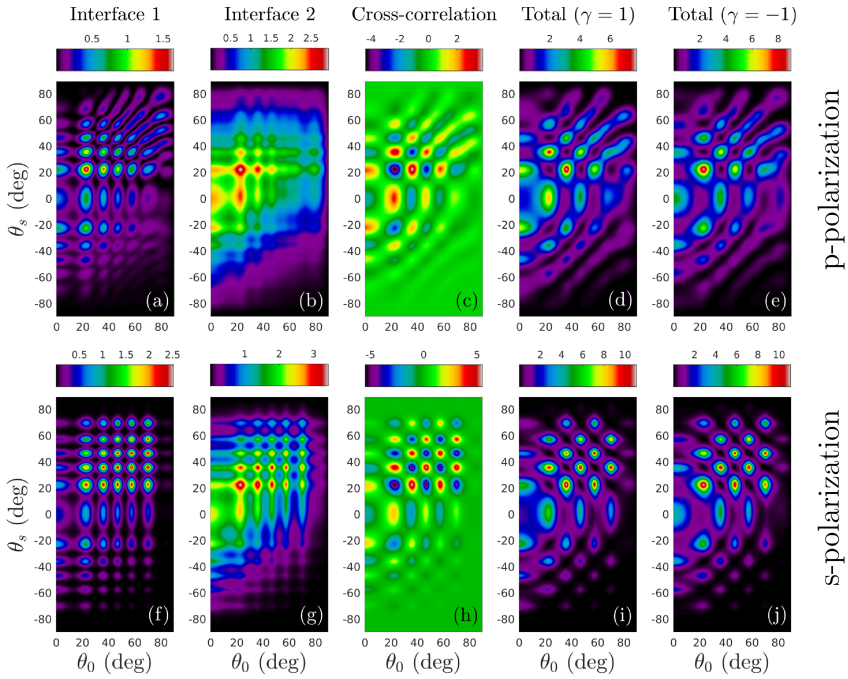


Figure 11.5: Scaled incoherent component of the mean DRCs for in-plane co-polarized scattering, $100 \times \langle \partial R_{\alpha\alpha}/\partial\Omega_s \rangle_{\text{incoh}}$, as functions of the polar angle of incidence θ_0 and the polar angle of scattering θ_s obtained on the basis of Eq. (11.7). The first row of sub-figures (Figs. 11.5(a)–(e)) corresponds to *p*-polarized light (as marked in the figure), while the second row (Figs. 11.5(f)–(j)) corresponds to *s*-polarized light. Apart from the varying angle of incidence, all physical parameters are those indicated in the caption of Fig. 11.3. The first column of sub-figures presents contour plots of the mean DRCs for a film geometry where only the top interface of the film is rough and the bottom interface planar. The second column shows similar results where the top film interface is planar and the bottom film interface is rough. In the third column, contour plots of *only* the cross-correlation term in Eq. (11.7) — that is, the contribution to the mean DRC produced by the last term in the square brackets of this equation — are depicted assuming a perfect correlation ($\gamma = 1$) between the rough top and rough bottom interface of the film. Finally, in the forth and fifth column, contour plots of the total mean DRCs obtained on the basis of Eq. (11.7) are presented for two-rough-interface film geometries characterized by $\gamma = 1$ and $\gamma = -1$, respectively. This figure is taken from **Paper** [3].

at this interface. Second, we observe that the interference fringes are located at the same angular positions in both configurations, and we will show in Section 11.2.1 that the positions of the local maxima and minima can be predicted quite accurately based on simple phase difference between optical paths arguments. Third, the contrast of the fringes is higher for the RP configuration than for the PR configuration. At first, one may think that the poor contrast in the PR configuration comes from the fact that the scattering event occurs on the interface between the media exhibiting a strong jump of refractive index. We will see that this is not really the case in Section 11.2.2, where we will discard this possible explanation by considering

a free standing dielectric film in vacuum and show that the contrast depends essentially on the side where the rough interface is with respect to the observed light. This curious effect will be explained in details based on a careful analysis of the types of optical path involved in each case. Finally, the expert eye will notice a small shift of the fringe position for p-polarized light compared with s-polarized light at large angles of scattering in Figs. 11.3(a-b). This effect is due to the Brewster scattering phenomenon and will be illustrated in a clearer way for a dielectric system without loss and with a substrate with a smaller dielectric constant in Section 11.2.3.

Figure 11.4 shows the full angular distributions of the incoherent component of the mean differential reflection coefficient corresponding to the simulations of Fig. 11.3. Note that the upper half of each contour plot presents first order SAPT data while the bottom half presents the Monte Carlo simulation in view of a visual comparison. We can appreciate that what we have called interference fringes in the scattering in the plane of incidence are cuts of interference rings in the full angular distribution.

Let us now study the behavior of the interference pattern as the angle of incidence is varied. Figures 11.5(a,b,f,g) show contour plots of the incoherent component of the MDRC as a function of the angles of incidence and scattering for co-polarized scattering in the plane of incidence for p and s-polarized light both in the case where the first interface is rough (PR) and the second interface is rough (RP) (see figure caption for the details). Each vertical cut in such a contour plot hence represents a plot such as the ones presented in Fig. 11.3 for a given angle of incidence. We can see that for s-polarized light the positions of the local maxima in the scattering distribution are independent of the angle of incidence. In addition, the overall scattered intensity is modulated with the angle of incidence in a similar interferential fashion. The same can be said for p-polarized light up to the region of large angles of scattering and angles of incidence where effects related to the Brewster scattering phenomenon must be taken into account and will be discarded at first in our discussion. The extrema of intensity in the scattering angular distribution are then independent of the angle of incidence, but the overall intensity is modulated by the latter.

11.2.1 Origin of the interference pattern

It is now time to explain our observations on the angular position of the fringes, or rings. To this end, we come back to the different types of paths found in Section 11.1 and discuss the difference in optical path length between consecutive paths of a given family. Consider first the RP configuration, and the family of paths denoted (n) (see Fig. 11.2(a)). The phase shift acquired by such a path after each bounce inside the film was found to be $2\varphi_s$. Equivalently, this means that the difference in optical paths length between two consecutive paths (n) and $(n+1)$ can be expressed as

$$\Delta = 2n_2\, d \cos\theta_s^{(2)}, \tag{11.15}$$

where θ_s in the vacuum is related to $\theta_s^{(2)}$ in the film by $n_2 \sin\theta_s^{(2)} = n_1 \sin\theta_s$ according to Snell's law. The polar angles of scattering for which the diffusely scattered intensity (associated with the paths family (n) i.e. proportional to $|r(\mathbf{p} \mid \mathbf{p}_0)|^2$) has local minima are given by

$$\frac{2\pi n_2 d}{\lambda}\, \cos\theta_s^{(2)} = \frac{2\pi d}{\lambda}\, \left(\epsilon_2 - \epsilon_1 \sin^2\theta_s\right)^{1/2} = (\nu + 1/2)\pi, \tag{11.16a}$$

while the positions of the maxima are determined from the relation

$$\frac{2\pi n_2 d}{\lambda}\,\cos\theta_s^{(2)} = \frac{2\pi d}{\lambda}\,\left(\epsilon_2 - \epsilon_1 \sin^2\theta_s\right)^{1/2} = \nu\pi, \tag{11.16b}$$

where[3] $\nu \in \mathbb{Z}$. It is then clear that a necessary condition to observe interference fringes in the diffusely scattered light, is that the thickness of the film must be large enough to contain a few wavelengths λ/n_2 in medium 2. The angular positions of the maxima and minima predicted by Eq. (11.16) are indicated by vertical dash-dotted and dotted vertical lines, respectively, in Fig. 11.3. The predicted positions agree well with the maxima and minima that can be observed in the intensity distributions. Equation (11.16) does not depend on the polar angle of incidence θ_0, which supports the observation that the positions of the maxima and minima of the incoherent components of the mean DRC do not move with the angle of incidence for weakly rough films. However, the modulation of the fringes with the angle of incidence cannot be explained if we consider solely the paths of type (n). Analyzing now the paths of types (n') (see Fig. 11.2(b)), the phase shift acquired by the wave after each bounce inside the film is $2\varphi_0$, i.e. that the difference of optical path length between two consecutive paths (n') and $(n+1')$ is given by

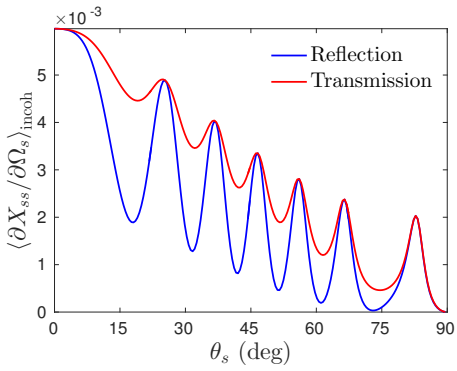
$$\Delta = 2n_2\, d\cos\theta_0^{(2)}, \tag{11.17}$$

where $n_2 \sin\theta_0^{(2)} = n_1 \sin\theta_0$ according to Snell's law. Hence, we again obtain a series of maxima and minima in the mean DRC if we replace $\theta_s^{(2)}$ by $\theta_0^{(2)}$ in Eq. (11.16), but this time the positions of the maxima and minima are indeed a function of the polar angle of incidence θ_0. This interference phenomenon serves to modulate the intensity of the Selényi interference patterns. Finally, the paths of type (n'') have the characteristic that they may exhibit both a dependence on the angle of incidence and the angle of scattering in the phase difference between two consecutive paths, depending on how one defines two *consecutive* paths in this case. This also holds for all paths involved in the PR configuration, and it is shown to exhibit the same grid of extrema as the RP configuration in Fig. 11.5, showing the intrinsic modulation as a function of both the angles of incidence and scattering. We will see in the next section the importance of these last type of paths for explaining the difference of contrast between the two configurations.

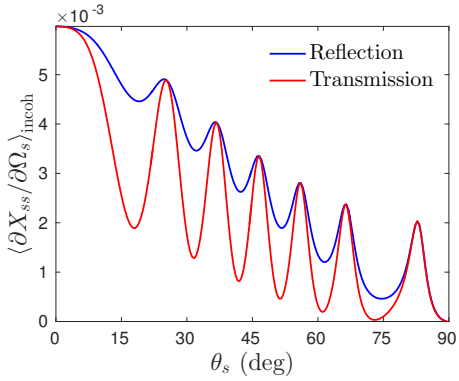
11.2.2 Ring contrast

Let us now explain the difference in ring contrast observed in Figs. 11.3 and 11.4 between the configurations for which either the top or bottom interface is rough. In order to clarify that this effect is *not* particularly due to an argument related to the dielectric contrast between the media around the rough interface, consider instead a dielectric film, of dielectric constant $\epsilon_2 = 2.25$, free standing inside vacuum $\epsilon_1 = \epsilon_3 = 1$. The remaining physical parameters are kept identical to those used previously in Figs. 11.3 and 11.4 and are given in the caption of Fig. 11.6. Figure 11.6 presents the angular distribution of the diffuse component of the mean reflection and transmission coefficients for *s*-co-polarized scattering in the plane of incidence for both configurations RP (Fig. 11.6(a)) and PR (Fig. 11.6(b)). What is striking by comparing Figs. 11.6(a) and 11.6(b), is that the behavior of the reflected intensity and transmitted intensity are interchanged for the two configurations. In particular, when the top interface is rough, the diffuse intensity of the reflected light exhibits a high fringe contrast while the diffuse intensity

[3] Of course only solutions for which the angle is real give the positions observed in the intensity.



(b) Top interface rough and bottom interface planar (RP).



(c) Top interface planar and bottom interface rough (PR).

Figure 11.6: Incoherent component of the mean differential reflection and transmission coefficients for s-co-polarized scattering in the plane of incidence as a function of scattering angle. The system is a free standing dielectric film in vacuum and the physical parameters are: $\epsilon_1 = \epsilon_3 = 1$, $\epsilon_2 = 2.25$, $a = \lambda/3$, $\sigma = \lambda/30$, $d = 8\lambda$, $\theta_0 = 0°$. (a) The top interface is rough and the bottom interface is planar. (b) The top interface is planar and the bottom interface is rough.

of the transmitted light exhibits a low fringe contrast. Contrarily, when the bottom interface is rough, it is the diffuse intensity of the transmitted light which exhibits a high fringe contrast while the diffuse intensity of the reflected light exhibits a low fringe contrast. We can then conjecture that the fringes will be seen with the highest contrast if they are observed for the light scattered in the medium which is on the side of the rough interface.

Let us now explain the reason behind this curious effect. For this let us focus on reflection and compare the type of paths involved respectively in the two configurations RP and PR. According to Section 11.1, there were essentially three types of paths involved in the scattering of light under the single scattering event approximation. These are depicted in Figs. 11.2(a-b) and denoted respectively in the form (n), (n'), (n''). We have seen that the paths of type (n) all acquire a phase shift which is a multiple of $2\varphi_s$ as the path bounces inside the film, and we have seen that these are essentially at the origin of the interference rings as a function of the angle of scattering. The paths of type (n') all acquire a phase shift which is multiple of $2\varphi_0$, and we have seen that these are essentially responsible for the modulation of the overall intensity as the angle of incidence varies. The paths of type (n'') all acquire a phase shift which is *an integer combination* of φ_0 and φ_s. Figure 11.2(e) shows the intensity corresponding to r'' as a function of the angle of scattering, i.e. as if only the paths of type (n'') where taken into account. We observe that compared with the full result (see Fig. 11.2(d)), the contrast is lower and is very similar to that observed for the PR configuration as seen in Fig. 11.2(f). Indeed, the paths of type (n'') in the RP configuration are very similar to those in the PR configuration. The poor contrast is therefore due to the nature of the paths. Why are not the paths of type (n'') degrading the contrast in the RP configuration then? In fact they do, but moderately since, as can be seen by comparing Figs.11.2(d) and 11.2(e), the contribution of r'' is weak compared with the other type of paths. What is the fundamental reason for the low contrast associated with the type of paths (n'') or those of the PR configuration? It is fundamentally the *phase mixing*, resulting from the integer combination of φ_0 and φ_s acquired during the propagation inside the

film which causes the poor contrast. Indeed allowing for combination of several phase-shift units has a tendency to degrade the coherence between the consecutive paths. This is similar to summing random phasors with uncorrelated phase, resulting in the absence of interference in average.

We can make the above assertion more precise in mathematical terms by deriving estimates for the contrast obtained for the different types of paths taken separately, and show that the type of paths involving phase mixing intrinsically leads to poorer contrast than the types of path involving only one type of phase. Consider then, as a prototypical reflection amplitude for a sum of paths that involves phase mixing and a sum of paths that does not (and will serve as reference), the following expressions

$$r_{\mathrm{mix}\varphi} = \frac{\tilde{r}}{[1 - r_0 \exp(2i\varphi_0)]\ [1 - r_s \exp(2i\varphi_s)]} \tag{11.18a}$$

$$r_{\mathrm{ref}} = \frac{\tilde{r}}{1 - r_s \exp(2i\varphi_s)}. \tag{11.18b}$$

These reflection amplitudes mimic the structures from Eqs. (11.14) and Eq. (11.10) respectively, but we will see that the precise expressions for the numerators do not matter for the contrast, and are hence denoted by the same symbol $\tilde{r}$. Note that all the reflection amplitudes in Eq. (11.18) depend on angles of incidence and scattering, but for clarity we drop these arguments. Since we are interested in the contrast for the fringes in the intensity, our first step consists in taking the square modulus of Eq. (11.18)

$$I_{\mathrm{mix}\varphi} = \frac{|\tilde{r}|^2}{|1 - r_0 \exp(2i\varphi_0)|^2\ |1 - r_s \exp(2i\varphi_s)|^2} \tag{11.19a}$$

$$I_{\mathrm{ref}} = \frac{|\tilde{r}|^2}{|1 - r_s \exp(2i\varphi_s)|^2}\,. \tag{11.19b}$$

We then bound the intensity by using the triangular inequality

$$\frac{|\tilde{r}|^2}{(1 + |r_0|)^2\,(1 + |r_s|)^2} \quad \leq \quad I_{\mathrm{mix}\varphi} \quad \leq \quad \frac{|\tilde{r}|^2}{(1 - |r_0|)^2\,(1 - |r_s|)^2} \tag{11.20a}$$

$$\frac{|\tilde{r}|^2}{(1 + |r_s|)^2} \quad \leq \quad I_{\mathrm{ref}} \quad \leq \quad \frac{|\tilde{r}|^2}{(1 - |r_s|)^2}. \tag{11.20b}$$

It is clear from Eq. (11.20) that the intensity lies between two bounding curves. A fair estimate for the trend, i.e. the intensity without the oscillations would be given by $|\tilde{r}|^2$, and we thus estimate, or rather define, the *inverse contrast* as

$$\eta^{-1}_{\mathrm{mix}\varphi} = (1 + |r_0|)^2\,(1 + |r_s|)^2 - (1 - |r_0|)^2\,(1 - |r_s|)^2 \tag{11.21a}$$

$$\eta^{-1}_{\mathrm{ref}} = (1 + |r_s|)^2 - (1 - |r_s|)^2. \tag{11.21b}$$

This may not be the most *natural* definition for the contrast, but we choose this one since it is easier to work with and will not change the conclusion. By recasting Eq. (11.21) by the use of straightforward algebra, we obtain

$$\eta^{-1}_{\mathrm{mix}\varphi} = 4|r_s| + 4|r_0| + 4|r_0||r_s| + 4|r_0|^2|r_s| \tag{11.22a}$$

$$\eta^{-1}_{\mathrm{ref}} = 4|r_s|. \tag{11.22b}$$

This shows that the inverse contrast for phase mixing is larger than that of the reference, i.e. that the *contrast* in the case of phase mixing is smaller than that of the reference. Indeed, the two last terms in Eq. (11.22a) are cross-terms resulting directly from the phase mixing nature of the paths encoded in the reflection amplitude. Note that the choice for the reference was arbitrary and one could choose to study paths of type (1'), (2'), *etc.*, in Fig. 11.2(b), and hence replace $r_s \exp(2i\varphi_s)$ in Eq. (11.18) by $r_0 \exp(2i\varphi_0)$, and the conclusion would still hold.

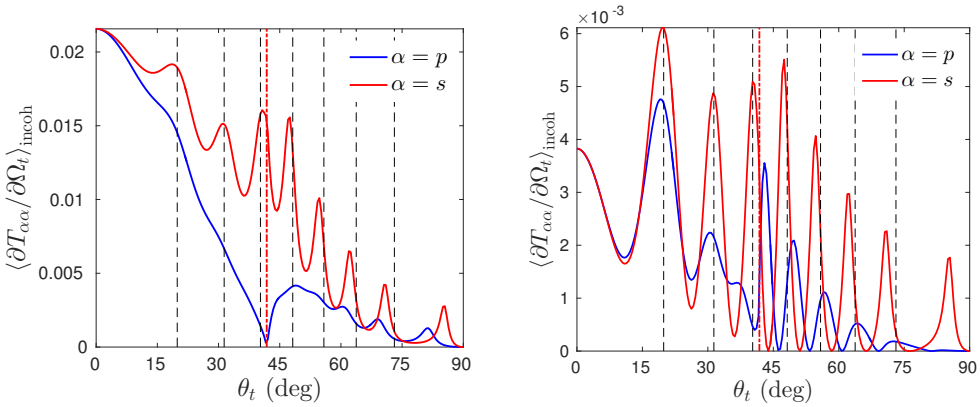
11.2.3 Influence of the Brewster and Yoneda effects on Selényi rings

We have seen in Chapter 10 that for the scattering of light by a rough interface between two dielectric media, the phenomena of Brewster scattering and of the Yoneda ring come with a phase jump at the Brewster scattering angle and a continuous phase shift above the critical angle. We illustrate now the influence these phase behaviors have on the positions of the Selényi rings. We restrict ourselves here to a simple illustration for co-polarized scattering in the plane of incidence, and consider the cases where either only the top interface is rough (RP) or only the bottom interface is rough (PR). The aim of this section is to observe these effects and to give an intuitive explanation. A more detailed analysis could be the object of a devoted research article.

Consider a film of dielectric constant $\epsilon_2 = 3.5$ deposited on a glass substrate of dielectric constant $\epsilon_3 = 2.25$ (and the host medium is vacuum $\epsilon_1 = 1$). The film thickness is chosen to be equal to $d = 5\lambda$, the rms-roughness and the correlation length of the rough interface are $\sigma = \lambda/30$ and $a = \lambda/3$ respectively. Figure 11.7 presents the incoherent component of the mean differential transmission coefficients as a function of the angle of scattering for the two configurations RP and PR. For s-polarized light, we can see that the fringes seem to be bounded by an eveloppe which exhibits a peak around the critical angle $\theta_c = \arcsin(n_1/n_3)$ as indicated by the vertical red dash-dotted line in Figs. 11.7(a) and 11.7(b). For p-polarized light, the intensity seems to go to zero at the critical angle in the RP configuration, while it remains positive in the PR configuration at the same angle. In the PR configuration, it is not clear whether there is a clear zero of intensity due to the Brewster scattering effect at a higher angle of scattering or if it is due to the interference mechanism. A more detailed analysis is then necessary to understand the difference between the two cases.

Let us now take a closer look at the position of the peaks. First, we can see that the local maxima of intensity for p- and s-polarized light coincide for angles of scattering smaller than the critical angle but are shifted with respect to each other for angles of scattering larger than the critical angle. This effect is particularly well visible in the PR configuration. We note that in this configuration, there is almost a flip of maximum and minimum for p-polarized light around the critical angle. We believe this to be due to the Brewster scattering phase jump (see Chapter 10). Indeed, when crossing the Brewster scattering angle, a path may acquire an additional phase shift compared to what would be expected with the difference of optical paths length argument, hence the corresponding flip of maxima and minima. Furthermore, focusing our attention on s-polarized light for clarity, we can see that the position of the local maxima agree well with the positions predicted by Eq. (11.15)[4], indicated by the vertical dashed lines, before the critical angle, but gradually shifts from the expected positions as the angle of scattering is increased beyond the critical angle. We believe this effect to be a signature of the continuous phase shift

[4]But one must change ϵ_1 by ϵ_3 since light is observed in transmission.



(a) RP (b) PR

Figure 11.7: Incoherent components of the mean differential transmission coefficients $\langle \partial T_{\alpha\alpha}/\partial\Omega_t \rangle_{\text{incoh}}$ as a function of the angle of scattering for in-plane co-polarized scattering from a film geometry for the polar angle of incidence $\theta_0 = 0°$. The wavelength of the incident light was $\lambda = 632\,\text{nm}$, the mean thickness of the film $d = 5\lambda$, and the dielectric constants of the media were $\epsilon_1 = 1.0$, $\epsilon_2 = 3.5$, $\epsilon_3 = 2.25$. The rms-roughness of the rough interface was $\sigma = \lambda/30$, and the correlation length was $a = \lambda/3$. (a) Top surface rough and bottom surface planar (RP). (b) Top surface planar and bottom surface rough (PR). The black dashed lines are the positions of the fringe maxima as predicted with a difference of optical paths length argument, Eq. (11.15), with ϵ_1 replaced by ϵ_3. The red dash-dotted line indicate the critical angle $\theta_c = \arcsin(n_1/n_3)$.

occurring upon the Yoneda phenomena as demonstrated in Chapter 10.

We would like to stress that these results illustrate that the Brewster phase jump and Yoneda phase shift influence the behavior of the Selényi rings. A more detailed analysis should be carried out to explain unambiguously the full mechanisms at play by combining the knowledge acquired both in Chapter 10 and the optical path analysis in Section 11.1. We can nevertheless conclude that a scattering experiment on such a sample could permit to measure the phase shifts predicted by the theory of the Brewster scattering and Yoneda phenomena based on the fact that such a rough film system acts both as a scattering sample and an interferometer. An experimental demonstration of these effects would be welcome as it could confirm or complement the theory developed in this thesis.

Moreover we would like to mentioned that more effects can be observed by playing with the dielectric constants in order to play with effects due to total internal reflection. Then effects due to guided modes can be observed too, and some configurations of dielectric constants can lead to intensity patterns with forbidden scattering or with regions with interference rings within a cone and without interference rings outside a cone. We leave the thorough discussion of the complete zoology of effects one may observe for Selényi rings to a devoted research article.

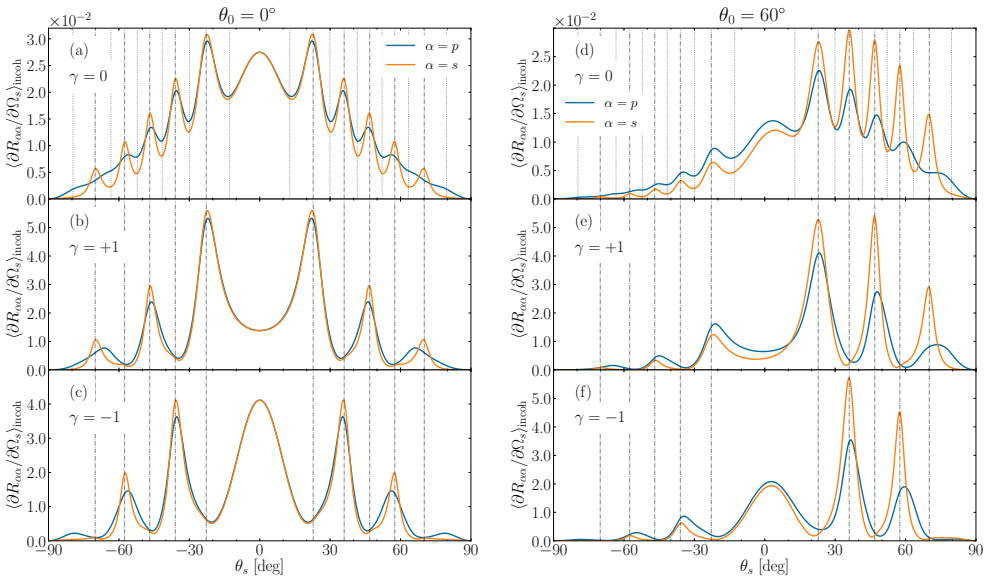


Figure 11.8: Incoherent components of the mean differential reflection coefficients $\langle \partial R_{\alpha\alpha}/\partial\Omega_s \rangle_{\text{incoh}}$ for in-plane co-polarized scattering from a two-rough-interface film geometry for the polar angle of incidence $\theta_0 = 0°$ (Figs. 11.8(a)–(c)) and $\theta_0 = 60°$ (Figs. 11.8(d)–(e)). The wavelength of the incident light was $\lambda = 632.8$ nm, the mean thickness of the film $d = 8\lambda$, and the dielectric constants of the media were $\epsilon_1 = 1.0$, $\epsilon_2 = 2.69$, $\epsilon_3 = 15.08 + 0.15i$. The rms-roughness of the interfaces were $\sigma_1 = \sigma_2 = \lambda/30$, and the Gaussian correlation functions of each of the surfaces were characterized by the correlation length $a = \lambda/3$. The cross-correlation function between the rough top and rough bottom interface of the film had the form Eq. (11.2) and was characterized by the parameter γ with values as indicated in each of the panels. The vertical dash-dotted and dotted lines indicate the expected angular positions of the maxima and minima of the scattered intensity as predicted by Eq. (11.16), respectively. For reasons of clarity only the expected positions of the maxima of the in-plane mean DRCs are indicated in Figs. 11.8(a) and 11.8(d). Figure taken from **Paper** [3]

11.3 Cross-correlation induced ring intensity enhancement and attenuation

So far we have restricted our study to the configurations RP and PR for the sake of simplicity and for illustrating the differences between the two configurations. It is now time to treat the case where the two interfaces are rough. We return to the system of the photo-resist film deposited on the silicon substrate studied previously. The two rough interfaces are then characterized by the same auto-correlation and with equal rms-roughness $\sigma_1 = \sigma_2 = \lambda/30$, all other parameters being kept identical as in the cases RP and PR studied previously.

For the case where the two interfaces are uncorrelated, i.e. when the cross-correlation parameter is $\gamma = 0$, we have seen in Section 11.1 that the intensity of the scattered light is the sum of the

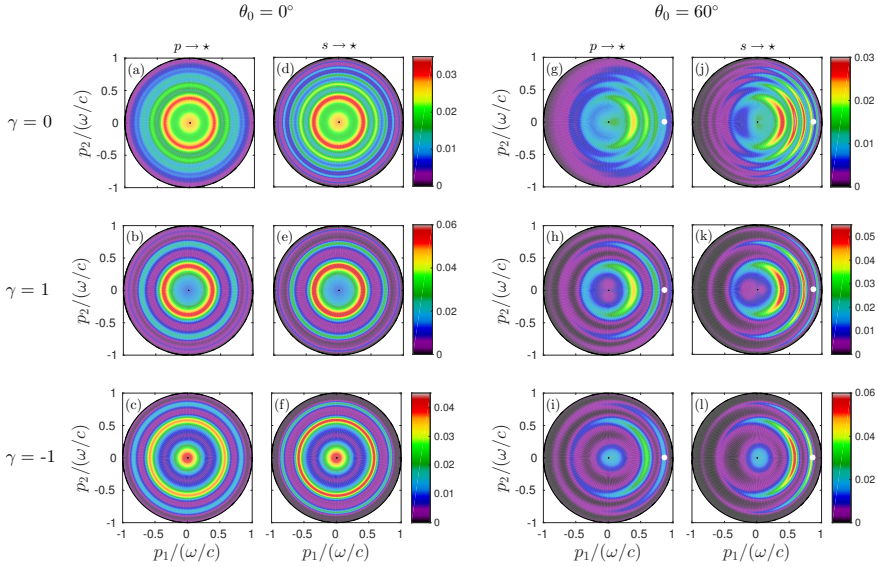


Figure 11.9: The full angular distribution of the incoherent component of the mean DRC, $\langle \partial R_{\alpha\beta}/\partial\Omega_s \rangle_{\text{incoh}}$, for incident β-polarized light that is scattered by a two-rough-interface film geometry into α-polarized light ($\alpha, \beta \in \{p, s\}$). When the polarization of the scattered light is not observed, the relevant mean DRC quantity is $\sum_{\alpha=p,s} \langle \partial R_{\alpha\beta}/\partial\Omega_s \rangle_{\text{incoh}}$ and this situation is labeled as $\beta \to \star$. The reported results were obtained on the basis of SAPT, Eq. (11.7), and the polar angles of incidence were $\theta_0 = 0°$ ((a)–(f)) and $\theta_0 = 60°$ ((g)–(l)). The incident in-plane wave vector is indicated by the white dot for non-normal incidence ((g)–(l)). The cross-correlation function between the rough top and rough bottom interface of the film had the form Eq. (11.2) and was characterized by the parameter γ as indicated in the figure (and constant for each row of sub-figure). The remaining roughness parameters are identical to those assumed in producing the results presented in Fig. 11.8. Figure adapted from **Paper** [3].

intensities of the two associated subsystems where only one of the interfaces is rough at a time[5]. The corresponding diffuse component of the mean differential reflection coefficients are shown for co-polarized scattering in the plane of incidence in Figs. 11.8(a) and 11.8(d) for normal incidence and $\theta_0 = 60°$ incidence respectively. We can see that, as the sum of two interference patterns of intensity with the same fringe positions, the resulting pattern also exhibits fringes at the same positions as for the two subsystems. In addition, the contrast is poor since one of the two subsystems has a poor fringe contrast, and, in this particular case, it happens that the subsystem with the rough interface on the substrate dominates the signal.

When now the interfaces are correlated, the last term in Eq. (11.7) may contribute either negatively or positively to the intensity. Note that the resulting intensity remains non-negative in virtue of $2\text{Re}(zz'^*) \leq |z|^2 + |z'|^2$. The contribution of this last term is plotted in the (θ_0, θ_s)-plane in Figs. 11.5(c, h), where we can appreciate that the extrema of the correlation induced

[5]Let us stress again that this is true only in the single scattering approximation.

term takes positive *and* negative values on the same grid as the maxima of the two independent subsystems (see Figs. 11.5(a, b, f, g)). This means that when summed to the intensities of the two subsystems, the cross-correlation term may enhance some fringes and attenuate some other with an efficiency that depends on the cross-correlation parameter γ. Figures 11.5(d, i) show the resulting total intensity when $\gamma = 1$, i.e. when the two surface profiles are perfectly positively correlated, while Figs. 11.5(e, j) show the resulting total intensity when $\gamma = -1$, i.e. when the two surface profiles are perfectly negatively correlated. When can see that by adding ($\gamma = 1$) or subtracting ($\gamma = -1$) the cross-correlation term to the sum of intensities of the two subsystems RP and PR, half of the fringes disappear while the other half are enhanced. The fringes being enhanced or attenuated depend on the sign of γ. This is illustrated for co-polarized scattering in the plane of incidence in Figs. 11.8(b, e) for $\gamma = 1$ and Figs. 11.8(c, f) for $\gamma = -1$. We observe that the cross-correlation induced interference selectively enhances and attenuates every other fringe, depending on the sign of the correlation parameter. Note that this effect also consequently enhances the contrast. Figure 11.9 shows the corresponding full angular distribution of the diffuse component of the mean differential reflection coefficients for scattering from incident *p*- or *s*-polarized light and detecting the intensity independently of polarization. The three cases of uncorrelated ($\gamma = 0$), positively correlated ($\gamma = 1$) and negatively correlated ($\gamma = -1$) interfaces are shown, and we can appreciate by comparison with the uncorrelated case the enhancement and attenuation of half of the rings, which are selectively attenuated or enhanced depending on the sign of γ. For intermediary values of $-1 < \gamma < 1$ the same behavior is observed but with a weaker enhancement and attenuation efficiency. For $0 < \gamma < 1$ the intensity pattern will be somewhat between those for $\gamma = 0$ and $\gamma = 1$.

11.4 Quételet rings and multiple scattering effects

A natural question after such a study is the following. What happens for rougher surfaces? When the surfaces become rougher, we expect that multiple scattering paths will contribute significantly to the scattering amplitudes. It was shown by Lu *et al.* based on one-dimensional simulations of the reflected light scattered by a dielectric film on a perfectly conducting surface, that the intensity experiences three regimes with increasing roughness [116]. In the weakly rough regime, Selényi fringes are observed with similar interference patterns as those presented in the present chapter. For slightly rougher surfaces, interference fringes are also observed but with a different behavior than the behavior of Selényi fringes with respect to the angle of incidence. It was shown that the position of the fringes in the scattering distribution vary with the angle of incidence. It was motivated that the relevant paths contributing to such a behavior was a single scattering path of the form shown in the PR configuration (see Fig. 11.2(c)), and mixing the phase φ_0 and φ_s. Assuming only one segment down and one segment up inside the film, the corresponding phase shift is $\varphi_0 + \varphi_s$, the associated optical path length shift is

$$\Delta = n_2\, d\big(\cos\theta_0^{(2)} + \cos\theta_s^{(2)}\big). \tag{11.23}$$

Such an optical path length shift could adequately describe the dependence of the fringe positions on the angle of incidence. Such interference fringes, or rings, are known as Quételet rings [116, 118, 119, 120]. We speculate that as the roughness is increased, the dominating single scattering paths in the weakly rough case, are those essentially giving either a phase shift being a multiple of φ_0 or φ_s and the paths mixing exactly an equal amount of φ_0 and φ_s are negligible. When multiple scattering starts to play a role, the single scattering paths have less chance of staying ... single scattering paths, or in other words their contribution decays. We believe that

the paths that mixes equally φ_0 and φ_s are less affected than other and becomes of the same order as the single scattering paths which were the dominant ones for weakly rough surfaces, and the Quételet rings are observed. For very rough surfaces, multiple scattering paths dominate inducing a loss of coherence and destroying the interference due to multiple reflections inside the film. There are no interference rings anymore. For metallic surfaces, the enhanced backscattering phenomenon can be observed, which is due to the excitation of surface plasmon polaritons at the dielectric metallic interface [116, 47]. Satellite peaks may also be observed in the multiple scattering regime if the film supports one or several guided modes [121, 110, 122, 52, 47].

11.5 Summary

After dealing with the physics of scattering by a single rough interface in Chapter 10, the present chapter was devoted to the study of the scattering by a rough dielectric film, and more precisely to the study of Selényi rings which are interference rings in the diffuse intensity. Based on small amplitude perturbation theory to first order, and on an equivalent picture based on optical paths allowing for a single scattering event, we have identified the physical mechanisms at play in the Selényi rings phenomenon. By a careful examination of the different possible single scattering paths, we were able to explain the difference of ring contrast observed when either a rough surface is on the side of the observed light (the remaining surface being planar) or the opposite configuration. The difference could be explained in terms of the nature of the optical paths involved, in particular the importance of paths with phase mixing related to both the angle of incidence and the angle of scattering. Moreover, we have illustrated the influence of the Brewster scattering effect and of the Yoneda effects on the positions of the rings due to additional phase jump and phase variations associated with these two phenomena as was explained in details in the previous chapter. In addition, a phenomenon of ring enhancement or attenuation was also observed and explained by tuning the cross-correlation between the two rough interfaces. This is an example of correlation induced interference.

Chapter 12

Inverse scattering problem

With four parameters I can fit an elephant, and with five I can make him wiggle his trunk.
John Von Neumann.

Can we deduce the shape of a surface from the knowledge of its scattering response? How to design a surface with a given scattering response? These two similar questions are admittedly more challenging to answer than that of the forward scattering problem, and will be referred to as the inverse scattering problem and the designer problem. In this chapter, we present methods to solve the two aforementioned problems, in some specific cases, by interpreting them as an optimization problem.

The aim of the present chapter is to explain and demonstrate how one can recover surface parameters of a system composed of one or several *periodic surfaces composed of elementary shapes*, whose Fourier moments are known analytically, and the *statistical properties of randomly rough surfaces*. This is, of course, a simpler problem than that of a general inverse scattering problem since we assume the knowledge of the materials the system is made of, and the shape of the surfaces but only ignore the scaling parameters that define the geometry. We will refer to this problem as the *parameter retrieval problem* or as *critical dimension metrology*. Nonetheless, this problem has some interesting practical applications. Assume one is manufacturing a surface, hence one has some knowledge of what the surface should look like but with some uncertainty. One would then be interested in verifying and assessing the quality of the manufactured surface. This could in principle be done at a low cost by a light scattering experiment and using the presented method to obtain quantitative estimates of the actual parameters of the surface. A second example is that of the designer problem. If one is able to construct some well pre-defined structures, how to optimize the parameters of these structures to obtain an optical response which is as close as possible to a desired optical response, given some constraints on what is practically realizable.

The basic idea for solving the aforementioned optimization problem is to construct a sequence of surfaces, or systems, defined by a set of geometrical parameters with an algorithm that minimizes an objective function which compares the scattering response of the simulated surfaces to a target scattering response. The present chapter is composed of three main parts. First, we will explain how the inverse problem is handled mathematically, viewed as a rather general optimization problem. We will define how the parameter space defining a system is described and introduce the concept of *measurement functional* which, we believe, is a rather elegant way of defining any objective function based on the reflection and transmission amplitudes, although the reader

may find it too abstract. Then we will present some classic methods for solving optimization problems which make use of the gradient or Jacobian of the objective function, as well as some techniques to evaluate the Jacobian numerically. Finally, the framework is applied to a concrete example of critical dimension metrology of a plasmonic photonic surface and we show that the method is able of retrieving geometrical parameters with great accuracy.

12.1 Position of the problem

Let a system be composed of a stack of $n+1$ layers separated by n interfaces as defined in Chapter 3. We assume that the set of surface profiles $(\zeta_i)_{i\in[\![1,n]\!]}$ is completely defined by a finite set of d parameters denoted $\mathbf{v} = (v_1, \cdots, v_d)$.

Example: for a stack of flat interfaces, only one parameter per interface is needed to encoded the geometry of the system, namely the distance H_i of the i^{th} surface from a given origin, hence $\mathbf{v} = (H_1, \cdots, H_n)$.

Example: for a single two-dimensional sinusoidal surface given by

$$\zeta_1(\mathbf{x}) = \frac{H}{2}\left[\sin\left(\frac{2\pi}{a_1}x_1\right) + \sin\left(\frac{2\pi}{a_2}x_2\right)\right],$$

three parameters are needed to define the surface, namely, the amplitude H, and the period in each direction a_1 and a_2, hence $\mathbf{v} = (H, a_1, a_2)$.

Example: for a randomly rough surface having a Gaussian height probability density with rms-roughness $\sigma > 0$ and a isotropic Gaussian auto-correlation function with correlation length $a > 0$, the parameter vector defining the *statistical* properties of the surface (or realizations of the surface) is then $\mathbf{v} = (\sigma, a)$. If we are studying a system having a film with randomly rough interfaces, we have $\mathbf{v} = (\sigma_1, \sigma_2, a_1, a_2, h)$ where h is the height offset of the second interface (the first being set to have zero average height) and the rms-roughness of the first and second interface are respectively σ_1 and σ_2, and their correlation lengths are a_1 and a_2. Note that we can also add cross-correlation between the interfaces, or consider anisotropic surfaces hence adding extra parameters.

Remark 12.1. In this chapter, we have chosen to work with a fixed set of materials for the different layers, and only the geometrical parameters will be part of the optimization variables. It is of course conceivable to add the material dielectric functions as part of the optimization parameters. We choose here to restrict ourselves to the case in which one has knowledge of the material composition of the system and one is interested in recovering the geometrical parameters.

The geometry of the whole system is then determined by the d-dimensional parameter vector $\mathbf{v} = (v_i)_{i\in[\![1,d]\!]}$. Let $\mathbf{R}(\mathbf{p}|\,\mathbf{p}_0, \omega, \mathbf{v})$ and $\mathbf{T}(\mathbf{p}|\,\mathbf{p}_0, \omega, \mathbf{v})$ respectively denote the reflection and transmission amplitudes of the whole system defined by $\mathbf{v}$. Let $\tilde{\mathbf{R}}(\mathbf{p}|\,\mathbf{p}_0, \omega)$ and $\tilde{\mathbf{T}}(\mathbf{p}|\,\mathbf{p}_0, \omega)$ denote the reflection and transmission amplitudes of a target system.

In practice, the reflection and transmission amplitudes of the target system are not known, but only a finite set of physical measurements is known. By measurement, we can understand either the outcome of an actual measurement from an experiment, or the desired outcome for a measurement in the case of the designer problem. Thus one might know only the intensity

scattered in a given set of directions given a set of incoming wave vectors, for a given set of photon energies, and for a given set of polarization states. These measurements will be mathematically described as *functionals* acting of the reflection or transmission amplitude distributions for a given system $\mathbf{R}(\,\cdot\,|\,\cdot\,,\,\cdot\,,\mathbf{v})$, $\mathbf{T}(\,\cdot\,|\,\cdot\,,\,\cdot\,,\mathbf{v})$. The action of a measurement $\mathcal{M}$ on the scattering amplitude $\mathbf{X}(\,\cdot\,|\,\cdot\,,\,\cdot\,,\mathbf{v})$, where $\mathbf{X}=\mathbf{R}$ or $\mathbf{T}$, is denoted as $\langle\mathcal{M},\mathbf{X}(\,\cdot\,|\,\cdot\,,\,\cdot\,,\mathbf{v})\rangle$. Thus only some $(\langle\mathcal{M}_j,\tilde{\mathbf{R}}\rangle)_{1\leq j\leq N}$ and/or $(\langle\mathcal{M}_j,\tilde{\mathbf{T}}\rangle)_{1\leq j\leq N}$, which represent a set of N measurements, are known and consequently only the corresponding measurements for the computed scattering response are allowed to be used as part of a cost function. Therefore a cost function for the set of parameters $\mathbf{v}$, $\chi(\mathbf{v})$ is of the form

$$\chi(\mathbf{v})=\chi\left(\langle\mathcal{M}_1,\mathbf{X}_1(\,\cdot\,|\,\cdot\,,\,\cdot\,,\mathbf{v})\rangle,\cdots,\langle\mathcal{M}_N,\mathbf{X}_N(\,\cdot\,|\,\cdot\,,\,\cdot\,,\mathbf{v})\rangle,\langle\mathcal{M}_1,\tilde{\mathbf{X}}_1\rangle,\cdots,\langle\mathcal{M}_N,\tilde{\mathbf{X}}_N\rangle\right). \tag{12.1}$$

Here the $\mathbf{X}_j$ denote either the reflection or transmission amplitude distributions, so that we could have mixed reflection and transmission data. A natural example of cost function that we will use in the following is a cost function quadratic in difference of measurements

$$\chi^2(\mathbf{v})=\frac{1}{2}\sum_{j=1}^{N}\frac{\left|\langle\mathcal{M}_j,\mathbf{X}_j(\,\cdot\,|\,\cdot\,,\,\cdot\,,\mathbf{v})\rangle-\langle\mathcal{M}_j,\tilde{\mathbf{X}}_j\rangle\right|^2}{\sigma_j^2}. \tag{12.2}$$

Here σ_j corresponds to an uncertainty or tolerance in the target data $\langle\mathcal{M}_j,\tilde{\mathbf{X}}_j\rangle$, so that each term in Eq. (12.2) is weighted with respect to the respective tolerance. A convenient feature of the above cost function is that it is (i) positive and (ii) if a system gives the exact same measurement outcomes as the ideal system, then the cost function vanishes.

Examples of measurement functional: as a first example of measurement functional, let us consider a single periodic interface whose scattering amplitude distributions are Dirac combs, as we have seen in Chapter 4 for example. In this case, only a discrete set of modes with in-plane wave vectors, up to the incident wave vector, lying on the reciprocal lattice of the interface are allowed. Among these modes, only those with in-plane wave vector within a disc of radius $\sqrt{\epsilon_1}\frac{\omega}{c}$ (resp. $\sqrt{\epsilon_2}\frac{\omega}{c}$) are propagating reflected (resp. transmitted) modes. Experimentally, for a given incident angular frequency ω, in-plane wave vector $\mathbf{p}_0$ and polarization state $\beta\in\{p,s\}$, one can measure the fraction of incident power scattered in the mode $\boldsymbol{\ell}\in\mathbb{Z}^2$ with polarization state $\alpha\in\{p,s\}$, also known as scattering efficiency (see Chapter 1). From the scattering amplitudes $\mathbf{X}^{(\boldsymbol{\ell})}(\mathbf{p}_0,\omega)$, one deduces the efficiency, which we will simply denote $e_{\alpha\beta}^{(X,\boldsymbol{\ell})}(\mathbf{p}_0,\omega)$, given by

$$e_{\alpha\beta}^{(X,\boldsymbol{\ell})}(\mathbf{p}_0,\omega|\,\mathbf{v})=\left\langle e_{\alpha\beta}^{(X,\boldsymbol{\ell})}(\mathbf{p}_0,\omega),\mathbf{X}(\,\cdot\,|\,\cdot\,,\,\cdot\,,\mathbf{v})\right\rangle=\begin{cases}\frac{\alpha_j(\mathbf{p}_0^{(\boldsymbol{\ell})})}{\alpha_1(\mathbf{p}_0)}\left|x_{\alpha\beta}^{(\boldsymbol{\ell})}(\mathbf{p}_0,\omega,\mathbf{v})\right|^2 & \text{if } |\,\mathbf{p}_0^{(\boldsymbol{\ell})}\,|^2\leq\epsilon_j\frac{\omega^2}{c^2}\\ 0 & \text{if } |\,\mathbf{p}_0^{(\boldsymbol{\ell})}\,|^2>\epsilon_j\frac{\omega^2}{c^2}\end{cases}, \tag{12.3}$$

where the subscript j denotes the medium in which the mode is scattered, i.e. $j=1$ for reflection $\mathbf{X}=\mathbf{R}$ and $j=2$ for transmission $\mathbf{X}=\mathbf{T}$. Note that in the following, we will only use the functional bracket notation when dealing with the abstract procedure, but we will use its result as a function of $\mathbf{v}$ when giving concrete example.

A second example of measurement functional, that will be particularly important in our concrete application (Section 12.5), is that giving the Mueller matrix elements (see Chapter 1).

As an example of measurement functional when dealing with randomly rough surfaces, we may consider that giving the *incoherent component of the mean differential reflection and*

transmission coefficients (see Chapter 1).

To sum up, for a given set of measurements $(\mathcal{M}_j)_{j\in[\![1,N]\!]}$ and their corresponding outcomes for the target, the cost function is only a function of the parameter vector $\mathbf{v}$. The last ingredient one needs to define the optimization problem is a set of constraints on the parameters. The set of constraints is represented by some set in $\mathbf{v}$-space. To fix the idea, and since all geometrical parameters considered here are real, we can see the parameter vector $\mathbf{v}$ as a vector in $\mathbb{R}^d$. Thus the constraints are represented abstractly by some set $\mathscr{C} \subset \mathbb{R}^d$. To use the vocabulary of the field of optimization, we will call an element $\mathbf{v} \in \mathscr{C}$ an *admissible solution* and an admissible solution which minimizes the cost function on $\mathscr{C}$ an *optimal solution* or *point of minimum*. With the presented set of notations, the minimization problem reads

$$\text{find } \bar{\mathbf{v}} \in \mathscr{C} \text{ such that} \qquad \chi(\bar{\mathbf{v}}) = \inf_{\mathbf{v}\in\mathscr{C}} \chi(\mathbf{v})\,. \tag{12.4}$$

We will now present algorithms that minimize the cost function under a set of constraints and yield an optimal solution.

12.2 Optimization methods based on the Jacobian

12.2.1 Conditions for a local minimum

Assuming the objective function to be twice differentiable and that $\mathscr{C} = \mathbb{R}^d$ (i.e. we consider a unconstrained problem) a necessary and sufficient condition for $\bar{\mathbf{v}}$ to be a local minimum read

$$\nabla\chi(\bar{\mathbf{v}}) = \mathbf{0}, \qquad \text{and} \qquad \mathbf{w}^{\mathrm{T}}\,\mathbf{H}(\mathbf{v})\,\mathbf{w} > 0 \tag{12.5}$$

for all $\mathbf{w} \in \mathbb{R}^d$ and all $\mathbf{v}$ in a neighborhood of $\bar{\mathbf{v}}$. Here $\nabla\chi(\mathbf{v})$ and $\mathbf{H}(\mathbf{v})$ denote respectively the Jacobian, $(\partial\chi/\partial v_i)_{1\le i\le d}$, and the Hessian of χ, $(\partial^2\chi/\partial v_i\partial v_j)_{1\le i,j\le d}$, at the point $\mathbf{v}$. This generalizes the well known result in one dimension that a local minimum is found at a point of vanishing slope and of positive curvature.

12.2.2 Gradient descent

The idea of the gradient descent or steepset descent method is, from an initial point in $\mathscr{C}$, to iteratively compute a sequence of points in $\mathscr{C}$ that converges to a local minimum or a saddle point by taking steps along the direction along which the cost function decreases the most, i.e. opposite to its gradient $\nabla\chi$, and this until a point where the gradient vanishes is found. Then more analysis is needed to decide whether the point of local minimum or saddle point found is a global point of minimum. The algorithm for the gradient decent reads

- Choose an initial point $\mathbf{v}_0 \in \mathscr{C}$.
- While the norm of the gradient is larger than some small number do
  $\mathbf{v}_{n+1} = \mathbf{v}_n - \mu_n\nabla\chi(\mathbf{v}_n)$
  where μ_n is a step size, which may be fixed or adaptive.
  If $\mathbf{v}_{n+1} \notin \mathscr{C}$, project $\mathbf{v}_{n+1}$ onto $\mathscr{C}$.

Remark 12.2. The projection onto $\mathscr{C}$ is not a trivial step in general as it depends on $\mathscr{C}$. In the cases we will deal with, $\mathscr{C}$ will be a d-dimensional box and the projection in that case is straightforward.

12.2.3 Gauss-Newton and Levenberg-Marquardt methods

We consider now the specific case of an objective function of the form Eq. (12.2). In order to lighten the notations, since each term $\langle \mathcal{M}_j, \mathbf{X}_j(\,\cdot\,|\,\cdot\,,\,\cdot\,,\mathbf{v})\rangle$ is only a function of the parameter vector $\mathbf{v}$ (all the data concerning angle of incidences, scattering, etc... are parameters of the measurement functional) we simply write $y_j(\mathbf{v}) = \langle \mathcal{M}_j, \mathbf{X}_j(\,\cdot\,|\,\cdot\,,\,\cdot\,,\mathbf{v})\rangle$ and the corresponding target data $\tilde{y}_j = \langle \mathcal{M}_j, \tilde{\mathbf{X}}_j\rangle$, so that Eq. (12.2) reads

$$\chi^2(\mathbf{v}) = \frac{1}{2}\sum_{j=1}^{N} \frac{|y_j(\mathbf{v}) - \tilde{y}_j|^2}{\sigma_j^2} = \frac{1}{2}\,|\boldsymbol{\rho}(\mathbf{v})|^2\,. \tag{12.6}$$

Here we have introduced the N-dimensional *residual vector* whose components are defined by $\rho_j = (y_j - \tilde{y}_j)/\sigma$. The main idea of Gauss-Newton methods is to approximate the objective function by a quadratic form (which should be a reasonable thing to do near a local minimum) and to take a leap to the stationary point (potentially a minimum) of this quadratic form. Let a quadratic approximation of χ^2 about a point $\mathbf{v}_n$ be

$$\chi^2(\mathbf{v}) \approx \chi^2(\mathbf{v}_n) + (\mathbf{v}-\mathbf{v}_n)\cdot\nabla\chi^2(\mathbf{v}_n) + \frac{1}{2}\,(\mathbf{v}-\mathbf{v}_n)\cdot\mathbf{H}(\mathbf{v}_n)(\mathbf{v}-\mathbf{v}_n)\,, \tag{12.7}$$

where $\nabla\chi^2$ and $\mathbf{H}$ are respectively the gradient and Hessian matrix of χ^2. In particular, this yields the following approximation for the gradient $\nabla\chi^2$

$$\nabla\chi^2(\mathbf{v}) \approx \nabla\chi^2(\mathbf{v}_n) + \mathbf{H}(\mathbf{v}_n)(\mathbf{v}-\mathbf{v}_n)\,. \tag{12.8}$$

We now want to take a leap to what could potentially be a local minimum of the quadratic approximation of χ^2, which means that we want to reach a point $\mathbf{v}_{n+1}$ where $\nabla\chi^2(\mathbf{v}_{n+1}) = 0$. Setting $\nabla\chi^2(\mathbf{v}_{n+1}) = 0$ in Eq. (12.8) yields

$$\mathbf{H}(\mathbf{v}_n)(\mathbf{v}_{n+1}-\mathbf{v}_n) = -\nabla\chi^2(\mathbf{v}_n)\,, \tag{12.9}$$

which is a linear system in the leap $\mathbf{v}_{n+1}-\mathbf{v}_n$. Solving Eq. (12.9) hence gives us a new point of approximation for the point of minimum. There are two difficulties to consider here. First, the quadratic form approximating best the objective function at some point $\mathbf{v}$ requires the computation of the Hessian matrix of χ^2, which is in principle a non trivial task since in our case the y_j are non-linear in $\mathbf{v}$ and not known in closed form. Indeed, recall that for evaluating y_j we need to solve numerically the RRE and then apply a measurement functional on the solution. We then need a simple way to approximate numerically the Hessian. Second, there is no guarantee that a point where the gradient of the quadratic form vanishes is a minimum. It could be a maximum or a saddle point too. The remedy to this issue will be to somewhat bias the new leap towards the gradient descent if that given by Eq. (12.9) actually increases the objective function. These considerations are at the core of the Gauss-Newton and Levenberg-Marquardt methods [123, 124, 61] that we now describe in details.

Let us then express the gradient and the Hessian matrix of the objective function, Eq. (12.6). The k^{th} component of the gradient reads ($k \in [\![1, d]\!]$)

$$\frac{\partial\chi^2}{\partial v_k} = \sum_{j=1}^{N} \frac{\partial y_j}{\partial v_k}\,\frac{y_j - \tilde{y}_j}{\sigma_j^2}\,, \tag{12.10}$$

or equivalently in vectorial notations using the residual vector

$$\nabla\chi^2 = \mathbf{J}^{\mathrm{T}}\,\boldsymbol{\rho}\,, \tag{12.11}$$

where $\mathbf{J}$ is the Jacobian matrix of the residual vector. The (k,l) element of the Hessian matrix reads ($k,l \in [\![1,d]\!]$)

$$\frac{\partial^2 \chi^2}{\partial v_k \partial v_l} = \sum_{j=1}^{N} \frac{1}{\sigma_j^2} \left[\frac{\partial y_j}{\partial v_k} \frac{\partial y_j}{\partial v_l} + \frac{\partial^2 y_j}{\partial v_k \partial v_l} (y_j - \tilde{y}_j) \right] . \tag{12.12}$$

The elements of the Hessian matrix requires in principle the evaluation of first and second order derivatives of the components y_j. A common approximation consists in ignoring the second term in the square bracket in Eq. (12.12). This can be motivated by the fact that this term will be negligible in practice compared with the first one. One way of seeing this is that near the minimum, we expect $(y_j - \tilde{y}_j)$ to be small. In fact, inclusion of the second term may lead to unstable schemes if outlier data points may not be fitted by the model [61]. Hence the Hessian is approximated by

$$\frac{\partial^2 \chi^2}{\partial v_k \partial v_l} \approx \sum_{j=1}^{N} \frac{1}{\sigma_j^2} \frac{\partial y_j}{\partial v_k} \frac{\partial y_j}{\partial v_l} . \tag{12.13}$$

which only requires the evaluation of first order derivatives, which we needed anyway for evaluation of the gradient.

Remark 12.3. Note that this approximation will not change the solution obtained when the iterates have converged but may only affect the speed at which the optimum is reached. Indeed, all we actually ask for is to have a good enough local approximation of the cost function by a quadratic form. Whether this quadratic approximation is the best one or is only close to it is not crucial [61].

The Gauss-Newton (or quasi Gauss-Newton) method then consists in solving Eq. (12.9) with $\mathbf{H}$ replaced by its estimate $\tilde{\mathbf{H}}$ defined for $k,l \in [\![1,d]\!]$ by

$$\tilde{H}_{kl} = \sum_{j=1}^{N} \frac{1}{\sigma_j^2} \frac{\partial y_j}{\partial v_k} \frac{\partial y_j}{\partial v_l} , \tag{12.14}$$

or in vectorial notation using the residual vector

$$\tilde{\mathbf{H}} = \mathbf{J}^{\mathrm{T}} \mathbf{J} . \tag{12.15}$$

In other words, a (quasi) Gauss-Newton steps consists in solving

$$\left[\mathbf{J}^{\mathrm{T}} \mathbf{J}\right](\mathbf{v}_n) \, (\mathbf{v}_{n+1} - \mathbf{v}_n) = - \, \mathbf{J}^{\mathrm{T}}(\mathbf{v}_n) \, \boldsymbol{\rho}(\mathbf{v}_n) . \tag{12.16}$$

Now if such a step does not decrease the objective function, one applies a regularization step, which consists in adding weight to the diagonal of $\tilde{\mathbf{H}}$. This is done in the following way. Let μ be a small positive scalar and set

$$\tilde{H}_{kl}^{(\mu)} = (1 + \mu \, \delta_{kl}) \, \tilde{H}_{kl} , \tag{12.17}$$

and repeat the Gauss-Newton step with the new matrix $\tilde{\mathbf{H}}^{(\mu)}$. If the new point $\mathbf{v}_{n+1}$ still does not decrease the objective function, repeat the regularization step by increasing the value μ, etc ... What is this actually doing? Let us first imagine that the regularization step instead reads

$$\tilde{H}_{kl}^{(\mu)} = \tilde{H}_{kl} + \mu \, \delta_{kl} , \tag{12.18}$$

which was Levenberg initial contribution. When μ becomes large the matrix $\tilde{\mathbf{H}}^{(\mu)}$ becomes diagonally dominant and we can write that it is essentially

$$\tilde{H}_{kl}^{(\mu)} \sim \mu\, \delta_{kl}\,, \tag{12.19}$$

hence solving Eq. (12.16) simply reduces to

$$\mathbf{v}_{n+1} = \mathbf{v}_n - \frac{1}{\mu}\, \nabla\chi^2(\mathbf{v}_n), \tag{12.20}$$

which is simply a gradient descent step with step size $1/\mu$. Such a regularization step then have a tendency to bias the leap towards the steepest descent for large μ. However, this may be inefficient since the step size may become small. In addition, Marquardt realized that the components of the gradient of the objective function may have different scales, since in fact they may have different units. Therefore, there should be a better way to weight each component of the gradient than by a common scaling factor, i.e. a pseudo gradient descent that would look, component wise, like

$$v_{n+1,k} = v_{n,k} - \frac{1}{\mu_k}\, \frac{\partial\chi^2}{\partial v_k}(\mathbf{v}_n)\,. \tag{12.21}$$

Marquardt's idea is that the Hessian matrix contains information about the scales for more appropriate step size on each gradient direction. Indeed, looking at the dimensions we have $[\partial\chi^2/\partial v_k] = [v_k]^{-1}$, and $[\partial^2\chi^2/\partial v_k^2] = [v_k]^{-2}$, hence we would rather like to have a proportionality constant $1/\mu_k$ between v_k and $\partial\chi^2/\partial v_k$ in Eq. (12.21), to have the dimension of $[v_k]^2$, i.e. $[\mu_k] = [v_k]^{-2}$ which is precisely the dimensions found in the diagonal elements of the Hessian matrix. Consequently, Marquardt suggested the regularization scheme we announced in Eq. (12.17).

To sum up the Levenberg-Marquardt method consists in approaching a local minimum of the objective function by solving

$$\big[\mathbf{J}^{\mathrm{T}}\,\mathbf{J} + \mu_n \mathrm{diag}(\mathbf{J}^{\mathrm{T}}\,\mathbf{J})\big](\mathbf{v}_n)\,(\mathbf{v}_{n+1} - \mathbf{v}_n) = -\,\mathbf{J}^{\mathrm{T}}(\mathbf{v}_n)\,\boldsymbol{\rho}(\mathbf{v}_n)\,, \tag{12.22}$$

with adaptive regularization parameter μ_n to ensure that the objective function decreases at each new iterate. It is clear from Eq. (12.22) that evaluating the Jacobian of the residual vector becomes a key step in the optimization process, and we devote the next section to this task.

12.3 Methods for computing the Jacobian for the RRE

In the methods presented in the previous section, the Jacobian of the cost function, or that of the residual vector, plays a key role. We expose now methods for computing the Jacobian of the residual vector and discuss their advantages and drawbacks in the case of a model being the solution of the reduced Rayleigh equations.

12.3.1 Finite difference

A rather simple way to evaluate the Jacobian of the residual vector, or that of the objective function χ, is through the use of finite difference methods. Assume that one has implemented a solver for the reduced Rayleigh equations giving the reflection and transmission amplitudes for a given set of scattering parameters (angles of incidence and scattering, wavelengths) and a given parameter vector $\mathbf{v}$ defining the scattering system. In addition, assume that one has implemented a function that takes as an input this reflection and transmission amplitudes (and

target data) and returns the residual vector. Then one can estimate the Jacobian of the residual vector by solving $d+1$ reduced Rayleigh equations, one for the current parameter $\mathbf{v}_n$ and one for each direction v_j (recall that $\mathbf{v} \in \mathbb{R}^d$) in parameter space with $\mathbf{v} = \mathbf{v}_n + \Delta v_j \, \hat{\mathbf{e}}_j = \mathbf{v}_n^{(j)}$, where $\hat{\mathbf{e}}_j$ is a unit vector in the j^{th} parameter. From the $d+1$ computed residual vectors, $\boldsymbol{\rho}_0 = \boldsymbol{\rho}(\mathbf{v}_n), \boldsymbol{\rho}_1 = \boldsymbol{\rho}(\mathbf{v}_n^{(1)}), \cdots, \boldsymbol{\rho}_d = \boldsymbol{\rho}(\mathbf{v}_n^{(d)})$, we can approximate the Jacobian matrix of the residual vector by

$$\boldsymbol{\rho}'(\mathbf{v}_n) = \left(\frac{\boldsymbol{\rho}_1 - \boldsymbol{\rho}_0}{\Delta v_1}, \cdots, \frac{\boldsymbol{\rho}_d - \boldsymbol{\rho}_0}{\Delta v_d} \right) . \tag{12.23}$$

The advantage of this technique resides in its simplicity of implementation. Although the derivatives are approximated by finite differences, the method is expected to work relative well if the residuals change smoothly with the parameters, which essentially means that the optical response changes smoothly with the parameters. The choice of the steps Δv_j must be taken carefully for each parameter. Indeed, note that all parameters, first, may not have all the same physical units (usually length, angles, unit-less scaling parameters between two lengths, etc ...), second, may vary in quite different ranges, and third, may not all have the same impact on the optical response. The user hence must carefully study the nature of each parameter, their expected range of admissibility and estimate their impact on the optical response to come up with a reasonably small but relevant step size in terms of change in the optical response.

12.3.2 Differentiation of the RRE

Is there a method to compute exactly the Jacobian of the scattering amplitudes solution of the reduced Rayleigh equations? Yes, there is. The idea is quite natural as it consists in differentiating the reduced Rayleigh equations with respect to the parameters in order to obtain an equation on the Jacobian of the scattering amplitudes. We will see that it is not much more costly than doing a finite difference approximation but it requires a bit more numerical implementation. Let us start by reminding the reader that the scattering amplitudes $\mathbf{X}(\mathbf{q} \,|\, \mathbf{p}_{\text{o}})$ are always given as the solution of (i) an integral equation of the form

$$\int \mathbf{M}(\mathbf{p} \,|\, \mathbf{q}) \; \mathbf{X}(\mathbf{q} \,|\, \mathbf{p}_{\text{o}}) \, \frac{\mathrm{d}^2 q}{(2\pi)^2} = -\,\mathbf{N}(\mathbf{p} \,|\, \mathbf{p}_{\text{o}}) \; , \tag{12.24}$$

for arbitrary surfaces or (ii) an infinite countable linear system

$$\sum_{\mathbf{m} \in \mathbb{Z}^2} \mathbf{M}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_{\mathbf{m}}) \; \mathbf{X}^{(\mathbf{m})}(\mathbf{p}_0) = -\,\mathbf{N}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_0) \, , \tag{12.25}$$

for periodic surfaces. We treat now the latter case since it takes a form which looks closest to numerical implementation. The former situation can be treated in a similar fashion. An interesting fact is that the periodic case contains an additional detail that one must pay attention to. The lattice constants being considered as parameters, taking derivatives with respect to the lattice constants induce extra factors due to the fact that the in-plane wave vectors of the form $\mathbf{p}_{\boldsymbol{\ell}}$ depend directly on the lattice parameters. We thus start by considering the case of a geometrical parameter which is not a lattice parameter.

Differentiation with respect to a non-lattice parameter

Consider a parameter v_i of the parameter vector $\mathbf{v} = (v_1, \cdots, v_d)$, such that v_i is not part of a lattice parametrisation. By taking a partial derivative of the reduced Rayleigh equations Eq. (12.25) with respect to v_i we obtain

$$\sum_{\mathbf{m}\in\mathbb{Z}^2} \left(\partial_{v_i} \mathbf{M}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_{\mathbf{m}})\ \mathbf{X}^{(\mathbf{m})}(\mathbf{p}_0) + \mathbf{M}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_{\mathbf{m}})\, \partial_{v_i} \mathbf{X}^{(\mathbf{m})}(\mathbf{p}_0)\right) = -\partial_{v_i} \mathbf{N}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_0)\,. \tag{12.26}$$

Here $\partial_{v_i}\mathbf{M}$ corresponds to a 2×2 matrix whose elements are $(\partial_{v_i} M_{\alpha,\beta})_{\alpha,\beta\in\{p,s\}}$ and similarly for $\partial_{v_i}\mathbf{N}$ and $\partial_{v_i}\mathbf{X}^{(\mathbf{m})}$. Assuming known the solution of Eq. (12.25), we obtain a linear system for $(\partial_{v_i}\mathbf{X}^{(\mathbf{m})})_{\mathbf{m}\in\mathbb{Z}^2}$, namely

$$\sum_{\mathbf{m}\in\mathbb{Z}^2} \mathbf{M}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_{\mathbf{m}})\, \partial_{v_i} \mathbf{X}^{(\mathbf{m})}(\mathbf{p}_0) = -\partial_{v_i} \mathbf{N}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_0) - \sum_{\mathbf{m}\in\mathbb{Z}^2} \partial_{v_i} \mathbf{M}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_{\mathbf{m}})\ \mathbf{X}^{(\mathbf{m})}(\mathbf{p}_0)\,. \tag{12.27}$$

The computation of the Jacobian of the scattering amplitudes thus reduces to solving a linear system similar to the one for computing the scattering amplitudes. In fact, the left-hand side matrix stays identical to that used for solving Eq. (12.25) but the right-hand side is composed of two terms. The first term is simply the partial derivative with respect to v_i of the right-hand side of Eq. (12.25) and the second is the product of the partial derivative of the left-hand side matrix of Eq. (12.25) and its solution. Numerically, this means that we need to set up the new matrix $\partial_{v_i}\mathbf{M}$ and vector $\partial_{v_i}\mathbf{N}$, and solve a system. It is therefore, roughly, as costly as doing a finite difference approximation. Indeed, for each partial derivative with respect to a parameter, the finite difference approximation requires to set up $\mathbf{M}$ and $\mathbf{N}$ for a system whose parameter v_i is perturbed by an amount Δv_i and solve the corresponding system.

Differentiation with respect to a lattice parameter

In the case of differentiation with respect to a lattice parameter, the same method applies but we have to remember that

$$\begin{aligned}
\partial_{v_i}\left[M_{\alpha\beta}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_{\mathbf{m}})\right] &= \partial_{v_i} M_{\alpha\beta}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_{\mathbf{m}}) + \partial_{v_i}\mathbf{p}_{\boldsymbol{\ell}}\cdot\nabla_{\mathbf{p}_{\boldsymbol{\ell}}} M_{\alpha\beta}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_{\mathbf{m}}) + \partial_{v_i}\mathbf{p}_{\mathbf{m}}\cdot\nabla_{\mathbf{p}_{\mathbf{m}}} M_{\alpha\beta}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_{\mathbf{m}}) \\
\partial_{v_i}\left[N_{\alpha\beta}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_0)\right] &= \partial_{v_i} N_{\alpha\beta}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_0) + \partial_{v_i}\mathbf{p}_{\boldsymbol{\ell}}\cdot\nabla_{\mathbf{p}_{\boldsymbol{\ell}}} N_{\alpha\beta}(\mathbf{p}_{\boldsymbol{\ell}} \mid \mathbf{p}_0)\,.
\end{aligned}$$

12.4 Other optimization methods

We briefly comment in this section on the potential use of other optimization methods with their advantages and drawbacks.

12.4.1 Genetic optimization

Genetic optimization, or more generally evolutionary algorithms, refers to a family of heuristic optimization methods inspired from biology. The parameters defining the system is seen as a genome and the measurements as a phenotype. The objective function then consists in having a phenotype which fits best the objective. The basic principles of a genetic algorithm read as follows (see e.g. [125, 126]):

1 Generate a set of systems (a generation of individuals) with randomly picked values of the parameters (genes). Compute the corresponding value of the objective function for each system.

2 Create a new generation of systems until some systems are found to have reached a small enough value of the objective function (loop). The rules for creating a new generation can combine the following tunable steps:

   2.1 Keep the first few individuals of the previous generation with the smallest value of the objective function. This avoid loosing potentially good solutions from a generation to the next.

   2.2 Create new individuals by "procreation", or cross-over of the genomes of several parents taken from the previous generations. This can be done by choosing randomly a certain number of parents and choosing randomly a certain number of genes (values of parameters) for each and making a new genome from them. The choice of parents can be done randomly with a tunable probability distribution for an individual to be picked as a parent. For example, an individual with a small value of the objective function may have better chances to be picked than an individual with a large value of the objective function.

   2.3 Mutate some of the new individuals by randomly changing the values of some of their genes. This ensures to explore more the parameter space and avoid that the individuals look too similar after a few generations, so that good "unexpected" solutions may be reached.

The above description is of course very basic and aims at giving the essential ideas behind genetic optimization. The art of building an efficient algorithm for a given problem consist in tuning appropriately the different steps and in particular the probability distribution for an individual to be a parent and for mutations. Indeed, giving to much weight to the currently best solutions may lead to an impoverishment of the genetic pool leading to similar solutions which do not explore sufficiently the parameter space. On the contrary, given even weight for all individuals or having too much mutations may lead to exploring too much the parameter space randomly without using the information on which individuals were potentially close to an optimal solution.

One may consider genetic algorithms for optimizing a structure where for example different type of particles with different shapes and sizes maybe placed on a substrate and optimized these parameters for achieving some targeted response. Implementing such an algorithm is relatively simple, and does not require the computation of the Jacobian of the objective function. However, one must keep in mind the cost associated with the evaluation of the objective function for each system, or individual. Indeed, genetic algorithms usually require that each generation contains many individuals, say a hundred or a thousand for the sake of the example. In other words, each generation requires to compute the response of a hundred or thousand systems, which is costly in the case of electromagnetic wave scattering. The method based on the reduced Rayleigh equations is relative efficient compared to other methods like Finite Elements Methods or Rigorous Coupled Waves Analysis, but nevertheless has a cost which does not allow for computing too many systems. In fact, the Levenberg-Marquardt used in Section 12.5 on a concrete example with experimental data, needed around 20 to 30 iterations before leading a trustful optimal solution which would be with the number of parameters for this problem around a hundred systems simulated in total. Therefore, genetic algorithms may not be the preferred optimization method in this case. A genetic algorithm may be preferred instead if the number of parameters becomes very large and/or if fast approximation methods for the computation of the objective function can be found.

12.4.2 Stochastic optimization

As mentioned earlier, methods based on the Jacobian of the objective function have essentially two drawbacks. First, such methods find local minima of the objective function, and the local minimum found by the algorithm depends in general on the initial condition. Second, for optimization problems involving a large number of parameters d the computation of the Jacobian may become very costly as essentially $d+1$ forward problems per iteration must be solved (by iteration we mean a new guess of the solution of the optimization problem). In such cases, where the parameter space is of high dimension and/or the objective function may feature many local minima, one may prefer stochastic methods. An advantage of stochastic optimization methods is that they usually do not require the objective function to be smooth as the Jacobian is not used. A second advantage is that stochastic methods are, in principle, made for finding the global minima of the objective function.

The basic ideas behind stochastic optimization methods are the following. Let us assume that the parameter space is discretized such that we allow the parameter vector $\mathbf{v}$ to take only discrete values. The number of possible discrete values of the parameter may potentially be huge, and evaluating the objective function for each possible parameter vector is of course out of the question. By analogy with problems of statistical physics where the state of a system is determined by a competition between thermal fluctuations and the minimization of a potential energy, we can define the Gibbs measure μ_T parametrized by an artificial temperature parameter T as

$$\mu_T(\mathbf{v}) = \frac{1}{Z_T} \exp\left(-\frac{\chi(\mathbf{v})}{T}\right) , \tag{12.29}$$

with the partition function Z_T defined such that μ_T is normalized, i.e.

$$Z_T = \sum_{\mathbf{v}} \exp\left(-\frac{\chi(\mathbf{v})}{T}\right) . \tag{12.30}$$

The objective function χ hence plays the role of a potential energy in our analogy with problems from statistical physics. Note that the partition function cannot be evaluated in practice since this would require to evaluate the objective function for each parameter vector. The interest of defining the Gibbs measure as in Eq. (12.29) is that it assigns a probability to each parameter vector $\mathbf{v}$ which becomes more and more concentrated on the points of minimum as the temperature goes towards zero. Indeed, let $\chi^\star$ be the minimum value of χ on the considered discrete space of $\mathbf{v}$, and assume that the minimum is reached for a set of n parameter vectors $\mathscr{C}^\star \overset{\text{def}}{=} \{\mathbf{v} \,|\chi(\mathbf{v}) = \chi^\star\} = \{\mathbf{v}_1^\star, \cdots, \mathbf{v}_n^\star\}$. Then by multiplying and dividing the Gibbs measure by $\exp\left(\frac{\chi^\star}{T}\right)$, it can be recast as

$$\mu_T(\mathbf{v}) = \frac{\exp\left(-\frac{\chi(\mathbf{v})-\chi^\star}{T}\right)}{\sum_{\mathbf{v}} \exp\left(-\frac{\chi(\mathbf{v})-\chi^\star}{T}\right)} . \tag{12.31}$$

It thus becomes clear that for all points $\mathbf{v} \notin \mathscr{C}^\star$, the numerator $\exp\left(-\frac{\chi(\mathbf{v})-\chi^\star}{T}\right) \to 0$ as $T \to 0$ while only the points of global minima, $\mathbf{v}_k^\star \in \mathscr{C}^\star$, retain a non-zero weight, since $\exp\left(-\frac{\chi(\mathbf{v}_k^\star)-\chi^\star}{T}\right) = 1$, which gives with the normalization factor a probability equal to $1/n$ for each point of minimum, i.e. the inverse of the number of points of minimum.

The art of stochastic algorithm is then to simulate the Gibbs measure for a decaying temperature. This is usually done by the use of Markov chains with transition probabilities based on conditional probability constructed from the Gibbs measure (in order to avoid to refer to the unknown partition function). Example of such techniques are the Metropolis-Hasting algorithm or simulated annealing [127, 128, 88] , which we choose not to discuss further here as the aim of this section is to only give a brief discussion on the basic ideas behind stochastic optimization.

More important to us is to see that stochastic optimization may be of interest for our inverse scattering problem if the parameter space becomes of high dimension. Stochastic algorithms usually require a relatively large number of evaluations of the objective function. Such methods would then be kept for problems for large d. An advantage of stochastic optimization compared with the genetic optimization is that it is not based on heuristics and is hence backed up with a mathematical theory providing theorems which (in some sense) guarantee the convergence of the algorithm towards a global optimum, with in some cases estimates of the convergence rates. A difficulty, however, is the choice of the decay rate for the artificial temperature, which intuitively should decay fast enough for the algorithm to be efficient but not to fast so that the system does not "freeze" in a region near a local minimum. This is an art in itself.

12.4.3 Adjoint method

We have seen previously that the methods based on the Jacobian require solving $d+1$ scattering problems at each iteration, d being the dimension of the parameter space. Solving a single reduced Rayleigh equation numerically, which consists in solving a linear system of equations, costs $\mathcal{O}(D^3)$ operations (assuming the use of a direct method) where D is the number of modes (see e.g. Chapter 5). Then each iteration of the optimization scheme thus requires $\mathcal{O}(dD^3)$ operations, which can become very costly if both D and d are large. If there are only a few parameters to be tuned the methods presented previously, such as the Levenberg-Marquardt, will be satisfactory. However, what to do if we wish to design a surface where all discretized points on the surface are left as free parameters to be optimized and d becomes significantly large (d could be several thousands or more)?

The adjoint method [129, 130] answers exactly this question, and we will show now how to obtain the Jacobian of the cost function in a number of operations *independent* of the number of parameters d, i.e. in $\mathcal{O}(D^3)$ operations. Assume for simplicity that we can write the cost function as

$$\chi(\mathbf{v}) = F(\mathbf{X}(\mathbf{v}))\,, \tag{12.32}$$

where $\mathbf{X}(\mathbf{v}) = (X_n(\mathbf{v}))_{1\leq n\leq N}$ formally represents the numerical solution of a reduced Rayleigh equation for some parameter $\mathbf{v}$, and that we denote formally in the following as the linear system

$$\mathbf{M}(\mathbf{v})\ \mathbf{X}(\mathbf{v}) = \mathbf{N}(\mathbf{v})\,, \tag{12.33}$$

and F is a known function of $\mathbf{X}$[1]. The Jacobian of χ at point $\mathbf{v}$ reads

$$\partial_{v_i}\chi(\mathbf{v}) = \sum_{n=1}^{N} \partial_{v_i} X_n(\mathbf{v})\, \partial_{X_n} F(\mathbf{X}(\mathbf{v})) = \partial_{v_i}\,\mathbf{X}(\mathbf{v}) \cdot \nabla_{\mathbf{X}} F(\mathbf{X}(\mathbf{v}))\,, \tag{12.34}$$

[1]The function F could also be a function of $\mathbf{X}$ *and* $\mathbf{v}$, and the presented method would also work in a similar fashion.

for each parameter v_i, or in matrix notations

$$\nabla_{\mathbf{v}}\chi(\mathbf{v}) = \nabla_{\mathbf{v}}\,\mathbf{X}(\mathbf{v})\,\nabla_{\mathbf{X}}F(\mathbf{X}(\mathbf{v}))\,. \tag{12.35}$$

Let us stress that in the above equation, $\nabla_{\mathbf{v}}\chi(\mathbf{v})$ is a column vector of size d, $\nabla_{\mathbf{X}}F(\mathbf{X}(\mathbf{v}))$ is a column vector of size N (size of $\mathbf{X}$) and $\nabla_{\mathbf{v}}\,\mathbf{X}(\mathbf{v})$ is the Jacobian matrix of $\mathbf{X}$ and is of size $d \times N$. We have seen in Section 12.3.2 that the differentiation of the reduced Rayleigh equation with respect to the parameters, or equivalently of the linear system Eq. (12.33), yields d linear systems of equations to be solved for computing $\nabla_{\mathbf{v}}\,\mathbf{X}(\mathbf{v})$, namely

$$\mathbf{M}(\mathbf{v})\,\partial_{v_i}\,\mathbf{X}(\mathbf{v}) = \partial_{v_i}\,\mathbf{N}(\mathbf{v}) - \partial_{v_i}\,\mathbf{M}(\mathbf{v})\;\mathbf{X}(\mathbf{v})\,. \tag{12.36}$$

for each v_i. Hence we can write

$$\partial_{v_i}\,\mathbf{X}(\mathbf{v}) = \mathbf{M}^{-1}(\mathbf{v})\,\left[\partial_{v_i}\,\mathbf{N}(\mathbf{v}) - \partial_{v_i}\,\mathbf{M}(\mathbf{v})\;\mathbf{X}(\mathbf{v})\right]\,. \tag{12.37}$$

Plugging the above equation into Eq. (12.34) we get

$$\partial_{v_i}\chi(\mathbf{v}) = \left(\mathbf{M}^{-1}(\mathbf{v})\,\left[\partial_{v_i}\,\mathbf{N}(\mathbf{v}) - \partial_{v_i}\,\mathbf{M}(\mathbf{v})\;\mathbf{X}(\mathbf{v})\right]\right)^{\mathrm{T}}\,\nabla_{\mathbf{X}}F(\mathbf{X}(\mathbf{v})) \tag{12.38}$$

$$= \left[\partial_{v_i}\,\mathbf{N}(\mathbf{v}) - \partial_{v_i}\,\mathbf{M}(\mathbf{v})\;\mathbf{X}(\mathbf{v})\right]^{\mathrm{T}}\,\left(\mathbf{M}^{-1,\mathrm{T}}(\mathbf{v})\,\nabla_{\mathbf{X}}F(\mathbf{X}(\mathbf{v}))\right)\,. \tag{12.39}$$

Here we have written the same expression but by stressing different orders of make the matrix products. In Eq. (12.38) the matrix multiplication $\mathbf{M}^{-1}(\mathbf{v})\,\left[\partial_{v_i}\,\mathbf{N}(\mathbf{v}) - \partial_{v_i}\,\mathbf{M}(\mathbf{v})\;\mathbf{X}(\mathbf{v})\right]$ is taken first and then the result (transposed) is multiplied by $\nabla_{\mathbf{X}}F(\mathbf{X}(\mathbf{v}))$. According to Eq. (12.38), the first factor thus requires to multiply a $N \times N$ matrix ($\mathbf{M}^{-1}(\mathbf{v})$) to d vectors (one for each v_i) of size N, resulting in the problematic complexity of $\mathcal{O}(dN^3)$. However, Eq. (12.39) suggests to rather take the matrix-vector product $\mathbf{Z}(\mathbf{v}) \stackrel{\text{def}}{=} \mathbf{M}^{-1,\mathrm{T}}(\mathbf{v})\,\nabla_{\mathbf{X}}F(\mathbf{X}(\mathbf{v}))$ only *once* and then take the d dot products with the vectors $\partial_{v_i}\,\mathbf{N}(\mathbf{v}) - \partial_{v_i}\,\mathbf{M}(\mathbf{v})\;\mathbf{X}(\mathbf{v})$. This reduces the asymptotic complexity to $\mathcal{O}(N^3)$ since then the most operation consuming step is to solve the single *adjoint* linear system

$$\mathbf{M}^{\mathrm{T}}(\mathbf{v})\;\mathbf{Z} = \nabla_{\mathbf{X}}F(\mathbf{X}(\mathbf{v}))\,. \tag{12.40}$$

Note that the adjoint method is particularly interesting when the function F and the matrices $\mathbf{M}$ and $\mathbf{N}$ are known analytically and so are the corresponding gradients with respect to their respective variables. We have thus replaced the conventional computation of the Jacobian of the cost function, which requires solving d linear systems of equations of size N, to solving a single adjoint linear system of equations of size N.

The adjoint method allows the use of optimization schemes based on the Jacobian with a complexity independent of the dimension of the parameter space. This feature opens a great path towards design of complex systems where a huge number of parameters can be taken into account. A typical example is the case of topological optimization where the shape of different components of a system can be optimized, i.e. that essentially each surface point is viewed as a parameter. The adjoint method has been used in several works for the optimization of optical components. A few examples are, the optimization of silicon photovoltaic cell front coating [131], the maximization of band gaps in two-dimensional photonic crystals [132, 133], the control of emission directionallity of a waveguide termination [134] or even for the design of full optical circuits [135].

We believe that the combination of the adjoint method together with the reduced Rayleigh equations solver should give a very power method for the design of surfaces with desired optical properties, and will be the topic of future research.

12.5 Critical dimension metrology of metasurfaces

We present in this section a technique for parameter retrieval, or critical dimensional metrology, based on Mueller matrix ellipsometry measurements. Such a technique has been used successfully in the past in conjunction with numerical simulations, for example based on Rigorous Coupled Waves Analysis, first for one-dimensional periodic gratings [136, 137, 138] and then for two-dimensional periodic gratings [139, 140]. We deal here with a concrete example of the use of the reduced Rayleigh equation for solving an inverse scattering problem for *critical dimension metrology* of a plasmonic photonic crystal based on *angle-resolved Mueller matrix ellipsometry* adapted from **Paper** [6]. Consider a sample composed of a lattice of gold hemispheroidal particles deposited on a fused silica substrate. The aim of the study is to use Mueller matrix ellipsometry measurements to assess the average size of the particles and lattice constants.

Sample – A scanning electron microscopy (SEM) image of the sample is shown in Fig. 12.1(a) [141, 6]. The in-plane and out-of-plane radii of the particles were estimated by SEM and atomic force microscopy (AFM) to be about $r_{\parallel} = 58 \pm 4$ nm and $r_{\perp} = 36 \pm 5$ nm, the lattice constants were estimated based on the SEM image to be about $a_1 = 208 \pm 2$ nm and $a_2 = 211 \pm 2$ nm. It was also observed that the particles were sitting on top of a silica mound as illustrated in Fig. 12.1(b). This is a defect that occurred during the production process, which is due to over-milling into the substrate. Indeed, the sample was produced by first depositing a planar 40 nm thick gold layer on the substrate and then the particles were shaped by the use of a Ga focused ion beam (FIB). The depth of the trenches was estimated based on AFM measurements to be about $h \approx 15 \pm 10$ nm.

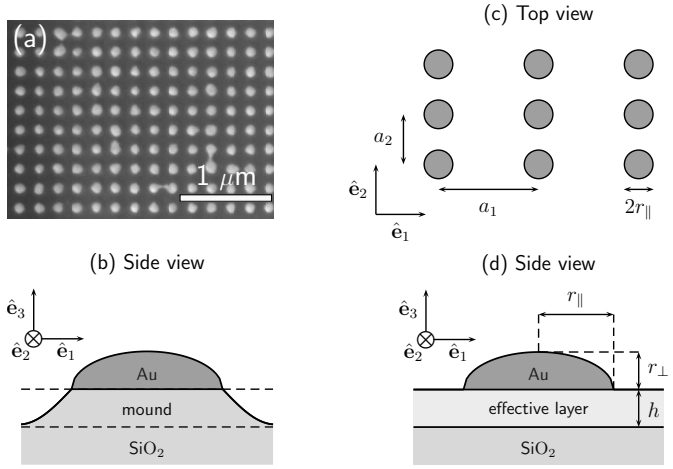


Figure 12.1: (a) Scanning electron microscopy image of the sample (courtesy of Thomas Brakstad). (b) Sketch of a side view of the sample illustrating the mound on top of which the gold particle is sitting. Top view (c) and side view (d) of the modeled geometry. Note that the mound is modeled as an effective layer. Figure adapted from **Paper** [6].

Experiment – The *experimental* normalized Mueller matrix elements measured in the reflected specular direction, *fixed* at $\theta_0 = 55°$, are plotted as functions of the photon energy and azimuthal angle of incidence in the lower halves of each contour plot in Fig. 12.2. The photon energies used in this experiment ranged from 0.7 eV to 5.9 eV and the azimuthal angle of incidence ranged from 0 to 360° in step of 5°. We have chosen to show only measurements for azimuthal angles of incidence from 0 to 180° in Fig. 12.2 due to the symmetry of sample and to make the comparison with the simulation easier (to be discussed soon). We can observe that the normalized Mueller matrix elements exhibit features characteristic of (i) a *localized surface plasmon resonance* around $\hbar\omega \approx 2.1$ eV and (ii) Rayleigh anomalies whose physical interpretation have been already discussed in Section 1.5.2.

Parameter retrieval – Now let us assume that we have a partial knowledge on the sample. We know which materials were used and that the manufacturer had intended to produce a lattice of hemispheroidal particles. However, we do not know the precise size of the particles nor the lattice constants. Given this information, our modeled system is thus parametrized by *a priori* four parameters, namely the lattice constants a_1, a_2 in both directions (assuming a rectangular lattice) and the in-plane and out-of-plane radii of the particles, $r_\parallel$, $r_\perp$ (see Fig. 12.1(c-d)). In addition, we may suspect that over-milling may occur during the production process. We choose to model the mound as a Bruggeman effective layer, mixing air and silica, whose thickness h and filling fraction of air f are parameters of the model (see Fig. 12.1(d)). The parameter vector defining the system is then $\mathbf{v} = (a_1, a_2, r_\parallel, r_\perp, h, f)$. In the following, we may choose to model the system with or without the effective layer. The parameter vector is then implicitly to be understood to be restricted to $\mathbf{v} = (a_1, a_2, r_\parallel, r_\perp)$ in the latter case. Moreover, the reduced Rayleigh equation used in either case needs to be that involving either three or four media (see Chapter 3).

We are now ready to proceed with the parameter retrieval. The strategy consists of two main steps. First, we will make use of our knowledge on the Rayleigh anomalies to deduce the lattice constants without having to solve any forward scattering problem. Then, once the lattice constants are determined, we proceed with the determination of the remaining parameters by minimizing a cost function based on the Mueller matrix elements. For the determination of the lattice constants, we track the position of the Rayleigh anomalies in the $(\phi_0, \hbar\omega)$-plane observed in the experimental data. We have experienced that tracking the anomalies associated with the *reflected* diffracted orders $\boldsymbol{\ell} = (-1, 0)$ and $\boldsymbol{\ell} = (0, -1)$ in the data for the ellipsometric angle ψ_{sp} is reasonably easy since it occurs as a rather sharp maximum. Note that this could be obtained from the Mueller matrix as well. One of the advantages of tracking a Rayleigh anomaly associated with a reflected diffractive order is that the position does not depend on the dielectric functions of the materials composing the sample, the latter being in principle wavelength dependent and may be known only within some accuracy. Indeed, if a Rayleigh anomaly associated with a transmitted diffractive order were tracked, one should be careful to correct for the possibly dispersive dielectric constant of the substrate. The experimentally tracked Rayleigh anomalies in the $(\phi_0, \hbar\omega)$-plane are shown as open circle in Fig. 12.3. Now we can fit the theoretical prediction of the Rayleigh lines to these data points. We recall that the Rayleigh anomalies occur in the $(\theta_0, \phi_0, \hbar\omega)$-space whenever

$$|\,\mathbf{p}_{\boldsymbol{\ell}}\,|^2 = \epsilon_j(\omega)\,\frac{\omega^2}{c^2}\,. \tag{12.41}$$

is satisfied, and we remind the reader that $\mathbf{p}_{\boldsymbol{\ell}} = \mathbf{p}_0 + \mathbf{G}^{(\boldsymbol{\ell})}$ and ϵ_j is either the dielectric function of the medium of incidence or that of the substrate. Applying this formula to the case of vacuum

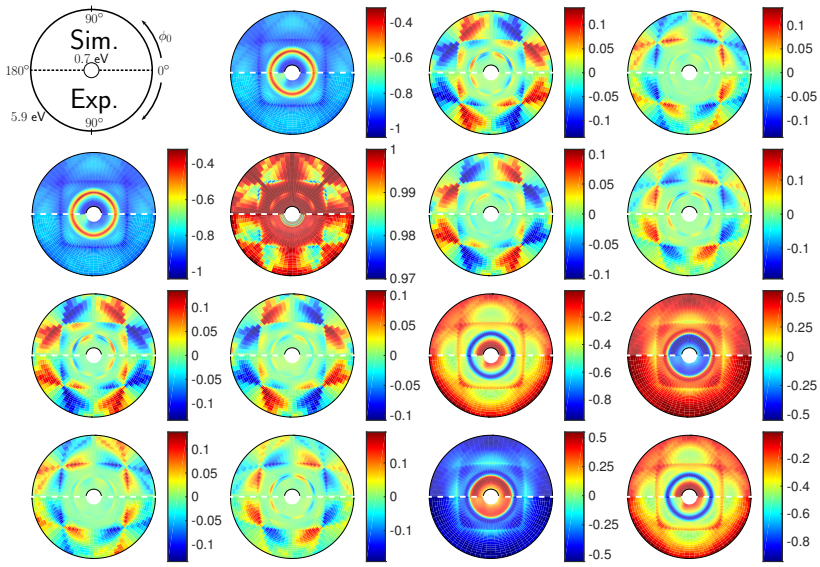


Figure 12.2: Normalized Mueller matrix elements as functions of the photon energy $\hbar\omega$ (radial variable) ranging from 0.7 eV to 5.9 eV, the azimuthal angle of incidence ϕ_0 (angular variable) ranging from 0° to 180° and for a fixed polar angle of incidence $\theta_0 = 55°$. The upper half of each map shows the result of the simulation for the best fit in Table 12.1 and the second half the experimental results. Note that the azimuthal angle positively spans from 0° to 180° in both cases, in such a way that comparison is made symmetrically with respect to the white dashed line. Figure adapted from **Paper** [6].

($\epsilon_j = \epsilon_1 = 1$) as the medium of incidence and assuming that we are dealing with a rectangular lattice we have

$$\begin{aligned}\frac{\omega^2}{c^2} &= \left|\mathbf{p}_0 + \ell_1 \frac{2\pi}{a_1}\,\hat{\mathbf{e}}_1 + \ell_2 \frac{2\pi}{a_2}\,\hat{\mathbf{e}}_2\right|^2 \\ &= \left(\frac{\omega}{c}\sin\theta_0\,\cos\phi_0 + \ell_1\frac{2\pi}{a_1}\right)^2 + \left(\frac{\omega}{c}\sin\theta_0\,\sin\phi_0 + \ell_2\frac{2\pi}{a_2}\right)^2 ,\end{aligned} \tag{12.42}$$

which by simplifying by $\omega^2/c^2 = 2\pi/\lambda$ can be re-written as

$$\left(\sin\theta_0\,\cos\phi_0 + \ell_1\frac{\lambda}{a_1}\right)^2 + \left(\sin\theta_0\,\sin\phi_0 + \ell_2\frac{\lambda}{a_2}\right)^2 = 1\,. \tag{12.43}$$

For a fixed polar angle of incidence θ_0 and order $\boldsymbol{\ell}$, this equation can be solved for ϕ_0 and $\hbar\omega$ (or equivalently λ) and fitted to the experimental data points in a least-mean square sense to obtain the optimal lattice constants a_1 and a_2. This method yields the following optimal lattice constants: $a_1^{\text{opt}} = 205.6$ nm and $a_2^{\text{opt}} = 210.9$ nm. These values agree remarkably well with the estimates obtained from SEM images. The corresponding Rayleigh lines are shown as red solid lines in Fig. 12.3 together with a region of uncertainty (dashed lines) defined as the Rayleigh lines one would obtain for a variation of ±2 nm from the optimal values of the lattice constants. The choice of $\Delta a = 2$ nm is arbitrary but serves the purpose of illustrating that it is indeed about the correct order of uncertainty one hopes to achieve in view of the experimental data

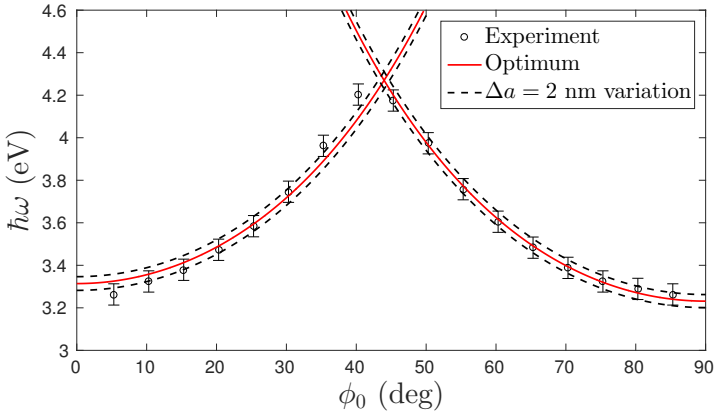


Figure 12.3: Location of the Rayleigh-Wood anomalies associated with the reflected diffractive orders $\boldsymbol{\ell} = (-1, 0)$ and $\boldsymbol{\ell} = (0, -1)$ in the $(\phi_0, \hbar\omega)$ plane for a fixed polar angle of incidence of $\theta_0 = 55°$. Open circles correspond to positions extracted from the experimental data, the red solid lines are the predicted Rayleigh lines for the optimum parameters: $a_1 = 205.6\,\text{nm}$ and $a_2 = 210.9\,\text{nm}$, and the black dashed lines delimit an uncertainty region for the predicted Rayleigh lines based on an uncertainty in lattice constants of $\Delta a = 2\,\text{nm}$ (i.e. $a_i = a_i^{(\text{opt})} \pm \Delta a$).

points. Note that here, the determination of the lattice constants is a fast process since it does not require to solve any forward scattering problem.

The lattice constants are now held fixed at their optimal values. The remaining parameters are obtained by minimizing the following objective function

$$\chi^2(\mathbf{v}) = \frac{1}{2} \sum_{\mathbf{p}_0,\omega} \sum_{i,j} \left[m_{ij}(\mathbf{p}_0, \omega|\mathbf{v}) - \widetilde{m}_{ij}(\mathbf{p}_0, \omega) \right]^2 . \tag{12.44}$$

Here, $m_{ij}(\mathbf{p}_0, \omega|\mathbf{v})$ denotes the ij-element of the normalized Mueller matrix obtained from the model for a *given* set of morphological parameters $\mathbf{v}$, that depends on the lateral wave vector of incidence $\mathbf{p}_0$ [or angles of incidence (θ_0, ϕ_0)] and the photon energy $\hbar\omega$; the corresponding measured normalized Mueller matrix element is denoted $\widetilde{m}_{ij}(\mathbf{p}_0, \omega)$. The outer sum that appears in Eq. (12.44) was performed over a set of 1501 values in the $(\phi_0, \hbar\omega)$-plane since $|\,\mathbf{p}_0\,|$ (or θ_0) was assumed constant in the experiment. This set was constructed by selecting 19 values for the azimuthal angle of incidence $\phi_0 \in [0°, 90°]$ with a step of 5°; and 79 photon energies chosen uniformly distributed between 1.5 eV to 5.9 eV. In the definition of $\chi^2(\mathbf{v})$ only the elements $(i, j) \in \{(1, 2), (3, 3), (3, 4)\}$ were taken into account in the inner sum of Eq. (12.44) in sets of block-diagonal Mueller matrix elements related by symmetries [142]. Note that the off-block-diagonal elements that show the polarization coupling were not included.

Remark 12.4. Keeping only these three terms is motivated by the following:

- First, notice that the Mueller matrix elements are normalized by the total intensity, hence $m_{11} = 1$, and is therefore irrelevant.
- The term m_{22} is close to unity here and features are rather poor and noisy in the experimental data. We have consequently chosen to discard this element.

Model	$r_{\parallel}$ (nm)	$r_{\perp}$ (nm)	h (nm)	f	χ^2
SEM, AFM	58 ± 4	36 ± 5	15 ± 10	–	–
Au/SiO_2	59.7	39.9	–	–	18.9
ELM -10	59.3	38.7	**10**	0.540	12.4
ELM -20	59.9	37.4	**20**	0.458	9.0
ELM -opt	59.6	36.7	33.5	0.517	6.4

Table 12.1: Microscopy and reconstructed morphological parameters for the different models assuming lattice parameters $a_1 = 205.6\,\text{nm}$ and $a_2 = 210.9\,\text{nm}$. Bold entries indicate parameters that were kept constant in the optimization.

- We have chosen to discard the off block-diagonal terms, related to cross-polarization, because they are weaker than the block diagonal ones. In addition, it is a common practice to focus on the block diagonal elements (or so called NCS elements) in the ellipsometry community, and so did we. However, we want to stress that keeping these elements would not change significantly the resulting parameters.
- Due to the symmetry of the system, it is expected that the following symmetries hold (or approximately hold): $m_{12} = m_{21}$, $m_{33} = m_{44}$, $m_{34} = -m_{43}$ for the block diagonal terms. Thus, to avoid redundancy in the data, we have preferred to only keep m_{12}, m_{33} and m_{34}.

We would like to stress that this choice is arbitrary, and that one may keep all terms and obtain similar results for the parameter retrieval problem within a couple of nanometers for the particle radii. As an interesting digression to this discussion, we have experienced that *only* keeping the off block-diagonal and *discarding* the block-diagonal elements leads to erroneous parameters. We speculate that the reason for such a behavior is that the off block diagonal elements are too sensitive to the precise shape of the particles and substrate corrugation. Thus one would try to fit for small details before fitting for the big ones. This is why we rather claimed that the block diagonal elements were the relevant elements for the problem at hand, and that using the off block diagonal elements could be useful for getting a more accurate idea of the shape, but then the experimental reproducibility of the particles shape over such a large sample area seems unrealistic at this stage and the assumption of incident plane wave instead of a finite size beam may also become a source of error at this point anyway.

Numerical details – The minimization of the cost function $\chi^2(\mathbf{v})$, for this and later models, was performed using the Levenberg-Marquardt algorithm where the Jacobian was calculated by a finite-difference approach as described in Section 12.2. The dielectric functions were obtained from oscillator fits to multiple data sets of SiO_2 [143] and inversion of ellipsometric measurements on the 40 nm thick uniform Au film performed prior to milling. The truncation in reciprocal space for the numerical RRE was made such that mode indices whose corresponding reciprocal lattice vectors lying within a circular domain of radius $G_{\max} = N\max(2\pi/a_1, 2\pi/a_2)$ with $N = 15$ (see Section 5.2.1). This choice for the truncation cut-off yields 729 diffractive modes per polarization state of the incident light and we experienced that such an amount of modes were enough to ensure convergence of the optical response on a wide range of photon energies. Only for energies below 2 eV the convergence of the numerical solution of the RRE with an increasing number of modes seems not to be guaranteed, as can be

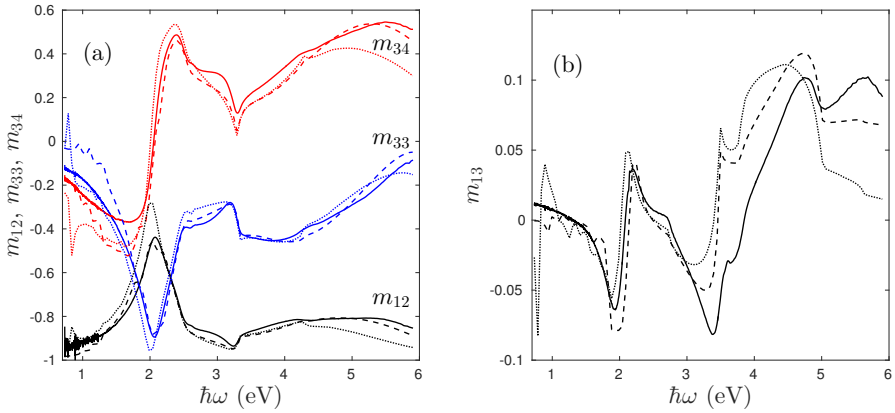


Figure 12.4: Mueller elements m_{12}, m_{33}, m_{34} (a) and m_{13} (b) as a function of photon energy for $\phi_0 = 0°$ (a), and $\phi_0 = 20°$ (b) (polar angle of incidence fixed at $\theta_0 = 55°$). Experimental data, simulations for the best fit of the model Au/SiO_2 (1st optimum in Table 12.1) and that of the ELM-opt (last row in Table 12.1) are plotted respectively as solid lines, dotted lines and dashed lines. Figure adapted from **Paper** [6].

seen in Fig. 12.4, with the presence of spurious oscillations. We identified that this range of energies, below 2 eV, corresponds to a range in which the ratio of the imaginary to real part of the dielectric function of gold becomes negatively large. Nevertheless, the number of points affected by this issue is small compared with the total number of points used for the optimization problem and, hence, the retrieved parameters are not assumed to be significantly altered. On the range of energies where convergent results were obtained, we have found that the optical response obtained by solving the RRE is, within insignificant numerical errors, identical to that obtained by the finite element method implemented in the commercial software COMSOL. The CPU time required to solve the RRE for a given photon energy and angles of incidence was 2.5 s on a desktop computer (Intel i7-5930K 3.5 GHz) for the system of interest. We have experienced that this corresponds to a speedup by *two orders of magnitude* compared with COMSOL to obtain equivalent results. Such a speedup is, of course, critical if one wants to achieve real-time growth monitoring for example. The memory footprint for the RRE method is also minor, as we found a ratio of 1:60 between the two methods.

Results – A first set of inversion results were obtained under the assumption that the Au hemi-spheroids were supported by a *planar* SiO_2 substrate. The parameters that one intends to retrieve are therefore, $\mathbf{v} = (r_\parallel, r_\perp)$. The reconstruction gave the values $r_\parallel = 59.7\,\text{nm}$ and $r_\perp = 39.9\,\text{nm}$ (see Table 12.1; row labeled Au/SiO_2). These values agree rather well with those estimated from microscopy. Note that the out-of-plane radius seems to be overestimated. The photon energy dependence of the resulting Mueller matrix elements that contributes to $\chi^2(\mathbf{v})$ are presented as dotted lines in Fig. 12.4(a) for $\phi_0 = 0°$, and they show good agreement with the corresponding measured data (solid lines). It is observed that the energy of the LSPR at 2.1 eV is well reproduced by the Au/SiO_2 model. The same is true for the location of Rayleigh-Wood anomalies, consistent with the proposed approach for the determination of the lattice parameters a_i. Figure 12.4(b) depicts the energy dependence of the off-block

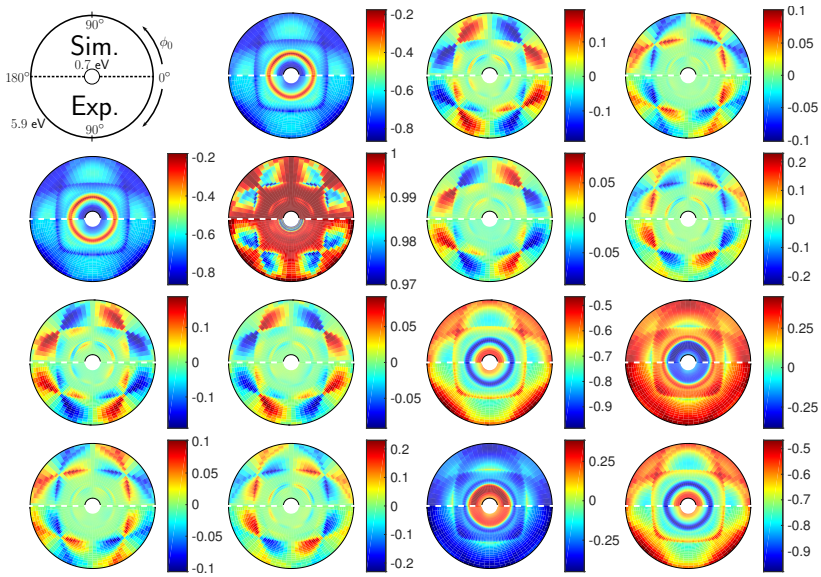


Figure 12.5: Same as Fig. 12.2 but for a fixed polar angle of incidence $\theta_0 = 45°$.

diagonal element m_{13} for $\phi_0 = 20°$. Although this element was not used in the optimization, a qualitatively good agreement is found between the experiment and the Au/SiO_2 model. Similar results were observed for the other Mueller matrix elements and/or other values of ϕ_0 and θ_0 (results not shown).

From the results presented in Fig. 12.4, it is observed that the agreement between the measured and the Au/SiO_2 model results are best in the low energy region. We speculate that the poorer agreement observed for high energies is mainly due to the non-planar features of the surface of the substrate that is caused by over-milling. This seems to also be the reason for the overestimation of the out-of-plane radius of the gold particles. Indeed, we can easily imagine that the mound, present in the experimental sample, acts at high energies as making an effective dielectric particles together with the gold particles. At high energies, the real part of the dielectric function of gold is negative but small, hence the metallic character of the particle is somewhat weak. Consequently, the gold particle and its mound become an effective dielectric particle bigger than the original particle. Hence the optimization scheme tries to catch this feature by increasing the out-of-plane radius of the gold particle. This speculation can be motivated by simulations of bigger dielectric particles on the silica substrate (not shown here), where a rather good agreement with the experiment can be found at high energy (but not at low energies since the LSPR would be absent for example).

As a remedy, a simple effective layer characterized by its thickness h and filling fraction f between air and glass was used to model the latter over-milling into the glass. We considered three such effective layer models (ELMs) corresponding to fixed thickness $h = 10\,\text{nm}$, $20\,\text{nm}$ and $\mathbf{v} = (r_\parallel, r_\perp, f)$, or variable thickness and $\mathbf{v} = (r_\parallel, r_\perp, h, f)$. Optimization performed on the basis of these models, resulted in the morphological parameters presented in Table 12.1. Of the three considered models, the ELM-opt represents best the measured data in terms of

the lowest value for χ^2. As the thickness of the effective layer is increased to 33.5 nm the value of $r_{\parallel}$ remains stable and the value of $r_{\perp}$ decreases, while the filling fraction f is rather stable. Note that for the best fit (last row in Table 12.1) the out-of-plane radius $r_{\perp}$ seems now to be correctly estimated in view of the value given by microscopy. Figure 12.4 presents as dashed lines the energy dependence of some of the m_{ij}-elements that were obtained from ELM-opt when assuming the parameters in Table 12.1. It is apparent from the results of this figure that ELM-opt better represents the measured data than the Au/SiO_2 model; this is in particular the case in the high energy region.

It is now time to go back to Fig. 12.2, where the full $(\phi_0, \hbar\omega)$ dependence of all normalized Mueller matrix elements obtained from the reduced Rayleigh equation (upper halves of the contour plots) are compared to the measurements (lower halves of the contour plots). The morphological parameters assumed in obtaining the simulation results were those of the ELM-opt model (see Table 12.1). The results of Fig. 12.2 show good agreement between *all* the measured and modeled Mueller matrix elements, and not only those used in the minimization. It is noted that similar results to those presented in Fig. 12.2 were obtained within the Au/SiO_2 model, except for larger discrepancies between measured and modeled data at high energies.

Prediction for other angles of incidence – We have now obtained a set of morphological parameters which give a good agreement between the simulation results and the measurements, but we have only considered experimental measurements for a fixed polar angle of incidence $\theta_0 = 55°$. Can we now predict accurately the optical response of the sample for a different polar angle of incidence θ_0? Yes, we can! This is illustrated in Fig. 12.5 where the optical response for a polar angle of incidence of $\theta_0 = 45°$ is simulated for the parameters obtained in the last row of Table 12.1, i.e. obtained by inverting the measured data for $\theta_0 = 55°$, and compared to measurements made for the same angle of incidence ($\theta_0 = 45°$). The comparison is to be made in the same way as in Fig. 12.2. The agreement is overall very good, thus showing the potential predictive power of the method.

12.6 Reconstruction of statistical properties of randomly rough surfaces

We could apply the method we have presented in Section 12.5 to the reconstruction of the statistical properties of a randomly rough surface. Consider a randomly rough surface separating two dielectric media. If the surface is weakly rough for some considered wavelength λ, we may test some hypothesis regarding the shape of the auto-correlation function, its characteristic correlation length and rms roughness. For example, we may assume that the auto-correlation function of the surface is Gaussian and find the values of the correlation length a and rms roughness σ by minimizing the error between some measured data and the corresponding simulated response. A significant difference compared to the previous case, is that the computation of the average optical response may require a high computational cost due to the fact that we will need to average the optical response over many realizations of the surface profile and that solving the optical response for each realization may be costly since in such a case of study, we need to discretize a surface whose size is of the order of many wavelength. If the surface is weakly rough, one may expect to obtain accurate enough statistical parameters by using analytically averaged closed form approximate expressions for the averaged optical response. Such approximations

like small amplitude perturbation theory for example have been discussed in Chapter 8, and will yield results for the statistical parameters with an error that will typically increase with increasing surface roughness. Such a method may be good enough for weakly rough surfaces and very computationally efficient, as the one presented for example in Refs. [144, 145, 146] based on a method known as phase perturbation theory and applied on synthetic data. For surfaces for which the deviation of these approximate methods to the correct optical response becomes significant, we need methods with better accuracy. A good strategy to treat this issue could be to implement a two-stage method. First, one may minimize an objective function by using small amplitude perturbation theory or the first iterate of the method of FIRES in order to obtain quickly a first approximation of the statistical parameters a and σ. Then, starting from this approximate optimum, one may continue the minimization based on an accurate modeling of the optical response by using the iterative Monte Carlo method together with the reduction of variance method based on a control variable as presented in Chapter 8.

12.7 Summary

The last chapter closes this thesis with a rather engineering touch. We have treated the problem of inverse scattering which is of interest both for optical surface characterization, surface growth monitoring during a fabrication process and for the design of surface with desired optical response. After posing the inverse problem in terms of an optimization problem, we have focused our attention on the Levenberg-Marquardt method of optimization which is well adapted to non-linear least-square problems and which makes use of the Jacobian of the residual vector. This method was implemented together with the reduced Rayleigh equations solver and shown to give a rather powerful method for inversion of experimental data. The method was successfully tested on experimental angle resolved spectroscopic Mueller matrix ellipsometry data for a plasmonic photonic crystal. The size of, and separation between the gold particles could be resolved within nanometric accuracy and the obtained simulated optical response is in very good agreement with the experimental data. The case of parameter retrieval of the statistical properties of randomly rough surface was also discussed. Other optimization methods were also discussed stressing their advantages and drawbacks. We have identified a rather new optimization method, the adjoint method, to be of great potential for future applications of sample design as it allows to compute the Jacobian of the objective function at a cost which is independent of the number of parameters.

Conclusions and outlook

It is now time to conclude this thesis by summarizing the main achievements of the presented work, but also the questions that remain unanswered and by suggesting potential research tracks.

In this work, we have built a versatile framework for the derivation of the reduced Rayleigh equations associated with the problem of electromagnetic wave scattering by multi-layer systems composed of stacks of linear, homogeneous, isotropic, non-magnetic materials bounded by arbitrary surfaces. The framework has been applied to study the scattering of electromagnetic waves by periodic and randomly rough surfaces. We have seen that in a number of scattering systems, the solution of the reduced Rayleigh equations gives numerical results and approximation formulae which can lead to the understanding of new scattering phenomena such as the optical Yoneda effect, the Brewster scattering effect, or the selective enhancement of Selényi rings for example. In addition, the reduced Rayleigh equations are not only useful for a qualitative description of scattering systems but yield quantitatively accurate optical response of complex systems such as photonic plasmonic surfaces, for which we have demonstrated that a method based on the reduced Rayleigh equations and an optimization scheme can be used for fast critical dimension metrology of metasurfaces within nanometric accuracy.

Despite its remarkable efficiency for systems which are not too rough, in some sense, the method based the reduced Rayleigh equations suffers from some limitations. First, there seems to be a lack of mathematical study of the domain of validity of the reduced Rayleigh equations for penetrable media, and on its numerical analysis in the literature. We have attempted such a numerical analysis in this work, which must be considered as a preliminary exploration and which we hope will motivate the development of deeper and rigorous study. In particular, important questions are those of the convergence rate of the numerical solution with the number of modes, and the consistency of the numerical solution with the initial scattering problem, where error bounds depending on the system parameters would be of particular interest.

All the limitations are not, however, purely theoretical. The framework that we have presented in this work may be very versatile (and maybe even elegant), the computational cost associated with the problem of the scattering of light by a stack containing more than two randomly rough surfaces remains a challenge. The computational cost of such a simulation makes such an approach unpractical and efficient approximation methods must be developed. The bottleneck for simulating such systems relies on the computation of the kernel which is defined as many successive integrals. We are thus facing the so-called curse of dimensionality. Monte Carlo methods are known to be among the best approaches for evaluating high-dimensional integrals and we therefore believe that such methods should be adapted to tackle the issue of the computation of the reduced Rayleigh equations kernel for multi-layer systems. In particular, we believe that Monte-Carlo methods based on the transition probability amplitudes as discussed

in Sections. 2.5 and 3.5.3 seem to be promising candidates.

In addition, the next step of generalization one can take concerning the reduced Rayleigh equations could be to include magnetic media, and/or anisotropic media. Allowing for the simulation of layers with different anisotropic media could be of particular interest for at least two reasons. First, it would help to understand physical mechanism involved in the coupling of ordinary and extraordinary waves at structured interfaces for example. Second, allowing for anisotropic material may reduce the complexity associated with the simulation of complex systems. Indeed, we could imagine having a system composed of a stack of layers with plasmonic and dielectric particles embedded in the layers. In the low frequency regime, one could derive the anisotropic dielectric tensor corresponding to each layer via homogenization methods and then use these homogenized anisotropic media as inputs for the reduced Rayleigh equations.

There are nevertheless interesting applications that can already be handled by the present methods. As illustrated in the present work, the method seems to be a promising candidate for fast optical characterization of surfaces and for the design of optical metasurfaces. We believe that combining the method based on the reduced Rayleigh equations with an optimization scheme based on the adjoint method should lead to a very efficient designer algorithm. Indeed, the adjoint method would remove the complexity associated with evaluating the Jacobian of the objective function, hence allowing for treating each discretized point of the surface as a parameter. One could also think of using the reduced Rayleigh equations to generate optical responses from random systems and use such a set of systems and optical responses to train a neural network for example. This could lead to fast optical characterization via machine learning.

Appendix A

The ξ-integral and Fourier moments for a collection of shapes

A.1 Introduction

Setting up the linear system of equations for solving numerically the reduced Rayleigh equation in the case of a surface composed of a periodic array of objects on a plane requires the evaluation of a significant number of the ξ-integrals (see e.g. Chapter 5). Analytic expressions of the ξ-integral are derived for particular surface profiles in this appendix. We first deal with the case of cylindrically and 'elliptically' symmetric shapes: cylinder, rings, cone, paraboloid, hemiellipsoid; and triangular prism which opens the possibility to approach any shape by polyhedra with triangular facets.

Before embarking in the derivation of the ξ-integral for this zoo of surface profiles, we discuss a few general properties, which are rather straightforward but nonetheless fundamental, and prove to be useful in practice when it comes to combining different shapes and the implementation of the perturbation theory.

A.2 A few fundamental observations

A.2.1 Translation property

We recall the definition of the ξ-integral, $\xi^{(\mathbf{m})}(\gamma)$, where $\mathbf{m} \in \mathbb{Z}^2$, $\gamma \in \mathbb{C}$. If ζ denotes a periodic surface profile, then $\xi^{(\mathbf{m})}(\gamma)$ is defined as

$$\xi^{(\mathbf{m})}(\gamma) = \frac{1}{a_c} \int_{a_c} e^{-i\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{x}} e^{-i\gamma\zeta(\mathbf{x})} \mathrm{d}^2x \,, \tag{A.1}$$

where $\mathbf{G}^{(\mathbf{m})}$ is a reciprocal lattice vector and a_c will denote both the unit cell and its area. Let us analyze the effect on $\xi^{(\mathbf{m})}$ of a translation of vector $\mathbf{x}_0$ of the surface profile. We will denote

$\xi^{(\mathbf{m})}_{\mathbf{x}_0}$ the ξ-integral of the translated profile $\zeta(\mathbf{x}-\mathbf{x}_0)$.

$$\begin{aligned}\xi^{(\mathbf{m})}_{\mathbf{x}_0}(\gamma) &= \frac{1}{a_c}\int_{a_c} e^{-i\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{x}} e^{-i\gamma\zeta(\mathbf{x}-\mathbf{x}_0)}\mathrm{d}^2x \\ &= \frac{1}{a_c}\int_{a_c} e^{-i\mathbf{G}^{(\mathbf{m})}\cdot(\mathbf{y}+\mathbf{x}_0)} e^{-i\gamma\zeta(\mathbf{y})}\mathrm{d}^2y \\ &= \frac{1}{a_c}\int_{a_c} e^{-i\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{y}} e^{-i\gamma\zeta(\mathbf{y})}\mathrm{d}^2y\, e^{-i\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{x}_0} \\ &= \xi^{(\mathbf{m})}(\gamma)\, e^{-i\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{x}_0}\,. \end{aligned} \tag{A.2}$$

The translation of the profile by a vector $\mathbf{x}_0$ induces a phase factor $e^{-i\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{x}_0}$. Therefore in the following it will be useful to choose a convenient origin to compute the ξ-integral, defined as a center of symmetry for instance, and then apply a phase factor if a translation of the profile is needed. This proves to be useful when combining shapes with disjoint support as discussed in the next section.

A.2.2 Union of shapes with disjoint supports

Consider a surface profile ζ defined in the unit cell as a sum of n functions $(\zeta_j)_{j\in[\![1,n]\!]}$ having disjoint compact supports included in the unit cell, i.e.

$$\begin{aligned} &\zeta|_{a_c} = \sum_{j=1}^{n}\zeta_j \\ &\operatorname{supp}\zeta_j \subset a_c\,, \qquad\qquad \forall j\in[\![1,n]\!] \\ &\operatorname{supp}\zeta_i \bigcap \operatorname{supp}\zeta_j = \varnothing\,, \quad \forall (i,j)\in[\![1,n]\!]^2,\ i\neq j\,. \end{aligned} \tag{A.3}$$

Using these properties in Eq. (A.1) yields

$$\begin{aligned} a_c\,\xi^{(\mathbf{m})}(\gamma) &= \int_{a_c} e^{-i\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{x}} e^{-i\gamma\sum_{j=1}^{n}\zeta_j(\mathbf{x})}\mathrm{d}^2x \\ &= \int_{a_c} e^{-i\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{x}} \prod_{j=1}^{n} e^{-i\gamma\zeta_j(\mathbf{x})}\mathrm{d}^2x \\ &= \int_{a_c\backslash\cup_j\operatorname{supp}\zeta_j} e^{-i\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{x}} \prod_{j=1}^{n} e^{-i\gamma\zeta_j(\mathbf{x})}\mathrm{d}^2x + \int_{\cup_j\operatorname{supp}\zeta_j} e^{-i\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{x}} \prod_{j=1}^{n} e^{-i\gamma\zeta_j(\mathbf{x})}\mathrm{d}^2x\,. \end{aligned}$$

Now by using that on one hand

$$\zeta_j(\mathbf{x}) = 0, \qquad \forall j\in[\![1,n]\!],\ \forall\mathbf{x}\in a_c\backslash\operatorname{supp}\zeta_j\,,$$

and on the other hand since the supports are disjoint we have

$$\int_{\cup_j\operatorname{supp}\zeta_j}\cdots\,\mathrm{d}^2x = \sum_{j=1}^{n}\int_{\operatorname{supp}\zeta_j}\cdots\,\mathrm{d}^2x\,,$$

we obtain

$$a_c\,\xi^{(\mathbf{m})}(\gamma) = \int_{a_c\backslash\cup_j\operatorname{supp}\zeta_j} e^{-i\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{x}}\,\mathrm{d}^2x + \sum_{j=1}^{n}\int_{\operatorname{supp}\zeta_j} e^{-i\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{x}}\, e^{-i\gamma\zeta_j(\mathbf{x})}\,\mathrm{d}^2x\,.$$

Now we add and subtract $\int_{\cup_j \operatorname{supp}\zeta_j} e^{-i\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{x}}\,\mathrm{d}^2x$ in order to complete the first term so that we get

$$a_c\,\xi^{(\mathbf{m})}(\gamma) = \int_{a_c} e^{-i\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{x}}\,\mathrm{d}^2x + \sum_{j=1}^{n}\int_{\operatorname{supp}\zeta_j} e^{-i\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{x}}\left(e^{-i\gamma\zeta_j(\mathbf{x})}-1\right)\mathrm{d}^2x\,.$$

The integral $a_c\,\xi^{(\mathbf{m})}(\gamma)$ is then the sum of two terms

$$a_c\,\xi^{(\mathbf{m})}(\gamma) = \mathcal{C}^{(\mathbf{m})} + \mathcal{G}^{(\mathbf{m})}(\gamma)\,, \tag{A.4}$$

where we define the *cell-term* $\mathcal{C}^{(\mathbf{m})}$ as

$$\mathcal{C}^{(\mathbf{m})} = \int_{a_c} e^{-i\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{x}}\,\mathrm{d}^2x\,, \tag{A.5}$$

and the *geometrical-term* $\mathcal{G}^{(\mathbf{m})}(\gamma)$ as

$$\mathcal{G}^{(\mathbf{m})}(\gamma) = \sum_{j=1}^{n}\int_{\operatorname{supp}\zeta_j} e^{-i\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{x}}\left(e^{-i\gamma\zeta_j\,(\mathbf{x})}-1\right)\mathrm{d}^2x\,. \tag{A.6}$$

Note that the cell-term is only dependent[1] on the cell and hence will always be present and identical for any choice of profile ζ. By decomposition of $\mathbf{x}$ on the direct lattice basis vectors $(\mathbf{a}_1,\mathbf{a}_2)$ and $\mathbf{G}^{(\mathbf{m})}$ on the reciprocal lattice basis vectors $(\mathbf{b}_1,\mathbf{b}_2)$, the scalar product becomes

$$\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{x} = (m_1\mathbf{b}_1 + m_2\mathbf{b}_2)\cdot(x_1\mathbf{a}_1 + x_2\mathbf{a}_2) = 2\pi(m_1x_1 + m_2x_2)\,. \tag{A.7}$$

Here the property of direct-reciprocal basis vectors $\mathbf{a}_i\cdot\mathbf{b}_j = 2\pi\delta_{ij}$ has been used. Substituting Eq. (A.7) in Eq. (A.5), we obtain

$$\mathcal{C}^{(\mathbf{m})} = a_c\int_{-1/2}^{1/2} e^{-2\pi i m_1x_1}\mathrm{d}x_1\int_{-1/2}^{1/2} e^{-2\pi i m_2x_2}\mathrm{d}x_2\,. \tag{A.8}$$

If $m_1 = m_2 = 0$, $\mathcal{C}$ reduces to the unit cell area a_c. If at least one of the component m_i of $\mathbf{m}$ is non zero, the corresponding integral vanishes

$$\int_{-1/2}^{1/2} e^{-2\pi i(m_ix_i)}\mathrm{d}x_i = \int_{-m_i\pi}^{m_i\pi} e^{iu}\mathrm{d}u = 0\,, \tag{A.9}$$

since, it corresponds to the integrations of cosine and sine over an integer number of periods. To sum up, the cell term is simply

$$\mathcal{C}^{(\mathbf{m})} = a_c\,\delta_{\mathbf{m},\mathbf{0}}\,, \tag{A.10}$$

where $\delta_{\mathbf{m},\mathbf{n}} = \delta_{m_1,n_1}\,\delta_{m_2,n_2}$ and $\delta_{i,j}$ is the Kronecker delta. Consequently the ξ-integral is of the form

$$\xi^{(\mathbf{m})}(\gamma) = \delta_{\mathbf{m},\mathbf{0}} + \frac{1}{a_c}\,\mathcal{G}^{(\mathbf{m})}(\gamma)\,. \tag{A.11}$$

We can re-write $\mathcal{G}^{(\mathbf{m})}(\gamma)$ as a sum the geometrical terms of each ζ_j as if there were alone

$$\mathcal{G}^{(\mathbf{m})}(\gamma) = \sum_{j=1}^{n}\mathcal{G}_j^{(\mathbf{m})}(\gamma). \tag{A.12}$$

[1] given $\mathbf{G}^{(\mathbf{m})}$

If we now assume that the profile ζ_j has a center of symmetry at $\mathbf{x}_j$ and that one knows the geometrical term of the same profile centered at the origin, $\mathcal{G}_{j,o}^{(\mathbf{m})}(\gamma)$, then using the translation property we have seen in the previous section (or adapting the change of variable $\mathbf{y} = \mathbf{x} - \mathbf{x}_j$ to $\mathcal{G}_j^{(\mathbf{m})}(\gamma)$) we get

$$\mathcal{G}^{(\mathbf{m})}(\gamma) = \sum_{j=1}^{n} \mathcal{G}_{j,o}^{(\mathbf{m})}(\gamma)\, e^{-i\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{x}_j}, \tag{A.13}$$

where

$$\mathcal{G}_{j,o}^{(\mathbf{m})}(\gamma) = \int_{\operatorname{supp}\zeta_{j,o}} e^{-i\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{x}} \left(e^{-i\gamma\zeta_{j,o}(\mathbf{x})} - 1\right) \mathrm{d}^2x\,. \tag{A.14}$$

Here we denote $\zeta_{j,o}(\mathbf{x}) = \zeta_j(\mathbf{x} - \mathbf{x}_j)$ and $\operatorname{supp}\zeta_{j,o} = \operatorname{supp}\zeta_\mathbf{j} - \mathbf{x}_\mathbf{j} = \{\mathbf{x} - \mathbf{x}_j | \mathbf{x} \in \operatorname{supp}\zeta_j\}$. In the next sections, we will thus focus on computing the geometrical terms for a collection of shapes centered at the origin.

A.2.3 ξ-integral as the Fourier moments generator

As we will see in the next sections, the expression of the ξ-integral will often be the sum of a series of the form

$$\xi^{(\mathbf{m})}(\gamma) = \sum_{n=0}^{\infty} a_n^{(\mathbf{m})}\,(-i\gamma)^n\,. \tag{A.15}$$

In Section 6.2.1, we saw that the resolution of the reduced Rayleigh equations by the use of the perturbation theory required the knowledge of the Fourier transforms of the powers of the surface profile restricted to the unit cell, so called Fourier moments. How do these relate to the a_n? To answer this question it suffices to expand the exponential in the definition of the ξ-integral and use the unicity of the analytical series representation. We have

$$\begin{aligned}\xi^{(\mathbf{m})}(\gamma) &= \frac{1}{a_c}\int_{a_c} e^{-i\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{x}} e^{-i\gamma\zeta(\mathbf{x})}\,\mathrm{d}^2x \\ &= \sum_{n=0}^{\infty} \frac{(-i\gamma)^n}{n!}\,\frac{1}{a_c}\int_{a_c} e^{-i\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{x}}\,\zeta^n(\mathbf{x})\,\mathrm{d}^2x\,.\end{aligned} \tag{A.16}$$

Therefore

$$\hat{\zeta}_{a_c}^{(n)}(\mathbf{G}^{(\mathbf{m})}) = n!\,a_n^{(\mathbf{m})}\,. \tag{A.17}$$

An equivalent way to look at it is to take the derivative of the ξ-integral with respect to γ and evaluate it at $\gamma = 0$. The conclusion of this straightforward observation is that the Fourier moments needed in the perturbation theory can be read directly from the terms of the series expansion of the ξ-integral.

A.3 One-dimensional surfaces

In this section, we derive the ξ-integral for one-dimensional surfaces. By one-dimensional surfaces we mean a two-dimensional surface whose profile is constant along the x_2-direction, i.e. invariant by any translation along the x_2-direction. Since we use the ξ-integral in the 2 dimensional periodic framework, one needs to define a period a_2 in the x_2-direction (this can be done

arbitrarily). However, these expressions can easily be adapted to the simpler one-dimensional reduced Rayleigh equations, by simply ignoring the $\delta_{m_2,0}$ factors.

A.3.1 Sinusoidal surface

We consider a one-dimensional sinusoidal surface of amplitude H, and period a_1 in the x_1-direction. The surface profile is constant along the x_2-direction. It can be described as

$$\zeta(\mathbf{x}) = H \sin\left(\frac{2\pi}{a_1} x_1\right) . \tag{A.18}$$

By plugging the above profile into Eq. (A.1) we have

$$\xi^{(\mathbf{m})}(\gamma) = \frac{1}{a_c} \int_{a_c} e^{-i\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{x}} e^{-i\gamma H \sin\left(\frac{2\pi}{a_1} x_1\right)} \mathrm{d}^2 x . \tag{A.19}$$

By making the linear change of variables $\hat{x}_i = 2\pi x_i / a_i$, $i \in \{1, 2\}$, we get

$$\begin{aligned} \xi^{(\mathbf{m})}(\gamma) &= \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} e^{-im_2\hat{x}_2} \,\mathrm{d}\hat{x}_2 \int_{-\pi}^{\pi} e^{-im_1\hat{x}_1 - i\gamma H \sin(\hat{x}_1)} \,\mathrm{d}\hat{x}_1 \\ &= \frac{\delta_{m_2,0}}{2\pi} \int_{-\pi}^{\pi} e^{-im_1\hat{x}_1 - i\gamma H \sin(\hat{x}_1)} \,\mathrm{d}\hat{x}_1 \\ &= \delta_{m_2,0} \, J_{-m_1}(\gamma H) . \end{aligned} \tag{A.20}$$

Here we have recognized the integral representation of a Bessel function of the first kind of order $-m_1$. Note that we can also use the identity $J_{-\nu} = (-1)^{\nu} J_{\nu}$ to obtain

$$\xi^{(\mathbf{m})}(\gamma) = (-1)^{m_1} \, \delta_{m_2,0} \, J_{m_1}(\gamma H) . \tag{A.21}$$

A.3.2 Rectangular box

A one-dimensional rectangular profile of full width $w < a_1$ and height H is now considered. The profile is represented as

$$\zeta(\mathbf{x}) = H \, \mathbb{1}_{[-w/2, w/2]}(x_1) , \tag{A.22}$$

where we recall that $\mathbb{1}_{[-w/2,w/2]}$ is the indicator function of the segment $[-w/2, w/2]$. For this profile, the ξ-integral reads

$$\xi^{(\mathbf{m})}(\gamma) = \frac{1}{a_c} \int_{a_c} e^{-i\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{x}} e^{-i\gamma H \, \mathbb{1}_{[-w/2,w/2]}(x_1)} \mathrm{d}^2 x . \tag{A.23}$$

As mentioned in Section A.2.2, the profile as a compact support along the x_1-direction. It is therefore practical to directly use the cell-term and geometrical-term splitting

$$\xi^{(\mathbf{m})}(\gamma) = \delta_{m_2,0} \left(\delta_{m_1,0} + \int_{-w/2a_1}^{w/2a_1} e^{-i2\pi m_1 \hat{x}_1} \left(e^{-i\gamma H} - 1 \right) \mathrm{d}\hat{x}_1 \right) . \tag{A.24}$$

Here we have also directly integrated over the x_2-direction and made the change of variable $\hat{x}_1 = x_1 / a_1$. The factor in parenthesis in the integrand of the above integral being constant, the integral becomes simple and gives

$$\xi^{(\mathbf{m})}(\gamma) = \delta_{m_2,0} \left[\delta_{m_1,0} + \left(e^{-i\gamma H} - 1 \right) \frac{w}{2a_1} \left(\delta_{m_1,0} + (1 - \delta_{m_1,0}) \operatorname{sinc}\left(\frac{m_1 w}{a_1} \right) \right) \right] . \tag{A.25}$$

A.3.3 Sawtooth

We are now interested in the asymmetric sawtooth of amplitude H and asymmetry factor $\alpha \in]-1/2, 1/2[$ defined as

$$\zeta(\mathbf{x}) = H\, s_\alpha\left(\frac{x_1}{a_1}\right), \tag{A.26}$$

where we define the asymmetric sawtooth function s_α for $x \in [-1/2, 1/2]$ by

$$s_\alpha(x) = \begin{cases} \frac{2(2x-\alpha)+1}{2\alpha+1} & \text{if } x < \alpha \\ \frac{2(2x-\alpha)-1}{2\alpha-1} & \text{if } x > \alpha \end{cases}. \tag{A.27}$$

With these definitions the ξ-integral reads

$$\xi^{(\mathbf{m})}(\gamma) = \frac{1}{a_c}\int_{a_c} e^{-i\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{x}} e^{-i\gamma H\, s_\alpha(x_1/a_1)}\mathrm{d}^2x\,. \tag{A.28}$$

By applying a change of variable $\hat{x}_i = x_i/a_i$, $i \in \{1,2\}$, and after the straightforward integration along x_2-direction, we obtain

$$\xi^{(\mathbf{m})}(\gamma) = \delta_{m_2,0}\int_{-1/2}^{1/2} e^{-i2\pi m_1\hat{x}_1 - i\gamma H\, s_\alpha(\hat{x}_1)}\mathrm{d}\hat{x}_1 \tag{A.29}$$

$$= \delta_{m_2,0}\left(\int_{-1/2}^{\alpha} e^{-i2\pi m_1\hat{x}_1 - i\gamma H\,\frac{2(2\hat{x}_1-\alpha)+1}{2\alpha+1}}\mathrm{d}\hat{x}_1 + \int_{\alpha}^{1/2} e^{-i2\pi m_1\hat{x}_1 - i\gamma H\,\frac{2(2\hat{x}_1-\alpha)-1}{2\alpha-1}}\mathrm{d}\hat{x}_1\right). \tag{A.30}$$

The two integrals are straightforward to perform and read

$$\int_{-1/2}^{\alpha} e^{-i2\pi m_1\hat{x}_1 - i\gamma H\,\frac{2(2\hat{x}_1-\alpha)+1}{2\alpha+1}}\mathrm{d}\hat{x}_1 = i\,\frac{e^{-i2\pi m_1\alpha - i\gamma H} - (-1)^{m_1}e^{i\gamma H}}{2\pi m_1 + \frac{4\gamma H}{2\alpha+1}}$$

$$\int_{\alpha}^{1/2} e^{-i2\pi m_1\hat{x}_1 - i\gamma H\,\frac{2(2\hat{x}_1-\alpha)-1}{2\alpha-1}}\mathrm{d}\hat{x}_1 = -i\,\frac{e^{-i2\pi m_1\alpha - i\gamma H} - (-1)^{m_1}e^{i\gamma H}}{2\pi m_1 + \frac{4\gamma H}{2\alpha-1}},$$

which yields after some simplifications

$$\xi^{(\mathbf{m})}(\gamma) = \delta_{m_2,0}\,\frac{2i\gamma H\left(e^{-i2\pi m_1\alpha - i\gamma H} - (-1)^{m_1}e^{i\gamma H}\right)}{(2\gamma H)^2 + 8\pi m_1\gamma H\alpha + (\pi m_1)^2(4\alpha^2-1)}. \tag{A.31}$$

In the case of a symmetric sawtooth, i.e. $\alpha = 0$, the above equation reduces to

$$\xi^{(\mathbf{m})}(\gamma) = \delta_{m_2,0}\,\frac{2i\gamma H\left(e^{-i\gamma H} - (-1)^{m_1}e^{i\gamma H}\right)}{(2\gamma H)^2 - (\pi m_1)^2}. \tag{A.32}$$

A.3.4 Gaussian profile

Assume that one approximates a bell-like profile by a Gaussian profile concentrated well within the unit cell boundary. In other words, we describe the profile in the unit cell by

$$\zeta(\mathbf{x}) \approx H\,\exp\left(-\frac{x_1^2}{\sigma^2}\right), \tag{A.33}$$

provided that $\sigma \ll a_1$. The ξ-integral reads

$$\xi^{(\mathbf{m})}(\gamma) = \frac{1}{a_c}\int_{a_c} e^{-i\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{x}} e^{-i\gamma H\,\exp\left(-\frac{x_1^2}{\sigma^2}\right)}\mathrm{d}^2x\,. \tag{A.34}$$

By applying a change of variable $\hat{x}_i = x_i/a_i$, $i \in \{1,2\}$, and after the straightforward integration along x_2-direction, we obtain

$$\xi^{(\mathbf{m})}(\gamma) = \delta_{m_2,0} \int_{-1/2}^{1/2} e^{-i2\pi m_1 \hat{x}_1} e^{-i\gamma H \exp\left(-\frac{\hat{x}_1^2}{\hat{\sigma}^2}\right)} \mathrm{d}\hat{x}_1 \,, \tag{A.35}$$

where $\hat{\sigma} = \sigma/a_1$. By expanding $e^{-i\gamma H \exp\left(-\frac{\hat{x}_1^2}{\hat{\sigma}^2}\right)}$ in a power series we get

$$\xi^{(\mathbf{m})}(\gamma) = \delta_{m_2,0} \sum_{n=0}^{\infty} \frac{(-i\gamma H)^n}{n!} \int_{-1/2}^{1/2} e^{-i2\pi m_1 \hat{x}_1} e^{-n\frac{\hat{x}_1^2}{\hat{\sigma}^2}} \mathrm{d}\hat{x}_1 \,. \tag{A.36}$$

For $n = 0$, the integral becomes $\delta_{m_1,0}$. For $n \neq 0$ we have

$$\begin{aligned}
\int_{-1/2}^{1/2} e^{-i2\pi m_1 \hat{x}_1} e^{-n\frac{\hat{x}_1^2}{\hat{\sigma}^2}} \mathrm{d}\hat{x}_1 &= e^{\frac{-\pi^2 m_1^2 \hat{\sigma}^2}{n}} \int_{-1/2}^{1/2} \exp\left[-\left(\frac{\sqrt{n}\hat{x}_1}{\hat{\sigma}} + \frac{i\pi m_1 \hat{\sigma}}{\sqrt{n}}\right)^2\right] \mathrm{d}\hat{x}_1 \\
&= \frac{\hat{\sigma}}{\sqrt{n}} e^{\frac{-\pi^2 m_1^2 \hat{\sigma}^2}{n}} \int_{-\frac{\sqrt{n}}{2\hat{\sigma}} + \frac{i\pi m_1 \hat{\sigma}}{\sqrt{n}}}^{\frac{\sqrt{n}}{2\hat{\sigma}} + \frac{i\pi m_1 \hat{\sigma}}{\sqrt{n}}} \exp\left(-u^2\right) \mathrm{d}u \\
&\approx \frac{\sqrt{\pi}\hat{\sigma}}{\sqrt{n}} e^{-\frac{\pi^2 m_1^2 \hat{\sigma}^2}{n}} = \frac{\sqrt{\pi}\sigma}{\sqrt{n}a_1} e^{-\frac{\pi^2 m_1^2 \sigma^2}{n a_1^2}} \,.
\end{aligned} \tag{A.37}$$

Here we approximated the last integral by the integral over the whole real line. Finally we obtain that

$$\xi^{(\mathbf{m})}(\gamma) \approx \delta_{m_2,0} \left[\delta_{m_1,0} + \frac{\sqrt{\pi}\sigma}{a_1} \sum_{n=1}^{\infty} \frac{(-i\gamma H)^n}{n!\sqrt{n}} \exp\left(-\frac{\pi^2 m_1^2 \sigma^2}{n a_1^2}\right)\right] . \tag{A.38}$$

A.4 Two-dimensional surfaces

A.4.1 Sinusoidal surface

We consider now a two-dimensional sinusoidal surface of amplitude H, of period a_1 in the x_1-direction and a_2 in the x_2-direction, defined as

$$\zeta(\mathbf{x}) = \frac{H}{2}\left[\sin\left(\frac{2\pi}{a_1} x_1\right) + \sin\left(\frac{2\pi}{a_2} x_2\right)\right] . \tag{A.39}$$

The ξ-integral reads

$$\xi^{(\mathbf{m})}(\gamma) = \frac{1}{a_c} \int_{a_c} e^{-i\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{x}} e^{-\frac{i\gamma H}{2}\left(\sin\left(\frac{2\pi}{a_1} x_1\right) + \sin\left(\frac{2\pi}{a_2} x_2\right)\right)} \mathrm{d}^2 x \,. \tag{A.40}$$

We note that the integrations are decoupled. By a change of variable $\hat{x}_i = \frac{2\pi}{a_i} x_i$, $i \in \{1,2\}$, we obtain

$$\xi^{(\mathbf{m})}(\gamma) = \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} e^{-im_1\hat{x}_1 - \frac{i\gamma H}{2}\sin\hat{x}_1} \mathrm{d}\hat{x}_1 \int_{-\pi}^{\pi} e^{-im_2\hat{x}_2 - \frac{i\gamma H}{2}\sin\hat{x}_2} \mathrm{d}\hat{x}_2 \,. \tag{A.41}$$

As encountered in the one-dimensional case, we recognize the integral representation of a Bessel function of the first kind of order $-m_1$ and $-m_2$. The ξ-integral finally reads as a product of Bessel functions

$$\xi^{(\mathbf{m})}(\gamma) = J_{-m_1}\left(\frac{\gamma H}{2}\right) J_{-m_2}\left(\frac{\gamma H}{2}\right) , \tag{A.42}$$

or equivalently

$$\xi^{(\mathbf{m})}(\gamma) = (-1)^{m_1+m_2}\, J_{m_1}\left(\frac{\gamma H}{2}\right)\, J_{m_2}\left(\frac{\gamma H}{2}\right)\,. \tag{A.43}$$

A.4.2 Rectangular box

Let the profile be a cylinder on a rectangular support, i.e. a box, defined by its height H, its full width $w_1 < a_1$ in the x_1-direction and its full width $w_2 < a_2$ in the x_2-direction. The profile function reads

$$\zeta\,(\mathbf{x}) = H\ \mathbb{1}_{[-w_1/2,w_1/2]\times[-w_2/2,w_2/2]}\,(\mathbf{x}) = H\ \mathbb{1}_{[-w_1/2,w_1/2]}(x_1)\ \mathbb{1}_{[-w_2/2,w_2/2]}(x_2)\,. \tag{A.44}$$

Since the profile has a compact support, we can use the cell-term - geometrical-term splitting from Section A.2.2 and the ξ-integral is thus given by

$$\xi^{(\mathbf{m})}(\gamma) = \delta_{\mathbf{m},\mathbf{0}} + \frac{1}{a_c}\int_{[-w_1/2,w_1/2]\times[-w_2/2,w_2/2]} e^{-i\mathbf{G}^{(\mathbf{m})}}\ \left(e^{-i\gamma H}-1\right)\mathrm{d}^2x \tag{A.45}$$

$$= \delta_{\mathbf{m},\mathbf{0}} + \left(e^{-i\gamma H}-1\right)\int_{-w_1/2a_1}^{w_1/2a_1} e^{-i2\pi m_1\hat{x}_1}\,\mathrm{d}\hat{x}_1 \int_{-w_2/2a_2}^{w_2/2a_2} e^{-i2\pi m_2\hat{x}_2}\mathrm{d}\hat{x}_2\,. \tag{A.46}$$

Here we have used the change of variable $\hat{x}_i = x_i/a_i$, $i \in \{1,2\}$ and observed that the integrals are decoupled. The integration is now straightforward and we obtain in a compact form

$$\xi^{(\mathbf{m})}(\gamma) = \delta_{\mathbf{m},\mathbf{0}} + \left(e^{-i\gamma H}-1\right)\ \frac{w_1\,w_2}{a_1\,a_2}\ \left[\delta_{m_1,0} + \frac{1-\delta_{m_1,0}}{2}\ \mathrm{sinc}\left(\frac{m_1 w_1}{a_1}\right)\right] \times \left[\delta_{m_2,0} + \frac{1-\delta_{m_2,0}}{2}\ \mathrm{sinc}\left(\frac{m_2 w_2}{a_2}\right)\right]\,, \tag{A.47}$$

or in a more detailed form

$$\xi^{(\mathbf{m})}(\gamma) = \begin{cases} 1 + \left(e^{-i\gamma H}-1\right)\ \frac{w_1\,w_2}{a_1\,a_2} & \text{if } m_1 = 0 \text{ and } m_2 = 0 \\ \left(e^{-i\gamma H}-1\right)\ \frac{w_1\,w_2}{a_1\,2a_2}\ \mathrm{sinc}\left(\frac{m_2 w_2}{a_2}\right) & \text{if } m_1 = 0 \text{ and } m_2 \neq 0 \\ \left(e^{-i\gamma H}-1\right)\ \frac{w_1\,w_2}{2a_1\,a_2}\ \mathrm{sinc}\left(\frac{m_1 w_1}{a_1}\right) & \text{if } m_1 \neq 0 \text{ and } m_2 = 0 \\ \left(e^{-i\gamma H}-1\right)\ \frac{w_1\,w_2}{2a_1\,2a_2}\ \mathrm{sinc}\left(\frac{m_1 w_1}{a_1}\right)\ \mathrm{sinc}\left(\frac{m_2 w_2}{a_2}\right) & \text{if } m_1 \neq 0 \text{ and } m_2 \neq 0 \end{cases}\,. \tag{A.48}$$

A.4.3 Gaussian profile

Similarly to the one-dimensional case, let us assume that a protuberance can be approximated by a two-dimensional Gaussian profile defined as

$$\zeta(\mathbf{x}) \approx H\ \exp\left(-\frac{x_1^2}{\sigma_1^2} - \frac{x_2^2}{\sigma_2^2}\right)\,, \tag{A.49}$$

provided $\sigma_i \ll a_i$ for $i \in \{1,2\}$. The ξ-integral reads

$$\xi^{(\mathbf{m})}(\gamma) = \frac{1}{a_c}\int_{a_c} e^{-i\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{x}} e^{-i\gamma H\ \exp\left(-\frac{x_1^2}{\sigma_1^2}-\frac{x_2^2}{\sigma_2^2}\right)}\mathrm{d}^2x\,. \tag{A.50}$$

By applying a change of variable $\hat{x}_i = x_i/a_i$, $i \in \{1,2\}$, we obtain

$$\xi^{(\mathbf{m})}(\gamma) = \int_{-1/2}^{1/2}\int_{-1/2}^{1/2} e^{-i2\pi m_1\hat{x}_1 - i2\pi m_2\hat{x}_2} e^{-i\gamma H\ \exp\left(-\frac{\hat{x}_1^2}{\hat{\sigma}_1^2}-\frac{\hat{x}_2^2}{\hat{\sigma}_2^2}\right)}\mathrm{d}\hat{x}_1\mathrm{d}\hat{x}_2\,, \tag{A.51}$$

where $\hat{\sigma}_i = \sigma_i/a_i$ for $i \in \{1,2\}$. By expanding $e^{-i\gamma H \exp\left(-\frac{\hat{x}_1^2}{\hat{\sigma}_1^2}-\frac{\hat{x}_2^2}{\hat{\sigma}_2^2}\right)}$ in a power series we get

$$\xi^{(\mathbf{m})}(\gamma) = \sum_{n=0}^{\infty} \frac{(-i\gamma H)^n}{n!} \int_{-1/2}^{1/2}\int_{-1/2}^{1/2} e^{-i2\pi m_1 \hat{x}_1 - i2\pi m_2 \hat{x}_2} \exp\left(-n\frac{\hat{x}_1^2}{\hat{\sigma}_2^2} - n\frac{\hat{x}_2^2}{\hat{\sigma}_2^2}\right) \mathrm{d}\hat{x}_1 \mathrm{d}\hat{x}_2 \,. \tag{A.52}$$

For $n = 0$, the integral becomes $\delta_{m_1,0}\delta_{m_2,0}$. For $n \neq 0$ we recognize the product of two similar integrals that we have already approximated in Eq. (A.37) and we obtain

$$\xi^{(\mathbf{m})}(\gamma) \approx \delta_{\mathbf{m},\mathbf{0}} + \frac{\pi\sigma_1\sigma_2}{a_1 a_2} \sum_{n=1}^{\infty} \frac{(-i\gamma H)^n}{n!\, n} \exp\left(-\frac{\pi^2 m_1^2 \sigma_1^2}{n a_1^2} - \frac{\pi^2 m_2^2 \sigma_2^2}{n a_2^2}\right) \,. \tag{A.53}$$

A.5 Cylindrically symmetric shapes

A.5.1 General framework

Cylindrically symmetric shapes are described by

$$\zeta(\mathbf{x}) = \begin{cases} 0 & \text{if } r > R \\ \zeta(r) & \text{if } r < R \end{cases} \,, \tag{A.54}$$

where $r = \sqrt{x_1^2 + x_2^2}$ and R is the radius of the disc outside which ζ vanishes. Notice that the disc must lie inside the unit cell. In order to compute the geometrical term

$$\mathcal{G}^{(\mathbf{m})}(\gamma) = \int_{\mathrm{supp}\,\zeta} e^{-i\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{x}} \left(e^{-i\gamma\zeta(\mathbf{x})} - 1\right) \mathrm{d}^2 x \,, \tag{A.55}$$

it is convenient to make a change of variable, from cartesian to polar coordinates. Equation (A.55) in polar coordinates reads

$$\mathcal{G}^{(\mathbf{m})}(\gamma) = \int_0^{2\pi}\int_0^R r e^{-iG^{(\mathbf{m})} r\cos\theta} \left(e^{-i\gamma\zeta(r)} - 1\right) \mathrm{d}r\mathrm{d}\theta \,, \tag{A.56}$$

where $G^{(\mathbf{m})} = |\mathbf{G}^{(\mathbf{m})}|$ and the origin of θ is taken from the direction of $\mathbf{G}^{(\mathbf{m})}$. The θ-integration can be carried on by recognizing the integral definition of the 0th order Bessel function of the first kind $\int_0^{2\pi} e^{iz\cos\theta}\mathrm{d}\theta = 2\pi J_0(z)$:

$$\mathcal{G}^{(\mathbf{m})}(\gamma) = \int_0^R r\left(e^{-i\gamma\zeta(r)} - 1\right)\int_0^{2\pi} e^{-iG^{(\mathbf{m})} r\cos\theta}\mathrm{d}\theta\mathrm{d}r = 2\pi\int_0^R r\left(e^{-i\gamma\zeta(r)} - 1\right) J_0(rG^{(\mathbf{m})})\mathrm{d}r \,. \tag{A.57}$$

Equation (A.57) will be the starting point for the derivation of the considered particular geometries.

A.5.2 Cylinder

Let us consider the case of a cylinder, $\zeta(r) = H$. For $\mathbf{m} \neq \mathbf{0}$, Eq. (A.57) reduces to

$$\mathcal{G}^{(\mathbf{m})}_{\mathrm{cyl}}(\gamma) = 2\pi\left(e^{-i\gamma H} - 1\right)\int_0^R r J_0(rG^{(\mathbf{m})})\mathrm{d}r = \frac{2\pi R}{G^{(\mathbf{m})}}\left(e^{-i\gamma H} - 1\right) J_1(RG^{(\mathbf{m})}) \,, \tag{A.58}$$

where the property $\frac{\mathrm{d}}{\mathrm{d}z}[z^n J_n] = z^n J_{n-1}$ of the Bessel functions has been used. For $\mathbf{m} = \mathbf{0}$, we have

$$\mathcal{G}^{(\mathbf{0})}_{\mathrm{cyl}}(\gamma) = \pi R^2\left(e^{-i\gamma H} - 1\right) \,, \tag{A.59}$$

which can be computed directly from Eq. (A.57) or by taking the limit of Eq. (A.58) when $G^{(\mathbf{m})} \to 0$. Consequently, we obtain

$$\xi^{(\mathbf{m})}_{\text{cyl}}(\gamma) = \begin{cases} 1 + \frac{\pi R^2}{a_c}\left(e^{-i\gamma H} - 1\right) & \text{if } \mathbf{m} = \mathbf{0} \\ \frac{2\pi R}{a_c G^{(\mathbf{m})}}\left(e^{-i\gamma H} - 1\right) J_1(RG^{(\mathbf{m})}) & \text{if } \mathbf{m} \neq \mathbf{0} \end{cases} . \tag{A.60}$$

A.5.3 Concentric rings

The case of concentric rings can be easily derived from the cylinder. Consider a set of n concentric rings of inner radii r^{i}_k, outer radii r^{o}_k and height H_k, $k \in [\![1, n]\!]$ indexing the rings. The rings are assumed to have disjoint supports and ordered by increasing radii

$$r^{\text{i}}_1 < r^{\text{o}}_1 < r^{\text{i}}_2 < r^{\text{o}}_2 < \cdots < r^{\text{i}}_n < r^{\text{o}}_n \, . \tag{A.61}$$

In that case the profile can be expressed by

$$\zeta(r) = \sum_{k=1}^{n} H_k \ \mathbb{1}_{[r^{\text{i}}_k, r^{\text{o}}_k]}(r) \, , \tag{A.62}$$

where we recall that $\mathbb{1}_{[a,b]}$ denotes the indicator function of the subset $[a, b]$ $(a < b)$ in $\mathbb{R}$. By plugging Eq. (A.62) into Eq. (A.57), we obtain, for $\mathbf{m} \neq \mathbf{0}$,

$$\mathcal{G}^{(\mathbf{m})}_{\text{rin}}(\gamma) = 2\pi \sum_{k=1}^{n} \left(e^{-i\gamma H_k} - 1\right) \int_{r^{\text{i}}_k}^{r^{\text{o}}_k} r J_0(rG^{(\mathbf{m})}) \mathrm{d}r \, . \tag{A.63}$$

By using the property of the Bessel functions $\frac{\mathrm{d}}{\mathrm{d}z}[z^n J_n] = z^n J_{n-1}$ as in the cylinder case gives

$$\mathcal{G}^{(\mathbf{m})}_{\text{rin}}(\gamma) = \frac{2\pi}{G^{(\mathbf{m})}} \sum_{k=1}^{n} \left(e^{-i\gamma H_k} - 1\right) \left(r^{\text{o}}_k J_1(r^{\text{o}}_k G^{(\mathbf{m})}) - r^{\text{i}}_k J_1(r^{\text{i}}_k G^{(\mathbf{m})})\right) . \tag{A.64}$$

For $\mathbf{m} = \mathbf{0}$, we get

$$\mathcal{G}^{(\mathbf{0})}_{\text{rin}}(\gamma) = \pi \sum_{k=1}^{n} \left(e^{-i\gamma H_k} - 1\right) \left([r^{\text{o}}_k]^2 - [r^{\text{i}}_k]^2\right) . \tag{A.65}$$

The result can be interpreted as a superposition of cylinders of radii corresponding to the rings outer radii subtracted by cylinders of radii corresponding the rings inner radii. Consequently, we obtain

$$\xi^{(\mathbf{m})}_{\text{rin}}(\gamma) = \begin{cases} 1 + \frac{\pi}{a_c} \sum_{k=1}^{n} \left(e^{-i\gamma H_k} - 1\right) \left([r^{\text{o}}_k]^2 - [r^{\text{i}}_k]^2\right) & \text{if } \mathbf{m} = \mathbf{0} \\ \frac{2\pi}{a_c G^{(\mathbf{m})}} \sum_{k=1}^{n} \left(e^{-i\gamma H_k} - 1\right) \left(r^{\text{o}}_k J_1(r^{\text{o}}_k G^{(\mathbf{m})}) - r^{\text{i}}_k J_1(r^{\text{i}}_k G^{(\mathbf{m})})\right) & \text{if } \mathbf{m} \neq \mathbf{0} \end{cases} . \tag{A.66}$$

A.5.4 Paraboloid

Consider the case of a paraboloid with circular section, $\zeta(r) = H\left(1 - \frac{r^2}{R^2}\right)$. For $\mathbf{m} = \mathbf{0}$, Eq. (A.57) gives directly

$$\mathcal{G}^{(\mathbf{0})}_{\text{par}}(\gamma) = 2\pi \int_0^R r \left(e^{-i\gamma H\left(1 - \frac{r^2}{R^2}\right)} - 1\right) \mathrm{d}r = \frac{\pi R^2}{\gamma H}\left(e^{-i\gamma H} - 1 + i\gamma H\right) . \tag{A.67}$$

For $\mathbf{m} \neq \mathbf{0}$, Eq. (A.57) becomes

$$\mathcal{G}^{(\mathbf{m})}_{\text{par}}(\gamma) = 2\pi \int_0^R r \left(e^{-i\gamma H\left(1-\frac{r^2}{R^2}\right)} - 1 \right) J_0(rG^{(\mathbf{m})}) \mathrm{d}r \tag{A.68}$$

$$= 2\pi \left(\int_0^R r e^{-i\gamma H\left(1-\frac{r^2}{R^2}\right)} J_0(rG^{(\mathbf{m})}) \mathrm{d}r - \frac{R}{G^{(\mathbf{m})}} J_1(RG^{(\mathbf{m})}) \right) . \tag{A.69}$$

Let us focus on the first term that we integrate by part

$$\int_0^R r J_0(rG^{(\mathbf{m})}) e^{-i\gamma H\left(1-\frac{r^2}{R^2}\right)} \mathrm{d}r = \frac{R}{G^{(\mathbf{m})}} J_1(RG^{(\mathbf{m})}) - \frac{2i\gamma H}{R^2 G^{(\mathbf{m})}} \int_0^R r^2 J_1(rG^{(\mathbf{m})}) e^{-i\gamma H\left(1-\frac{r^2}{R^2}\right)} \mathrm{d}r \, . \tag{A.70}$$

By successive integrations by part, we obtain that the n^{th} integration by part terms gives

$$\int_0^R r^n J_{n-1}(rG^{(\mathbf{m})}) e^{-i\gamma H\left(1-\frac{r^2}{R^2}\right)} \mathrm{d}r = \frac{R^n}{G^{(\mathbf{m})}} J_n(RG^{(\mathbf{m})}) - \frac{2i\gamma H}{R^2 G^{(\mathbf{m})}} \int_0^R r^{n+1} J_n(rG^{(\mathbf{m})}) e^{-i\gamma H\left(1-\frac{r^2}{R^2}\right)} \mathrm{d}r \, , \tag{A.71}$$

which is an iterative relation of the form $I_n = \frac{R^n}{G^{(\mathbf{m})}} J_n(RG^{(\mathbf{m})}) - \frac{2i\gamma H}{R^2 G^{(\mathbf{m})}} I_{n+1}$, where $I_n = \int_0^R r^n J_{n-1}(rG^{(\mathbf{m})}) e^{-i\gamma H\left(1-\frac{r^2}{R^2}\right)} \mathrm{d}r$. Then Eq. (A.70) yields

$$\int_0^R r J_0(rG^{(\mathbf{m})}) e^{-i\gamma H\left(1-\frac{r^2}{R^2}\right)} \mathrm{d}r = R^2 \sum_{n=1}^{\infty} \frac{(-2i\gamma H)^{n-1} J_n(RG^{(\mathbf{m})})}{(RG^{(\mathbf{m})})^n} \, . \tag{A.72}$$

Substituting the previous expression in Eq. (A.69) and noticing that the first term in the sum cancels with $\frac{R}{G^{(\mathbf{m})}} J_1(RG^{(\mathbf{m})})$ in Eq. (A.69), we finally get

$$\mathcal{G}^{(\mathbf{m})}_{\text{par}}(\gamma) = 2\pi R^2 \sum_{n=2}^{\infty} \frac{(-2i\gamma H)^{n-1} J_n(RG^{(\mathbf{m})})}{(RG^{(\mathbf{m})})^n} = 2\pi R^2 \sum_{n=1}^{\infty} \frac{(-2i\gamma H)^{n} J_{n+1}(RG^{(\mathbf{m})})}{(RG^{(\mathbf{m})})^{n+1}} \, . \tag{A.73}$$

Consequently:

$$\xi^{(\mathbf{m})}_{\text{par}}(\gamma) = \begin{cases} 1 + \frac{\pi R^2}{\gamma H \, a_c} \left(e^{-i\gamma H} - 1 + i\gamma H \right) & \text{if } \mathbf{m} = \mathbf{0} \\ \frac{2\pi R^2}{a_c} \sum_{n=1}^{\infty} \frac{(-2i\gamma H)^n J_{n+1}(RG^{(\mathbf{m})})}{(RG^{(\mathbf{m})})^{n+1}} & \text{if } \mathbf{m} \neq \mathbf{0} \end{cases} . \tag{A.74}$$

Equation (A.73) takes the form of a sum of a series in powers of $-2i\gamma H/RG^{(\mathbf{m})}$. Such a series will converge quickly for values of $-2i\gamma H/RG^{(\mathbf{m})}$ such that $|2\gamma H/RG^{(\mathbf{m})}| < 1$, and somewhat more slowly for $|2\gamma H/RG^{(\mathbf{m})}| > 1$. This is a general feature of all the series expansion presented in this appendix. However, in the case of the paraboloid, an alternative power series can be constructed easily in power of $RG^{(\mathbf{m})}/2i\gamma H$ instead. The method to construct such a series expansion is to choose to integrate by part the integral in Eq. (A.69) in the following way instead

$$\int_0^R J_0(rG^{(\mathbf{m})}) \, r \, e^{-i\gamma H\left(1-\frac{r^2}{R^2}\right)} \mathrm{d}r = \left[J_0(RG^{(\mathbf{m})}) - e^{-i\gamma H} \right] \frac{R^2}{2i\gamma H} - \frac{R^2 G^{(\mathbf{m})}}{2i\gamma H} \int_0^R J_{-1}(rG^{(\mathbf{m})}) \, e^{-i\gamma H\left(1-\frac{r^2}{R^2}\right)} \mathrm{d}r \, , \tag{A.75}$$

and keep doing similar successive integration by parts, i.e. taking derivatives of $r^{-n}J_{-n}$ and integrating $r\exp\left[-i\gamma H\big(1-\frac{r^2}{R^2}\big)\right]$. This yields the following power series expansion

$$\mathcal{G}^{(\mathbf{m})}_{\rm par}(\gamma)=\frac{2\pi R^2}{RG^{(\mathbf{m})}}\,\exp\left[\left(\frac{RG^{(\mathbf{m})}}{2i\gamma H}-\frac{2i\gamma H}{RG^{(\mathbf{m})}}\right)\,\frac{RG^{(\mathbf{m})}}{2}\right]-\frac{2\pi R^2}{RG^{(\mathbf{m})}}\,\sum_{n=0}^{\infty}J_n(RG^{(\mathbf{m})})\,\left[\frac{RG^{(\mathbf{m})}}{2i\gamma H}\right]^n\,. \quad \text{(A.76)}$$

The latter can also be obtained directly from Eq. (A.73) by the use of the identity

$$\sum_{n=-\infty}^{\infty}J_n(x)\,t^n=\exp\left[\left(t-\frac{1}{t}\right)\,\frac{x}{2}\right]\,. \quad \text{(A.77)}$$

Indeed, applying the above identity for $x=RG^{(\mathbf{m})}$ and $t=\frac{-2i\gamma H}{RG^{(\mathbf{m})}}$ and splitting the sum for $n<1$ and $n\geq 1$ we have

$$\begin{aligned}\sum_{n=1}^{\infty}J_n(RG^{(\mathbf{m})})\,\left[\frac{-2i\gamma H}{RG^{(\mathbf{m})}}\right]^n&=\exp\left[\left(\frac{RG^{(\mathbf{m})}}{2i\gamma H}-\frac{2i\gamma H}{RG^{(\mathbf{m})}}\right)\,\frac{RG^{(\mathbf{m})}}{2}\right]\\&\quad-\sum_{n=0}^{\infty}J_n(RG^{(\mathbf{m})})\,\left[\frac{RG^{(\mathbf{m})}}{2i\gamma H}\right]^n\,, \quad \text{(A.78)}\end{aligned}$$

hence

$$\mathcal{G}^{(\mathbf{m})}_{\rm par}(\gamma)=\frac{2\pi R^2}{RG^{(\mathbf{m})}}\,\exp\left[\left(\frac{RG^{(\mathbf{m})}}{2i\gamma H}-\frac{2i\gamma H}{RG^{(\mathbf{m})}}\right)\,\frac{RG^{(\mathbf{m})}}{2}\right]-\frac{2\pi R^2}{RG^{(\mathbf{m})}}\,\sum_{n=0}^{\infty}J_n(RG^{(\mathbf{m})})\,\left[\frac{RG^{(\mathbf{m})}}{2i\gamma H}\right]^n\,,$$

as announced. The two alternative expressions Eqs. (A.73) and (A.76) are of particular from a numerical point of view. Indeed, we can make an alogrithm that, depending on the arguments γ and $\mathbf{m}$, chooses to use Eq. (A.73) or Eq. (A.76) depending on whether $|2\gamma H/RG^{(\mathbf{m})}|<1$ or $|2\gamma H/RG^{(\mathbf{m})}|>1$. This way the algorithm chooses automatically the power series that converges fastest for the given values of the arguments.

A.5.5 Cone

Let us consider the case of a cone, $\zeta(r)=H(1-\frac{r}{R})$. For $\mathbf{m}\neq\mathbf{0}$, Eq. (A.57) becomes

$$\mathcal{G}^{(\mathbf{m})}_{\rm con}(\gamma)=2\pi\left(\int_0^R re^{-i\gamma H(1-\frac{r}{R})}J_0(rG^{(\mathbf{m})})\mathrm{d}r-\frac{R}{G^{(\mathbf{m})}}J_1(RG^{(\mathbf{m})})\right)\,. \quad \text{(A.79)}$$

The integral in Eq. (A.79) can be expanded as

$$\int_0^R re^{-i\gamma H(1-\frac{r}{R})}J_0(rG^{(\mathbf{m})})\mathrm{d}r=\sum_{n=0}^{\infty}\frac{(-i\gamma H)^n}{n!}\int_0^R r\left(1-\frac{r}{R}\right)^n J_0(rG^{(\mathbf{m})})\mathrm{d}r\,. \quad \text{(A.80)}$$

Using the change of variable $u=r/R$, Eq. (A.80) becomes

$$\int_0^R re^{-i\gamma H(1-\frac{r}{R})}J_0(rG^{(\mathbf{m})})\mathrm{d}r=R^2\sum_{n=0}^{\infty}\frac{(-i\gamma H)^n}{n!}\int_0^1 u\,(1-u)^n\,J_0(RG^{(\mathbf{m})}\,u)\mathrm{d}u\,. \quad \text{(A.81)}$$

The integral in Eq. (A.81) is a particular case of the integral given in Eq. (6.569) p.690 in [147]

$$\begin{aligned}\int_0^1 u^\lambda\,(1-u)^{\mu-1}\,J_\nu(au)\mathrm{d}u&=\frac{\Gamma(\mu)\,\Gamma(1+\lambda+\nu)2^{-\nu}a^\nu}{\Gamma(\nu+1)\,\Gamma(1+\lambda+\mu+\nu)}\times\\&\quad{}_2\mathrm{F}_3\left(\frac{\lambda+1+\nu}{2},\frac{\lambda+2+\nu}{2};\nu+1,\frac{\lambda+1+\mu+\nu}{2},\frac{\lambda+2+\mu+\nu}{2};\frac{-a^2}{4}\right)\,, \quad \text{(A.82)}\end{aligned}$$

under the condition $\mathrm{Re}(\mu) > 0$, $\mathrm{Re}(\lambda+\nu) > -1$. The function ${}_p\mathrm{F}_q$ is a generalized hypergeometric function. Applying Eq. (A.82) with $\lambda = 1$, $\mu = n+1$, $\nu = 0$, $a = RG^{(\mathbf{m})}$, Eq. (A.81) yields

$$\int_0^R re^{-i\gamma H(1-\frac{r}{R})} J_0(rG^{(\mathbf{m})})\mathrm{d}r = R^2 \sum_{n=0}^{\infty} \frac{(-i\gamma H)^n}{(n+2)!} \times {}_2\mathrm{F}_3\left(1, \frac{3}{2}; 1, \frac{n+3}{2}, \frac{n+4}{2}; \frac{-R^2G^{(\mathbf{m})2}}{4}\right) , \quad \text{(A.83)}$$

where we have simplified $\frac{\Gamma(n+1)\,\Gamma(2)}{\Gamma(1)\,\Gamma(n+3)} = \frac{n!}{(n+2)!} = \frac{1}{(n+1)(n+2)}$. It can be shown that the term of the sum for $n = 0$ cancels the second term in Eq. (A.79). Finally, we obtain

$$\mathcal{G}^{(\mathbf{m})}_{\mathrm{con}}(\gamma) = 2\pi R^2 \sum_{n=1}^{\infty} \frac{(-i\gamma H)^n}{(n+2)!} \times {}_2\mathrm{F}_3\left(1, \frac{3}{2}; 1, \frac{n+3}{2}, \frac{n+4}{2}; \frac{-R^2G^{(\mathbf{m})2}}{4}\right) . \quad \text{(A.84)}$$

For $\mathbf{m} = \mathbf{0}$, Eq. (A.57) gives directly

$$\begin{aligned}
\mathcal{G}^{(\mathbf{0})}_{\mathrm{con}}(\gamma) &= 2\pi \int_0^R r\left(e^{-i\gamma H(1-\frac{r}{R})} - 1\right)\mathrm{d}r & \text{(A.85)}\\
&= 2\pi \int_0^R re^{-i\gamma H(1-\frac{r}{R})}\mathrm{d}r - \pi R^2 & \text{(A.86)}\\
&= \frac{2\pi R^2}{i\gamma H} - \frac{2\pi R}{i\gamma H}\int_0^R e^{-i\gamma H(1-\frac{r}{R})}\mathrm{d}r - \pi R^2 & \text{(A.87)}\\
&= \frac{2\pi R^2}{(\gamma H)^2}\left(1 - i\gamma H - \frac{(\gamma H)^2}{2} - e^{-i\gamma H}\right) . & \text{(A.88)}
\end{aligned}$$

Consequently, we have

$$\xi^{(\mathbf{m})}_{\mathrm{con}}(\gamma) = \begin{cases} 1 - \frac{2\pi R^2}{(\gamma H)^2 a_c}\left(e^{-i\gamma H} - 1 + i\gamma H + \frac{(\gamma H)^2}{2}\right) & \text{if } \mathbf{m} = \mathbf{0} \\ \frac{2\pi R^2}{a_c}\sum_{n=1}^{\infty} \frac{(-i\gamma H)^n}{(n+2)!} \times {}_2\mathrm{F}_3\left(1, \frac{3}{2}; 1, \frac{n+3}{2}, \frac{n+4}{2}; \frac{-R^2G^{(\mathbf{m})2}}{4}\right) & \text{if } \mathbf{m} \neq \mathbf{0} \end{cases} . \quad \text{(A.89)}$$

A.5.6 Hemiellipsoid

Let us consider the case of a hemiellipsoid, $\zeta(r) = H\sqrt{1 - \frac{r^2}{R^2}}$. For $\mathbf{m} \neq \mathbf{0}$, Eq. (A.57) becomes

$$\mathcal{G}^{(\mathbf{m})}_{\mathrm{hem}}(\gamma) = 2\pi \left(\int_0^R re^{-i\gamma H\sqrt{1-\frac{r^2}{R^2}}} J_0(rG^{(\mathbf{m})})\mathrm{d}r - \frac{R}{G^{(\mathbf{m})}} J_1(RG^{(\mathbf{m})})\right) . \quad \text{(A.90)}$$

By Taylor expansion of the exponential, we get

$$\mathcal{G}^{(\mathbf{m})}_{\mathrm{hem}}(\gamma) = 2\pi \left(\int_0^R rJ_0(rG^{(\mathbf{m})}) \sum_{n=0}^{\infty} \frac{(-i\gamma H)^n \left(1 - \frac{r^2}{R^2}\right)^{n/2}}{n!}\mathrm{d}r - \frac{R}{G^{(\mathbf{m})}} J_1(RG^{(\mathbf{m})})\right) . \quad \text{(A.91)}$$

After switching the sum and integration, the first term of the sum cancels the last term in the above equation, and we have

$$\mathcal{G}^{(\mathbf{m})}_{\mathrm{hem}}(\gamma) = 2\pi \sum_{n=1}^{\infty} \frac{(-i\gamma H)^n}{n!} \int_0^R r\left(1 - \frac{r^2}{R^2}\right)^{n/2} J_0(rG^{(\mathbf{m})})\mathrm{d}r . \quad \text{(A.92)}$$

We apply a change of variable $u = r/R$ and the integral in Eq. (A.92) reads

$$\int_0^R r\left(1 - \frac{r^2}{R^2}\right)^{n/2} J_0(rG^{(\mathbf{m})})\mathrm{d}r = R^2 \int_0^1 u(1-u^2)^{n/2} J_0(RG^{(\mathbf{m})}u)\mathrm{d}u . \quad \text{(A.93)}$$

The integral in the right hand side of the above equation is a particular case of the integral Eq. (6.567(1)) p.688 in [147] (under the condition $\alpha > 0$, $\mathrm{Re}\nu > -1$, $\mathrm{Re}\mu > -1$):

$$\int_0^1 u^{\nu+1}(1-u^2)^{\mu} J_\nu(\alpha u)\mathrm{d}u = 2^\mu \Gamma(\mu+1)\alpha^{-(\mu+1)} J_{\nu+\mu+1}(\alpha)\,. \tag{A.94}$$

Using the above result for $\alpha = RG^{(\mathbf{m})}$, $\nu = 0$, $\mu = n/2$, Eq. (A.92) yields

$$\mathcal{G}^{(\mathbf{m})}_{\mathrm{hem}}(\gamma) = 2\pi R^2 \sum_{n=1}^{\infty} \frac{(-i\gamma H)^n}{n!} \frac{2^{n/2}\Gamma(n/2+1) J_{n/2+1}(RG^{(\mathbf{m})})}{(RG^{(\mathbf{m})})^{n/2+1}}\,. \tag{A.95}$$

This is the result obtained by Kretschmann and Maradudin in [48]. For $\mathbf{m} = \mathbf{0}$, the geometrical term is $\mathcal{G}^{(\mathbf{0})}_{\mathrm{hem}}(\gamma) = 2\pi \int_0^R r e^{-i\gamma H\sqrt{1-\frac{r^2}{R^2}}}\mathrm{d}r - \pi R^2$ and its computation does not require a Taylor expansion of the exponential. Indeed, by using the substitution $u = \sqrt{1-r^2/R^2}$, $u\mathrm{d}u = -r/R^2\mathrm{d}r$, we get

$$\mathcal{G}^{(\mathbf{0})}_{\mathrm{hem}}(\gamma) = 2\pi R^2 \left(\int_0^1 u e^{-i\gamma H u}\mathrm{d}u - \frac{1}{2} \right)\,. \tag{A.96}$$

Then an integration by part gives

$$\int_0^1 u e^{-i\gamma H u}\mathrm{d}u = \frac{1}{-i\gamma H}\left(e^{-i\gamma H} - \int_0^1 e^{-i\gamma H u}\mathrm{d}u \right) = \frac{1}{(\gamma H)^2}\left(e^{-i\gamma H}(1+i\gamma H) - 1 \right)\,. \tag{A.97}$$

Hence

$$\mathcal{G}^{(\mathbf{0})}_{\mathrm{hem}}(\gamma) = \frac{2\pi R^2}{(\gamma H)^2}\left(e^{-i\gamma H}(1+i\gamma H) - 1 - \frac{(\gamma H)^2}{2} \right)\,. \tag{A.98}$$

Consequently for $\mathbf{m} = \mathbf{0}$, we have

$$\xi^{(\mathbf{0})}_{\mathrm{hem}}(\gamma) = 1 + \frac{2\pi R^2}{a_c(\gamma H)^2}\left(e^{-i\gamma H}(1+i\gamma H) - 1 - \frac{(\gamma H)^2}{2} \right)\,. \tag{A.99}$$

The Taylor expansion method gives (either by careful re-derivation or by taking the limit $G^{(\mathbf{m})} \to 0$ in Eq. (A.101) or Taylor expansion of the previous result)

$$\xi^{(\mathbf{0})}_{\mathrm{hem}}(\gamma) = 1 + \frac{2\pi R^2}{a_c} \sum_{n=1}^{\infty} \frac{(-i\gamma H)^n}{n!(n+2)}\,, \tag{A.100}$$

which is the expression found by Kretschmann and Maradudin in [48]. Consequently, we obtain

$$\xi^{(\mathbf{m})}_{\mathrm{hem}}(\gamma) = \begin{cases} 1 + \frac{2\pi R^2}{a_c(\gamma H)^2}\left(e^{-i\gamma H}(1+i\gamma H) - 1 - \frac{(\gamma H)^2}{2} \right) = 1 + \frac{2\pi R^2}{a_c}\sum_{n=1}^{\infty}\frac{(-i\gamma H)^n}{n!(n+2)} & \text{if } \mathbf{m} = \mathbf{0} \\ \frac{2\pi R^2}{a_c}\sum_{n=1}^{\infty}\frac{(-i\gamma H)^n}{n!}\frac{2^{n/2}\Gamma(n/2+1)J_{n/2+1}(RG^{(\mathbf{m})})}{(RG^{(\mathbf{m})})^{n/2+1}} & \text{if } \mathbf{m} \neq \mathbf{0} \end{cases}\,. \tag{A.101}$$

A.6 Shapes with elliptic section

The shapes considered in Section A.5 have the property to be cylindrically symmetric along the x_3-axis, i.e. their section in a plane of constant x_3 is a circle. In practice, one can be interested in studying the influence of an asymmetry of the shape. Indeed, in reality it may (or always) happen that when manufacturing such structures, the particles result in being asymmetric, whether this is done on purpose or not. In the present section, we derive a simple formula that

gives the ξ-integral for shapes with elliptic section in a constant x_3-plane using the result known for cylindrically symmetric shapes. In other words, this formula can be seen as a correction to be applied to the results derived in Section A.5 to handle shapes with elliptic section.

A.6.1 Ellipse's axes collinear to the direct lattice's axes

Consider a shape with elliptic sections in planes of constant x_3 (it may or course reduce to a point or the empty set also for some value of x_3). The section in the plane $x_3 = 0$ is defined by the ellipse equation

$$\left(\frac{x_1}{a}\right)^2 + \left(\frac{x_2}{b}\right)^2 = 1 \tag{A.102}$$

where $a, b \in \mathbb{R}_+^*$. By definition of the elliptic section, the geometry profile ζ^{ell} has the following property

$$\zeta^{\text{ell}}(\mathbf{x}) = \begin{cases} 0 & \text{if } \mathbf{x} \notin \mathcal{D}_{\text{ell}} \\ \zeta(\rho) & \text{if } \mathbf{x} \in \mathcal{D}_{\text{ell}} \end{cases}, \tag{A.103}$$

where $\mathcal{D}_{\text{ell}} = \left\{(x_1, x_2) \in \mathbb{R}^2, \left(\frac{x_1}{a}\right)^2 + \left(\frac{x_2}{b}\right)^2 < 1\right\}$ is the elliptic disc support of the profile. Here $\rho = \sqrt{\left(\frac{x_1}{a}\right)^2 + \left(\frac{x_2}{b}\right)^2}$ and then denotes elliptic contours for constant height level, $\rho \in [0, 1[$. The geometrical term Eq. (A.55) expressed in the cartesian coordinate system reads

$$\mathcal{G}^{\text{ell},(\mathbf{m})}(\gamma) = \int_{\mathcal{D}_{\text{ell}}} \left(e^{-i\gamma\zeta^{\text{ell}}(\mathbf{x})} - 1\right) e^{-i\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{x}} \mathrm{d}^2x \,. \tag{A.104}$$

Consider the unit circular disc $\mathcal{D}_{\text{circ}} = \left\{(u_1, u_2) \in \mathbb{R}^2, u_1^2 + u_2^2 < 1\right\}$ and the linear mapping, ϕ, that maps the unit circular disc $\mathcal{D}_{\text{circ}}$ onto the elliptic disc $\mathcal{D}_{\text{ell}}$, defined as

$$\phi : \begin{cases} \mathcal{D}_{\text{circ}} & \mapsto & \mathcal{D}_{\text{ell}} \\ (u_1, u_2) & \mapsto & (au_1, bu_2) \end{cases}. \tag{A.105}$$

By using the change of variable $\mathbf{x} = \phi(\mathbf{u})$, Eq. (A.104) becomes

$$\mathcal{G}^{\text{ell},(\mathbf{m})}(\gamma) = ab \int_{\mathcal{D}_{\text{circ}}} \left(e^{-i\gamma\zeta^{\text{ell}}(\mathbf{x}(\mathbf{u}))} - 1\right) e^{-i\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{x}(\mathbf{u})} \mathrm{d}^2u \,. \tag{A.106}$$

Let analyze the term $\zeta^{\text{ell}}(\mathbf{x}(\mathbf{u}))$.

$$\zeta^{\text{ell}}(\mathbf{x}(\mathbf{u})) = \begin{cases} 0 & \text{if } \left(\frac{x_1(u_1,u_2)}{a}\right)^2 + \left(\frac{x_2(u_1,u_2)}{b}\right)^2 = u_1^2 + u_2^2 > 1 \\ \zeta(\rho(\mathbf{x}(\mathbf{u}))) & \text{if } \left(\frac{x_1(u_1,u_2)}{a}\right)^2 + \left(\frac{x_2(u_1,u_2)}{b}\right)^2 = u_1^2 + u_2^2 < 1 \end{cases}, \tag{A.107}$$

with $\rho(\mathbf{x}(\mathbf{u})) = \sqrt{u_1^2 + u_2^2}$. This means that $\zeta^{\text{ell}}(\mathbf{x}(\mathbf{u}))$ can be replaced by the corresponding shape with circular section:

$$\zeta^{\text{circ}}(\mathbf{u}) = \begin{cases} 0 & \text{if } r > 1 \\ \zeta(r) & \text{if } r < 1 \end{cases}, \tag{A.108}$$

with $r = \sqrt{u_1^2 + u_2^2}$. This is the way we defined cylindrically symmetric shapes in Section A.5. Let us analyze now the term

$$\mathbf{G}^{(\mathbf{m})} \cdot \mathbf{x}(\mathbf{u}) = (m_1\mathbf{b}_1 + m_2\mathbf{b}_2) \cdot (au_1\mathbf{a}_1 + bu_2\mathbf{a}_2) = 2\pi\,(a\,m_1u_1 + b\,m_2u_2) \,. \tag{A.109}$$

Here we have used the property Eq. (A.7). Then an equivalent way of writing the above scalar product would be

$$\mathbf{G}^{(\mathbf{m})} \cdot \mathbf{x}\,(\mathbf{u}) = \tilde{\mathbf{G}}^{(\mathbf{m})} \cdot \mathbf{u}\,, \tag{A.110}$$

where $\tilde{\mathbf{G}}^{(\mathbf{m})} = (m_1\, a\, \mathbf{b}_1 + m_2\, b\, \mathbf{b}_2)$, which can be interpreted as a re-scaling of the reciprocal lattice basis vectors. Then Eq. (A.106) becomes

$$\mathcal{G}^{\mathrm{ell},(\mathbf{m})}(\gamma) = ab \int_{\mathcal{D}_{\mathrm{circ}}} \left(e^{-i\gamma\zeta^{\mathrm{circ}}(\mathbf{u})} - 1\right) e^{-i\tilde{\mathbf{G}}^{(\mathbf{m})}\cdot\mathbf{u}} \mathrm{d}^2u\,. \tag{A.111}$$

where we recognize the geometrical term of the cylindrically symmetric shape (with unit disc support) associated with the elliptic one, but corrected by a factor ab and taking argument $\tilde{\mathbf{G}}^{(\mathbf{m})}$. To sum up

$$\mathcal{G}^{\mathrm{ell},(\mathbf{m})} = ab\, \mathcal{G}^{\mathrm{circ},(\tilde{\mathbf{m}})}\,, \tag{A.112}$$

where the notation circ denotes *the cylindrically symmetric shape with unit disc support associated with the elliptic shape* and $\tilde{\mathbf{m}}$ denotes the modified reciprocal vector.

A.6.2 General cases: a priori non-collinear axes

In the previous section we have treated the case where the ellipse's axes are collinear with the direct lattice's axes to explain clearly the reasoning step by step. We shall treat the general case below in the same fashion but passing quickly the straightforward check in the derivation, since the method is identical to what has been done before.

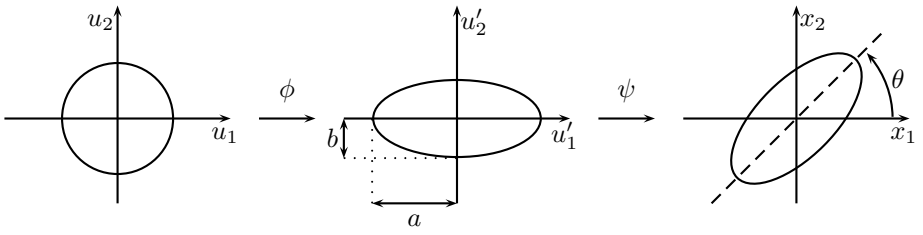


Figure A.1: Mappings from unit circle to tilted ellipse.

Consider an ellipse whose first axis is tilted by an angle θ from the x_1-axis. The equation of the ellipse in the (x_1, x_2) plane is

$$\left(\frac{\cos\theta\, x_1 + \sin\theta\, x_2}{a}\right)^2 + \left(\frac{-\sin\theta\, x_1 + \cos\theta\, x_2}{b}\right)^2 = 1\,. \tag{A.113}$$

This is immediate by considering the rotation of center the origin, and angle θ of the ellipse of the previous section. We define the tilted elliptic disc $\mathcal{D}^{\theta}_{\mathrm{ell}} = \left\{(x_1, x_2) \in \mathbb{R}^2, \left(\frac{\cos\theta\, x_1 + \sin\theta\, x_2}{a}\right)^2 + \left(\frac{-\sin\theta\, x_1 + \cos\theta\, x_2}{b}\right)^2 < 1\right\}$, where the superscript θ will

help us distinguishing it from the non tilted one. As before, the geometry profile $\zeta^{\mathrm{ell},\theta}$ has the following property

$$\zeta^{\mathrm{ell},\theta}(\mathbf{x}) = \begin{cases} 0 & \text{if } \mathbf{x} \notin \mathcal{D}^{\theta}_{\mathrm{ell}} \\ \zeta(\rho) & \text{if } \mathbf{x} \in \mathcal{D}^{\theta}_{\mathrm{ell}} \end{cases} , \tag{A.114}$$

where $\rho = \sqrt{\left(\frac{\cos\theta\, x_1 + \sin\theta\, x_2}{a}\right)^2 + \left(\frac{-\sin\theta\, x_1 + \cos\theta\, x_2}{b}\right)^2}$ and still denotes elliptic contours for constant height level, $\rho \in [0,1[$. The geometrical term Eq. (A.55) expressed in the cartesian coordinate system reads

$$\mathcal{G}^{\mathrm{ell},\theta,(\mathbf{m})}(\gamma) = \int_{\mathcal{D}^{\theta}_{\mathrm{ell}}} \left(e^{-i\gamma\zeta^{\mathrm{ell},\theta}(\mathbf{x})} - 1\right) e^{-i\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{x}} \mathrm{d}^2x \,. \tag{A.115}$$

We will re-use the linear mapping ϕ, that maps the unit circular disc $\mathcal{D}_{\mathrm{circ}}$ on the non-tilted elliptic disc $\mathcal{D}_{\mathrm{ell}}$, but this time in composition with a rotation of angle θ afterwards, denoted ψ and defined as

$$\psi : \begin{cases} \mathcal{D}_{\mathrm{ell}} & \mapsto & \mathcal{D}^{\theta}_{\mathrm{ell}} \\ (u_1', u_2') & \mapsto & (\cos\theta\, u_1' - \sin\theta\, u_2', \sin\theta\, u_1' + \cos\theta\, u_2') \end{cases} . \tag{A.116}$$

For more clarity we will associate to the linear mappings ϕ and ψ their representative matrices Φ and Ψ (in the canonical basis of the $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2)$-plane) defined as

$$\Phi = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \quad \text{and} \quad \Psi = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} . \tag{A.117}$$

Then the unit circular disc $\mathcal{D}_{\mathrm{circ}}$ is mapped on the tilted elliptic disc $\mathcal{D}^{\theta}_{\mathrm{ell}}$ through $\psi\circ\phi$ (Fig. A.1) and we denote the corresponding matrix

$$\mathbf{M} = \Psi\Phi = \begin{pmatrix} a\cos\theta & -b\sin\theta \\ a\sin\theta & b\cos\theta \end{pmatrix} . \tag{A.118}$$

By using the change of variable $\mathbf{x} = \psi\circ\phi(\mathbf{u})$, Eq. (A.115) becomes

$$\mathcal{G}^{\mathrm{ell},\theta,(\mathbf{m})}(\gamma) = ab \int_{\mathcal{D}_{\mathrm{circ}}} \left(e^{-i\gamma\zeta^{\mathrm{ell},\theta}(\mathbf{x}(\mathbf{u}))} - 1\right) e^{-i\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{x}(\mathbf{u})} \mathrm{d}^2u \,. \tag{A.119}$$

As previously, it is straightforward to verify that $\zeta^{\mathrm{ell},\theta}\circ\psi\circ\phi = \zeta^{\mathrm{circ}}$. The scalar product can be written as

$$\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{M}\mathbf{u} = \tilde{\mathbf{G}}^{(\mathbf{m})}\cdot\mathbf{u} \,, \tag{A.120}$$

where $\tilde{\mathbf{G}}^{(\mathbf{m})} = \mathbf{M}^{\mathrm{T}}\mathbf{G}^{(\mathbf{m})}$. We finally obtain

$$\mathcal{G}^{\mathrm{ell},\theta,(\mathbf{m})} = ab\, \mathcal{G}^{\mathrm{circ},(\tilde{\mathbf{m}})} \,, \tag{A.121}$$

where $\tilde{\mathbf{m}}$ is a short hand notation to denote the modified reciprocal vector $\tilde{\mathbf{G}}^{(\mathbf{m})} = \mathbf{M}^{\mathrm{T}}\mathbf{G}^{(\mathbf{m})}$.

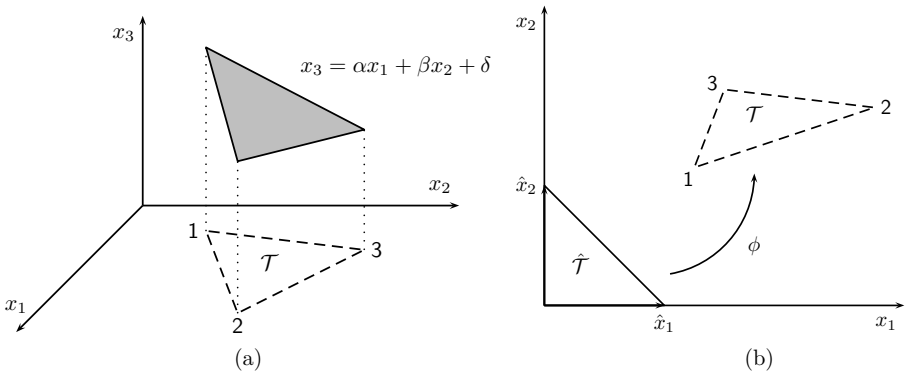


Figure A.2: (a) Triangular facet. (b) Transformation of the canonical triangle through the affine mapping ϕ.

A.7 Plane on a triangular support

Let us consider the case of plane defined on a triangular support $\mathcal{T}$, $\zeta(\mathbf{x}) = \alpha x_1 + \beta x_2 + \delta$ (Fig. A.2(a)). The geometrical term Eq. (A.55) is expressed in the cartesian coordinate system as

$$\mathcal{G}^{(\mathbf{m})}(\gamma) = \int_{\mathcal{T}} \left(e^{-i\gamma\zeta(\mathbf{x})} - 1 \right) e^{-i\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{x}} \mathrm{d}^2x \,. \tag{A.122}$$

We will denote by a super-script - taken in $\{1, 2, 3\}$ - the coordinates of the vertices of $\mathcal{T}$ taken in a direct (trigonometric) order, i.e. $(x_1^{(i)}, x_2^{(i)})$ are the coordinates of the vertex i (Fig. A.2). We note that $\mathcal{G}^{(\mathbf{m})}$ can be recast as

$$\mathcal{G}^{(\mathbf{m})}(\gamma) = \mathcal{I}^{(\mathbf{m})}(\gamma) - \mathcal{I}^{(\mathbf{m})}(0) \,, \tag{A.123}$$

where we have introduced

$$\mathcal{I}^{(\mathbf{m})}(\gamma) = \int_{\mathcal{T}} e^{-i\gamma\zeta(\mathbf{x})} e^{-i\mathbf{G}^{(\mathbf{m})}\cdot\mathbf{x}} \mathrm{d}^2x \tag{A.124}$$

$$= e^{-i\gamma\delta} \int_{\mathcal{T}} e^{\alpha' x_1 + \beta' x_2} \mathrm{d}x_1 \mathrm{d}x_2 \,, \tag{A.125}$$

with

$$\alpha' = -i\gamma\alpha - i\,\mathbf{G}^{(\mathbf{m})}\cdot\hat{\mathbf{e}}_1 \,, \qquad \beta' = -i\gamma\beta - i\,\mathbf{G}^{(\mathbf{m})}\cdot\hat{\mathbf{e}}_2 \,. \tag{A.126}$$

Even though it is straight forward to find a primitive of the integrand in Eq. (A.125), the boundary for the integration for a general triangle $\mathcal{T}$ is not trivial. It is convenient to consider an affine change of variables that maps together a generic triangle $\mathcal{T}$ and what we will call the canonical triangle $\hat{\mathcal{T}}$ (i.e. constructed on the unit vectors of the canonical basis). As depicted in Fig. A.2(b), consider the affine mapping:

$$\phi : \begin{cases} \mathbb{R}^2 & \mapsto & \mathbb{R}^2 \\ (x_1, x_2) & \mapsto & (x_1^{(1)}, x_2^{(1)}) + A(x_1, x_2) \end{cases} , \tag{A.127}$$

where A is the automorphism of $\mathbb{R}^2$ defined by the following matrix in the canonical basis (we will denote both the automorphism and its representative matrix by the same letter):

$$A = \begin{pmatrix} x_1^{(2)} - x_1^{(1)} & x_1^{(3)} - x_1^{(1)} \\ x_2^{(2)} - x_2^{(1)} & x_2^{(3)} - x_2^{(1)} \end{pmatrix} . \tag{A.128}$$

We show without trouble that A is an automorphism if and only if the triangle $\mathcal{T}$ has a non-zero area, which is the case of interest. Indeed we have

$$|\det A| = \left|(x_1^{(2)} - x_1^{(1)})(x_2^{(3)} - x_2^{(1)}) - (x_1^{(3)} - x_1^{(1)})(x_2^{(2)} - x_2^{(1)})\right| = 2\mathcal{A}_{\mathcal{T}} , \tag{A.129}$$

where $\mathcal{A}_{\mathcal{T}}$ is the area of the triangle $\mathcal{T}$. Furthermore, we can verify without effort that the vertices of $\hat{\mathcal{T}}$ map onto those of $\mathcal{T}$, and conserve the orientation. Considering now the change of variables $(x, y) = \phi^{-1}(x_1, x_2)$ in Eq. (A.125), we obtain

$$\mathcal{I}^{(\mathbf{m})}(\gamma) = 2\mathcal{A}_{\mathcal{T}}\, e^{\alpha' x_1^{(1)} + \beta' x_2^{(1)} - i\gamma\delta} \int_{\hat{\mathcal{T}}} e^{(\alpha' a_{11} + \beta' a_{21})x + (\alpha' a_{12} + \beta' a_{22})y} \mathrm{d}x\mathrm{d}y . \tag{A.130}$$

The coefficients a_{ij} denote the elements of A, and we have used Eq. (A.129) to express the determinant of the Jacobian of ϕ, $\det J_\phi = \det A$. We are then able to integrate without difficulty. By denoting $\alpha'' = \alpha' a_{11} + \beta' a_{21}$ and $\beta'' = \alpha' a_{12} + \beta' a_{22}$, and if $\alpha'' \neq 0$, $\beta'' \neq 0$

$$\begin{aligned} \int_{\hat{\mathcal{T}}} e^{\alpha'' x + \beta'' y} \mathrm{d}x\mathrm{d}y &= \int_0^1 \int_0^{1-x} e^{\alpha'' x + \beta'' y} \mathrm{d}y\mathrm{d}x \\ &= \frac{\beta''\left(e^{\alpha''} - 1\right) - \alpha''\left(e^{\beta''} - 1\right)}{\alpha''\beta''(\alpha'' - \beta'')} \end{aligned}$$

We can treat the cases $\alpha'', \beta'' = 0$, either by starting from the first line or by taking the limit in the second. This yields the following cases:

$$\int_{\hat{\mathcal{T}}} e^{\alpha'' x + \beta'' y} \mathrm{d}x\mathrm{d}y = \begin{cases} \frac{\beta''\left(e^{\alpha''} - 1\right) - \alpha''\left(e^{\beta''} - 1\right)}{\alpha''\beta''(\alpha'' - \beta'')} & \text{if } \alpha'' \neq 0\,, \beta'' \neq 0 \\ \frac{e^{\beta''} - 1 - \beta''}{\beta''^2} & \text{if } \alpha'' = 0\,, \beta'' \neq 0 \\ \frac{e^{\alpha''} - 1 - \alpha''}{\alpha''^2} & \text{if } \alpha'' \neq 0\,, \beta'' = 0 \\ 1/2 & \text{if } \alpha'' = 0\,, \beta'' = 0 \end{cases} . \tag{A.131}$$

To sum up, $\mathcal{I}^{(\mathbf{m})}(\gamma)$ is given by Eq. (A.130) and Eq. (A.131) plugging in the adequate parameters defined through the reasoning. The geometrical term, $\mathcal{G}^{(\mathbf{m})}(\gamma)$, for a polyhedral shape with triangular facets is then the sum of the geometrical terms for each triangular facet. This is useful for treating triangulated surfaces.

Appendix B

Algorithm for determining the Isserlis-Wick coefficients

"One and one and one is three."
Come together,
The Beatles (1969).

B.1 Listing all pair-partitions

The first step in determining the coefficients of the Isserlis-Wick polynomials, is to be able to make a list of all the possible distinct pair-partitioning of $[\![1, N]\!]$ where $N = 2k > 0$ is a even integer. First, it is instructive to ask ourselves how many distinct pair-partitioning there are for a given k. Let us denote this number ν_k. We will reason recursively on the number of such partitioning. First, for $k = 1$, there is only one way of paring the two elements 1 and 2, hence $\nu_1 = 1$. Assuming ν_k known, what is ν_{k+1}? To make a pair-partitioning of $[\![1, 2(k+1)]\!]$ we need to pick two distinct elements in $[\![1, 2(k+1)]\!]$ and make a pair-partitioning of the $2k$ remaining elements. There are $2(k+1) \times (2k+1)$ ways of choosing the first pair of the partitioning, and ν_k possible pair-partitions of the remaining elements. However, we are now counting too many. Indeed, choosing first the pair (i, j) and a remaining pair-partition can be written in the form $\{(i, j), (a, b), \cdots, (\alpha, \beta)\}$, but this pair-partition would also occur if we chose first (a, b) and then the remaining same pairs, or (α, β) and then the remaining same pairs, etc ... since the order of the pairs does not matter. A way to avoid this issue is to force a certain order of the pairs so that each pair-partitions are only counted once. Let us consider the following reasoning: choose the element 1, there is only one way of doing so, then choose one of the remaining elements, say i, there are $2k + 1$ ways of doing so and finally there are ν_k possible partitioning for the remaining elements. This gives the correct recursion relation for ν_k, namely

$$\nu_{k+1} = (2k+1)\nu_k \,. \tag{B.1}$$

We then easily deduce that

$$\nu_k = (2k-1)(2k-3)\cdots 3 \cdot 1 = \frac{(2k)!}{2^k k!} \,. \tag{B.2}$$

The reasoning for counting the number of pair-partitions now gives a concrete idea for an algorithm for listing all these pair-partitions. Indeed, a pair-partition can be viewed as a specific permutation of $(1, \cdots, 2k)$. Let us construct a matrix, $\mathbf{M}$, with ν_k rows and $2k$ columns that

we want to fill with all the possible pair-partition, one per row. Let us now revisit the counting reasoning stated above and extend it to make a recursive algorithm as follows.

0 For a given $k > 1$, start with an empty matrix $\mathbf{M}^{(k)}$ and the array $\mathbf{v}^{(k)} = (1, 2, \cdots, 2k)$.

- Execute recursively the following instructions:

  1.1 if $k > 1$ then pick the first (which is also the smallest) element in $\mathbf{v}^{(k)}$ and set it in the first column of $\mathbf{M}^{(k)}$: $\forall i \in [\![1, \nu_k]\!]$, $M^{(k)}_{i,1} \leftarrow v^{(k)}_1$.

  2.1 initialize a counter: $c \leftarrow 0$;

  2.2 for i between 2 and $2k$

  · set: $\forall j \in [\![c+1, c+\nu_k]\!]$, $M^{(k)}_{j,2} \leftarrow v^{(k)}_i$

  · make a new array $\mathbf{v}^{(k-1)}$ of dimension $2k-2$ composed of the ordered elements of $\mathbf{v}^{(k)}$ distinct from $v^{(k)}_1$ and $v^{(k)}_i$.

  · select the matrix block $\mathbf{M}^{(k-1)} = (M^{(k)}_{m,n})$ for $m \in [\![c+1, c+\nu_k]\!]$ and $n \in [\![3, 2k]\!]$.

  · increment the counter $c \leftarrow c + \nu_k$.

  · repeat the recursive instruction with $\mathbf{v}^{(k-1)}$, $\mathbf{M}^{(k-1)}$, $k \leftarrow k-1$.

1.2 if $k = 1$ then set $\mathbf{v}^{(1)}$ in $\mathbf{M}^{(1)}$: $\mathbf{M}^{(1)}_{1,1} \leftarrow \mathbf{v}^{(1)}_1$ and $\mathbf{M}^{(1)}_{1,2} \leftarrow \mathbf{v}^{(1)}_2$.

Let us illustrate what the algorithm is doing in a more visual way for a starting $k = 2$. We start with an empty matrix, that we fill with zeros for the sake of illustration, and the vector $\mathbf{v}^{(2)}$ with components from 1 to 4

$$\mathbf{M}^{(2)} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad \mathbf{v}^{(2)} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$$

Step 1.1 sets the first column of $\mathbf{M}^{(2)}$ to 1, and let us slash the corresponding element in $\mathbf{v}^{(2)}$ to denote that it will not be used anymore.

$$\mathbf{M}^{(2)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \qquad \mathbf{v}^{(2)} = \begin{bmatrix} \not{1} \\ 2 \\ 3 \\ 4 \end{bmatrix}$$

Then the counter c in step 2.1 is just a convenience that helps selecting sub-matrices as we will soon see. Now step 2.2, for i from 2 to 4 the algorithm actually splits into three branches. The first one will take care of all pair-partitions starting by $(1, 2)$, the second branch will take care of all pair-partitions starting by $(1, 3)$ and the third branch will take care of all pair-partitions starting by $(1, 4)$. Indeed, let us follow the first branch $i = 2$. In this branch, we set $M^{(2)}_{1,2}$ to

2 (colored in red), we slash it from $\mathbf{v}^{(2)}$, and repeat the recursive function with the final 1×2 block elementary block which will be filled with the remaining element of $\mathbf{v}^{(2)}$ according to step 1.2 (colored in pink).

$$\mathbf{M}^{(2)} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \qquad \mathbf{v}^{(2)} = \begin{bmatrix} \not{1} \\ \not{2} \\ 3 \\ 4 \end{bmatrix}$$

First branch: the gray elements are not affected.

Then for the second branch, $i = 3$, we set $M^{(2)}_{2,2}$ to 3 (colored in blue), we slash it from $\mathbf{v}^{(2)}$, and repeat the recursive function with the final 1×2 block elementary block which will be filled with the remaining element of $\mathbf{v}^{(2)}$ according to step 1.2 (colored in cyan). Note that in this branch the element 2 were not slashed and is still available for filling the last elementary block.

$$\mathbf{M}^{(2)} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 2 & 4 \\ 1 & 0 & 0 & 0 \end{bmatrix} \qquad \mathbf{v}^{(2)} = \begin{bmatrix} \not{1} \\ 2 \\ \not{3} \\ 4 \end{bmatrix}$$

Second branch: the gray elements are not affected.

Finally, for the last branch, $i = 4$, we set $M^{(2)}_{3,2}$ to 4 (colored in green), we slash it from $\mathbf{v}^{(2)}$, and repeat the recursive function with the final 1×2 block elementary block which will be filled with the remaining element of $\mathbf{v}^{(2)}$ according to step 1.2 (colored in orange). Note that in this branch the element 2 were not slashed and is still available for filling the last elementary block.

$$\mathbf{M}^{(2)} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 2 & 4 \\ 1 & 4 & 2 & 3 \end{bmatrix} \qquad \mathbf{v}^{(2)} = \begin{bmatrix} \not{1} \\ 2 \\ 3 \\ \not{4} \end{bmatrix} \tag{B.3}$$

Third branch: the gray elements are not affected.

We have now filled the matrix with all permutations of $(1, 2, 3, 4)$ that correspond to a pair-partitions. We encourage the reader to repeat the exercise for $k = 3$ and verify that the algorithm yields

$$\mathbf{M}^{(3)\mathrm{T}} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 3 & 3 & 3 & 4 & 4 & 4 & 5 & 5 & 5 & 6 & 6 & 6 \\ 3 & 3 & 3 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 \\ 4 & 5 & 6 & 4 & 5 & 6 & 3 & 5 & 6 & 3 & 4 & 6 & 3 & 4 & 5 \\ 5 & 4 & 4 & 5 & 4 & 4 & 5 & 3 & 3 & 4 & 3 & 3 & 4 & 3 & 3 \\ 6 & 6 & 5 & 6 & 6 & 5 & 6 & 6 & 5 & 6 & 6 & 4 & 5 & 5 & 4 \end{bmatrix}$$

which contains all the permutations of $(1, 2, 3, 4, 5, 6)$ that correspond to a pair-partition.

B.2 Counting monomes

Now that we have a list of all the pair-partitions available, we can proceed to determine the coefficients of the Isserlis-Wick polynomials. Recall that we have found in Section 8.3 that a covariance of the form $\langle \zeta^n(\mathbf{x}+\mathbf{u})\zeta^m(\mathbf{x})\rangle / \sigma^{n+m}$ can be expressed as an integer polynomial in the auto-correlation $W(\mathbf{u})$ of degree at most k, where n and m are two positive integers such that $n+m=2k$ is an even integer (see Eq. 8.10). The reasoning that lead to this conclusion, presented in Section 8.3, was based on the fact that the covariance of a pair of random variable $\langle X_i X_j \rangle$, for $X_1 = \cdots = X_n = \zeta(\mathbf{x}+\mathbf{u})$ and $X_{n+1} = \cdots = X_{n+m} = \zeta(\mathbf{x})$, could possibly take only two values, σ^2 or $\sigma^2 W(\mathbf{u})$. The algorithm to obtain the polynomial coefficients $(p^{(j)}_{n,m})_{0\leq j\leq k}$ is essentially constructed on the very same observation. The idea is the following. For a given pair (n, m) and for a given term in the Isserlis sum, i.e. a given pair-partitioning of $[\![1, 2k]\!]$, we can count the degree of the resulting monome in $W(\mathbf{u})$, i.e. it will be some power j. Such a monome will hence contribute to one unit in $p^{(j)}_{n,m}$. Therefore, it suffices to loop over all pair-partitioning (which are now available thanks to the previous algorithm) and keep track of how many of each monome are generated. But how to we determine the degree of a monome resulting from a given pair-partitioning? Simply by going through each pair of the partition and checking whether the corresponding random variables are the same or not. The algorithm reads as follows.

1 Start with a one-dimensional array of size $k+1$ initialized with zeros. We will denote this array $\mathbf{P}$ and we choose to index its components with j from 0 to k. The component p_j in this array will keep the count of how many terms in the Isserlis sum leads to a monome of degree j.

2 Generate the list of all pair-partitions of $[\![1, n+m]\!]$ with the previous algorithm. We then have $\mathbf{M}^{(k)}$ available (recall $n+m=2k$ here). We will drop the superscript (k) for clarity since we work at k given.

3 Loop over each pair-partition: for i from 1 to ν_k

3.1 Start a counter to be incremented $j \leftarrow 0$.

3.2 Loop over pairs in the partition i: for q from 1 to $2k$ in step of 2

* Check the whether the covariance of $X_{M_{i,q}}$ and $X_{M_{i,q+1}}$ is σ^2 or $\sigma^2 W(\mathbf{u})$ and increment the power index accordingly:
if ($M_{i,q} < n+1/2$ and $M_{i,q+1} > n+1/2$) or ($M_{i,q} > n+1/2$ and $M_{i,q+1} < n+1/2$) then $j \leftarrow j+1$.

3.3 Increment p_j by one unit: $\mathrm{p}_j \leftarrow \mathrm{p}_j + 1$

4 All the $(p^{(j)}_{n,m})_{0\leq j\leq k}$ are now stored in $\mathbf{P}$.

Let us visualize our algorithm working on a simple example. Assume we are given $\mathbf{M}^{(2)}$ from Eq. (B.3), and we wish to compute the $(p^{(j)}_{3,1})_{0\leq j\leq 2}$, and the $(p^{(j)}_{2,2})_{0\leq j\leq 2}$. Let us start with $(n, m) = (3, 1)$, hence $n + 1/2 = 3.5$. We initialize $\mathbf{P} = (\mathrm{p}_0, \mathrm{p}_1, \mathrm{p}_2) = (0, 0, 0)$.

- Let us set j to zero and go through the pairs of the *first* row of $\mathbf{M}^{(2)}$. The first pair is $(1, 2)$, both elements are smaller than 3.5 hence nothing happens to j. The next pair is $(3, 4)$ where one element is smaller than 3.5 and the second is larger, hence we increment j by one. This is it for the first row, then since $j = 1$, we increment p_1 by one, i.e. that now $\mathbf{P} = (0, 1, 0)$.
- Let us set j to zero and go through the pairs of the *second* row of $\mathbf{M}^{(2)}$. The first pair is $(1, 3)$, both elements are smaller than 3.5 hence nothing happens to j. The next pair is $(2, 4)$ where one element is smaller than 3.5 and the second is larger, hence we increment j by one. This is it for the second row, then since $j = 1$, we increment p_1 by one, i.e. that now $\mathbf{P} = (0, 2, 0)$.
- Let us set j to zero and go through the pairs of the *third* row of $\mathbf{M}^{(2)}$. The first pair is $(1, 4)$ where one element is smaller than 3.5 and the second is larger, hence we increment j by one. The next pair is $(2, 3)$, both elements are smaller than 3.5 hence nothing more happens to j. This is it for the third row, then since $j = 1$, we increment p_1 by one, i.e. that now $\mathbf{P} = (0, 3, 0)$.

Let us continue with $(n, m) = (2, 2)$, hence $n + 1/2 = 2.5$. We initialize $\mathbf{P} = (\mathrm{p}_0, \mathrm{p}_1, \mathrm{p}_2) = (0, 0, 0)$.

- Let us set j to zero and go through the pairs of the *first* row of $\mathbf{M}^{(2)}$. The first pair is $(1, 2)$, both elements are smaller than 2.5 hence nothing happens to j. The next pair is $(3, 4)$, both elements are smaller than 2.5 hence nothing happens to j. This is it for the first row, then since $j = 0$, we increment p_0 by one, i.e. that now $\mathbf{P} = (1, 0, 0)$.
- Let us set j to zero and go through the pairs of the *second* row of $\mathbf{M}^{(2)}$. The first pair is $(1, 3)$ where one element is smaller than 2.5 and the second is larger, hence we increment j by one. The next pair is $(2, 4)$ where one element is smaller than 2.5 and the second is larger, hence we increment j by one. This is it for the second row, then since $j = 2$, we increment p_2 by one, i.e. that now $\mathbf{P} = (1, 0, 1)$.
- Let us set j to zero and go through the pairs of the *third* row of $\mathbf{M}^{(2)}$. The first pair is $(1, 4)$ where one element is smaller than 2.5 and the second is larger, hence we increment j by one. The next pair is $(2, 3)$ where one element is smaller than 2.5 and the second is larger, hence we increment j by one. This is it for the third row, then since $j = 2$, we increment p_2 by one, i.e. that now $\mathbf{P} = (1, 0, 2)$.

We see with this two simple examples that we indeed recover the coefficients $(p^{(j)}_{3,1})_{0\leq j\leq 2}$, and $(p^{(j)}_{2,2})_{0\leq j\leq 2}$ found by hand and reported in Table. 8.1.

Bibliography

[1] A. K. González-Alcalde, J.-P. Banon, Ø. S. Hetland, A. A. Maradudin, E. R. Méndez, T. Nordam, and I. Simonsen. Experimental and numerical studies of the scattering of light from a two-dimensional randomly rough interface in the presence of total internal reflection: optical Yoneda peaks. *Opt. Express*, 24(23):25995–26005, 2016.

[2] J.-P. Banon, Ø. S. Hetland, and I. Simonsen. On the physics of polarized light scattering from weakly rough dielectric surfaces: Yoneda and Brewster scattering phenomena. *Submitted to Phys. Rev. A, arXiv:1804.07507*, 2018.

[3] J.-P. Banon, Ø. S. Hetland, and I. Simonsen. Selective enhancement of Selényi rings induced by the cross-correlation between the interfaces of a two-dimensional randomly rough dielectric film. *Ann. Phys.*, 389:352–382, 2018.

[4] J.-P. Banon, Ø. S. Hetland, and I. Simonsen. A reduction of variance technique applied to the simulation of light scattering from randomly rough surfaces. *Manuscript in preparation*, 2018.

[5] M. Kildemo, J.-P. Banon, A. Baron, B. B. Svendsen, T. Brakstad, and I. Simonsen. Optical response of gold hemispheroidal lattices on transparent substrates. *App. Surf. Sci.*, 421:593 – 600, 2017.

[6] J.-P. Banon, T. Nesse, Z. Ghadyani, M. Kildemo, and I. Simonsen. Critical dimension metrology of a plasmonic photonic crystal based on Mueller matrix ellipsometry and the reduced Rayleigh equation. *Opt. Lett.*, 42(13):2631–2634, 2017.

[7] S. Enoch and N. Bonod. *Plasmonics, From Basics to Advanced Topics*. Springer Series in Optical Sciences, 2012.

[8] S. Maier. *Plasmonics, Fundamentals and Applications*. Springer NY, 2007.

[9] B. Liedberg, C. Nylander, and I. Lunström. Surface plasmon resonance for gas detection and biosensing. *Sensors and Actuators*, 4:299 – 304, 1983.

[10] J.-J. Greffet, R. Carminati, K. Joulain, J.-P. Mulet, S. Mainguy, and Y. Chen. Coherent emission of light by thermal sources. *Nature*, 416:61–67, 2002.

[11] J.-P. Mulet, K. Joulain, R. Carminati, and J.-J. Greffet. Enhanced radiative heat transfer at nanometric distances. *Microscale Thermophysical Engineering*, 6(3):209–222, 2002.

[12] M. Laroche, C. Arnold, F. Marquier, R. Carminati, J.-J. Greffet, S. Collin, N. Bardou, and J.-L. Pelouard. Highly directional radiation generated by a tungsten thermal source. *Opt. Lett.*, 30(19):2623–2625, 2005.

[13] Y. De Wilde, F. Formanek, R. Carminati, B. Gralak, P. A. Lemoine, K. Joulain, J.-P. Mulet, Y. Chen, and J.-J. Greffet. Thermal radiation scanning tunnelling microscopy. *Nature*, 444:740–743, 2006.

[14] X. Liu, T. Tyler, T. Starr, A. F. Starr, N. M. Jokerst, and W. J. Padilla. Taming the Blackbody with Infrared Metamaterials as Selective Thermal Emitters. *Phys. Rev. Lett.*, 107:045901, 2011.

[15] Lord Rayleigh. On the Dynamical Theory of Gratings. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 79(532):399–416, 1907.

[16] D. B. Melrose and R. C. McPhedran. *Electromagnetic Processes in Dispersive Media: A Treatment Based on the Dielectric Tensor*. Cambridge University Press, 1991.

[17] J. D. Jackson. *Classical Electrodynamics*. John Wiley & Sons, New York, 3rd edition, 1999.

[18] M. Joffre. *Optique non-linéaire en régimes continu et femtoseconde*. Lecture notes at École Polytechnique, 2014.

[19] P. Drude. Zur Elektronentheorie der Metalle. *Annalen der Physik*, 306(3):566–613, 1900.

[20] P. B. Johnson and R. W. Christy. Optical Constants of the Noble Metals. *Phys. Rev. B*, 6:4370–4379, 1972.

[21] M. A. Ordal, R. J. Bell, R. W. Alexander, L. L. Long, and M. R. Querry. Optical properties of fourteen metals in the infrared and far infrared: Al, Co, Cu, Au, Fe, Pb, Mo, Ni, Pd, Pt, Ag, Ti, V, and W. *Appl. Opt.*, 24(24):4493–4499, 1985.

[22] A. D. Rakić, A. B. Djurišić, J. M. Elazar, and M. L. Majewski. Optical properties of metallic films for vertical-cavity optoelectronic devices. *Appl. Opt.*, 37(22):5271–5283, 1998.

[23] J.-C. Nédélec. *Acoustic and Electromagnetic Equations: Integral Representations for Harmonic Problems*. Springer, Applied mathematical sciences 144, 2001.

[24] C. M. Bender and S. A. Orszag. *Advanced Mathematical Methods for Scientists and Engineers I: Asymptotic Methods and Perturbation Theory*. Springer, 1978.

[25] G. S. Agarwal. Interaction of electromagnetic waves at rough dielectric surfaces. *Phys. Rev. B*, 15(4):2371–2383, 1977.

[26] K. Miyamoto and E. Wolf. Generalization of the Maggi-Rubinowicz Theory of the Boundary Diffraction Wave Part I. *J. Opt. Soc. Am.*, 52(6):615–625, 1962.

[27] W. T. Doyle. Scattering approach to Fresnel's equations and Brewster's law. *Am. J. Phys.*, 53(5):463–468, 1985.

[28] P. P. Ewald. Zur Begründung der Kristalloptik. *Annalen der Physik*, 354(1):1–38, 1916.

[29] R. W. Wood. On a remarkable case of uneven distribution of light in a diffraction grating spectrum. *Philosophical Magazine*, 4(21):396–402, 1902.

[30] Lord Rayleigh. Note on the remarkable case of diffraction spectra described by Prof. Wood. *Philosophical Magazine*, 14(79):60–65, 1907.

[31] J. Strong. Effect of Evaporated Films on Energy Distribution in Grating Spectra. *Phys. Rev.*, 49:291–296, 1936.

[32] U. Fano. The Theory of Anomalous Diffraction Gratings and of Quasi-Stationary Waves on Metallic Surfaces (Sommerfeld's Waves). *J. Opt. Soc. Am.*, 31(3):213–222, 1941.

[33] A. Hessel and A. A. Oliner. A New Theory of Wood's Anomalies on Optical Gratings. *Appl. Opt.*, 4(10):1275–1297, 1965.

[34] M. C. Hutley and V. M. Bird. A detailed experimental study of the anomalies of a sinusoidal diffraction grating. *Opt. Acta*, 20:771–782, 1973.

[35] R.C. McPhedran and D. Maystre. A Detailed Theoretical Study of the Anomalies of a Sinusoidal Diffraction Grating. *Optica Acta: International Journal of Optics*, 21(5):413–421, 1974.

[36] A. A. Maradudin, T. Michel, A. R. McGurn, and E. R. Méndez. Enhanced backscattering of light from a random grating. *Ann. Phys. (N.Y.)*, 203(2):255 – 307, 1990.

[37] F. Toigo, A. Marvin, V. Celli, and N. R. Hill. Optical properties of rough surfaces: General theory and the small roughness limit. *Phys. Rev. B*, 15:5618–5626, 1977.

[38] G.C. Brown, V. Celli, M. Coopersmith, and M. Haller. Unitary and reciprocal expansions in the theory of light scattering from a grating. *Surface Science*, 129(2):507 – 515, 1983.

[39] G.C. Brown, V. Celli, M. Haller, and A. Marvin. Vector theory of light scattering from a rough surface: Unitary and reciprocal expansions. *Surface Science*, 136(2):381 – 397, 1984.

[40] G. Brown, V. Celli, M. Haller, A. A. Maradudin, and A. Marvin. Resonant light scattering from a randomly rough surface. *Phys. Rev. B*, 31:4993–5005, 1985.

[41] T. A. Leskova, P. A. Letnes, A. A. Maradudin, T. Nordam, and I. Simonsen. The scattering of light from two-dimensional randomly rough surfaces. *Proc. SPIE*, 8172:817209–817209–20, 2011.

[42] T. Nordam, P. A. Letnes, I. Simonsen, and A. A. Maradudin. Numerical solutions of the Rayleigh equations for the scattering of light from a two-dimensional randomly rough perfectly conducting surface. *J. Opt. Soc. Am. A*, 31(5):1126–1134, May 2014.

[43] T. Nordam, P. A. Letnes, and I. Simonsen. Numerical simulations of scattering of light from two-dimensional rough surfaces using the reduced Rayleigh equation. *Frontiers in Physics*, 1, 2013.

[44] P. A. Letnes, T. Nordam, and I. Simonsen. Coherent effects in the scattering of light from two-dimensional rough metal surfaces. *J. Opt. Soc. Am. A*, 30(6):1136–1145, 2013.

[45] Ø. S. Hetland, A. A. Maradudin, T. Nordam, and I. Simonsen. Numerical studies of the scattering of light from a two-dimensional randomly rough interface between two dielectric media. *Phys. Rev. A*, 93:053819, 2016.

[46] Ø. S. Hetland, A. A. Maradudin, T. Nordam, P. A. Letnes, and I. Simonsen. Numerical studies of the transmission of light through a two-dimensional randomly rough interface. *Phys. Rev. A*, 95:043808, 2017.

[47] T. Nordam, P. A. Letnes, I. Simonsen, and A. A. Maradudin. Satellite peaks in the scattering of light from the two-dimensional randomly rough surface of a dielectric film on a planar metal surface. *Opt. Express*, 20(10):11336–11350, 2012.

[48] M. Kretschmann and A. A. Maradudin. Band structures of two-dimensional surface-plasmon polaritonic crystals. *Phys. Rev. B*, 66:245408, 2002.

[49] A. V. Shchegrov, I. V. Novikov, and A. A. Maradudin. Scattering of Surface Plasmon Polaritons by a Circularly Symmetric Surface Defect. *Phys. Rev. Lett.*, 78:4269–4272, 1997.

[50] J. A. Sánchez-Gil and A. A. Maradudin. Surface-plasmon polariton scattering from a finite array of nanogrooves / ridges: Efficient mirrors. *Applied Physics Letters*, 86(25):251106, 2005.

[51] R. E. Arias and A. A. Maradudin. Scattering of a surface plasmon polariton by a localized dielectric surface defect. *Opt. Express*, 21(8):9734, 2013.

[52] A. Soubret, G. Berginc, and C. Bourrely. Backscattering enhancement of an electromagnetic wave scattered by two-dimensional rough layers. *J. Opt. Soc. Am. A*, 18:2778–2788, 2001.

[53] A. Soubret, G. Berginc, and C. Bourrely. Application of reduced Rayleigh equations to electromagnetic wave scattering by two-dimensional randomly rough surfaces. *Phys. Rev. B*, 63:245411, 2001.

[54] R. F. Millar. On the Rayleigh assumption in scattering by a periodic surface. *Math. Proc. Camb.*, 65:773–791, 1969.

[55] R. F. Millar. On the Rayleigh assumption in scattering by a periodic surface. II. *Math. Proc. Camb.*, 69:217–225, 1971.

[56] A. A. Maradudin. Iterative solutions for electromagnetic scattering by gratings. *J. Opt. Soc. Am.*, 73(6):759–764, 1983.

[57] F. G. Tricomi. *Integral Equations*. Dover Books on Mathematics. Dover Publications, revised edition, 1985.

[58] V. N. G. Kampen. *Stochastic Processes in Physics and Chemistry*. North-Holland Personal Library. Elsevier, Amsterdam, third edition, 2007.

[59] Max Born and Emil Wolf. *Principles of Optics*. Pergamon, 6th edition, 1980.

[60] G. Allaire and S. M. Kaber. *Numerical Linear Algebra*. Text in applied mathematics, Springer, 2008.

[61] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes*. Cambridge University Press, 3rd edition, 2007.

[62] J.-P. Grivet. *Méthodes Numériques Appliquées pour le scientifique et l'ingénieur*. EDP Science, Collection Grenoble Sciences, 2009.

[63] D. E. Amos. Algorithm 644, A Portable Package For Bessel Functions of a Complex Argument and Nonnegative Order. *ACM Transactions on Mathematical Software*, 12:265–273, 1986.

[64] D. E. Amos. Remark on Algorithm 644. *ACM Transactions on Mathematical Software*, 16:404, 1990.

[65] J. A. Sánchez-Gil, A. A. Maradudin, and E. R. Méndez. Limits of validity of three perturbation theories of the specular scattering of light from one-dimensional, randomly rough, dielectric surfaces. *J. Opt. Soc. Am. A*, 12(7):1547–1558, 1995.

[66] T. Nordam, P. A. Letnes, and I. Simonsen. Validity of the Rayleigh hypothesis for two-dimensional randomly rough metal surfaces. *24th IUPAP Conference on Computational Physics (IUPAP-CCP 2012), Journal of Physics: Conference Series*, 454:012033, 2013.

[67] J.-J. Greffet and Z. Maassarani. Scattering of electromagnetic waves by a grating: a numerical evaluation of the iterative-series solution. *J. Opt. Soc. Am. A*, 7(8):1483–1493, 1990.

[68] J.-J. Greffet, C. Baylard, and P. Versaevel. Diffraction of electromagnetic waves by crossed gratings: a series solution. *Opt. Lett.*, 17(24):1740–1742, 1992.

[69] B. A. Lippmann. Note on the Theory of Gratings∗. *J. Opt. Soc. Am.*, 43(5):408–408, May 1953.

[70] R. Petit and M. Cadilhac. Sur la diffraction d'une onde plane par un réseau infiniment conducteur. *C.R. Acad. Sci. Paris, Sér. A.-B.*, 262:468–471, 1966.

[71] R. Petit. Electromagnetic grating theories: limitations and successes. *Nouv. Rev. Opt.*, 6:129–135, 1975.

[72] Wirgin A. Reflection by a grating: Rayleigh methods: comments. *JOSA Letters*, 72(7):963–969, 1982.

[73] A. V. Tishchenko. Numerical demonstration of the validity of the Rayleigh hypothesis. *Optics Express*, 17(19):17102–17117, 2009.

[74] Wauer J. and Rother T. Considerations to Rayleigh's hypothesis. *Optics Communications*, 282:339–350, 1982.

[75] P. M. van den Berg. Reflection by a grating: Rayleigh methods. *J. Opt. Soc. Am.*, 71(10):1224–1229, 1981.

[76] P. M. van den Berg. Reflection by a grating: Rayleigh methods: author's reply to comments by A. Wirgin. *J. Opt. Soc. Am.*, 22(7):970–971, 1982.

[77] P. M. van den Berg and J. T. Fokkema. The Rayleigh hypothesis in the theory of reflection by a grating. *J. Opt. Soc. Am.*, 69(1):27–31, 1978.

[78] P. M. van den Berg and J. T. Fokkema. The Rayleigh hypothesis in the theory of diffraction by a perturbation in a plane surface. *Radio Science*, 15(4):723–732, 1980.

[79] B. Ruffing. Application of speckle-correlation methods to surface-roughness measurement: a theoretical study. *J. Opt. Soc. Am. A*, 3(8):1297–1304, 1986.

[80] K. D. Hinsch, T. Fricke-Begemann, G. Gülker, and K. Wolff. Speckle correlation for the analysis of random processes at rough surfaces. *Optics and Lasers in Engineering*, 33(2):87 – 105, 2000.

[81] J. Goodman. *Statistical Optics*. Wiley, New York, 1985.

[82] J. Goodman. *Statistical Phenomena in Optics*. Roberts and Company, Greenwood, 2007.

[83] N. Fayard, A. Goetschy, R. Pierrat, and R. Carminati. Mutual Information between Reflected and Transmitted Speckle Images. *Phys. Rev. Lett.*, 120:073901, 2018.

[84] L. Isserlis. On certain probable errors and correlation coefficients of multiple frequency distributions with skew regression. *Biometrika*, 11(3):185, 1916.

[85] L. Isserlis. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12(1-2):134, 1918.

[86] G. C. Wick. The Evaluation of the Collision Matrix. *Phys. Rev.*, 80:268–272, 1950.

[87] E. Gobet. *Méthodes de Monte-Carlo et processus stochastiques: du linéaire au non-linéaire*. Éditions de l'École Polytechnique, 2013.

[88] S. Asmussen and P. W. Glynn. *Stochastic Simulation; Algorithms and Analysis*. Springer, 2007.

[89] P. Lorrain and D. R. Corson. *Electromagnetic Fields and Waves*. Freeman, San Francisco, 2nd edition, 1970.

[90] Y. Yoneda. Anomalous Surface Reflection of X Rays. *Phys. Rev.*, 131:2010–2013, 1963.

[91] G. H. Vineyard. Grazing-incidence diffraction and the distorted-wave approximation for the study of surfaces. *Phys. Rev. B*, 26:4146–4159, 1982.

[92] S. K. Sinha, E. B. Sirota, S. Garoff, and H. B. Stanley. X-ray and neutron scattering from rough surfaces. *Phys. Rev. B*, 38:2297–2311, 1988.

[93] E. E. Gorodnichev, S. L. Dudarev, D. B. Rogozkin, and M. I. Ryazanov. Nature of anomalous x-ray reflection from a surface. *Sov. Phys. JETP Lett.*, 48(3):147–150, 1988.

[94] T. A. Leskova and A. A. Maradudin. X-ray scattering from a randomly rough surface. *Waves in Random Media*, 7(3):395–434, 1997.

[95] G. Renaud, R. Lazzari, and F. Leroy. Probing surface and interface morphology with Grazing Incidence Small Angle X-Ray Scattering . *Surf. Sci. Rep.*, 64(8):255 – 380, 2009.

[96] H. Dosch. Evanescent absorption in kinematic surface Bragg diffraction. *Phys. Rev. B*, 35:2137–2143, 1987.

[97] S. Stepanov. X-ray Diffuse Scattering from Interfaces in Semiconductor Multilayers. In N.Q. Lam, C.A. Melendres, and S.K. Sinha, editors, *Exploration of Subsurface Phenomena by Particle Scattering*, pages 119–137. International Advanced Studies Institute Press, North East, Maryland, 2000.

[98] Amane Kitahara, Koji Inoue, Hiroyuki Kikkawa, Kiyohiko Matsushita, and Isao Takahashi. X-ray Reflectivity study on polymeric surfaces near glass transition temperature. Photon factory activity report no. 20, part b, 2002 (unpublished). p. 83.

[99] J-C Gasse, D Lützenkirchen-Hecht, R Wagner, and R Frahm. Yoneda-XAFS with Area X-Ray Detectors. *J. Phys.: Conf. Ser.*, 712(1):012028, 2016.

[100] M. Nieto-Vesperinas and J. A. Sánchez-Gil. Light scattering from a random rough interface with total internal reflection. *J. Opt. Soc. Am. A*, 9(3):424–436, 1992.

[101] T Kawanishi, H Ogura, and Z L Wang. Scattering of an electromagnetic wave from a slightly random dielectric surface: Yoneda peak and Brewster angle in incoherent scattering. *Wave. Random Media*, 7(3):351–384, 1997.

[102] W. T. Doyle. Graphical approach to Fresnel's equations for reflection and refraction of light. *Am. J. Phys.*, 48(8):643–647, 1980.

[103] M. Saillard and D. Maystre. Scattering from metallic and dielectric rough surfaces. *J. Opt. Soc. Am. A*, 7(6):982–990, 1990.

[104] M. Nieto-Vesperinas, J. A. Sánchez-Gil, and A. A. Maradudin. Multiple light scattering from metal and dielectric rough surfaces. *Waves in Random Media*, 3:157–163, 1991.

[105] J.-J. Greffet. Theoretical model of the shift of the Brewster angle on a rough surface. *Opt. Lett.*, 17(4):238–240, 1992.

[106] K. A. O'Donnell and E. R. Méndez. Enhanced specular peaks in diffuse light scattering from weakly rough metal surfaces. *J. Opt. Soc. Am. A*, 20(12):2338–2346, 2003.

[107] A. R. McGurn, A. A. Maradudin, and V. Celli. Localization effects in the scattering of light from a randomly rough grating. *Phys. Rev. B*, 31:4866–4871, 1985.

[108] A. A. Maradudin and E. R. Méndez. Enhanced backscattering of light from weakly rough, random metal surfaces. *Appl. Opt.*, 32(19):3335–3343, 1993.

[109] C. S. West and K. A. O'Donnell. Observations of backscattering enhancement from polaritons on a rough metal surface. *J. Opt. Soc. Am. A*, 12(2):390–397, 1995.

[110] V. Freilikher, E. Kanzieper, and A.A. Maradudin. Coherent scattering enhancement in systems bounded by rough surfaces. *Physics Reports*, 288(1):127 – 204, 1997. I.M. Lifshitz and Condensed Matter Theory.

[111] E.R. Mendez and K.A. O'Donnell. Observation of depolarization and backscattering enhancement in light scattering from gaussian random surfaces. *Optics Communications*, 61(2):91 – 95, 1987.

[112] K. A. O'Donnell and E. R. Mendez. Experimental study of scattering from characterized random surfaces. *J. Opt. Soc. Am. A*, 4(7):1194–1205, 1987.

[113] M. Nieto-Vesperinas and J. M. Soto-Crespo. Monte Carlo simulations for scattering of electromagnetic waves from perfectly conductive random rough surfaces. *Opt. Lett.*, 12(12):979–981, 1987.

[114] Paul Selényi. Über Lichtzerstreuung im Raume Wienerscher Interferenzen und neue, diesen reziproke Interferenzerscheinungen. *Ann. Phys.*, 340(8):444–460, 1911.

[115] V. Freilikher, V. I. Tatarskii, M. Pustilnik, and I. Yurkevich. Polarization of light scattered from slightly rough dielectric film. *Opt. Lett.*, 19(18):1382, 1994.

[116] J. Q. Lu, J. A. Sánchez-Gil, E. R. Méndez, Z.-H. Gu, and A. A. Maradudin. Scattering of light from a rough dielectric film on a reflecting substrate: diffuse fringes. *J. Opt. Soc. Am. A*, 15(1):185, 1998.

[117] O. Calvo-Perez, A. Sentenac, and J. J. Greffet. Light scattering by a two-dimensional, rough penetrable medium : A mean-field theory. *Radio Sci.*, 34(2):311–335, 1999.

[118] W. Suhr and H. J. Schlichting. Quételet's fringes due to scattering by small spheres just above a reflecting surface. *Appl. Opt.*, 48(26):4978–84, 2009.

[119] A. J. de Witte. Interference in Scattered Light. *Am. J. Phys.*, 35(4):301–313, 1967.

[120] C.V. Raman and Goverdhan Lal Datta. On Quételet's rings and other allied phenomena. *Philos. Mag.*, 42(251):826–840, 1921.

[121] V. Freilikher, M. Pustilnik, and I. Yurkevich. Wave scattering from a bounded medium with disorder. *Physics Letters A*, 193(5):467 – 470, 1994.

[122] E. R. Méndez, E. I. Chaikina, and H. M. Escamilla. Observation of satellite peaks and dips in the scattering of light in a double-pass geometry. *Opt. Lett.*, 24(11):705–707, 1999.

[123] K. Levenberg. A Method for the Solution of Certain Problems in Least Squares. *Quart. Appl. Math.*, 2:164–168, 1944.

[124] D. Marquardt. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *J. Appl. Math.*, 11:431–441, 1963.

[125] D. E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning.* Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1st edition, 1989.

[126] M. Mitchell. *An Introduction to Genetic Algorithms.* Cambridge, MA: MIT Press, 1996.

[127] N. Metropolis, A. W. Rosenbluth, Rosenbluth M. N., A. H. Teller, and E. Teller. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[128] W. K. Hastings. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, 57(1):97–109, 1970.

[129] Y. Cao, S. Li, L. Petzold, and R. Serban. Adjoint Sensitivity Analysis for Differential-Algebraic Equations: The Adjoint DAE System and Its Numerical Solution. *SIAM Journal on Scientific Computing*, 24(3):1076–1089, 2003.

[130] G. Strang. *Computational Science and Engineering.* Wellesley, MA: Wellesley-Cambridge Press, 2007.

[131] M. Ghebrebrhan, P. Bermel, Y. Avniel, J. D. Joannopoulos, and S. G. Johnson. Global optimization of silicon photovoltaic cell front coatings. *Opt. Express*, 17(9):7505–7518, 2009.

[132] D. C. Dobson and S. J. Cox. Maximizing Band Gaps in Two-Dimensional Photonic Crystals. *SIAM Journal on Applied Mathematics*, 59(6):2108–2120, 1999.

[133] C. Y. Kao, S. Osher, and E. Yablonovitch. Maximizing band gaps in two-dimensional photonic crystals by using level set methods. *Appl. Phys. B*, 81((2-3)):235–244, 2005.

[134] W. R. Frei, D. A. Tortorelli, and H. T. Johnson. Topology optimization of a photonic crystal waveguide termination to maximize directional emission. *Applied Physics Letters*, 86(11):111114, 2005.

[135] Y. Tsuji and K. Hirayama. Design of Optical Circuit Devices Using Topology Optimization Method With Function-Expansion- Based Refractive Index Distribution. *IEEE Photonics Technology Letters*, 20(12):982–984, 2008.

[136] B. K. Minhas, S. A. Coulombe, S. Sohail, H. Naqvi, and J. R. McNeil. Ellipsometric scatterometry for the metrology of sub-0.1-μm-linewidth structures. *Appl. Opt.*, 37:5112–5115, 1998.

[137] X.-T. Huang and Jr. Terry, F. L. Erratum to Spectroscopic ellipsometry and reflectrometry from grating (scatterometry) for critical dimension measurement and *in situ*, real-time process monitoring [thin solid films 455-456, 828-836 (2004)]. *Thin Solid Films*, 486:339–346, 2004.

[138] T. Novikova, A. De Martino, S. Ben Hatit, and B. Drévillon. Application of Mueller polarimetry in conical diffraction for critical dimension measurements in microelectronics. *Appl. Opt.*, 45:3688–3697, 2006.

[139] B. Kaplan, T. Novikova, A. De Martino, and B. Drévillon. Characterization of bidimensional gratings by spectroscopic ellipsometry and angle-resolved Mueller polarimetry. *Appl. Opt.*, 43:1233–1240, 2004.

[140] M. Foldyna, A. De Martino, E. Garcia-Caurel, R. Ossikovski, C. Licitra, F. Bertin, K. Postava, and B. Drévillon. Critical dimension of biperiodic gratings determined by spectral ellipsometry and Mueller matrix polarimetry. *Eur. Phys. J. Appl. Phys.*, 42:351–359, 2008.

[141] T. Brakstad, M. Kildemo, Z. Ghadyani, and I. Simonsen. Dispersion of polarization coupling, localized and collective plasmon modes in metallic photonic crystal mapped out by Mueller Matrix Ellipsometry. *Opt. Express*, 23:22800–22815, 2015.

[142] A. Schönhofer and H.-G. Kuball. Symmetry properties of the Mueller matrix. *Chem. Phys.*, 115(2):159 – 167, 1987.

[143] C. M. Herzinger, B. Johs, W. A. McGahan, J. A. Woollam, and W. Paulson. Ellipsometric determination of optical constants for silicon and thermally grown silicon dioxide via a multi-sample, multi-wavelength, multi-angle investigation. *J. Appl. Phys.*, 83(6):3323–3336, 1998.

[144] I. Simonsen, Ø. S. Hetland, J. B. Kryvi, and A. A. Maradudin. Determination of the normalized-surface-height autocorrelation function of a two-dimensional randomly rough dielectric surface by the inversion of light-scattering data. *Phys. Rev. A*, 93:043829, Apr 2016.

[145] J. B. Kryvi, I. Simonsen, and A. A. Maradudin. Determination of the normalized surface height autocorrelation function of a two-dimensional randomly rough dielectric surface by the inversion of light scattering data in p-polarization. *Proc.SPIE*, 9961:9961 – 9961 – 14, 2016.

[146] S. Chakrabarti, A. A. Maradudin, I. Simonsen, and E. I. Chaikina. The inversion of inchoherent light scattering data to obtain statistical and optical properties of a two-dimensional randomly rough dielectric surface. *Proc.SPIE*, 9205:9205 – 9205 – 7, 2014.

[147] I. S. Gradshteyn and I. M. Ryzhik. *Table of integrals, series, and products.* Academic Press, 1965.

Part IV

Papers

Paper 1:

A. K. González-Alcalde, J.-P. Banon, Ø. S. Hetland, A. A. Maradudin, E. R. Méndez, T. Nordam, and I. Simonsen. Experimental and numerical studies of the scattering of light from a two-dimensional randomly rough interface in the presence of total internal reflection: optical Yoneda peaks. *Opt. Express*, 24(23):25995–26005, 2016

Is not included due to copyright

Paper 2:

J.-P. Banon, Ø. S. Hetland, and I. Simonsen. On the physics of polarized light scattering from weakly rough dielectric surfaces: Yoneda and Brewster scattering phenomena. *Submitted to Phys. Rev. A, arXiv:1804.07507*, 2018

On the physics of polarized light scattering from weakly rough dielectric surfaces: Yoneda and Brewster scattering phenomena

J.-P. Banon[1], Ø. S. Hetland[1], and I. Simonsen[1,2]
[1] *Department of Physics, NTNU Norwegian University of Science and Technology, NO-7491 Trondheim, Norway and*
[2] *Surface du Verre et Interfaces, UMR 125 CNRS/Saint-Gobain, F-93303 Aubervilliers, France*
(Dated: April 20, 2018)

The optical Yoneda and Brewster scattering phenomena are studied theoretically based on pertubative solutions of the reduced Rayleigh equations. The Yoneda phenomenon is characterized as an enhancement of the intensity of the diffuse light scattered by a randomly rough interface between two dielectric media when the light is observed in the optically denser medium. The intensity enhancement occurs above a critical angle of scattering which is independent of the angle of incidence of the excitation. The Brewster scattering phenomenon is characterized by a zero scattered intensity either in the reflected or transmitted light for an angle of scattering which depends on the angle of incidence. We also describe a generalization of the Brewster scattering phenomenon for outgoing evanescent waves and circularly-polarized waves. The physical mechanisms responsible for these phenomena are described in terms of simple notions such as scalar waves, oscillating and rotating dipoles and geometrical arguments, and are valid in a regime of weakly rough interfaces.

I. INTRODUCTION

When light is scattered in either reflection or transmission from or through a weakly rough interface, two phenomena of interest can be observed in the scattered intensity distributions. These are the *Yoneda phenomenon*, relatable to the idea of total internal reflection, and the *Brewster scattering phenomenon*, relatable to the polarizing angle.

The Yoneda phenomenon is characterized as an enhancement of the intensity of the light scattered diffusely by a randomly rough interface between two dielectric media when the light is observed in the optically denser medium. The intensity enhancement occurs above a critical angle of scattering which is independent of the angle of incidence of the excitation. This critical angle is always the polar angle, in the denser medium, for which the wavenumber of a plane wave turns non-propagating in the less dense medium [1, 2]. Although well known in the scattering of x-rays from both metallic [3–8] and non-metallic [9–12] surfaces, a paper by Kawanishi *et al.* [13] marks their first explicit appearance in optics[14]. Kawanishi *et al.*, by the use of the stochastic functional approach, studied the case where a two-dimensional randomly rough interface between two dielectric media is illuminated by p- or s-polarized light from either medium. They obtained several interesting properties of the reflected and transmitted light that are associated with the phenomenon of total internal reflection when the medium of observation is the optically denser medium. These include the appearance of Yoneda peaks, which were described by the authors as "quasi-anomalous scattering peaks." As an interpretation of their results, the authors suggested that the Yoneda peaks may be associated with the presence of lateral waves [15] propagating along the interface in the optically less dense medium. Although the mathematical origin of the Yoneda effect has been shown through various perturbative approaches based on the reduced Rayleigh equations (RRE), a physical interpretation of the effect is still under discussion; a summary of which can be found in Ref. 1. Optical Yoneda peaks were recently observed experimentally for a configuration of reflection from a randomly rough dielectric interface, when the medium of incidence was the optically denser medium [16].

The *Brewster angle* is maybe the best known planar surface reflection effect where the polarization of light plays a major role. Proposed as a *polarizing angle* by Sir David Brewster in 1812 [17], its exact definition has been a slight matter of debate in modern times [18]. For isotropic dielectric non-magnetic materials, however, it may be defined to be the angle of incidence, onto a planar dielectric surface, for which the reflection amplitude for p-polarized light (light polarized in the plane given by the incident light and the surface normal) is zero.

A complete physical understanding of the Brewster phenomenon is, at best, non trivial. The most common explanation for the gradual disappearance of the reflection amplitude is based on the radiation pattern of dipoles induced in the scattering substrate [19, 20]. This idea is not new, and can be traced back to investigations by e.g. Sommerfeld [21]. Modelling the scattering from a rough surface as a layer of polarizable spheres led Greffet and Sentenac [22] to the same conclusion. In a later collaboration with Calvo-Perez, this point of view was reinforced through the development of, and the results given by, the Mean Field Theory (MFT) [23, 24]. However, amongst others Lekner [25] argues that even if the dipole argument holds great explanatory power for a wide range of scattering systems, he challenges the argument for the case of the Brewster angle for dielectric media. His main issue with the argument is that the accelerated electrons cannot oscillate as dipoles in the transmitted medium in the case of the wave approaching an interface with vacuum on the opposing side, since the argument goes that the dipoles are oriented according to the field in the refracted wave. Also, there is an analogue to Brew-

ster's angle for longitudinal acoustic waves called Green's angle, and in this case the radiation from each scatterer does not have dipole character [25].

These and other conceptual issues in the explanatory model for the Brewster angle are attempted reconciled by Doyle [26] in his work with a factored form of the Fresnel equations. Inspired by the work of Sein [27], and Pattanayak and Wolf [28] on the interpretation and generality of the extinction theorem, Doyle claims that the proper understanding of the Brewster phenomenon has been hampered by the attention given to surface sources through a slightly misunderstood interpretation of the Ewald-Oseen extinction theorem [29]. Doyle emphatizes the participation of the entire media in the creation of the reflected wave, and makes use of Ewald's original concept of "wave triads". Doyle's factored form of the Fresnel equations separately expresses the scattering pattern from individual dipoles and the coherent scattering function of the dipole array, and manages in this way to explain the polarizing angles for any combination of transparent media.

Kawanishi *et al.* [13] observed angles of zero scattering intensity to first order in their approach in the distributions of the intensity of the incoherently scattered light when the incident light was p-polarized. Due to their resemblance to the Brewster angle in the reflectivity from a flat interface, they dubbed these angles the "Brewster scattering angles". These angles were observed in both reflection *and* transmission, for light incident from either medium.

Both the Brewster scattering angles and Yoneda peaks were recently observed and discussed in numerical simulations of scattering in both reflection and transmission from weakly rough surfaces [1, 2], and also in a film geometry [30] where it was claimed that the phase shifts associated with these phenomena impact the angular positions of interference rings of diffusely scattered light, known as Selényi rings.

In this paper we seek to further illuminate the phenomena of Brewster scattering angles and Yoneda peaks and more generally identify the fundamental mechanisms at play in the scattering of polarized light by a weakly rough surface. After describing the statistical properties of the interface in Sec. II, we derive, in Sec. III, a set of reduced Rayleigh equations (RREs) for the case of electromagnetic scattering inspired by the work of Soubret *et al.* [31] and give the corresponding RRE for scalar waves subjected to the continuity of the scalar field and its normal derivatives with respect to the interface. Furthermore, we give an approximate solution of the RREs to first order in the surface profile function in a series expansion of the reflection and transmission amplitudes. The first order perturbative solution will be our main tool of investigation in Sec. IV. Section IV A is devoted to summarizing some phenomenological observations which have been obtained in the literature before embarking in Sec. IV B into a more in-depth analysis of the reflection and transmission amplitudes with special care given to their physical interpretation. In particular, we show how the response can be factorized as a product of a term reminiscent of a scalar wave response and a term encoding the component of the response specific to polarization. Such a factorization is a clear signature of two aspects of scattering by arrays of dipoles; the radiated power is controlled both by the interference of the spherical-like waves emitted by each atomic source and their individual characteristic dipolar radiation. Once the general physical interpretation of the equations is clarified, we explain in detail the origin of the Yoneda phenomenon in Sec. IV C, and show that it is fundamentally a single scattering, scalar wave phenomenon. The Brewster scattering phenomenon, and more generally all polarization induced effects, are then discussed thoroughly in Secs. IV D-IV F. We first restrict the analysis of the Brewster scattering phenomenon to scattering in the plane of incidence and derive a one-line criterion for predicting the Brewster scattering angle which allows for a simple geometrical interpretation. A detour via the analysis of the polarization properties of the radiation of oscillating and rotating dipoles in free space is made in Sec. IV E in order to facilitate the intuitive understanding of the full angular distribution of scattering by a rough surface discussed in Sec. IV F. Finally, Sec. V summarizes the conclusions we have drawn from this study and suggests experimental setups to test some interesting predictions made by the theory.

II. SCATTERING SYSTEMS

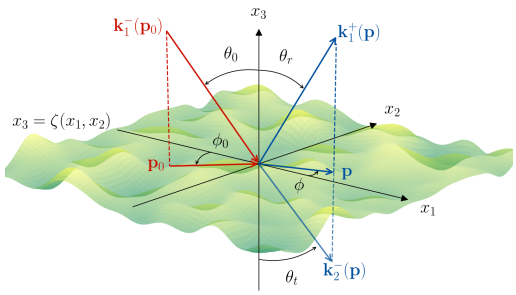


FIG. 1. Definitions of the angles of incidence and scattering, together with the relevant wave vectors.

The system we study in this work consists of a non-magnetic dielectric medium (medium 1), whose dielectric constant is $\epsilon_1 > 0$ (refractive index $n_1 = \sqrt{\epsilon_1}$), in the region $x_3 > \zeta(\mathbf{x}_\parallel)$, and a non-magnetic dielectric medium (medium 2), whose dielectric constant is $\epsilon_2 > 0$ (refractive index $n_2 = \sqrt{\epsilon_2}$), in the region $x_3 < \zeta(\mathbf{x}_\parallel)$ [Fig. 1]. The definition of the geometry is set in the three-dimensional space endowed with a Cartesian coordinate system $(O, \hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3)$, with the vector plane $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2)$ parallel to the mean plane of the interface. The origin, O,

of the coordinate system can be arbitrarily chosen, only affecting the complex reflection and transmission amplitudes by an overall phase factor which plays no role in the intensity of the scattered light. A point is then represented as $\mathbf{x} = \sum_{i=1}^{3} x_i \hat{\mathbf{e}}_i = \mathbf{x}_\parallel + x_3\,\hat{\mathbf{e}}_3$. An overview of a typical system geometry is provided in Fig. 1. The surface profile function ζ will be assumed to be a realization of a continuous, differentiable, single-valued, stationary, isotropic, Gaussian random process with zero mean and given auto-correlation. More specifically, the surface profile function is assumed to satisfy the following properties:

$$\left\langle \zeta(\mathbf{x}_\parallel) \right\rangle = 0\,, \tag{1a}$$

$$\left\langle \zeta(\mathbf{x}_\parallel)\zeta(\mathbf{x}'_\parallel) \right\rangle = \sigma^2\, W(\mathbf{x}_\parallel - \mathbf{x}'_\parallel)\,. \tag{1b}$$

Here and in the following, the angle brackets denote an average over an ensemble of realizations of the stochastic process, σ denotes the rms roughness and W the height auto-correlation function normalized so that $W(\mathbf{0}) = 1$. In particular, we will deal with the special case of a Gaussian auto-correlation function defined by

$$W(\mathbf{x}_\parallel) = \exp\left(-\frac{|\mathbf{x}_\parallel|^2}{a^2}\right)\,, \tag{2}$$

where a is the correlation length. The corresponding power spectrum (defined as the Fourier transform of W) is then

$$g(\mathbf{p}) = \pi a^2 \exp\left(-\frac{|\,\mathbf{p}\,|^2 a^2}{4}\right)\,, \tag{3}$$

with $\mathbf{p} = p_1\,\hat{\mathbf{e}}_1 + p_2\,\hat{\mathbf{e}}_2$.

III. THEORY

The theoretical approach used in this work to study the scattering of light from the systems of interest is based on the so-called reduced Rayleigh equations. A reduced Rayleigh equation is an inhomogeneous integral equation in which the integral kernel encodes the materials and geometry of the scattering system, and the unknowns are the reflection or transmission amplitudes for each polarization. First derived by Brown *et al.* [32], the reduced Rayleigh equation is obtained from the Rayleigh solution to the electromagnetic boundary problem. Using inspiration drawn from the extinction theorem it is possible to "reduce" the full Rayleigh equations through the elimination of either the reflected or transmitted field. In the following, in order to establish the notation and highlight the main assumptions of the method, we will briefly recall the key ideas of the derivation of the reduced Rayleigh equations for a system composed of two media separated by a rough interface. We will use, to our knowledge, the most general form of the reduced Rayleigh equations for a single interface derived by Soubret *et al.* in Ref. 31 and used by these authors in Refs. 31 and 33 in the case of a single interface system and a film geometry.

A. The reduced Rayleigh equations

In this work we assume the electromagnetic response of the media to be modeled by non-magnetic, homogeneous, isotropic, linear constitutive relations in the frequency domain. We consider the presence of an electromagnetic field $(\mathbf{E}, \mathbf{H})$ in the whole space, where their restriction will be denoted by a subscript j depending on the medium in which they are evaluated. As an example, the electric field evaluated at a point $\mathbf{x}$ in medium 1 at time t is denoted $\mathbf{E}_1(\mathbf{x}, t)$. The source free Maxwell equations, together with homogeneous, linear and isotropic constitutive relations in the frequency domain, result in the fact that the electric and magnetic fields satisfy the Helmholtz equation in each region. Namely, for $j \in \{1, 2\}$,

$$\nabla^2 \mathbf{E}_j(\mathbf{x}, \omega) + \epsilon_j(\omega)\,\left(\frac{\omega}{c}\right)^2\,\mathbf{E}_j(\mathbf{x}, \omega) = \mathbf{0}\,, \tag{4}$$

and similarly for $\mathbf{H}$. Here, ∇^2 denotes the vector Laplace operator and c is the speed of light in vacuum. Here onward, we will drop the time, or frequency, dependence, since we assume a stationary regime at a fixed frequency where time contributes only by an overall phase factor $\exp(-i\omega t)$. It is well known that a solution to the Helmholtz equation can be written as a linear combination of plane waves, thus the electric field in each region can be represented as

$$\mathbf{E}_j(\mathbf{x}) = \sum_{a=\pm} \int_{\mathbb{R}^2} \left[\mathcal{E}^a_{j,p}(\mathbf{q})\,\hat{\mathbf{e}}^a_{p,j}(\mathbf{q}) + \mathcal{E}^a_{j,s}(\mathbf{q})\,\hat{\mathbf{e}}_s(\mathbf{q})\right] \times \exp\left(i\;\mathbf{k}^a_j(\mathbf{q})\cdot\;\mathbf{x}\right)\frac{\mathrm{d}^2 q}{(2\pi)^2}\,, \tag{5}$$

where we have defined

$$\mathbf{k}^\pm_j(\mathbf{q}) = \mathbf{q} \pm \alpha_j(\mathbf{q})\,\hat{\mathbf{e}}_3\,, \tag{6a}$$

$$\alpha_j(\mathbf{q}) = \sqrt{k_j^2 - \mathbf{q}^2}, \quad \mathrm{Re}\,(\alpha_j), \mathrm{Im}\,(\alpha_j) \geq 0\,, \tag{6b}$$

$$\hat{\mathbf{e}}_s(\mathbf{q}) = \hat{\mathbf{e}}_3 \times \hat{\mathbf{q}}\,, \tag{6c}$$

$$\hat{\mathbf{e}}^\pm_{p,j}(\mathbf{q}) = k_j^{-1}\,(\pm\alpha_j(\mathbf{q})\,\hat{\mathbf{q}} - |\mathbf{q}|\,\hat{\mathbf{e}}_3) \tag{6d}$$

$$k_j = n_j\frac{\omega}{c} = |\,\mathbf{k}^\pm_j(\mathbf{q})|\,. \tag{6e}$$

In other words, the wave vector $\mathbf{k}^\pm_j(\mathbf{q})$ of an elementary plane wave is decomposed into its projection $\mathbf{q}$ in the lateral vector plane $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2)$ and the component $\pm\alpha_j(\mathbf{q})$ along $\hat{\mathbf{e}}_3$. The sum over $a = \pm$ takes into account both upwards $(+)$ and downwards $(-)$ propagating and evanescent (and possibly growing) waves. The field amplitude is decomposed in the *local polarization basis* $(\hat{\mathbf{e}}^a_{p,j}(\mathbf{q}), \hat{\mathbf{e}}_s(\mathbf{q}))$, hence $\mathcal{E}^a_{j,\alpha}(\mathbf{q})$ denotes the component of the field amplitude in the polarization state α of the mode characterized by a and $\mathbf{q}$. In this basis, the directions given by $\hat{\mathbf{e}}^\pm_{p,j}(\mathbf{q})$, and $\hat{\mathbf{e}}_s(\mathbf{q})$ are the directions of the p- and s-polarization of the electric field amplitude, respectively. Furthermore, the electromagnetic fields have to

satisfy the boundary conditions

$$\mathbf{n}(\mathbf{x}_\parallel) \times \Big[\mathbf{E}_2(\mathbf{s}(\mathbf{x}_\parallel)) - \mathbf{E}_1(\mathbf{s}(\mathbf{x}_\parallel))\Big] = \mathbf{0} \tag{7a}$$

$$\mathbf{n}(\mathbf{x}_\parallel) \times \Big[\mathbf{H}_2(\mathbf{s}(\mathbf{x}_\parallel)) - \mathbf{H}_1(\mathbf{s}(\mathbf{x}_\parallel))\Big] = \mathbf{0}\,, \tag{7b}$$

where $\mathbf{n}(\mathbf{x}_\parallel)$ is a vector normal to the interface at the surface point $\mathbf{s}(\mathbf{x}_\parallel) = \mathbf{x}_\parallel + \zeta(\mathbf{x}_\parallel)\hat{\mathbf{e}}_3$, given by

$$\mathbf{n}(\mathbf{x}_\parallel) = \hat{\mathbf{e}}_3 - \frac{\partial \zeta}{\partial x_1}(\mathbf{x}_\parallel)\,\hat{\mathbf{e}}_1 - \frac{\partial \zeta}{\partial x_2}(\mathbf{x}_\parallel)\,\hat{\mathbf{e}}_2\,. \tag{8}$$

Here, $\partial\cdot/\partial x_k$ denotes the partial derivative along the direction $\hat{\mathbf{e}}_k$. Following Soubret *et al.* [31], by substituting the field expansion Eq. (5) into Eq. (7) and by a clever linear integral combination of the boundary conditions inspired by the extinction theorem [28], one can show that the upward or downward field amplitudes in medium 2 can be linked to the upward and downward field amplitudes in medium 1 via the following integral equation defined for $a_2 = \pm$, and $\mathbf{p}$ in the vector plane $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2)$:

$$\sum_{a_1=\pm} \int \mathcal{J}_{2,1}^{a_2,a_1}(\mathbf{p}\,|\,\mathbf{q})\; \mathbf{M}_{2,1}^{a_2,a_1}(\mathbf{p}\,|\,\mathbf{q})\; \boldsymbol{\mathcal{E}}_1^{a_1}(\mathbf{q})\, \frac{\mathrm{d}^2 q}{(2\pi)^2} = \frac{2\,a_2\,n_1 n_2\,\alpha_2(\mathbf{p})}{\epsilon_2 - \epsilon_1}\; \boldsymbol{\mathcal{E}}_2^{a_2}(\mathbf{p}). \tag{9}$$

Here $\boldsymbol{\mathcal{E}}_j^a(\mathbf{q}) = (\mathcal{E}_{j,p}^a(\mathbf{q}), \mathcal{E}_{j,s}^a(\mathbf{q}))^{\mathrm{T}}$ denotes a column vector of the polarization components of the field amplitude in medium j. Moreover, $\mathbf{M}_{l,m}^{b,a}(\mathbf{p}\,|\,\mathbf{q})$ is the 2×2 matrix

$$\mathbf{M}_{l,m}^{b,a}(\mathbf{p}\,|\,\mathbf{q}) = k_1 k_2 \begin{pmatrix} \hat{\mathbf{e}}_{p,l}^b(\mathbf{p})\cdot\hat{\mathbf{e}}_{p,m}^a(\mathbf{q}) & \hat{\mathbf{e}}_{p,l}^b(\mathbf{p})\cdot\hat{\mathbf{e}}_s(\mathbf{q}) \\ \hat{\mathbf{e}}_s(\mathbf{p})\cdot\hat{\mathbf{e}}_{p,m}^a(\mathbf{q}) & \hat{\mathbf{e}}_s(\mathbf{p})\cdot\hat{\mathbf{e}}_s(\mathbf{q}) \end{pmatrix}, \tag{10}$$

which originates from a change of coordinate system between the local polarization basis $(\hat{\mathbf{e}}_{p,l}^b(\mathbf{p}), \hat{\mathbf{e}}_s(\mathbf{p}))$ and $(\hat{\mathbf{e}}_{p,m}^a(\mathbf{q}), \hat{\mathbf{e}}_s(\mathbf{q}))$, defined for $a = \pm$, $b = \pm$, and $l, m \in \{1, 2\}$ with $l \neq m$. The kernel scalar factor $\mathcal{J}_{l,m}^{b,a}(\mathbf{p}\,|\,\mathbf{q})$ encodes the surface geometry and is defined as

$$\mathcal{J}_{l,m}^{b,a}(\mathbf{p}\,|\,\mathbf{q}) = [b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q})]^{-1} \times \int \exp\left[-i(\mathbf{k}_l^b(\mathbf{p}) - \mathbf{k}_m^a(\mathbf{q}))\cdot\mathbf{s}(\mathbf{x}_\parallel)\right] \mathrm{d}^2 x_\parallel. \tag{11}$$

Notice that, as already pointed out in Ref. 31, due to the symmetry of the boundary conditions, one may also show in the same way that

$$\sum_{a_2=\pm} \int \mathcal{J}_{1,2}^{a_1,a_2}(\mathbf{p}\,|\,\mathbf{q})\; \mathbf{M}_{1,2}^{a_1,a_2}(\mathbf{p}\,|\,\mathbf{q})\; \boldsymbol{\mathcal{E}}_2^{a_2}(\mathbf{q})\, \frac{\mathrm{d}^2 q}{(2\pi)^2} = \frac{2\,a_1\,n_1 n_2\,\alpha_1(\mathbf{p})}{\epsilon_1 - \epsilon_2}\; \boldsymbol{\mathcal{E}}_1^{a_1}(\mathbf{p}), \tag{12}$$

which can be obtained from Eq. (9) by interchanging the subscripts 1 and 2. Typically, Eq. (9) is appropriate to solve the problem of reflection whereas Eq. (12) is appropriate to solve the problem of transmission, as we will see later.

So far, we have stayed general and simply assumed the presence of an electromagnetic field decomposed in propagating and non-propagating waves in each region. Therefore, there is no uniqueness in the solutions to the transfer equations, Eqs. (9) and (12). To ensure a unique solution, one needs to impose some constraints on the field. First, we need to introduce an incident field to our model. This will split the field expansion into a sum of an incident field, which is given by our model of the problem, and a scattered field. Note that within this framework, the incident field may be chosen to be in either medium, or to be a combination of excitations incident from different media.

In our case, the incident field will be taken as a plane wave incident from medium 1 and defined as

$$\mathbf{E}_0(\mathbf{x}) = \big[\mathcal{E}_{0,p}\;\hat{\mathbf{e}}_{p,1}^-(\mathbf{p}_0) + \mathcal{E}_{0,s}\;\hat{\mathbf{e}}_s(\mathbf{p}_0)\big] \times \exp\left(i\,\mathbf{k}_1^-(\mathbf{p}_0)\cdot\mathbf{x}\right), \tag{13}$$

where $\mathbf{p}_0$ is the projection of the wave vector of the incident wave onto the $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2)$ plane, with the property $|\,\mathbf{p}_0\,| \leq k_1$, i.e. we consider an incident propagating wave. The fact that this is the only incident wave considered, together with the Sommerfeld radiation condition at infinity, gives that the only elementary waves allowed in the scattered field are those with wave vectors of the form $\mathbf{k}_1^+(\mathbf{p})$ and $\mathbf{k}_2^-(\mathbf{p})$ in medium 1 and 2, respectively. This property can be expressed by defining the field amplitudes

$$\boldsymbol{\mathcal{E}}_1^-(\mathbf{q}) = (2\pi)^2\,\delta(\mathbf{q} - \mathbf{p}_0)\;\boldsymbol{\mathcal{E}}_0\,, \tag{14a}$$

$$\boldsymbol{\mathcal{E}}_2^+(\mathbf{q}) = \mathbf{0}\,, \tag{14b}$$

where $\boldsymbol{\mathcal{E}}_0 = (\mathcal{E}_{0,p}, \mathcal{E}_{0,s})^{\mathrm{T}}$. Next, we assume that the scattered field amplitudes are linearly related to the incident field amplitude $\boldsymbol{\mathcal{E}}_0$ via the reflection and transmission amplitudes, $\mathbf{R}(\mathbf{q}\,|\,\mathbf{p}_0)$ and $\mathbf{T}(\mathbf{q}\,|\,\mathbf{p}_0)$, defined as

$$\boldsymbol{\mathcal{E}}_1^+(\mathbf{q}) = \mathbf{R}(\mathbf{q}\,|\,\mathbf{p}_0)\,\boldsymbol{\mathcal{E}}_0, \tag{15a}$$

$$\boldsymbol{\mathcal{E}}_2^-(\mathbf{q}) = \mathbf{T}(\mathbf{q}\,|\,\mathbf{p}_0)\,\boldsymbol{\mathcal{E}}_0\,. \tag{15b}$$

The reflection and transmission amplitudes are therefore described by 2×2 matrices of the form,

$$\mathbf{X} = \begin{pmatrix} X_{pp} & X_{ps} \\ X_{sp} & X_{ss} \end{pmatrix}, \tag{16}$$

with $\mathbf{X} = \mathbf{R}$ or $\mathbf{T}$. From a physical point of view, the complex amplitude $R_{\alpha\beta}(\mathbf{q}\,|\,\mathbf{p}_0)$ (resp. $T_{\alpha\beta}(\mathbf{q}\,|\,\mathbf{p}_0)$) for $\alpha, \beta \in \{p, s\}$ is the field amplitude for the reflected light (resp. transmitted) with lateral wave vector $\mathbf{q}$ in the polarization state α from a unit incident field with lateral wave vector $\mathbf{p}_0$ in the polarization state β. The reflection and transmission amplitudes are then the unknowns in our scattering problem. The equations we need to

solve are deduced from the transfer equations, Eqs. (9) and (12), by applying them respectively at $a_2 = +$ and $a_1 = -$ and by using Eqs. (14) and (15) for the model of the field expansion. This yields the following two decoupled integral equations for the reflection and transmission amplitudes, the so-called reduced Rayleigh equations, that can be written in the following general form, for $\mathbf{X} = \mathbf{R}$ or $\mathbf{T}$:

$$\int \mathbf{M}_{\mathbf{X}}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{X}(\mathbf{q}\,|\,\mathbf{p}_0)\,\frac{\mathrm{d}^2q}{(2\pi)^2} = -\mathbf{N}_{\mathbf{X}}(\mathbf{p}\,|\,\mathbf{p}_0)\,, \quad (17)$$

where the matrices $\mathbf{M}_{\mathbf{X}}$ and $\mathbf{N}_{\mathbf{X}}$ are given by

$$\mathbf{M}_{\mathbf{R}}(\mathbf{p}\,|\,\mathbf{q}) = \mathcal{J}^{+,+}_{2,1}(\mathbf{p}\,|\,\mathbf{q})\;\mathbf{M}^{+,+}_{2,1}(\mathbf{p}\,|\,\mathbf{q}) \quad (18a)$$

$$\mathbf{M}_{\mathbf{T}}(\mathbf{p}\,|\,\mathbf{q}) = \mathcal{J}^{-,-}_{1,2}(\mathbf{p}\,|\,\mathbf{q})\;\mathbf{M}^{-,-}_{1,2}(\mathbf{p}\,|\,\mathbf{q}) \quad (18b)$$

$$\mathbf{N}_{\mathbf{R}}(\mathbf{p}\,|\,\mathbf{q}) = \mathcal{J}^{+,-}_{2,1}(\mathbf{p}\,|\,\mathbf{q})\;\mathbf{M}^{+,-}_{2,1}(\mathbf{p}\,|\,\mathbf{q}) \quad (18c)$$

$$\mathbf{N}_{\mathbf{T}}(\mathbf{p}\,|\,\mathbf{q}) = \frac{2n_1n_2\alpha_1(\mathbf{p})}{\epsilon_2-\epsilon_1}(2\pi)^2\delta(\mathbf{p}-\mathbf{q})\,\mathbf{I}_2, \quad (18d)$$

with $\mathbf{I}_2$ denoting the 2×2 identity matrix.

B. RRE for scalar waves

The reduced Rayleigh equations can also be derived for scalar waves satisfying the scalar Helmholtz equation and subjected to various boundary conditions at the interfaces. Here, we focus on scalar waves subjected to the continuity of the field and its normal derivative at the interface. Under these hypotheses, one can derive the corresponding reduced Rayleigh equations which read

$$\int M_X(\mathbf{p}\,|\,\mathbf{q})\,X(\mathbf{q}\,|\,\mathbf{p}_0)\frac{\mathrm{d}^2q}{(2\pi)^2} = -N_X(\mathbf{p}\,|\,\mathbf{p}_0)\,, \quad (19)$$

where $X = R$ or T is either the scalar reflection or transmission amplitude, and the scalar kernels and right-hand-sides are given by Eq. (18) where all the $\mathbf{M}^{b,a}_{l,m}$ matrices are replaced by the scalar constant $k_1k_2 = n_1n_2\,\omega^2/c^2$ and $\mathbf{I}_2$ is replaced by the scalar constant 1. We would like to stress that the fact that one can go from the electromagnetic RRE to the scalar RRE by simply replacing all the aforementioned matrices by 1 is only true for the case where the scalar field is subjected to the continuity of the field and its normal derivative at the surface. For other types of boundary conditions, as for the case of acoustic waves for example, one would obtain different expressions [25]. The obtained equations could in principle be used for modeling the scattering of a quantum particle by a surface between two regions of constant potential. In this paper, we will use the presented scalar RRE, for which the analysis is simplified compared to the case for electromagnetic waves, to explain the fundamental mechanism of the Yoneda effect. We will show that the Yoneda effect is present for scalar waves (subjected to the aforementioned boundary conditions) and can be decoupled from additional effects induced by the polarization of electromagnetic waves, such as the Brewster scattering effect. The identified mechanism for scalar waves will then be extended to electromagnetic waves.

C. Perturbative method

Probably the most common approximate solution to Eq. (17) is based on a perturbative expansion of the reflection and transmission amplitudes in powers of the interface profile function. This approach, often called "small amplitude perturbation theory" (SAPT) or "small perturbation method" (SPM), has shown that it is capable of obtaining solutions of the RRE of high qualitative and quantitative predictive power, for interfaces with sufficiently small slopes and amplitudes. To first order in ζ for the reflection and transmission amplitudes, the method is often interpreted as a *single scattering approximation*. When implemented to the complete fourth order in the surface profile function for the intensity, i.e. involving terms up to third order in the amplitude, the method has been used to obtain reliable results that also correctly include multiple scattering effects, most notably the backscattering peaks observed in reflection from metallic surfaces [34–37].

To first order in the interface profile function ζ, we have for $\mathbf{X} = \mathbf{R}$ or $\mathbf{T}$ that

$$\mathbf{X}\,(\mathbf{p}\,|\,\mathbf{p}_0) \approx \mathbf{X}^{(0)}\,(\mathbf{p}\,|\,\mathbf{p}_0) - i\,\mathbf{X}^{(1)}\,(\mathbf{p}\,|\,\mathbf{p}_0)\,, \quad (20)$$

where

$$\mathbf{R}^{(0)}\,(\mathbf{p}\,|\,\mathbf{p}_0) = (2\pi)^2\delta(\mathbf{p}-\mathbf{p}_0)\,\boldsymbol{\rho}^{(0)}(\mathbf{p}_0)\,, \quad (21a)$$

$$\mathbf{T}^{(0)}\,(\mathbf{p}\,|\,\mathbf{p}_0) = (2\pi)^2\delta(\mathbf{p}-\mathbf{p}_0)\,\boldsymbol{\tau}^{(0)}(\mathbf{p}_0)\,, \quad (21b)$$

$$\mathbf{R}^{(1)}\,(\mathbf{p}\,|\,\mathbf{p}_0) = \hat{\zeta}(\mathbf{p}-\mathbf{p}_0)\,\boldsymbol{\rho}^{(1)}\,(\mathbf{p}\,|\,\mathbf{p}_0)$$
$$= [\alpha_1(\mathbf{p})-\alpha_2(\mathbf{p})]\,\hat{\zeta}(\mathbf{p}-\mathbf{p}_0)\,\hat{\boldsymbol{\rho}}^{(1)}\,(\mathbf{p}\,|\,\mathbf{p}_0)\,, \quad (21c)$$

$$\mathbf{T}^{(1)}\,(\mathbf{p}\,|\,\mathbf{p}_0) = \hat{\zeta}(\mathbf{p}-\mathbf{p}_0)\,\boldsymbol{\tau}^{(1)}\,(\mathbf{p}\,|\,\mathbf{p}_0)$$
$$= [\alpha_1(\mathbf{p})-\alpha_2(\mathbf{p})]\,\hat{\zeta}(\mathbf{p}-\mathbf{p}_0)\,\hat{\boldsymbol{\tau}}^{(1)}\,(\mathbf{p}\,|\,\mathbf{p}_0)\;. \quad (21d)$$

Here $\hat{\zeta}$ denotes the Fourier transform of ζ, and $\boldsymbol{\rho}^{(0)}(\mathbf{p}_0)$ and $\boldsymbol{\tau}^{(0)}(\mathbf{p}_0)$ are matrix-valued amplitudes for the zero order reflection and transmission amplitudes, respectively. The matrix-valued amplitudes $\boldsymbol{\rho}^{(1)}\,(\mathbf{p}\,|\,\mathbf{p}_0)$, $\hat{\boldsymbol{\rho}}^{(1)}\,(\mathbf{p}\,|\,\mathbf{p}_0)$, $\boldsymbol{\tau}^{(1)}\,(\mathbf{p}\,|\,\mathbf{p}_0)$, and $\hat{\boldsymbol{\tau}}^{(1)}\,(\mathbf{p}\,|\,\mathbf{p}_0)$ for the first order terms are derived in A. In Eqs. (21c) and (21d), we have given two alternative factorizations of the first order reflection and transmission amplitudes. The factorization including the caret amplitudes is the most appropriate for physical interpretation, while the factorization including the non-caret amplitudes simply aims at separating $\hat{\zeta}$, which is the only factor depending on the specific realization of the surface profile, from the remaining profile-independent amplitude factor.

D. Observables

The observables of interest in this study are the so-called *diffuse* or *incoherent component of the mean differential reflection and transmission coefficients* (MDRC and MDTC) denoted $\langle \partial R_{\alpha\beta}(\mathbf{p}|\mathbf{p}_0)/\partial\Omega_r \rangle_{\text{incoh}}$ and $\langle \partial T_{\alpha\beta}(\mathbf{p}|\mathbf{p}_0)/\partial\Omega_t \rangle_{\text{incoh}}$, respectively. They are both defined as the ensemble average over realizations of the surface profile of the incoherent component of the radiated reflected/transmitted flux of an α-polarized wave around a direction given by $\mathbf{k}_1^+(\mathbf{p})/\mathbf{k}_2^-(\mathbf{p})$ per unit incident flux of a β-polarized plane wave with wave vector $\mathbf{k}_1^-(\mathbf{p}_0)$, per unit solid angle. Based on the reflection and transmission amplitudes found to first order in ζ, the incoherent component of the MDRC and MDTC can be expressed as

$$\left\langle \frac{\partial R_{\alpha\beta}}{\partial\Omega_r}(\mathbf{p}\,|\,\mathbf{p}_0) \right\rangle_{\text{incoh}} = \epsilon_1 \left(\frac{\omega}{2\pi c}\right)^2 \frac{\cos^2\theta_r}{\cos\theta_0} \times \sigma^2\, g(\mathbf{p}-\mathbf{p}_0) \left|\rho_{\alpha\beta}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_0)\right|^2, \tag{22}$$

and

$$\left\langle \frac{\partial T_{\alpha\beta}}{\partial\Omega_t}(\mathbf{p}\,|\,\mathbf{p}_0) \right\rangle_{\text{incoh}} = \frac{\epsilon_2^{3/2}}{\epsilon_1^{1/2}} \left(\frac{\omega}{2\pi c}\right)^2 \frac{\cos^2\theta_t}{\cos\theta_0} \times \sigma^2\, g(\mathbf{p}-\mathbf{p}_0) \left|\tau_{\alpha\beta}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_0)\right|^2. \tag{23}$$

The detailed derivation of the Eqs. (22) and (23) can be found in B. The definition of the angles of incidence and scattering can be deduced from Figure 1.

IV. RESULTS AND DISCUSSION

In order to study the phenomena observed in the scattering of light from weakly rough dielectric interfaces, we choose to base our discussion on results obtained through small amplitude perturbation theory (SAPT) to lowest non-zero order in the interface profile function, Eqs. (21). For sufficiently smooth interfaces this approximation has previously been compared to numerical non-perturbative solutions to the reduced Rayleigh equations, where it has been shown to adequately model the phenomena of both the Brewster scattering angles and the Yoneda peaks [1, 2]. We will start our investigations with a summary of the features observed in the main physical observables, the MDRC and MDTC [Eqs. (22) and (23)], followed by more in-depth analyses and discussions from a physics point of view.

A. Phenomenology of the Yoneda and Brewster scattering effects

The top panel of each subfigure in Fig. 2 presents results based on Eqs. (22) and (23) for the contribution to the co-polarized diffuse component of the MDRC and MDTC in the plane of incidence ($\hat{\mathbf{p}} \parallel \hat{\mathbf{p}}_0$), for a configuration where light is incident from vacuum [$\epsilon_1 = 1$] onto a two-dimensional randomly rough interface with glass [$\epsilon_2 = 2.25$]. The incident light was assumed to be a p- or s-polarized plane wave of wavelength $\lambda = 632.8\,\text{nm}$ in vacuum. In the current work all results presented for randomly rough interfaces consist of interfaces defined by an isotropic Gaussian height distribution with rms height $\sigma = 32\,\text{nm} = \lambda/20$ and an isotropic Gaussian correlation function of transverse correlation length $a = 211\,\text{nm} = \lambda/3$.

For normal incidence [$\theta_0 = 0°$, Fig. 2(a)] the MDRC distributions are nearly featureless. The differences in the scattered intensities observed for p- and s-polarized incident light are very small. Note that the scattered intensity is zero beyond the limit of propagation in the medium of reflection ($|p_1| > k_1$). The overall bell-shape of the distributions can be attributed in part to the Gaussian correlation function for the transverse correlation length in the interface profile together with the $\cos^2\theta_r$ factor of the MDRC, as seen in Eq. (22). The corresponding transmitted intensity (MDTC) shown in Fig. 2(d), however, shows several interesting features. As is detailed in Ref. 2, we now observe pronounced peaks in s-polarization and narrow dips to zero in p-polarization around $|p_1| = k_1$. For normal incidence these features are independent of the azimuthal angle of transmission ϕ. The peaks have become known as "Yoneda peaks", and are always found at the parallel wavevectors $\mathbf{p}$ along the propagation limit in the less dense medium (i.e. $|\mathbf{p}| = \min(k_1, k_2)$). The polar angles corresponding to the dips to zero in the MDTC have been called the "Brewster scattering angles"[13], and are unique to scattered light which is p-polarized. As the polar angle of incidence is increased [$\theta_0 = 35°$ or $70°$ in Fig. 2], we observe that a Brewster scattering angle also appears in the MDRC. In transmission, the distributions of the MDTC behave very predictably in the s-polarized case as the weight of the distribution is shifted to higher polar scattering angles. However, in the case of p-polarization the Brewster scattering angle in the direction of $\phi = 180°$ (negative values of θ_t in Fig. 2(d)) shifts to positions closer to $\theta_t = 0°$ as the angle of incidence is increased, and the dip to zero in the forward scattering direction [$\phi = 0°$] first becomes a non-zero local minimum and is gradually replaced with a Yoneda peak similar to the one found for s-polarization.

Figure 3 presents results similar to those in Fig. 2, but for the situation where the media are interchanged; the light is now incident from glass [$\epsilon = 2.25$] onto a two-dimensional randomly rough interface with vacuum [$\epsilon = 1$]. A closer inspection of the distributions of the MDRC for normal incidence reveals that the distributions are reminiscent of the distributions seen in transmission for the MDTC in Fig. 2, and vice versa. This similarity between intensity distributions for which the media of propagation is the same is an expected symmetry, but as the angle of incidence increases these sim-

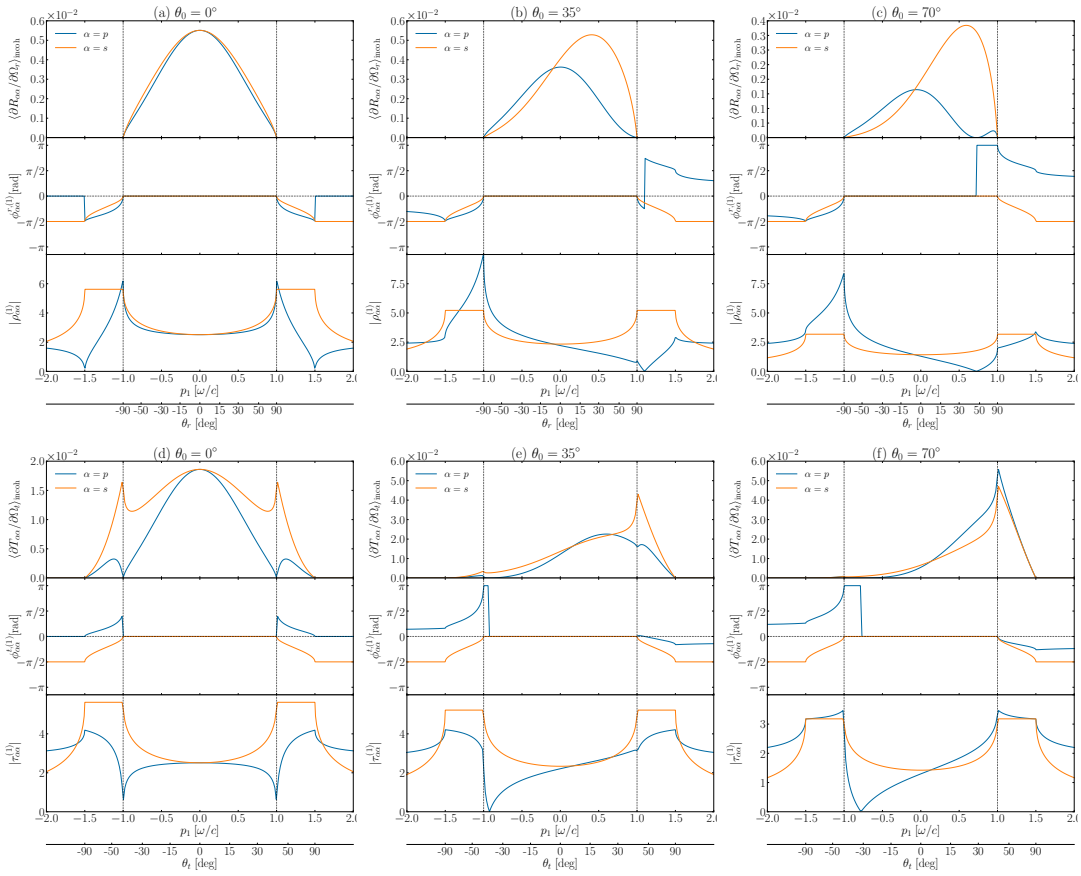


FIG. 2. The incoherent component of the MDRC (top row) and MDTC (bottom row) for light incident from vacuum [$\epsilon_1 = 1.0$] onto a randomly rough interface with glass [$\epsilon_2 = 2.25$], for in-plane co-polarized scattering, as a function of the lateral component of the wave vector of scattering p_1 or polar angle of scattering $\theta_{r,t}$. The polar angle of incidence is indicated on top of each subfigure. The argument and the modulus of $\rho_{\alpha\alpha}^{(1)}$ and $\tau_{\alpha\alpha}^{(1)}$ are indicated in the middle and bottom section of each subfigure respectively. Note that we have adopted here the convention that negative $\theta_{r,t}$ values correspond to $\theta_{r,t} > 0$ according Fig. 1 but for $\phi = 180°$. The vertical lines indicate $|\,\mathbf{p}\,| = k_{\rm min}$.

ilarities gradually fade. For light impinging on the interface at $\theta_0 = 35°$, the Brewster scattering angle for the MDRC is now in the forward scattering direction, and as documented in Ref. 1 it shifts closer to $\theta_r = 0°$ as the angle of incidence increases towards the critical angle given by $\theta_c = \sin^{-1}(n_2/n_1)$. Results for an angle of incidence equal to the critical angle, $\theta_0 = \theta_c$, are presented in Figs. 3(c) and 3(f). For the same system but for polar angles of incidence larger than the critical angle, presented in Fig. 4, the dip to zero MDRC in the forward scattering direction is gradually overtaken by a Yoneda peak for p-polarized light. Contrary to the case for transmission in Fig. 2 however, the peak in p-polarization never grows beyond the peak in s-polarization. For the intensity distributions of the transmitted light we again observe a gradual shift of the weight of the distributions into the forward scattering direction, but the Brewster scattering angle is now only visible (strictly speaking) for $\theta_0 = 35°$ and $\theta_0 = 41.81°$, where it is now found in the backward scattering direction and at $\theta_t = 0°$, respectively.

As an aid in understanding the Brewster scattering angles in Figs. 2, 3 and 4, and also as a support to the further discussion of these angles in both reflection and transmission, Fig. 5 presents an overview of the Brewster scattering angles found in the MDRC/MDTC as derived from first order SAPT. Figure 5 is based on the following

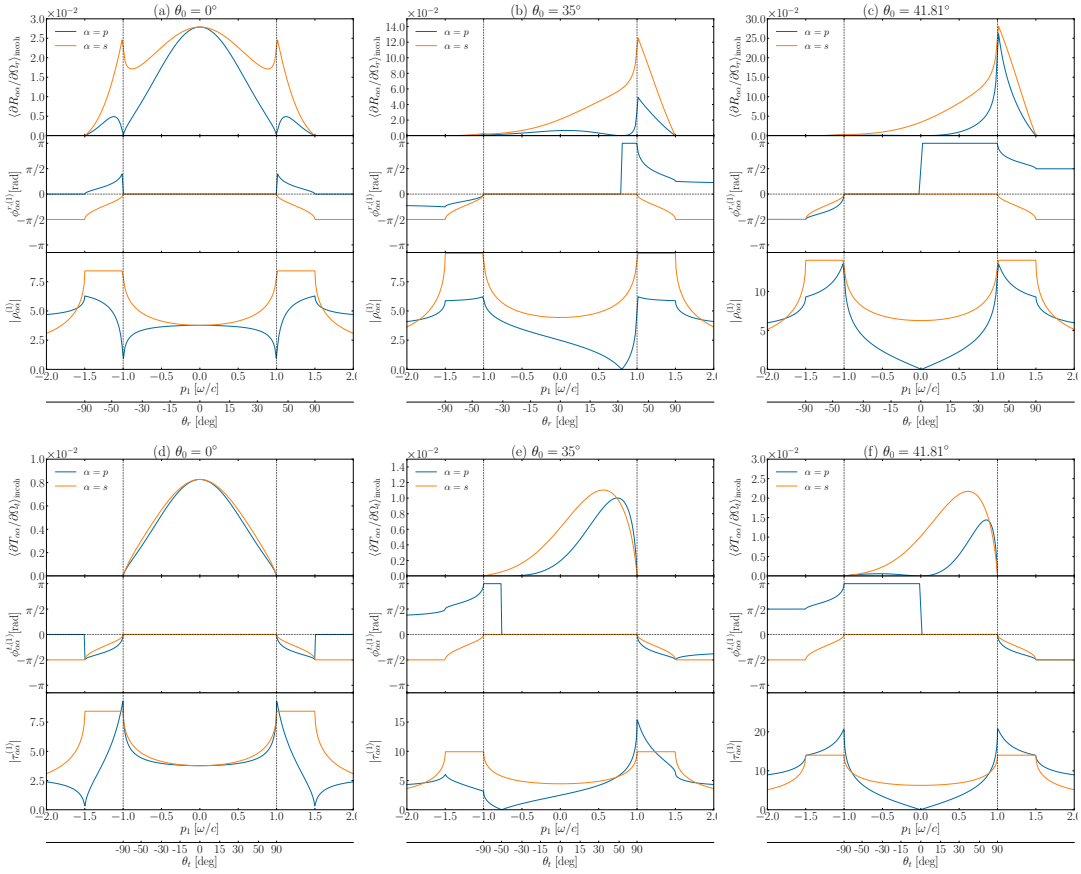


FIG. 3. Same as Fig. 2, but for light incident from glass [$\epsilon_1 = 2.25$] onto a randomly rough interface with vacuum [$\epsilon_2 = 1.0$].

result, obtained from SAPT [1, 2]:

$$\Theta_B(\theta_0) = \sin^{-1}\left(\sqrt{\frac{\epsilon_2(\epsilon_2 - \epsilon_1 \sin^2\theta_0)}{(\epsilon_2^2 - \epsilon_1^2)\sin^2\theta_0 + \epsilon_1\epsilon_2}}\right). \quad (24)$$

for reflection and

$$\Theta_B(\theta_0) = \sin^{-1}\left(\frac{\epsilon_1}{\epsilon_2}\sqrt{\frac{\epsilon_2}{\epsilon_1} - \sin^2\theta_0}\right), \quad (25)$$

for transmission. Here Θ_B indicates the Brewster scattering angle. As can be observed in both Fig. 5 and from Eqs. (24) and (25), Θ_B is not well-defined for all angles of incidence for all ϵ_1 and ϵ_2. These limits will be further explored in Sec. IV D where a clear geometrical interpretation will be given.

The scattering in both reflection and transmission from such a randomly rough interface has been thoroughly studied in the past, and the distributions of the MDRC and MDTC presented in Figs. 2, 3 and 4 were partially explained based on the components of the perturbative approximation in Refs. 1 and 2. However, these publications stopped short of presenting a full physical interpretation of the features seen in these distributions. In the current work we aim to finalize this analysis, and to that end we expand the investigation to include results for the complex amplitudes on which the MDRC and MDTC are based. The center panel of each subfigure in Figs. 2, 3 and 4 presents the average phase, $\phi_{\alpha\alpha}^{r,(1)}$, $\phi_{\alpha\alpha}^{t,(1)}$ of the co-polarized scattered light, obtained from the argument of the complex amplitudes $\rho_{\alpha\alpha}^{(1)}$ or $\tau_{\alpha\alpha}^{(1)}$ for $\alpha \in \{p, s\}$ given in Eqs. (21 d) and (21 e), respectively. The lower panel of each subfigure shows the modulus of $\rho_{\alpha\alpha}^{(1)}$ and $\tau_{\alpha\alpha}^{(1)}$.

In passing we emphasize that even if the results presented are based on a perturbation method to lowest nonzero order in the interface profile function, previous studies have demonstrated their validity for the parameters and dielectric constants assumed in obtaining them. In

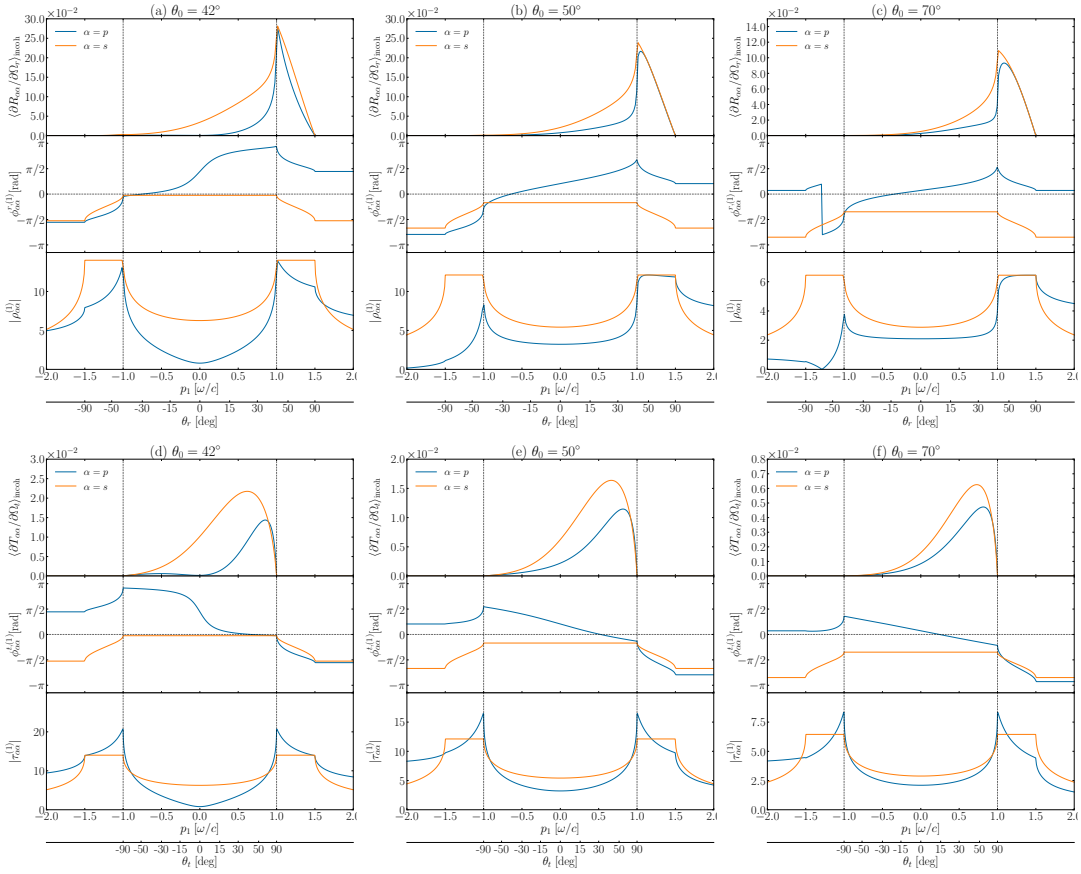


FIG. 4. Same as Fig. 3, but for additional polar angles of incidence θ_0.

addition, the results for both MDRC and MDTC have been compared against numerical results based on the extinction theorem based method described in Ref. 37 for a 1D system, a method known to be rigorous.

[19]

$$r_s^{\mathrm{F}}(\mathbf{p}_0) = \frac{\alpha_1(\mathbf{p}_0) - \alpha_2(\mathbf{p}_0)}{\alpha_1(\mathbf{p}_0) + \alpha_2(\mathbf{p}_0)} \tag{26a}$$

$$r_p^{\mathrm{F}}(\mathbf{p}_0) = \frac{\epsilon_2\alpha_1(\mathbf{p}_0) - \epsilon_1\alpha_2(\mathbf{p}_0)}{\epsilon_2\alpha_1(\mathbf{p}_0) + \epsilon_1\alpha_2(\mathbf{p}_0)}\,, \tag{26b}$$

B. Physical interpretation of SAPT to first order

Order zero, Fresnel amplitudes — First we revisit the interpretation of the Fresnel coefficients which are encoded in the amplitudes $\boldsymbol{\rho}^{(0)}(\mathbf{p}_0)$ and $\boldsymbol{\tau}^{(0)}(\mathbf{p}_0)$ [Eqs. (21a) and (21b)].

We start our analysis looking at the case of reflection. The Fresnel amplitudes for s- and p-polarized waves reflected by a planar surface between two dielectrics read which we have here written in a common form in terms of the components of the wave vectors along $\hat{\mathbf{e}}_3$. It is easy to show by using straightforward algebra that these expressions are equivalent to $\rho_{ss}^{(0)}(\mathbf{p}_0)$ and $\rho_{pp}^{(0)}(\mathbf{p}_0)$ respectively, given by perturbation theory to zero order. An equivalent way of writing the Fresnel amplitudes which follows directly from Eq. (A6) and the definition of the $\mathbf{M}_{l,m}^{b,a}(\mathbf{p}\,|\,\mathbf{q})$ matrix in terms of the polarization vectors,

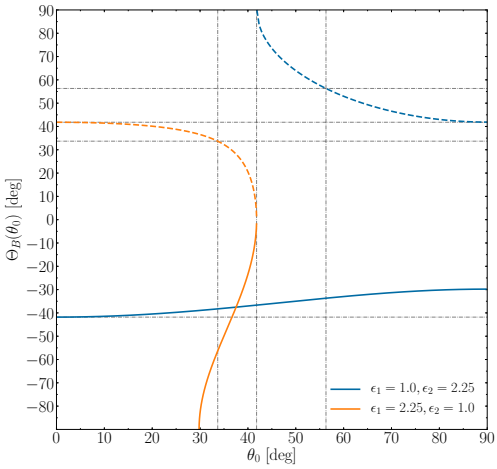


FIG. 5. Dependence of the in-plane Brewster scattering angle Θ_B on the polar angle of incidence θ_0, as given by Eq. (25). $\phi = 180°$ is indicated by negative values of Θ_B and θ_0. Solid lines indicate Θ_B in transmission, while the dashed lines with color corresponding to the values of ϵ indicate Θ_B in reflection. The critical angles and the regular Brewster angles for a corresponding system with a planar interface have been indicated on both axes as black dash-dotted lines.

Eq. (10), is

$$\rho^{(0)}_{ss}(\mathbf{p}_0) = \rho^{(0)}(\mathbf{p}_0)\,\frac{\hat{\mathbf{e}}_s(\mathbf{p}_0)\cdot\hat{\mathbf{e}}_s(\mathbf{p}_0)}{\hat{\mathbf{e}}_s(\mathbf{p}_0)\cdot\hat{\mathbf{e}}_s(\mathbf{p}_0)} \tag{27a}$$

$$\rho^{(0)}_{pp}(\mathbf{p}_0) = \rho^{(0)}(\mathbf{p}_0)\,\frac{\hat{\mathbf{e}}^+_{p,2}(\mathbf{p}_0)\cdot\hat{\mathbf{e}}^-_{p,1}(\mathbf{p}_0)}{\hat{\mathbf{e}}^+_{p,2}(\mathbf{p}_0)\cdot\hat{\mathbf{e}}^+_{p,1}(\mathbf{p}_0)} \tag{27b}$$

$$\rho^{(0)}(\mathbf{p}_0) = \frac{\alpha_1(\mathbf{p}_0)-\alpha_2(\mathbf{p}_0)}{\alpha_1(\mathbf{p}_0)+\alpha_2(\mathbf{p}_0)}\,. \tag{27c}$$

In Eq. (27a), we have intentionally chosen not to simplify the dot products (all equal to 1) to illustrate that the Fresnel amplitudes expressed in the form given by Eqs. (27a) and (27b) exhibit a remarkable factorization which reveals two facets of the physics of scattering from a microscopic point of view. First, both Fresnel amplitudes in Eqs. (27a) and (27b) share the same first factor, $\rho^{(0)}(\mathbf{p}_0)$ defined in Eq. (27c), which corresponds to the reflection amplitude for a scalar plane wave subjected to the continuity of the scalar field and its normal derivative at the surface. In other words, this first factor can be interpreted as the coherent response of arrays of individual scatterers (at the atomic level) which scatter the incident wave as *spherical waves*. The second factor, which differs for each polarization, is the signature of the dipolar nature of the radiation of each individual scatterer. Indeed, for an s-polarized incident wave, the scattering dipoles are excited along the $\hat{\mathbf{e}}_2$-axis and hence re-emit isotropically in the plane of incidence $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_3)$. We argue that this is the reason why the second factor is identically equal to 1 for s-polarized light. For p-polarization, the scattering dipoles are excited along some direction in the plane of incidence $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_3)$ and therefore the reflection amplitude given by the scattering of spherical waves must be weighted with the second factor in Eq. (27b) in order to take into account the dipole radiation pattern. Such a factorization and interpretation of the Fresnel amplitudes were given and thoroughly discussed by Doyle [26] in light of the Ewald-Oseen extinction theorem and its original derivation by Ewald based on microscopic optics [29].

For a planar surface all scattered waves interfere destructively in all directions but the specular, as indicated by the Dirac distribution in Eqs. (21a) and (21b). This is not the case when the surface is non-planar, and the above interpretation suggests that the spherical-like waves scattered away from the specular direction are then to be weighted by the appropriate dipole factor, even for s-polarized light (as will be the case for the first order term).

From Eq. (27), we can deduce two properties well known for the reflection of a plane wave at a planar interface between two dielectric media. First, for $\epsilon_1 > \epsilon_2$ there exists a critical polar angle $\theta_c = \sin^{-1}(n_2/n_1)$, or equivalently a critical norm of the lateral wave vector $p_c = n_2\,\omega/c$, such that for all angles of incidence larger than θ_c (equivalently for all lateral wave vectors where $|\,\mathbf{p}_0\,| > p_c$), all the incident power is reflected. The phenomenon of total internal reflection is entirely controlled by the factor $\rho^{(0)}(\mathbf{p}_0)$ present for both polarizations, and hence can be analyzed from a scalar wave picture decoupled from polarization effects. From a physical point of view, total internal reflection occurs whenever the refracted wave is evanescent in the medium of transmission, and therefore it cannot transport energy away from the surface. It is instructive to analyze the behavior of the reflection amplitude $\rho^{(0)}(\mathbf{p}_0)$ as the refracted wave turns evanescent in the second medium as one varies the incident lateral wave vector $\mathbf{p}_0$. For $|\,\mathbf{p}_0\,| < p_c$, both $\alpha_1(\mathbf{p}_0)$ and $\alpha_2(\mathbf{p}_0)$ are real. As $|\,\mathbf{p}_0\,| \to p_c$, $\rho^{(0)}(\mathbf{p}_0)$ moves on the real line towards 1 when α_2 vanishes, $\alpha_1(\mathbf{p}_0) = \sqrt{\epsilon_1-\epsilon_2}\;\omega/c$. When $|\,\mathbf{p}_0\,| > p_c$, $\alpha_2(\mathbf{p}_0)$ becomes pure imaginary and $\rho^{(0)}(\mathbf{p}_0)$ starts to trace a circular arc in the lower half of the complex plane (negative imaginary part) with unit modulus (the fact that $|\rho^{(0)}(\mathbf{p}_0)| = 1$ for $|\,\mathbf{p}_0\,| > p_c$ is immediate since then $\rho^{(0)}(\mathbf{p}_0)$ is of the form z^*/z where z is a non-zero complex number). As $|\,\mathbf{p}_0\,| \to n_1\omega/c$, the reflected wave (and the incident wave) reaches the limit of propagation in the first medium and $\alpha_1(\mathbf{p}_0) \to 0$ which makes the reflection amplitude real, negative, equal to -1. Thus, as we go from the critical point to grazing incidence the reflection amplitude traces a half circle in the complex plane with unit modulus. The argument of the reflection amplitude, the phase, hence varies from 0 to $-\pi$ rad. This gradual phase shift is known as the Goos-Hänchen phase shift and can be interpreted as follows. If we regard the reflected

and refracted waves as two components of a single mode, then as the wave enters the second medium the wave oscillates while propagating along the surface before it eventually goes back into the first medium where it can continue to propagate to infinity. As the wave propagates along the surface, while being evanescent in medium 2, it acquires a temporal delay which depends on its wave vector. This delay is translated into a phase shift as the wave oscillates back into medium 1. Geometrically this process is often interpreted as if the wave is reflected from the second medium only after a slight penetration into it [19].

The second phenomenon of interest is that of the polarization angle, or Brewster's angle, which, as the name indicates, requires us to analyze the polarization dependent factor in the reflection amplitudes. For an s-polarized wave the polarization factor is identically equal to 1 and no polarization angle is observed. However, the Fresnel amplitude for p-polarized light, Eq. (27b), is shown to be proportional to $\hat{\mathbf{e}}^+_{p,2}(\mathbf{p}_0) \cdot \hat{\mathbf{e}}^-_{p,1}(\mathbf{p}_0)$ i.e. it is proportional to the component of the incident electric field (given by $E_0\, \hat{\mathbf{e}}^-_{p,1}(\mathbf{p}_0)$) along the direction given by $\hat{\mathbf{e}}^+_{p,2}(\mathbf{p}_0)$. We recall that the direction given by $\hat{\mathbf{e}}^+_{p,2}(\mathbf{p}_0)$ corresponds to the local p-polarization direction for a wave whose wave vector is given by $\mathbf{k}^+_2(\mathbf{p}_0)$, in other words, a wave which propagates upwards in the second medium. These factors of the dot products in Eq. (27b) therefore seem to indicate that the reflected amplitude for p-polarized light depends on a projection of the incident field along the polarization vector of a seemingly *nonexisting* wave, propagating along the wave vector $\mathbf{k}^+_2(\mathbf{p}_0)$. However, such a seemingly virtual wave does have a physical interpretation, based on the mutual interaction between waves propagating in dielectric media. Doyle [26] provided an explanation based on the concept of the *wave triad* originally introduced by Ewald [29]. Ewald considered a dense array of dipole scatterers (the entire dielectric medium) situated in a half space and excited by an incident plane wave incident from the vacuum half-space, filling the whole space between the scatterers. He showed that the dipole scatterers would respond to the excitation in such a way that there exist planes of scatterers of coherent response, meaning that all dipoles within such a plane oscillate in phase. As a consequence of this fact and that the array of scatterers is bounded within a half space, the superposition of all elementary wavelets emitted by each individual scatterer results in the propagation of three plane waves: two waves called *vacuum waves* propagating with a phase velocity equal to c and one wave propagating with phase velocity c/n called *polarization wave*, where n corresponds to the refractive medium made of scatterers within the macroscopic picture. The wave propagating with phase velocity c/n corresponds to the transmitted wave in the macroscopic picture, while one of the waves propagating with phase velocity c serves to exactly cancel the incident wave within the dielectric medium. The other wave propagating with phase velocity c exits the medium and corresponds to the reflected wave. The wave vectors of the different waves are naturally given by Snell's law, and Ewald's derivation can be viewed as a microscopic validation of Snell's law.

When two half-spaces are filled with dipole scatterers of different dipole moments, similar arguments apply with the difference that the superposition of all wavelets emitted by all scatterers (i.e. from both sides of the interface) must be taken into account. This results in three wave triads: one triad associated with the incident wave, one for the reflected wave and one for the refracted wave, which all satisfy the so-called dynamical conditions [26]. To the incident polarization wave in medium 1, propagating with phase velocity c/n_1 and wave vector $\mathbf{k}^-_1(\mathbf{p}_0)$ are associated two waves propagating with phase velocity c/n_2 with wave vectors $\mathbf{k}^\pm_2(\mathbf{p}_0)$. Similarly, to the reflected (resp. refracted) polarization wave, whose wave vector is given by $\mathbf{k}^+_1(\mathbf{p}_0)$ (resp. $\mathbf{k}^-_2(\mathbf{p}_0)$), are associated two waves propagating with phase velocity c/n_2 (resp. c/n_1) and wave vectors $\mathbf{k}^\pm_2(\mathbf{p}_0)$ (resp. $\mathbf{k}^\pm_1(\mathbf{p}_0)$). The dynamical conditions are state that the amplitudes of the different waves are such that (i) the wave associated with the refracted polarization wave and propagating along $\mathbf{k}^-_1(\mathbf{p}_0)$ in medium 2 cancels the incident wave and (ii) that the superposition of waves associated with the incident and reflected polarization waves and propagating along $\mathbf{k}^+_2(\mathbf{p}_0)$ vanishes (more details can be found in Refs. 26 and 29).

In the following, we will refer to the wave vectors $\mathbf{k}^a_1(\mathbf{p})$ and $\mathbf{k}^a_2(\mathbf{p})$, i.e. wave vectors sharing the same projection in the $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2)$-plane and pointing either both upward or downward, as *Snell-conjugate* wave vectors. When it comes to the polarization dependence of the reflection amplitudes, the fact that $\rho^{(0)}_{pp}(\mathbf{p}_0)$ is proportional to $\hat{\mathbf{e}}^+_{p,2}(\mathbf{p}_0) \cdot \hat{\mathbf{e}}^-_{p,1}(\mathbf{p}_0)$ indicates that the amplitude of the reflected wave is controlled by the component of the incident field along the p-polarization vector associated with the Snell-conjugate wave vector of the wave vector of the reflected wave. This indicates that the direction of the dipole oscillation is intimately linked to waves in the aforementioned triad.

Equation (27b) provides an interesting condition for the well known Brewster's angle. The Brewster phenomenon for dielectric media is commonly defined as the extinction of the p-polarized reflected wave in the case of a planar interface. From Eq. (27b), it is clear that the Fresnel amplitude vanishes if and only if [38] $\hat{\mathbf{e}}^+_{p,2}(\mathbf{p}_0) \cdot \hat{\mathbf{e}}^-_{p,1}(\mathbf{p}_0) = 0$. Since $\hat{\mathbf{e}}^+_{p,2}(\mathbf{p}_0)$ is orthogonal to $\mathbf{k}^+_2(\mathbf{p}_0)$, we can restate the condition for Brewster's angle as the polar angle where $\mathbf{k}^+_2(\mathbf{p}_0) \parallel \hat{\mathbf{e}}^-_{p,1}(\mathbf{p}_0)$. This means that we can define Brewster's angle as *the angle of reflection ensuring colinearity between the incident field amplitude and the wave vector which is Snell-conjugate to that of the reflected wave.* Note that we take here a slight change of point of view compared to the common phrasing. One usually refer to Brewster's angle as an angle of incidence, while we prefer to refer to the angle of reflection. Obviously, the two are the same for a planar interface, but the latter point of view is the one which

will hold true for non-planar interfaces. Nevertheless, it is convenient to use the term Brewster's incidence for a planar interface and we can define it as the angle of incidence $\theta_{\rm B} = \arctan(n_2/n_1)$ which yields a Brewster (non-)reflected wave. We will see that this angle of incidence, $\theta_0 = \theta_{\rm B}$, has a remarkable property in the case of scattering by a non-planar interface.

Brewster's angle in the case of non-magnetic media is often said to be the angle of incidence that results in a right angle (90°) between the wave vector of the transmitted wave and that of the (non-)reflected wave. In the case of a planar interface, our new definition of Brewster angle agrees with this explanation. Indeed, if $\hat{\mathbf{e}}^-_{p,1}(\mathbf{p}_0) \cdot \hat{\mathbf{e}}^+_{p,2}(\mathbf{p}_0) = 0$ it is immediate that $\mathbf{k}^+_1(\mathbf{p}_0) \cdot \mathbf{k}^-_2(\mathbf{p}_0) = 0$. However, we will see below that the new geometrical criterion proposed in the above holds when applied with Snell-conjugate wave vector associated with a non-specularly scattered wave, while the "right angle" criterion between wave vectors breaks down.

The Fresnel amplitudes for the *refracted* wave for s- and p-polarized light expressed in terms of polarization vectors, presented in a similar fashion as Eq. (27), read

$$\tau^{(0)}_{ss}(\mathbf{p}_0) = \frac{\tau^{(0)}(\mathbf{p}_0)}{\hat{\mathbf{e}}_s(\mathbf{p}_0) \cdot \hat{\mathbf{e}}_s(\mathbf{p}_0)} \tag{28a}$$

$$\tau^{(0)}_{pp}(\mathbf{p}_0) = \frac{\tau^{(0)}(\mathbf{p}_0)}{\hat{\mathbf{e}}^-_{p,1}(\mathbf{p}_0) \cdot \hat{\mathbf{e}}^-_{p,2}(\mathbf{p}_0)} \tag{28b}$$

$$\tau^{(0)}(\mathbf{p}_0) = \frac{c^2}{\omega^2} \frac{2\alpha_1(\mathbf{p}_0)}{(\epsilon_1 - \epsilon_2)} [\alpha_1(\mathbf{p}_0) - \alpha_2(\mathbf{p}_0)] \,. \tag{28c}$$

From Eq. (28), it is readily observed that neither the s- nor p-polarized zero order transmitted wave vanishes in general, which relates to the common experience that no Brewster angle is known for transmission through a planar interface. This fact does not, however, prevent the existence of Brewster *scattering* angles in the diffusely transmitted light. Equation (28) will therefore be important in the remainder of this paper. Note the presence of the factor $\alpha_1(\mathbf{p}_0) - \alpha_2(\mathbf{p}_0)$ in the transmission amplitude of the scalar wave in Eq. (28c) which is identical to the numerator of the reflection amplitude in Eq. (27c). The analysis of this term on total internal reflection hence leads to a similar behavior for the transmission amplitude, in the sense that $\tau^{(0)}(\mathbf{p}_0)$ leaves the real line and traces a path in the complex plane when total internal reflection occurs. This fact illustrates the coupling between the reflected and the transmitted waves, which may be interpreted as two components of the same mode.

First order — We now turn to the first order amplitudes $\mathbf{R}^{(1)}(\mathbf{p} \,|\, \mathbf{p}_0)$ and $\mathbf{T}^{(1)}(\mathbf{p} \,|\, \mathbf{p}_0)$. The first remarkable point to notice is that, when using the caret amplitudes, *both* the first order reflection and transmission amplitudes are proportional to $[\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p})]\hat{\zeta}(\mathbf{p} - \mathbf{p}_0)$. By a careful examination of the derivation of Eqs. (21c) and (21d) we note that this factor originates from a Taylor expansion of the term $\exp[-i(b\alpha_l(\cdot) - a\alpha_m(\cdot))\zeta(\cdot)]$ appearing in the $\mathcal{J}^{b,a}_{l,m}$ integral (see A). Intuitively, the $\mathcal{J}^{b,a}_{l,m}$ integral encodes information about the sum of amplitudes of scattering events occurring near the surface. A comparison of the expression for the $\mathcal{J}^{b,a}_{l,m}$ integral and the expression obtained by summing complex amplitudes for single scattering paths scattered at the surface of a perfect conductor supports this analogy. In other words, the integral encodes the resulting interference due to the phase difference between any scattering path occurring along the surface. To be more accurate, if one has the microscopic picture in mind one might say that it corresponds to summing scattering events occurring anywhere in the bulk, but in virtue of the extinction theorem the summation reduces to a sum over the surface [28]. With this picture in mind, the factor $[\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p})]\hat{\zeta}(\mathbf{p} - \mathbf{p}_0)$ corresponds to the resulting *interference* pattern when the phase factor is linearly approximated. Note that this factor does not contain any information about polarization. It is therefore instructive to consider the first order reflection and transmission amplitudes for scalar waves as a first step in order to obtain a better understanding of the full amplitudes. For scalar waves we set all the $\mathbf{M}^{b,a}_{l,m}$ matrices to unity in Eqs. (21c) and (21d) to obtain

$$R^{(1)}(\mathbf{p} \,|\, \mathbf{p}_0) = [\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p})]\, \hat{\zeta}(\mathbf{p} - \mathbf{p}_0) \times [1 + \rho^{(0)}(\mathbf{p}_0)] \tag{29a}$$

$$T^{(1)}(\mathbf{p} \,|\, \mathbf{p}_0) = [\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p})]\, \hat{\zeta}(\mathbf{p} - \mathbf{p}_0)\, \tau^{(0)}(\mathbf{p}_0) \,. \tag{29b}$$

From Eq. (29) it is apparent that the first order amplitudes are equal to the aforementioned interference factor, $[\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p})]\, \hat{\zeta}(\mathbf{p} - \mathbf{p}_0)$, weighted by the *total scalar zero order field amplitude* in either medium 1 or 2. Indeed, the factor $[1 + \rho^{(0)}(\mathbf{p}_0)]$ in Eq. (29a) represents the sum of the unit incident field amplitude and the amplitude of the reflected zero order field, while $\tau^{(0)}(\mathbf{p}_0)$ is simply the refracted zero order field amplitude; hence the denomination total zero order field.

One may think of the total zero order field as characterizing the state of a background field to which the interference pattern induced by the surface corrugation is superimposed (in a multiplicative sense). One could say that the interference pattern allows to probe the state of the zero order field away from the specular direction. In addition, note that the dependence on the outgoing wave vector $\mathbf{p}$ is entirely encoded in the term $[\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p})]\hat{\zeta}(\mathbf{p} - \mathbf{p}_0)$ while the state of the zero order field only depends on the incident wave vector $\mathbf{p}_0$. This indicates that the factor $[\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p})]\, \hat{\zeta}(\mathbf{p} - \mathbf{p}_0)$ can be thought to define a probability measure (by taking its modulus square) for a change of lateral wave vector from $\mathbf{p}_0$ to $\mathbf{p}$ (or its corresponding lateral momentum). Moreover, since the amplitude for a wave scattered with lateral wave vector $\mathbf{p}$ does not, to first order, involve any other wave than the incident one through its wave vector $\mathbf{p}_0$, we can consider a wave with lateral wave vector $\mathbf{p}$ as completely decoupled from a wave with lateral wave vector $\mathbf{q} \neq \mathbf{p}$. However, the wave reflected with lateral wave

vector $\mathbf{p}$ is considered to be coupled to the wave transmitted with the same lateral wave vector. In fact, this point of view is strengthened by the fact that both the reflected and transmitted waves sharing the same lateral wave vector also share exactly the same $\mathbf{p}$-dependence for their amplitude according to Eq. (29). We can interpret this fact by saying that the reflected and transmitted waves sharing the same lateral wave vector are two pieces of *one* mode defined in the whole space, and not solely on a single side of the interface. Note that this point of view is not new. It was adopted by Fano in the early 1940s as a general interpretation of the first order perturbative solution of the Rayleigh equation and used to further deduce the origin of the Wood anomalies [39].

What is now the physical interpretation of the polarization factors $\hat{\boldsymbol{\rho}}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_0)$ and $\hat{\boldsymbol{\tau}}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_0)$? It is clear from the definition of the amplitudes $\hat{\boldsymbol{\rho}}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_0)$ and $\hat{\boldsymbol{\tau}}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_0)$, given in Eqs. (A9) and (A16) respectively, that they are identical for different realizations of the surface profile. This suggests that these amplitudes are, in some sense, rather a signature of the bulk polarization response than the surface scattering properties. In using their definitions together with Eqs. (A9) and (A16) and by expressing the $\mathbf{M}^{b,a}_{l,m}$ matrices as functions of the polarization vectors in the local polarization basis as given in Eq. (10), we obtain the following expressions:

$$\hat{\rho}^{(1)}_{ss}(\mathbf{p}\,|\,\mathbf{p}_0) = \frac{\hat{\mathbf{e}}_s(\mathbf{p})\cdot\hat{\mathbf{e}}_s(\mathbf{p}_0) + \rho^{(0)}_{ss}(\mathbf{p}_0)\;\hat{\mathbf{e}}_s(\mathbf{p})\cdot\hat{\mathbf{e}}_s(\mathbf{p}_0)}{\hat{\mathbf{e}}_s(\mathbf{p})\cdot\hat{\mathbf{e}}_s(\mathbf{p})} = \hat{\mathbf{e}}_s(\mathbf{p})\cdot\mathbf{E}^{(0)}_{1,s}(\mathbf{p}_0) \tag{30a}$$

$$\hat{\rho}^{(1)}_{ps}(\mathbf{p}\,|\,\mathbf{p}_0) = \frac{\hat{\mathbf{e}}^{+}_{p,2}(\mathbf{p})\cdot\hat{\mathbf{e}}_s(\mathbf{p}_0) + \rho^{(0)}_{ss}(\mathbf{p}_0)\,\hat{\mathbf{e}}^{+}_{p,2}(\mathbf{p})\cdot\hat{\mathbf{e}}_s(\mathbf{p}_0)}{\hat{\mathbf{e}}^{+}_{p,2}(\mathbf{p})\cdot\hat{\mathbf{e}}^{+}_{p,1}(\mathbf{p})} = \frac{\hat{\mathbf{e}}^{+}_{p,2}(\mathbf{p})\cdot\mathbf{E}^{(0)}_{1,s}(\mathbf{p}_0)}{\hat{\mathbf{e}}^{+}_{p,2}(\mathbf{p})\cdot\hat{\mathbf{e}}^{+}_{p,1}(\mathbf{p})} \tag{30b}$$

$$\hat{\rho}^{(1)}_{sp}(\mathbf{p}\,|\,\mathbf{p}_0) = \frac{\hat{\mathbf{e}}_s(\mathbf{p})\cdot\hat{\mathbf{e}}^{-}_{p,1}(\mathbf{p}_0) + \hat{\mathbf{e}}_s(\mathbf{p})\cdot\hat{\mathbf{e}}^{+}_{p,1}(\mathbf{p}_0)\rho^{(0)}_{pp}(\mathbf{p}_0)}{\hat{\mathbf{e}}_s(\mathbf{p})\cdot\hat{\mathbf{e}}_s(\mathbf{p})} = \hat{\mathbf{e}}_s(\mathbf{p})\cdot\mathbf{E}^{(0)}_{1,p}(\mathbf{p}_0) \tag{30c}$$

$$\hat{\rho}^{(1)}_{pp}(\mathbf{p}\,|\,\mathbf{p}_0) = \frac{\hat{\mathbf{e}}^{+}_{p,2}(\mathbf{p})\cdot\hat{\mathbf{e}}^{-}_{p,1}(\mathbf{p}_0) + \hat{\mathbf{e}}^{+}_{p,2}(\mathbf{p})\cdot\hat{\mathbf{e}}^{+}_{p,1}(\mathbf{p}_0)\rho^{(0)}_{pp}(\mathbf{p}_0)}{\hat{\mathbf{e}}^{+}_{p,2}(\mathbf{p})\cdot\hat{\mathbf{e}}^{+}_{p,1}(\mathbf{p})} = \frac{\hat{\mathbf{e}}^{+}_{p,2}(\mathbf{p})\cdot\mathbf{E}^{(0)}_{1,p}(\mathbf{p}_0)}{\hat{\mathbf{e}}^{+}_{p,2}(\mathbf{p})\cdot\hat{\mathbf{e}}^{+}_{p,1}(\mathbf{p})} \tag{30d}$$

for the reflection amplitudes and

$$\hat{\tau}^{(1)}_{ss}(\mathbf{p}\,|\,\mathbf{p}_0) = \tau^{(0)}_{ss}(\mathbf{p}_0)\,\frac{\hat{\mathbf{e}}_s(\mathbf{p})\cdot\hat{\mathbf{e}}_s(\mathbf{p}_0)}{\hat{\mathbf{e}}_s(\mathbf{p})\cdot\hat{\mathbf{e}}_s(\mathbf{p})} = \frac{\hat{\mathbf{e}}_s(\mathbf{p})\cdot\mathbf{E}^{(0)}_{2,s}(\mathbf{p}_0)}{\hat{\mathbf{e}}_s(\mathbf{p})\cdot\hat{\mathbf{e}}_s(\mathbf{p})} \tag{31a}$$

$$\hat{\tau}^{(1)}_{ps}(\mathbf{p}\,|\,\mathbf{p}_0) = \tau^{(0)}_{ss}(\mathbf{p}_0)\,\frac{\hat{\mathbf{e}}^{-}_{p,1}(\mathbf{p})\cdot\hat{\mathbf{e}}_s(\mathbf{p}_0)}{\hat{\mathbf{e}}^{-}_{p,1}(\mathbf{p})\cdot\hat{\mathbf{e}}^{-}_{p,2}(\mathbf{p})} = \frac{\hat{\mathbf{e}}^{-}_{p,1}(\mathbf{p})\cdot\mathbf{E}^{(0)}_{2,s}(\mathbf{p}_0)}{\hat{\mathbf{e}}^{-}_{p,1}(\mathbf{p})\cdot\hat{\mathbf{e}}^{-}_{p,2}(\mathbf{p})} \tag{31b}$$

$$\hat{\tau}^{(1)}_{sp}(\mathbf{p}\,|\,\mathbf{p}_0) = \tau^{(0)}_{pp}(\mathbf{p}_0)\,\frac{\hat{\mathbf{e}}_s(\mathbf{p})\cdot\hat{\mathbf{e}}^{-}_{p,2}(\mathbf{p}_0)}{\hat{\mathbf{e}}_s(\mathbf{p})\cdot\hat{\mathbf{e}}_s(\mathbf{p})} = \frac{\hat{\mathbf{e}}_s(\mathbf{p})\cdot\mathbf{E}^{(0)}_{2,p}(\mathbf{p}_0)}{\hat{\mathbf{e}}_s(\mathbf{p})\cdot\hat{\mathbf{e}}_s(\mathbf{p})} \tag{31c}$$

$$\hat{\tau}^{(1)}_{pp}(\mathbf{p}\,|\,\mathbf{p}_0) = \tau^{(0)}_{pp}(\mathbf{p}_0)\,\frac{\hat{\mathbf{e}}^{-}_{p,1}(\mathbf{p})\cdot\hat{\mathbf{e}}^{-}_{p,2}(\mathbf{p}_0)}{\hat{\mathbf{e}}^{-}_{p,1}(\mathbf{p})\cdot\hat{\mathbf{e}}^{-}_{p,2}(\mathbf{p})} = \frac{\hat{\mathbf{e}}^{-}_{p,1}(\mathbf{p})\cdot\mathbf{E}^{(0)}_{2,p}(\mathbf{p}_0)}{\hat{\mathbf{e}}^{-}_{p,1}(\mathbf{p})\cdot\hat{\mathbf{e}}^{-}_{p,2}(\mathbf{p})} \tag{31d}$$

for the transmission amplitudes. Here we have defined the total zero order field amplitudes in media 1 and 2, for s- and p-polarized incident light, as

$$\mathbf{E}^{(0)}_{1,s}(\mathbf{p}_0) = \left[1+\rho^{(0)}_{ss}(\mathbf{p}_0)\right]\,\hat{\mathbf{e}}_s(\mathbf{p}_0) \tag{32a}$$

$$\mathbf{E}^{(0)}_{1,p}(\mathbf{p}_0) = \hat{\mathbf{e}}^{-}_{p,1}(\mathbf{p}_0) + \rho^{(0)}_{pp}(\mathbf{p}_0)\,\hat{\mathbf{e}}^{+}_{p,1}(\mathbf{p}_0) \tag{32b}$$

$$\mathbf{E}^{(0)}_{2,s}(\mathbf{p}_0) = \tau^{(0)}_{ss}(\mathbf{p}_0)\;\hat{\mathbf{e}}_s(\mathbf{p}_0) \tag{32c}$$

$$\mathbf{E}^{(0)}_{2,p}(\mathbf{p}_0) = \tau^{(0)}_{pp}(\mathbf{p}_0)\,\hat{\mathbf{e}}^{-}_{p,2}(\mathbf{p}_0)\,. \tag{32d}$$

The amplitudes given in Eqs. (32) correspond to the sum of the unit incident field and the reflected or transmitted zero order field amplitudes. In other words, they characterize the state of the field given by the superposition of the incident wave and the zero order response of the media. This is analogous to what we found for scalar waves in Eq. (29), with the difference that due to the dipolar nature of the scatterers the state of this zero order background is anisotropic, as indicated by the dependence on $\mathbf{p}$.

An interesting point to notice from Eqs. (30) and (31) is that the amplitude of the first order α-polarized wave scattered with lateral wave vector $\mathbf{p}$ in medium j is proportional to the projection of the total zero order field amplitude in medium j, induced by an incident β-polarized wave with lateral wave vector $\mathbf{p}_0$, on the direction of the polarization vector associated with its Snell-conjugate wave vector (with lateral wave vector $\mathbf{p}$). For s-polarized scattered waves the Snell-conjugate wave vector is not apparent since all s-polarization vectors are in the $(\hat{\mathbf{e}}_1,\hat{\mathbf{e}}_2)$-plane. For p-polarized scattered waves this is clear as is indicated by the vectors $\hat{\mathbf{e}}^{+}_{p,2}(\mathbf{p})$ and $\hat{\mathbf{e}}^{-}_{p,1}(\mathbf{p})$, for the first order amplitudes in reflection and

transmission, respectively. One way of interpreting these equations is to imagine that the wave scattered in one medium is induced by the roughness of the surface in the sense that, contrary to the case of the planar surface, the path issued from a scattering event on the surface will *not* destructively interfere anymore when summed over the surface, and hence probe the underlying polarized radiation pattern induced by arrays of dipole emitters in the bulk. The scattering amplitude of such a scattered wave will be controlled by the polarized state of the media, in such a way that the amplitude is proportional to the total zero order field amplitude in the medium of scattering but projected on the polarization vector the scattered wave would have had, keeping its lateral wave vector and according to Snell's law, if it were incoming from the opposite medium. This is a generalization of Doyle's analysis based on the Ewald triad, although we now have a triad associated with each observed scattered wave.

A more detailed microscopic understanding of this result would require an analysis of the way arrays of dipoles emit when excited by a primary incident wave, and by summing the elementary wavelets emitted by all the dipoles while also taking into account the geometry of the surface. Such an analysis was done by Ewald in his derivation of the extinction theorem for a planar surface [29], and later used by Doyle [26] in interpreting the Fresnel coefficients at a microscopic level, as explained earlier. A Snell-conjugate wave vector can be viewed as the wave vector of one of the waves of the Ewald triad associated to a (first order) scattered wave, in a similar fashion as what we have already encountered when factorizing the Fresnel coefficients for a p-polarized wave in Eq. (27b). Modern derivations of the extinction theorem are directly based on the macroscopic properties of the media, and what we obtain from such formalism can only be the integrated contribution of all the scatterers, as is also the case with the solution of the reduced Rayleigh equations, which can be shown to derive from the extinction theorem. Consequently, we must take our Snell-conjugate wave interpretation as a signature of a more fundamental microscopic view, and consider this concept as a useful short cut for reasoning directly on the integrated response, just as we did for interpreting the Fresnel coefficients following Doyle's interpretation.

We can summarize the interpretations of the first order amplitudes discussed in this section as a factorization of two main mechanisms. All the first order amplitudes can be written as the product of a polarization independent factor, $[\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p})]\ \hat{\zeta}(\mathbf{p}-\mathbf{p}_0)$, common to all amplitudes sharing the same outgoing lateral wave vector, and a polarization dependent factor. The polarization independent factor encodes the interference pattern of all spherical waves scattered in the vicinity of the non-planar surface, and can be viewed as the scattering or probing mechanism. The polarization dependent factor is controlled by the state of the zero order field, and is always proportional to the projection of the total zero order field amplitude onto the polarization vector of a Snell-conjugate wave associated to the observed scattered wave. This mechanism ensures that the polarization and amplitude of the observed scattered wave is consistent with the polarized dipole radiation characterized by the state of the zero order field. We are now ready for a more in-depth analysis of the Yoneda and Brewster scattering effects based on the physical interpretation and concepts we have developed in the present section.

C. The physical origin of the Yoneda effect

Our observations on the Yoneda and Brewster scattering effects in Section IV A led us to the conclusion that the two effects can be explained independently. The fact that the Brewster scattering angle coincides with the Yoneda critical angle for normal incidence can, for the time being, be considered a simple coincidence. Since the Yoneda phenomenon seems to be independent of polarization we can attempt an explanation solely based on scalar waves and consider Eq. (29) a relevant simplified model, in an analogous fashion as Eq. (27c) was sufficient to explain total internal reflection from a planar surface. In fact, for the scattering of s-polarized waves restricted to the plane of incidence ($\mathbf{p} \parallel \mathbf{p}_0$) the reflection and transmission amplitudes are exactly given by Eq. (29). We will therefore keep to scalar waves for the main analysis, but we will also illustrate our conclusions with results obtained for s-polarized waves. In the following, it will be convenient to refer to the smallest and largest dielectric constant by $\epsilon_{\min}$ and $\epsilon_{\max}$ respectively, and more generally we will index by min and max the quantities corresponding to these media. Our analysis will be independent of the configurations of the media but will require us to distinguish the optically denser medium from the less dense medium for the scattered waves.

We can view the scattering mechanism responsible for the Yoneda peak phenomenon as a two step process for the sake of clarity. First, the incident wave impinges on the surface with an in-plane lateral wave vector $\mathbf{p}_0$ and, within a single scattering point of view, gives rise to a scattered elementary wave reflected with the in-plane lateral wave vector $\mathbf{p}$ and a scattered elementary wave transmitted with the in-plane lateral wave vector $\mathbf{p}$. These are arbitrarily chosen wave vectors; the total scattered field will have components *a priori* for all wave vectors but in our analysis we consider just these two arbitrary waves. One may have in mind the picture of an optical path allowed to be scattered only once in our single scattering view and the total field will be obtained by summing the probability amplitude of all optical paths. As argued in Section **??**, the probability for a change of lateral wave vector from $\mathbf{p}_0$ to $\mathbf{p}$ is controlled by the factor $[\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p})]\ \hat{\zeta}(\mathbf{p}-\mathbf{p}_0)$ (or its modulus square). This is not sufficient to obtain the complete probability

amplitude, which in general and for polarized waves will depend on the wave vectors, dielectric constants and polarizations of the waves involved in the process. However, we will show that we do not need to analyze the details of these amplitudes in order to investigate the Yoneda phenomenon.

The second step determines whether the intensity is enhanced in the optically denser medium for a given elementary scattered wave. We can view the elementary reflected and transmitted waves as coupled into a single mode as explained in the previous section. We can take this one step further, and interpret the argument based on the probability measure for a change of momentum as a way to allocate part of the energy from the incident wave, to be shared between, and radiated away by, the two scattered waves gathered in a coupled mode with a shared lateral wave vector $\mathbf{p}$.

Let us first consider the situation where the shared lateral wave vector of the scattered waves is restricted to $|\,\mathbf{p}\,| < n_{\min}\omega/c = p_c$, which means that both waves are allowed to propagate to infinity in their respective medium. Under this assumption, the total energy of the two waves will be shared *a priori* non-trivially between the two waves. In particular, for a given $\mathbf{p}$ it is apparent from Eq. (29) that the energy will be split according to the relative amplitudes given by the zero order state since the factor $[\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p})]\,\hat{\zeta}(\mathbf{p}-\mathbf{p}_0)$ is common to both the reflected and transmitted wave. However, if now the shared lateral wave vector is such that $n_{\min}\omega/c < |\,\mathbf{p}\,| < n_{\max}\omega/c$, the wave scattered in the optically less dense medium will be evanescent. Therefore the total energy for the coupled mode will be carried away solely by the wave which can propagate, namely the one scattered into the dense medium, resulting in the apparent sudden increase of intensity at the transition between propagation and evanescence of the wave scattered in the optically less dense medium. An illustrative way of seeing that the intensity needs to be enhanced is by analyzing the factor $\alpha_1(\mathbf{p})-\alpha_2(\mathbf{p})$ assuming $|\hat{\zeta}|$ to vary slowly. For $|\,\mathbf{p}\,| < p_c$ both $\alpha_1(\mathbf{p})$ and $\alpha_2(\mathbf{p})$ are real. As $|\,\mathbf{p}\,| \to p_c$ from below, $\alpha_{\min}(\mathbf{p}) \to 0$ and $\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p}) \to \pm\alpha_c$, with $\alpha_c = \sqrt{\epsilon_{\max} - \epsilon_{\min}}\,\omega/c$. By writing $p = |\,\mathbf{p}\,| = p_c - \Delta p$, with $\Delta p > 0$, we can make an asymptotic analysis of $|\alpha_1(\mathbf{p})-\alpha_2(\mathbf{p})|$ as $p \to p_c$ from below. In this way we obtain the following result

$$
\begin{aligned}
&|\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p})|\frac{c}{\omega} = [\alpha_{\max}(\mathbf{p}) - \alpha_{\min}(\mathbf{p})]\frac{c}{\omega} \\
&= [\epsilon_{\max} - (\tilde{p}_c - \Delta\tilde{p})^2]^{1/2} - [\epsilon_{\min} - (\tilde{p}_c - \Delta\tilde{p})^2]^{1/2} \\
&= [\epsilon_{\max} - \epsilon_{\min} + 2\epsilon_{\min}^{1/2}\Delta\tilde{p} - \Delta\tilde{p}^2]^{1/2} - [2\epsilon_{\min}^{1/2}\Delta\tilde{p} - \Delta\tilde{p}^2]^{1/2} \\
&= \alpha_c\,\frac{c}{\omega} - [2\epsilon_{\min}^{1/2}\Delta\tilde{p}]^{1/2} + o(\Delta\tilde{p}^{1/2})\,. \qquad (33)
\end{aligned}
$$

Here we have chosen to work with unit-less quantities and denoted $\tilde{p} = p\,c/\omega$ for conciseness. From Eq. (33) it then follows that as $\Delta p \to 0$, $|\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p})|$ must increase towards α_c in an inner-neighborhood of the circle $p = p_c$. Furthermore, the asymptotic expansion reveals that the critical point will be reached with a sharp edge (infinite slope) for $p < p_c$ as can be deduced from the square root behavior in Δp. Note that both the reflection and transmission amplitudes exhibit the same behavior independently of which medium is denser. This is due to the fact that the two waves are part of the same mode. However, as the wave propagating in the less dense medium becomes a grazing wave, the corresponding differential scattering coefficient is forced to vanish due to the angular dependence in $\cos^2\theta_s$ ($\theta_s = \theta_r$ or θ_t depending on the context). The complex amplitude is nevertheless enhanced for both the reflected and transmitted wave. This is illustrated for example in Figs. 2(a) and 2(d), which corresponds to a case for which the medium of incidence is vacuum. From the results presented in these figures, we can see that while the incoherent component of the MDRC is forced to go to zero when $p_1 \to p_c = \omega/c$, the surface-independent part of the reflection amplitude $\rho_{ss}^{(1)}$ exhibits a sharp increase in modulus. Simultaneously, the surface-independent part of the transmission amplitude $\tau_{ss}^{(1)}$ also exhibits a similar sharp increase in modulus as p_1 approaches p_c. Consequently, since the wave can propagate away from the surface in the second medium (which consists of glass in this specific case), the corresponding incoherent component of the MDTC exhibits a similar increase. Note that both the phases associated with $\rho_{ss}^{(1)}$ and $\tau_{ss}^{(1)}$ remain constant and equal to 0 for $p_1 < p_c$ for all θ_0 in Fig. 2, since the complex amplitude stays on the real line in the case where $\epsilon_1 < \epsilon_2$ independent of the angle of incidence.

Figures 3 and 4 support the same conclusion but by interchanging the role of the media. The only difference worth noting is that the phases $\phi_{ss}^{r,(1)}$ and $\phi_{ss}^{t,(1)}$ have a constant plateau for $p_1 < p_c$ which is equal to 0 only for $\theta_0 < \theta_c$. The plateau is offset for $\theta_0 > \theta_c$. This overall phase offset is due to the Goos-Hänchen phase shift associated with total internal reflection of the zero order wave. Indeed, recall that the first order amplitudes are proportional to the total zero order field amplitudes. As a consequence, if the zero order waves exhibit a phase shift, it will affect the first order amplitudes in the form of a constant phase offset for all $\mathbf{p}$.

When $|\,\mathbf{p}\,| > p_c$, $\alpha_{\min}$ becomes purely imaginary and $\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p})$ thus moves off the real line. For $p_c < |\,\mathbf{p}\,| < n_{\max}\omega/c$, we find that in this regime $\alpha_1(\mathbf{p})-\alpha_2(\mathbf{p})$ keeps a constant modulus equal to α_c. Indeed, by writing $\alpha_{\min}(\mathbf{p}) = i\beta_{\min}(\mathbf{p})$ we have

$$
\begin{aligned}
|\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p})| &= |\alpha_{\max}(\mathbf{p}) - i\beta_{\min}(\mathbf{p})| \\
&= \left[\alpha_{\max}^2(\mathbf{p}) + \beta_{\min}^2(\mathbf{p})\right]^{1/2} \\
&= \left[\epsilon_{\max} - \tilde{p}^2 + \tilde{p}^2 - \epsilon_{\min}\right]^{1/2}\omega/c \\
&= \alpha_c\,. \qquad (34)
\end{aligned}
$$

The complex number $\alpha_1(\mathbf{p})-\alpha_2(\mathbf{p})$ thus traces a circular arc of radius α_c in the complex plane. Finally, when

$|\,\mathbf{p}\,| > n_{\max}\,\omega/c$, both the reflected and transmitted waves are evanescent, $\alpha_{\max}$ becomes pure imaginary and hence $\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p})$ moves along the imaginary axis. The constant value of $|\rho_{ss}^{(1)}|$ and $|\tau_{ss}^{(1)}|$ in the regime $n_{\min}\,\omega/c < |\,\mathbf{p}\,| < n_{\max}\,\omega/c$ can be appreciated for all angles of incidence illustrated in Figs. 2 – 4, while the phases exhibit a smooth variation from their plateau value and decay by a total amount of $-\pi/2$ when reaching $p_1 = n_{\max}\,\omega/c$. Once the threshold of $n_{\max}\,\omega/c$ has been passed, the phases remain constant and the modulii decay towards zero as $|\,\mathbf{p}\,| \to \infty$ (which can easily be deduced from a straightforward asymptotic analysis leading to $\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p}) \sim i(\epsilon_{\max} - \epsilon_{\min})\omega^2/(2c^2 p)$). The phase change associated with the transition from the real line to the imaginary line in the complex plane is therefore $-\pi/2$. This gradual phase change is similar to that of the Goos-Hänchen phase shift discussed for the reflection by a planar surface. The difference of absolute total phase change, of π for the case of the Fresnel amplitude and $\pi/2$ in the case of the scattered waves, comes mathematically from the fact that in the former case the amplitude is written as the ratio of a complex number and its complex conjugate, while in the latter case there is no such ratio. The phase consequently turns twice as fast in the former case than in the latter. A physical interpretation of this difference is that for the Fresnel amplitude both the incident and outgoing wave vector must vary simultaneously (since they are the same), while in the case of a scattered wave the incident wave vector is fixed while only the outgoing wave vector is allowed to vary. In fact, we have only analyzed the phase associated with the factor $\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p})$ in Eq. (29). The phase of the overall complex amplitude will be the sum of the aforementioned phase, that given by the argument of $\hat{\zeta}(\mathbf{p} - \mathbf{p}_0)$, and the phase given by the argument of the total zero order amplitude $[1 + \rho^{(0)}(\mathbf{p}_0)]$ or $\tau^{(0)}(\mathbf{p}_0)$. In particular, if the angle of incidence is such that total internal reflection occurs for the zero order field, the overall phase of the scattered amplitude will contain a signature of the Goos-Hänchen phase shift associated with the total internal reflection of the zero order field in addition to the corresponding Goos-Hänchen phase shift associated with the Yoneda effect. Note that when averaged over surface realizations, the phase contribution coming from $\hat{\zeta}$ averages to zero. This supports our choice of limiting the detailed investigation to the surface-independent factors in Eq. (29).

To summarize, let us gather some important results and answer some of the questions which were left unanswered in previous studies. First, we would like to stress that the above analysis predicts a critical angle for the Yoneda phenomenon which is independent both of the angle of incidence and of which medium the incident wave came from. The Yoneda transition will therefore always occur at the same polar angle of scattering: the one given by $|\,\mathbf{p}_c\,| = n_{\min}\,\omega/c$. We also want to emphasize that the approximate solution of the reduced Rayleigh equations obtained via SAPT to first order in the surface profile is commonly accepted as a single scattering approximation. In light of our analysis of the Yoneda phenomenon, it is clear that the analogy of the Yoneda phenomenon with that of total internal reflection put forward in the literature may seem a valid one. There are, however, some comments to be made about this analogy. It is important to emphasize the underlying cause of total internal reflection, namely the impossibility of an evanescent wave to carry energy away from the surface, given the assumed scattering system. Indeed, trying to directly and naively apply the total internal reflection argument would lead one to expect an absence of the Yoneda effect in transmission into the dense medium based on a single scattering picture, as this would require multiple scattering events. Indeed, one could imagine that the incident wave would need to scatter once to a transmitted grazing or evanescent wave and then a second time to be scattered in reflection in the dense medium and therefore follow the rule of total internal reflection. Such a naive picture would be in contradiction with results from numerical experiments based on first order perturbation theory [1, 2], or at least contradict the common single-scattering picture associated with it, and we believe that our interpretation resolves this issue. The results presented in Refs. 1 and 2 validated the qualitative use of SAPT in describing the Yoneda phenomenon, for the roughness parameters assumed in these studies, when compared to numerical results obtained through a non-perturbative solution of the reduced Rayleigh equations. Similar non-perturbative solutions were found to match experimental results showing the Yoneda phenomenon in Ref. 16.

In fact, the Yoneda phenomenon for weakly rough surfaces originates from the same physical mechanism as the Rayleigh anamolies for periodic dielectric gratings. The continuous set of scattered wave vectors in the case of a randomly rough surface can be viewed as probing a diffracted order scattered from a periodic surface with continuously changing lattice constant. It is easy to show numerically and with SAPT to first order, that the behavior of the efficiency of a given diffractive order as the lattice constant is changed exhibits the same characteristic peak as the Yoneda peak when its counter part in the less dense medium becomes evanescent. The perturbative analysis in the case of a periodic grating is exactly the same as in the case of a randomly rough surface with the only difference being that $\mathbf{p}$ must be replaced by the in-plane wave vector of the diffractive order of interest and make the lattice constant vary instead.

As a remark, we would like to point out that since the analysis was carried out for the scattering of a scalar wave subjected to the continuity of the field and its normal derivative with respect to the surface, we predict that the Yoneda phenomenon should also be observed for the scattering of a quantum particle by a rough interface between two regions of constant potential.

In studying the results from Figs. 2 – 4 we avoided

a direct discussion for p-polarized waves, for which the results put forward by the scalar wave analysis seem to be invalidated. The analysis done for scalar waves is, in fact, still valid but must be complemented with additional effects, due to polarization, not only for p-polarized light but also for s-polarized light when the scattering direction is out of the plane of incidence as suggested by Eqs. (30, 31). This is the subject of the following section.

D. Physical and geometrical explanations of the Brewster scattering effect

For a randomly rough surface, we have seen in Figs. 2 – 4 that we may find a Brewster scattering angle for a wide range of angles of incidence if we look at both the reflected and transmitted light (the MDRC and MDTC). We will now see that the general Brewster scattering phenomenon, roughly defined as a wave scattered with zero amplitude in a single scattering approximation, also extends to scattered waves in the evanescent regime. To this end we will continue our dissection of the phenomenon through perturbative theory.

In-plane reflection — Let us focus first on the case of co-polarized scattering in the plane of incidence to fix the ideas. Equation (30a) shows that $\hat{\rho}_{ss}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_0)$ is proportional to $\hat{\mathbf{e}}_s(\mathbf{p})\cdot\mathbf{E}_{s,1}^{(0)}(\mathbf{p}_0)$, where $\mathbf{E}_{s,1}^{(0)}(\mathbf{p}_0)$ is the *total zero order field amplitude* in medium 1 given by the sum of the unit incident field amplitude and the reflected field amplitude given by the Fresnel coefficient for an s-polarized wave. This relation indicates that the field amplitude of the *first order* reflected amplitude for the wave scattered with lateral wave vector $\mathbf{p}$ is proportional to the projection of its polarization vector on the total zero order field. For scattering in the plane of incidence $\hat{\mathbf{e}}_s(\mathbf{p}) = \hat{\mathbf{e}}_s(\mathbf{p}_0)$ and therefore the first order reflection amplitude reduces to that of the scalar wave Eq. (29a). Consequently, there is no extinction for $s \to s$ scattering in the plane of incidence for any angle of incidence. The same analysis and conclusion hold for the transmitted s-polarized wave.

Similarly, for p-polarized light, Eq. (30d) shows that the first order reflection amplitude is proportional to $\hat{\mathbf{e}}_{p,2}^{+}(\mathbf{p})\cdot\mathbf{E}_{1,p}^{(0)}(\mathbf{p}_0)$, where we recall that $\mathbf{E}_{1,p}^{(0)}(\mathbf{p}_0)$ is the total zero order field amplitude given by the sum of the unit incident field amplitude and the reflected field amplitude given by the Fresnel coefficient for p-polarized waves. Equation (30d) states that the *first order* field amplitude is proportional to the projection of the Snell-conjugate wave's polarization vector $\hat{\mathbf{e}}_{p,2}^{+}(\mathbf{p})$ along the direction of the *total zero order field*. Note the similarity with what was found for the Fresnel coefficient for p-polarized light in Eq. (27b). From Eq. (30d) we can deduce a simple geometrical criterion for Brewster scattering within first order perturbation theory: *The lateral wave vector(s)* $\mathbf{p}_{\rm B}$ *of the elementary Brewster scattered wave(s), for which the reflection amplitude for a p-polarized reflected wave vanishes given a p-polarized incident wave with lateral wave vector* $\mathbf{p}_0$ *is given by the condition of orthogonality between the p-polarization vector of the Snell-conjugate scattered wave(s) and the total zero order field in medium 1, i.e.*

$$\hat{\mathbf{e}}_{p,2}^{+}(\mathbf{p}_{\rm B})\cdot\mathbf{E}_{p,1}^{(0)}(\mathbf{p}_0) = 0\,. \tag{35}$$

As a direct consequence, in the case of co-polarized scattering in the plane of incidence, the geometrical condition can be re-stated as a requirement on the colinearity between the Snell-conjugate wave vector and the total zero order field, which is exactly the same geometrical criterion found in the case of reflection from a planar interface. A second corollary is that for in-plane scattering $\Theta_{\rm B}(\theta_{\rm B}) = \theta_{\rm B}$: the *Brewster scattering angle* is equal to the *Brewster angle for a planar interface* when the angle of incidence is equal to the Brewster angle for a planar interface, $\theta_0 = \theta_{\rm B}$ (or so-called Brewster incidence). In other words, the Brewster angle for a planar interface, $\theta_{\rm B}$, is a fixed point for the mapping which associates the angle of incidence to the Brewster scattering angle: $\Theta_{\rm B} : \theta_0 \mapsto \Theta_{\rm B}(\theta_0)$. This is readily understood from the geometrical criterion expressed by Eq. (35). At Brewster incidence the zero order reflected wave vanishes (by definition of Brewster incidence). Thus the total zero field amplitude is simply the incident field amplitude, $\mathbf{E}_{p,1}^{(0)}(\mathbf{p}_0) = \hat{\mathbf{e}}_{p,1}^{-}(\mathbf{p}_0)$, and consequently, the Brewster scattering angle is necessarily equal to $\theta_{\rm B}$.

Let us now apply the above criterion for tracking the Brewster scattering direction while the angle of incidence varies. We start with the case where the incident plane wave is approaching the rough interface from vacuum, and is reflected from a glass substrate [$\epsilon_1 = 1$ and $\epsilon_2 = 2.25$]. Figure 6 presents selected wave vectors for different polar angles of incidence θ_0, highlighting the geometrical relations leading to the Brewster scattering direction. The dashed circles in this figure represent the dispersion relations ($|\,\mathbf{k}_j^{\pm}\,| = k_j = n_j\omega/c$) by indicating the norm of wave vectors allowed to propagate in the two media. The incident wave vector $\mathbf{k}_1^{-}(\mathbf{p}_0)$ is represented pointing towards the origin for clarity while the wave vector for the reflected zero order wave, $\mathbf{k}_1^{+}(\mathbf{p}_0)$, is represented as pointing outwards. The red dashed line corresponds to the direction of the reflected wave vector for Brewster scattering in the case of a planar interface with the purpose of illustrating the aforementioned fixed point property of the Brewster incident angle $\theta_0 = \theta_{\rm B}$. The general construction rules go as follows. First, the wave vectors of the incident and the reflected zero order waves are drawn in black. Second, the direction of the total zero order field given by Eq. (32b) is determined and the wave vector of the virtual wave, which is colinear to the total zero order field (not represented), is drawn as the blue wave vector $\mathbf{k}_2^{+}(\mathbf{p}_{\rm B})$. Note that $\mathbf{k}_2^{+}(\mathbf{p}_{\rm B})$ lies on the circle of radius $n_2\omega/c$. The projection of $\mathbf{k}_2^{+}(\mathbf{p}_{\rm B})$ along $\hat{\mathbf{e}}_1$ gives the Brewster lateral wave vector $\mathbf{p}_{\rm B}$ from which we deduce $\mathbf{k}_1^{+}(\mathbf{p}_{\rm B})$ in red. Note that the reflected wave associated with $\mathbf{k}_1^{+}(\mathbf{p}_{\rm B})$ may be evanescent, and in that case we simply represent its lateral component $\mathbf{p}_{\rm B}$ as its component along $\hat{\mathbf{e}}_3$ is pure imaginary.

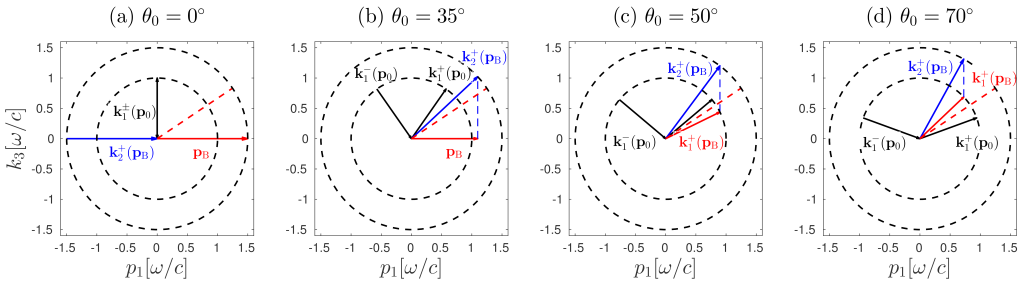


FIG. 6. Illustration of the geometrical criterion for in-plane Brewster scattering for different polar angles of incidence: (a) $\theta_0 = 0°$, (b) $\theta_0 = 35°$, (c) $\theta_0 = 50°$ and (d) $\theta_0 = 70°$. The dashed circles represent the norm of the full wave vectors, given by the dispersion relations, in vacuum ($\epsilon_1 = 1$ inner circle) and glass ($\epsilon_2 = 2.25$ outer circle). The black arrows represent respectively the incident wave vector $\mathbf{k}_1^-(\mathbf{p}_0)$, which is drawn as pointing towards the origin for clarity, and the wave vector of the reflected zero order wave, $\mathbf{k}_1^+(\mathbf{p}_0)$. The blue arrow represents the wave vector of the virtual wave, $\mathbf{k}_2^+(\mathbf{p}_B)$, from which the lateral wave vector of the Brewster wave, $\mathbf{p}_B$, is deduced by projection along $\hat{\mathbf{e}}_1$. From $\mathbf{p}_B$, the full wave vector for the Brewster wave, $\mathbf{k}_1^+(\mathbf{p}_B)$, can be drawn (provided propagation in medium 1) as a red arrow. Note that if the Brewster wave is evanescent, only $\mathbf{p}_B$ is draw in red as the out-of-plane component of $\mathbf{k}_1^+(\mathbf{p}_B)$ is purely imaginary. The red dashed line indicates the Brewster angle for a planar surface approximately equal to 56.3° in this case.

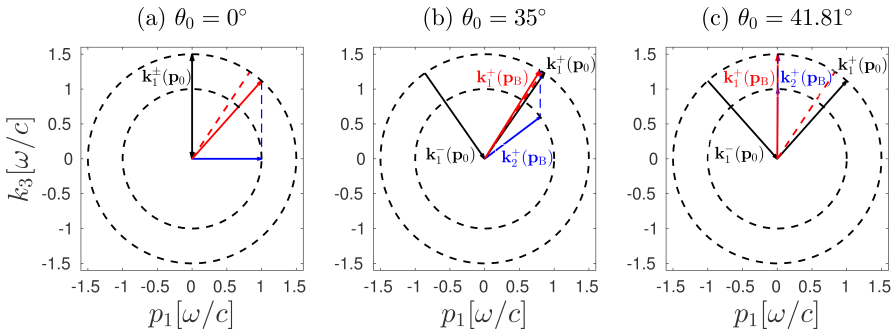


FIG. 7. Illustration of the geometrical criterion for in-plane Brewster scattering for different polar angles of incidence: (a) $\theta_0 = 0°$, (b) $\theta_0 = 35°$ and (c) $\theta_0 = 41.81°$. The dashed circles represent the norm of the full wave vectors, given by the dispersion relations, in glass ($\epsilon_1 = 2.25$ inner circle) and vacuum ($\epsilon_2 = 1$ outer circle). The black arrows represent respectively the incident wave vector $\mathbf{k}_1^-(\mathbf{p}_0)$, which is drawn as pointing towards the origin for clarity, and the wave vector of the reflected zero order wave, $\mathbf{k}_1^+(\mathbf{p}_0)$. The blue arrow represents the wave vector of the virtual wave, $\mathbf{k}_2^+(\mathbf{p}_B)$, from which the lateral wave vector of the Brewster wave, $\mathbf{p}_B$, is deduced by projection along $\hat{\mathbf{e}}_1$. From $\mathbf{p}_B$, the full wave vector for the Brewster wave, $\mathbf{k}_1^+(\mathbf{p}_B)$, can be drawn (provided propagation in medium 1) as a red arrow. Note that if the Brewster wave is evanescent, only $\mathbf{p}_B$ is drawn in red as the out-of-plane component of $\mathbf{k}_1^+(\mathbf{p}_B)$ is purely imaginary. The red dashed line indicates the Brewster angle for a planar surface, approximately equal to 33.7° in this case.

For normal incidence [Fig. 6(a)] the total zero order electric field lies along $\hat{\mathbf{e}}_1$, and consequently, so does $\mathbf{k}_2^+(\mathbf{p}_B)$. In fact, for normal incidence, due to the fact that the total zero order field lies along $\hat{\mathbf{e}}_1$, there are *two* Brewster waves in the plane of incidence with opposite wave vectors $p_1 = \pm n_2\,\omega/c$, but we focus on the one pointing to the right for clarity in Fig. 6(a). It follows from $\mathbf{k}_2^+(\mathbf{p}_B)$ that $|\,\mathbf{p}_B\,| > n_1\,\omega/c$ and the corresponding Brewster (non-) reflected wave is therefore evanescent. Such a case could not be revealed in previous work which focused on the diffusely scattered intensity radiated away from the surface. Nevertheless, the theory suggests that the notion of Brewster scattering should be extended to evanescent waves. This effect is indeed visible by inspection of the modulus of the amplitude $\rho_{pp}^{(1)}$ in Fig. 2(a). Indeed, we observe that for $p_1 = \pm n_2\,\omega/c$, $\rho_{pp}^{(1)}$ vanishes. The corresponding phase $\phi_{pp}^{\mathrm{r},(1)}$ exhibits a jump which is characteristic of the Brewster effect. The phase jump is equal to $\pi/2$ in this case, while in general the phase jump associated with the Brewster effect is equal to π. The $\pi/2$ jump seems to happen only when two Brewster waves with opposite lateral wave vectors are solutions of the criterion Eq. (35), which as far as we can see only occurs at normal incidence for the sys-

tems studied in this paper. It is tempting to interpret the $\pi/2$ jump as actually a π jump evenly shared by the two Brewster waves (although this is over interpreted as we will see later). By progressively increasing the polar angle of incidence, the direction of the total zero order field changes, and so does the wave vector of the Brewster Snell-conjugate wave (which now is unique). For a polar angle of incidence equal to 35°, as sketched in Fig. 6(b), we can observe that the projection of $\mathbf{k}_2^+(\mathbf{p}_\mathrm{B})$ along $\hat{\mathbf{e}}_1$ still yields an evanescent Brewster wave, but the lateral wave vector is now closer to the propagation limit. This case corresponds to the parameters assumed in obtaining the results in Fig. 2(b) and we can observe that $\rho_{pp}^{(1)}$ vanishes indeed for p_1 just above ω/c, and that the corresponding phase exhibits a π jump. By further increasing the polar angle of incidence the Brewster wave is found in the propagating region as $|\,\mathbf{p}_\mathrm{B}\,| < n_1\,\omega/c$, and its full wave vector can now be represented as following the inner dashed circle. As the polar angle of incidence increases towards the Brewster angle for a planar surface, the wave vector associated with the reflected zero order wave $\mathbf{k}_1^+(\mathbf{p}_0)$, drawn in black in Fig. 6, and the wave vector of the Brewster scattered wave both approach the red dashed line from either sides and cross it at the same angle of incidence, namely the Brewster angle for a planar interface, $\theta_0 = \theta_\mathrm{B}$. Figure 6(c) shows the case where $\theta_0 = 50°$ at a slightly lower angle than the Brewster angle of incidence (approximately equal to 56.3°), i.e. just before the cross-over. When θ_0 is further increased the lateral component of the Brewster wave vector continues to decrease. Figure 6(d) assumes $\theta_0 = 70°$ which corresponds to Fig. 2(c) where we now observe that the Brewster wave is indeed in the propagating region as can be seen both from $\rho_{pp}^{(1)}$ and the extinction of the incoherent component of the MDRC. Note also the π jump in the phase. Finally, as the polar angle of incidence approaches 90°, the virtual wave does *not* approach the vertical direction as one might naively expect. Indeed, the total zero order field does not become oriented along $\hat{\mathbf{e}}_3$ but along the direction given by the critical angle for total internal reflection. The change in the expected Brewster scattering angle $\Theta_B(\theta_0)$ for a range of angles of incidence and for the currently discussed system is visible as blue dashed lines in Fig. 5.

We now repeat the analysis but for an incident wave approaching the surface in the denser medium [$\epsilon_1 = 2.25$, $\epsilon_2 = 1.0$]. For normal incidence, the total zero order field is along $\hat{\mathbf{e}}_1$, and yet again we recover two Brewster waves. However, since now the magnitudes of the Snell-conjugate waves are situated on the inner circle ($\epsilon_2 = 1$), the wave vectors $\mathbf{k}_1^+(\pm\,\mathbf{p}_\mathrm{B})$ correspond to propagating waves in glass, and coincide with the Yoneda threshold. This situation is illustrated in Fig. 7(a) and Fig. 3(a). Due to the presence of two Brewster waves, the phase jump is $\pi/2$ [see Fig. 3(a)]. The coincidence of the Yoneda threshold and the Brewster scattering angle for internal reflection for normal incidence is now explained, and we see that although the two effects are of different nature and decoupled, they occur simultaneously in this case simply as a consequence of the geometry imposed by the dispersion relations. As the polar angle of incidence is increased, only one Brewster wave remains, and the corresponding lateral wave vector shrinks [see Fig. 7(b) and Fig. 3(b)]. The wave vectors of the reflected zero order wave and of the Brewster wave cross each other at the Brewster angle of incidence ($\approx 33.7°$). Now comes an interesting effect which was not present when the wave was incident from the less dense medium. As the polar angle of incidence approaches the critical angle of total internal reflection of the zero order reflected wave, the Snell-conjugate wave vector and that of the Brewster wave approach the vertical direction and reach it for $\theta_0 = \theta_c$, as displayed in Fig. 3(c). Then a sudden transition occurs when θ_0 is increased beyond θ_c. In Fig. 4(a), which shows results for θ_0 just above the critical angle, it seems that the Brewster scattering angle is nowhere to be found. However, the Brewster scattering angle now comes back from the left (backscattering) side, visible in the evanescent region of Fig. 4(c) where the polar angle of incidence is 70°. What happened? The overall behaviour of the phase in Figs. 4(a)–(c) gives us a good indication. We have mentioned earlier that for s-polarized light, when the zero order reflected wave undergoes total internal reflection, the central phase plateau must undergo a Goos-Hänchen shift with θ_0 (in fact it is the whole graph which undergoes the shift). Similarly, the p-polarized zero order reflected wave undergoes a Goos-Hänchen shift and, as a consequence, the two terms in Eq. (32b) are *not* any longer in phase. In the case where $\epsilon_1 < \epsilon_2$, the arguments of the two complex amplitudes in Eq. (32b) are always either in phase or are separated by a phase difference of π. Therefore, as time progresses, the *real* total zero order field keeps a fixed direction. When $\epsilon_1 > \epsilon_2$, the Goos-Hänchen phase shift makes the real total zero order field change direction and turn in the plane of incidence as time progresses (it describes an ellipse). Intuitively, this seems to indicate that the corresponding dipole radiation is not expected to be that of an oscillating dipole anymore but that of a *rotating* dipole. It is therefore understandable that the measurement of a propagating p-polarized wave does not yield any direction of extinction when the radiation is emitted from a rotating dipole. Stated in an equivalent way, $\rho_{pp}^{(0)}$ now draws a lower half circle in the complex plane from 1 to -1 as the angle of incidence is varied from the critical angle to 90°, while it previously stayed on the real line. It follows that $\mathbf{E}_{p,1}^{(0)}(\mathbf{p}_0)$ has a complex amplitude with non-zero imaginary part. It is therefore not possible for a propagating Snell-conjugate wave to satisfy the requirement of Eq. (35) since its p-polarization vector would be real. Hence, in order to satisfy Eq. (35) the polarization vector $\hat{\mathbf{e}}_{p,2}^+(\mathbf{p}_\mathrm{B})$ must itself be complex, and the Snell-conjugate wave is now found in the *evanescent* region of medium 2. This is the reason why the Brewster scattering lateral wave vector seems to disappear at the transition $\theta_0 = \theta_c + \Delta\theta_0$ and then come back from the negative p_1

side as the angle of incidence is increased, which reveals the evanescent nature of the Snell-conjugate wave.

Note that what we have defined as a p-polarized wave, according to the polarization vector $\hat{\mathbf{e}}^{\pm}_{p,j}(\mathbf{p})$ given in Eq. (6d), takes a rather interesting structure when it is evanescent. For an evanescent wave, $\alpha_j(\mathbf{p})$ is pure imaginary and the polarization vector $\hat{\mathbf{e}}^{\pm}_{p,j}(\mathbf{p})$ hence has a real component along $\hat{\mathbf{e}}_3$ and a pure imaginary component along the transverse wave vector direction. This means that the corresponding real electric field is the sum of a wave polarized along $\hat{\mathbf{e}}_3$ and a longitudinal wave (*longitudinal* with respect to the lateral wave vector) dephased by $\pi/2$ radians with respect to the first wave. The resulting field therefore describes an ellipse in the $(\hat{\mathbf{e}}_3, \hat{\mathbf{p}})$-plane.

In-plane transmission — The analysis for the Brewster scattering effect in the transmitted light is similar to that of the reflected light and will not be analyzed in details. One difference worth mentioning, however, is that the Brewster scattering direction is generally found in the backscattering region when the corresponding Brewster scattering for reflection is found in the forward scattering direction. Intuitively, this effect can be related to the emission of an oscillating dipole which yields zero emitted power along the direction of oscillation, hence producing two antipodal zero intensity points when the intensity is mapped onto a sphere. This fact is better illustrated in the next section.

E. Polarization of the radiation of oscillating and rotating dipoles in free space

Before treating the full angular distribution of the light scattered diffusely by a randomly rough surface, we allow ourselves a detour via the analysis of the polarization properties of the radiation emitted by an oscillating dipole or a rotating dipole in free space. The study of the polarization of the radiation in these two cases gives remarkable insight and intuition into the qualitative physical mechanisms at play for the case of the scattering from a random interface, for which a more quantitative analysis requiring Snell-conjugate waves will be given in the next section.

Polarization of the radiation from an oscillating dipole in free space with respect to the local $(\hat{\mathbf{e}}_p, \hat{\mathbf{e}}_s)$ *basis* – We consider first the radiation emitted by a single oscillating dipole in free space. We let this dipole, of dipole moment $\mathbf{D}(\vartheta) = d(\sin\vartheta\,\hat{\mathbf{e}}_1 + \cos\vartheta\,\hat{\mathbf{e}}_3)/2 = d/2\,\hat{\mathbf{e}}_\vartheta$, be tilted from the x_3-axis by an angle of $\vartheta \in [0, \pi/2]$ radians. The dipole is placed in free space at the origin of the coordinate system, where it oscillates with angular frequency ω and radiates the following electric field in the far-field [19]:

$$\mathbf{E}_{\rm dip}(\mathbf{r}, t) = -\frac{\omega^2}{4\pi\varepsilon_0 c^2}\,\frac{\hat{\mathbf{e}}_r \times [\hat{\mathbf{e}}_r \times \mathbf{D}(\vartheta)]}{r}\,e^{-i\omega(t-r/c)}\,, \quad (36)$$

where $\mathbf{r} = r\,\hat{\mathbf{e}}_r = r(\sin\theta\cos\phi\,\hat{\mathbf{e}}_1 + \sin\theta\sin\phi\,\hat{\mathbf{e}}_2 + \cos\theta\,\hat{\mathbf{e}}_3)$ is the point of observation, and $r = |\,\mathbf{r}\,|$. It is well known that no power is radiated along the axis of oscillation of the dipole ($\hat{\mathbf{e}}_r \times \mathbf{D}(\vartheta)$ vanishes in Eq. (36) when $\hat{\mathbf{e}}_r \parallel \hat{\mathbf{e}}_\vartheta$) and that the radiation is polarized in accordance with the cross products in Eq. (36). The electric field is polarized along the vector $\hat{\mathbf{e}}_\theta{}'$, which is the basis vector tangent to a meridian in a spherical coordinate system (r, θ', ϕ') attached to the dipole direction. We are, however, interested in analyzing the polarization of the dipole radiation with respect to the local polarization basis given in Eq. (6), which is defined with respect to the propagation direction of the radiation *and* the plane $x_3 = 0$. Thus we study the following dot products: $\hat{\mathbf{e}}_r \times[\hat{\mathbf{e}}_r \times \hat{\mathbf{e}}_\vartheta] \cdot \hat{\mathbf{e}}_s$ and $\hat{\mathbf{e}}_r \times[\hat{\mathbf{e}}_r \times \hat{\mathbf{e}}_\vartheta] \cdot \hat{\mathbf{e}}_p$ where $\hat{\mathbf{e}}_s = \frac{\hat{\mathbf{e}}_3 \times \hat{\mathbf{e}}_r}{|\,\hat{\mathbf{e}}_3 \times \hat{\mathbf{e}}_r\,|}$ and $\hat{\mathbf{e}}_p = \frac{\hat{\mathbf{e}}_s \times \hat{\mathbf{e}}_r}{|\,\hat{\mathbf{e}}_s \times \hat{\mathbf{e}}_r\,|}$ are defined with respect to $\hat{\mathbf{e}}_r$ in order to mimic the local s- and p-polarization vectors attached to a scattering direction along $\hat{\mathbf{e}}_r$. The unit vectors $\hat{\mathbf{e}}_p = \hat{\mathbf{e}}_\theta = \mathrm{d}\,\hat{\mathbf{e}}_r\,/\mathrm{d}\theta$ and $\hat{\mathbf{e}}_s = \hat{\mathbf{e}}_\phi = 1/\sin\theta\,\mathrm{d}\,\hat{\mathbf{e}}_r\,/\mathrm{d}\phi$ are also the conventional basis vectors in spherical coordinates. First we observe that $\hat{\mathbf{e}}_r \times[\hat{\mathbf{e}}_r \times \hat{\mathbf{e}}_\vartheta] \cdot \hat{\mathbf{e}}_s$ and $\hat{\mathbf{e}}_r \times[\hat{\mathbf{e}}_r \times \hat{\mathbf{e}}_\vartheta] \cdot \hat{\mathbf{e}}_p$ are invariant under the transformation $\hat{\mathbf{e}}_r \mapsto -\,\hat{\mathbf{e}}_r$, and so the s- and p-polarized distributions of the dipole radiation are symmetric with respect to the origin as $\hat{\mathbf{e}}_r$ runs over the unit sphere. Second, for $\vartheta \in (0, \pi/2]$ radians the identity $\mathbf{a} \times[\mathbf{b} \times \mathbf{c}] = (\mathbf{a}\cdot\mathbf{c})\,\mathbf{b} - (\mathbf{a}\cdot\mathbf{b})\,\mathbf{c}$ leads to

$$\hat{\mathbf{e}}_r \times[\hat{\mathbf{e}}_r \times \hat{\mathbf{e}}_\vartheta] = (\hat{\mathbf{e}}_r \cdot \hat{\mathbf{e}}_\vartheta)\,\hat{\mathbf{e}}_r - \hat{\mathbf{e}}_\vartheta\,, \quad (37)$$

hence the projection of the dipole radiation on the local s-polarization basis reads

$$\hat{\mathbf{e}}_r \times[\hat{\mathbf{e}}_r \times \hat{\mathbf{e}}_\vartheta] \cdot \hat{\mathbf{e}}_s = -\,\hat{\mathbf{e}}_\vartheta \cdot \hat{\mathbf{e}}_\phi = -\sin\vartheta\sin\phi\,. \quad (38)$$

A direct consequence of Eq. (38) is that $\hat{\mathbf{e}}_r \times[\hat{\mathbf{e}}_r \times \hat{\mathbf{e}}_\vartheta] \cdot \hat{\mathbf{e}}_s$ vanishes for all $\hat{\mathbf{e}}_r$ in the $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_3)$-plane [see Fig. 8(d)]. The corresponding projection on the local p-polarization basis reads

$$\hat{\mathbf{e}}_r \times[\hat{\mathbf{e}}_r \times \hat{\mathbf{e}}_\vartheta] \cdot \hat{\mathbf{e}}_p = -\,\hat{\mathbf{e}}_\vartheta \cdot \hat{\mathbf{e}}_\theta\,, \quad (39)$$

which is a quantity that depends on ϑ, θ, and ϕ. In the particular case where $\hat{\mathbf{e}}_\theta$ belongs to the $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_3)$-plane, there are two solutions for Eq.(39) equal to zero: $\hat{\mathbf{e}}_r = \pm\,\hat{\mathbf{e}}_\vartheta$, which correspond to the two intersections of the dipole moment direction with the unit sphere. This is not surprising since we already know that no power is emitted along the direction of oscillation of the dipole, independent of polarization. More interesting are cases for which $\hat{\mathbf{e}}_\theta$, and hence $\hat{\mathbf{e}}_r$, does not belong to the $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_3)$-plane. Expanding the dot product in Eq. (39) in terms of the angles ϑ, θ, and ϕ we obtain the following implicit equation for the set of points on the unit sphere where the p-polarization component of the dipole radiation vanishes:

$$\sin\vartheta\,\cos\theta\,\cos\phi - \cos\vartheta\,\sin\theta = 0\,, \quad (40)$$

or equivalently for non-pathologic cases

$$\frac{\tan\vartheta}{\tan\theta} = \frac{1}{\cos\phi}\,. \quad (41)$$

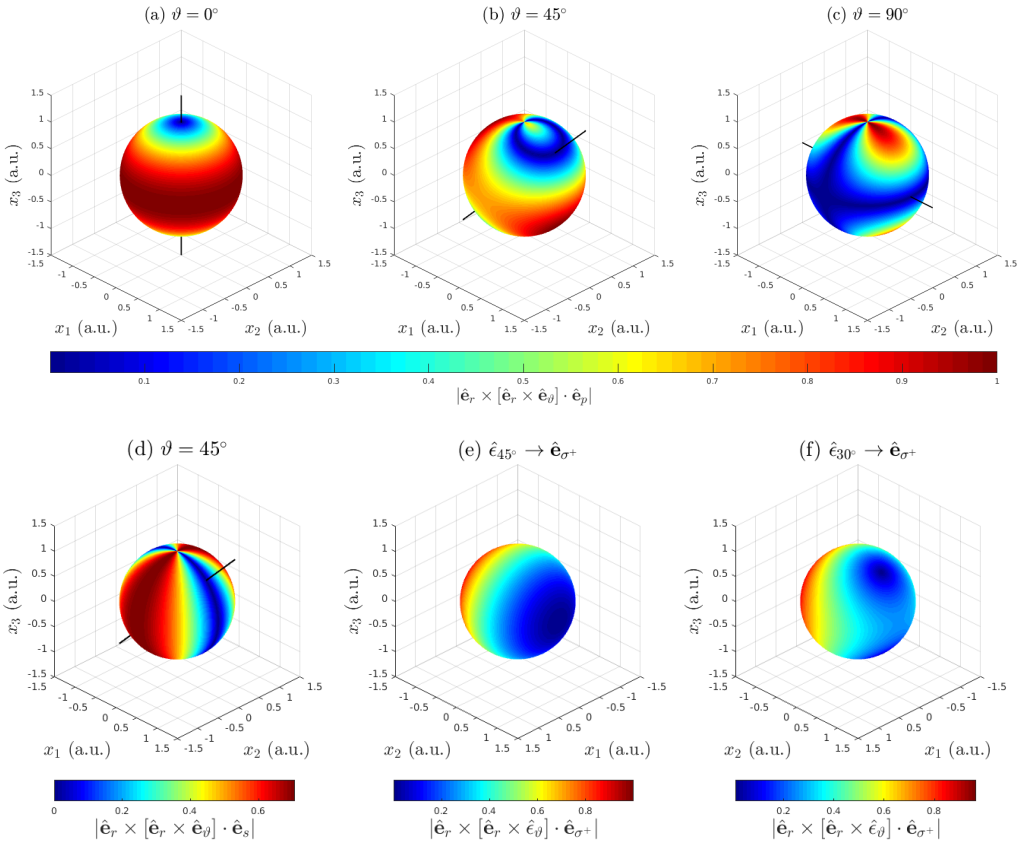


FIG. 8. (a-c) Dependence of the p-polarized radiation of a tilted dipole in free space, $|\,\hat{\mathbf{e}}_r \times [\hat{\mathbf{e}}_r \times \hat{\mathbf{e}}_\vartheta] \cdot \hat{\mathbf{e}}_p\,|$, on the direction of $\hat{\mathbf{e}}_r$ as it runs over the unit sphere for different dipole tilting angles $\vartheta \in \{0°, 45°, 90°\}$. (d) Similar dependence of the s-polarized radiation of a tilted dipole in free space on $\hat{\mathbf{e}}_r$ for $\vartheta = 45°$. The black line in panels (a-d) indicates the direction of the dipole moment. (e-f) Dependence of the σ^+-polarized radiation of a rotating dipole in free space parametrized by $\vartheta = 45°$ and $\vartheta = 30°$ respectively (note the orientation of the coordinate system).

We verify that for the cases $\phi = 0$ and $\phi = \pi$ radians, we recover that $\theta = \vartheta$ and $\theta = \pi - \vartheta$, i.e. the points of intersection of the dipole moment direction and the unit sphere. For $\phi \in (-\pi/2, \pi/2)$, $\cos\phi > 0$, which implies that $\tan\theta > 0$ (recall that $0 < \vartheta < \pi/2$ hence $\tan\vartheta > 0$) and $\tan\vartheta > \tan\theta$. By the monotony of the tangent function, and the continuity of Eq. (41) with respect to the variables, we thus deduce that when ϕ varies in $(-\pi/2, \pi/2)$ the set of the points of zero traces a curve on the unit sphere latitude-bounded by $\theta < \vartheta$. By the aforementioned symmetry of the polarization dependence of the dipole radiation with respect to the origin we immediately deduce that when ϕ varies in $(\pi/2, 3\pi/2)$ the set of the points of zero traces a curve on the unit sphere latitude-bounded by $\theta > \pi - \vartheta$. This is well illustrated in Fig. 8(b) where $|\,\hat{\mathbf{e}}_r \times [\hat{\mathbf{e}}_r \times \hat{\mathbf{e}}_\vartheta] \cdot \hat{\mathbf{e}}_p\,|$ is shown as a color map on a unit sphere. For $\vartheta = 45°$ we here observe that the curve of zero p-polarized radiation passes through both the north pole and the intersection point of the dipole moment direction on the northern hemisphere. The degenerate cases $\vartheta = 0°$ [Fig. 8(a)] and $\vartheta = 90°$ [Fig. 8(c)] are also illustrated. For these cases the curves of zero p-polarized radiation reduces to two points (the poles) in the former case, and the equator ($\theta = \pi/2$) and meridians $\phi = \pm\pi/2$ in the latter. Indeed, θ must go to zero when $\vartheta \to 0$ as $\tan\vartheta$ vanishes, and, *either* ϕ must go towards $\pm\pi/2$ *or* θ must go towards $\pi/2$ when $\vartheta \to \pi/2$ as $\tan\vartheta$ diverges.

Polarization of the radiation from a rotating dipole in free space with respect to the local $(\hat{\mathbf{e}}_{\sigma^+}, \hat{\mathbf{e}}_{\sigma^-})$ *basis* – We now consider the radiation of a dipole rotating in the $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_3)$-plane. Equation (36) still holds, but we need to

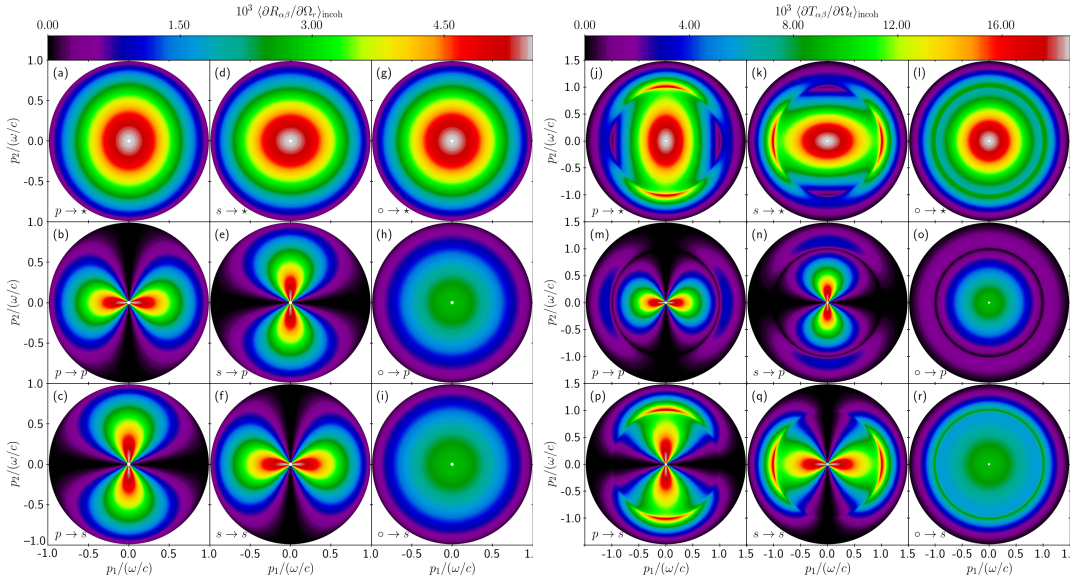


FIG. 9. The full angular distribution of the incoherent component of the MDRC/MDTC, $\langle \partial X_{\alpha\beta}/\partial\Omega_r \rangle_{\rm incoh}$ for $X = R$ or T, as function of the lateral wave vector $\mathbf{q}$ of the light that is scattered from a rough interface where the angle of incidence $\theta_0 = 0°$. The positions of the specular directions in reflection and transmission are indicated by white dots. The parameters assumed for the scattering geometry and used in performing the numerical calculation had values that are identical to those assumed in obtaining the results of Fig. 2. The sub-figures in Figs. 9(a)–(i) and 9(j)–(r) are both organized in the same manner and show how incident β-polarized light is scattered by the one-rough-interface film geometry into α-polarized light [with $\alpha = p, s$ and $\beta = p, s$] and denoted $\beta \to \alpha$. Moreover, the notation $\circ \to \star$ is taken to mean that the incident light was unpolarized while the polarization of the scattered light was not recorded. For instance, this means that the data shown in Fig. 9(a) are obtained by adding the data sets presented in Figs. 9(b)–(c); similarly, the data shown in Fig. 9(g) result from the addition and division by a factor two of the the data sets presented in Figs. 9(a) and 9(d); *etc.* Finally, the in-plane intensity variations from Figs. 9(b, f) and 9(k, o) are the curves depicted in Figs. 2(a) and Figs. 2(d), respectively.

modify the dipole moment which now reads

$$\mathbf{D}(\vartheta) = d\,(\sin\vartheta\ \hat{\mathbf{e}}_1 + i\cos\vartheta\ \hat{\mathbf{e}}_3)/2 = d/2\,\hat{\boldsymbol{\epsilon}}_\vartheta\ . \tag{42}$$

The real vector $\mathrm{Re}[\hat{\boldsymbol{\epsilon}}_\vartheta \exp(-i\omega t)]$ hence describes an ellipse in the $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_3)$-plane whose excentricity is parametrized by ϑ. In the limiting cases $\vartheta = 0$ and $\vartheta = \pi/2$ radians we obtain an oscillating dipole along $\hat{\mathbf{e}}_3$ and $\hat{\mathbf{e}}_1$ respectively. For $\vartheta = \pi/4$ we obtain a circularly rotating dipole. We now consider the polarization of the radiation from such an elliptically rotating dipole with respect to the local left and right circularly polarized basis $\hat{\mathbf{e}}_{\sigma^+}$ and $\hat{\mathbf{e}}_{\sigma^-}$ defined as

$$\hat{\mathbf{e}}_{\sigma^\pm} = \frac{1}{\sqrt{2}}\,(\hat{\mathbf{e}}_p \pm i\,\hat{\mathbf{e}}_s)\,. \tag{43}$$

The σ^+ polarization component of the rotating dipole radiation is then measured by

$$\hat{\mathbf{e}}_r \times[\hat{\mathbf{e}}_r \times\hat{\boldsymbol{\epsilon}}_\vartheta]\cdot\hat{\mathbf{e}}_{\sigma^+} = -\hat{\boldsymbol{\epsilon}}_\vartheta\cdot\hat{\mathbf{e}}_{\sigma^+}\ , \tag{44}$$

which when expressed in terms of the angles reads[40]

$$\hat{\mathbf{e}}_r \times[\hat{\mathbf{e}}_r \times\hat{\boldsymbol{\epsilon}}_\vartheta]\cdot\hat{\mathbf{e}}_{\sigma^+} = -\frac{1}{\sqrt{2}}\ \sin\vartheta\cos\theta\cos\phi - \frac{i}{\sqrt{2}}\,(\cos\vartheta\sin\theta - \sin\vartheta\sin\phi)\,. \tag{45}$$

The modulus square of Eq. (45) yields

$$|\,\hat{\mathbf{e}}_r \times[\hat{\mathbf{e}}_r \times\hat{\boldsymbol{\epsilon}}_\vartheta]\cdot\hat{\mathbf{e}}_{\sigma^+}\,|^2 = \frac{1}{2}\ \sin^2\vartheta\cos^2\theta\cos^2\phi + \frac{1}{2}\,(\cos\vartheta\sin\theta - \sin\vartheta\sin\phi)^2\,. \tag{46}$$

The directions of zero σ^+-polarized light radiation are obtained if and only if both terms on the right-hand side of Eq. (46) are zero. The first term vanishes if at least $\sin\vartheta$, $\cos\theta$ or $\sin\phi$ is zero. If we first assume that $\vartheta = 0$, then the second term is zero if and only if the condition $\sin\theta = 0$ is satisfied. Such a case corresponds to a dipole oscillating along the x_3-axis and its radiation vanishes at the poles of the unit sphere. More interesting are

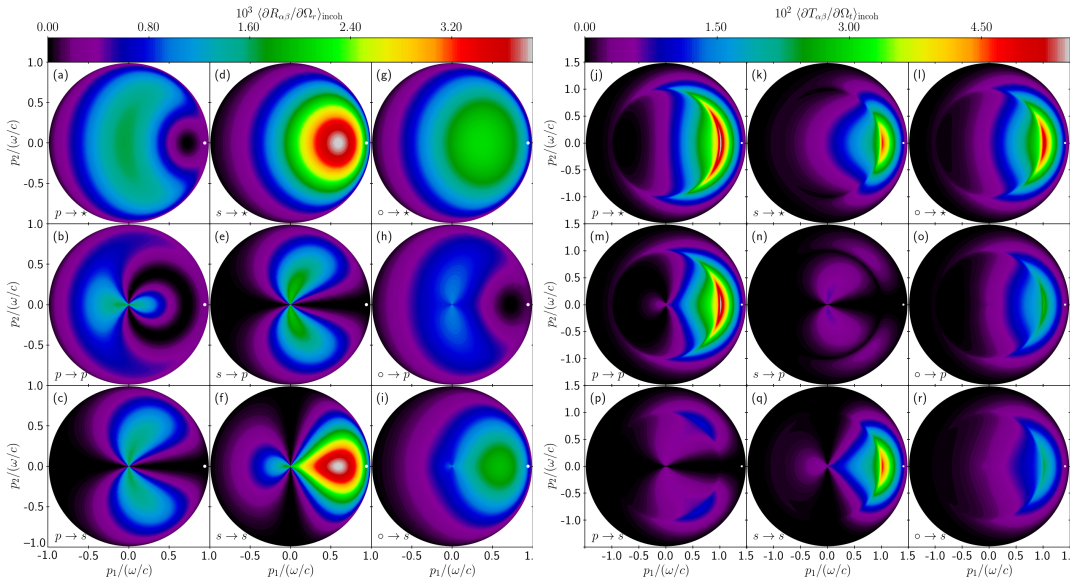


FIG. 10. Same as Fig. 9, but now for the angle of incidence $\theta_0 = 70°$.

the cases for which $\vartheta \neq 0$ and either $\theta = \pi/2$ (recall that $\theta \in (0, \pi)$) or $\phi = \pm\pi/2$. Let us first assume that $\phi = \pm\pi/2$. The second term in Eq. (46) then vanishes if and only if

$$\sin\theta = \pm \tan\vartheta \,. \tag{47}$$

This last condition imposes a constraint on ϑ, which must then have a value between 0 and $\pi/4$ in order for $\tan\vartheta$ (and hence $\sin\theta$) to be less than unity. Since $\theta \in (0, \pi)$, only the case $\phi = \pi/2$ yields two solutions, θ_1 and θ_2, that are symmetric with respect to $\theta = \pi/2$. This is illustrated in Fig. 8(f). Assuming now that $\theta = \pi/2$, the second term in Eq. (46) vanishes if and only if

$$\sin\phi = \operatorname{cotan}\vartheta \,. \tag{48}$$

Since $\sin\phi$ requires $\operatorname{cotan}\vartheta$ to be less than unity the above condition can only be satisfied if $\vartheta \in (\pi/4, \pi/2)$. There are then two solutions for ϕ between 0 and π (since $\sin\phi > 0$ for $\vartheta \in (\pi/4, \pi/2)$), which are symmetric with respect to $\pi/2$. In fact, it can be shown that the polarization of the radiation of the rotating dipole for $\vartheta \in (\pi/4, \pi/2)$ corresponds to that of a rotating dipole for which $\vartheta' = \pi/2 - \vartheta$ (as in Fig. 8(f)) but rotated by 90° with respect to the x_2-axis. These different cases where Eq. (46) vanishes for a given circular polarization have a very simple geometrical interpretation. For $\vartheta < \pi/4$ the rotating dipole describes an ellipse whose long axis is oriented along the x_3-axis. The two directions of zero σ^+-polarized radiation correspond to the two directions from which the ellipse is observed as a circle, with the orientation of the dipole rotation opposite to that of the σ^+ polarization. For these two directions the radiation is therefore purely σ^--polarized, explaining why the zeros of radiation are found on the meridian where $\phi = \pi/2$. For $\vartheta > \pi/4$ the long axis of the ellipse is along the x_1-direction, which explains the fact that the directions where one circular polarization is zero are found at the equator. By symmetry the directions where the σ^--polarized radiation vanishes are symmetric to those of the σ^+-polarized radiation with respect to the $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_3)$-plane.

These results, obtained for the polarization of the dipole radiation in free space, will prove to be useful for the qualitative understanding of the full angular distribution of the incoherent component of the MDRC and MDTC in the case of the scattering by a randomly rough dielectric surface.

F. Full angular distributions of the MDRC/MDTC

P-polarized Brewster scattering — Figure 9 presents the full angular distributions of the diffuse contribution to the MDRC and MDTC for $\theta_0 = 0°$ and parameters equivalent to those assumed in Figs. 2(a) and 2(d), respectively. The overall dipole-like appearance of the lower left 2×2 panels in each collection of panels in Fig. 9 is reminiscent of the polarization pattern of the

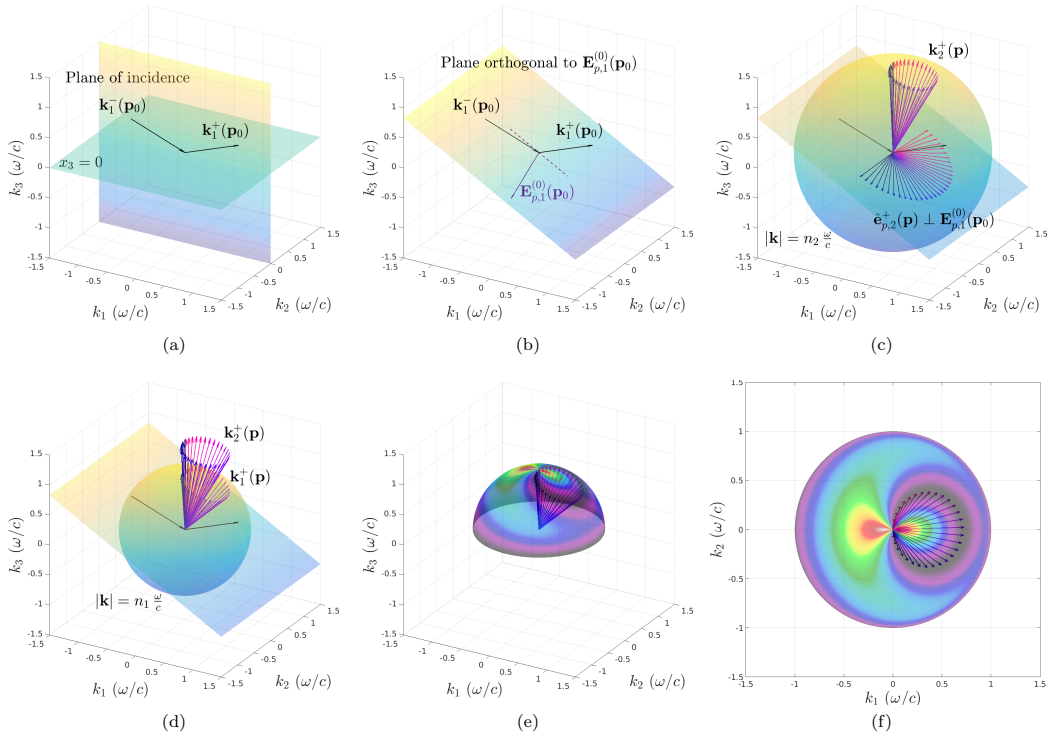


FIG. 11. Illustration of the construction steps leading to the set of directions of zero $p \to p$ reflection for an angle of incidence of $\theta_0 = 70°$. (a) Sketch of the average surface, the plane of incidence and the considered wave vectors of the incident and reflected zero order waves ($\mathbf{k}_1^-(\mathbf{p}_0)$ and $\mathbf{k}_1^-(\mathbf{p}_0)$). (b) Construction of the total zero order field amplitude $\mathbf{E}_{p,1}^{(0)}(\mathbf{p}_0)$ and the plane orthogonal to it. Note that the incident wave vector does not in general belong in this plane as illustrated with the dashed indigo line indicating the intersection of the plane of incidence and the plane $\mathbf{E}_{p,1}^{(0)}(\mathbf{p}_0)^\perp$. (c) Unit vectors belonging to the lower half $\mathbf{E}_{p,1}^{(0)}(\mathbf{p}_0)^\perp$-plane. They correspond to the possible polarization vectors $\hat{\mathbf{e}}_{p,2}^+(\mathbf{p})$ of Eq. (35). The wave vectors $\mathbf{k}_2^+(\mathbf{p})$ associated to the polarization vectors $\hat{\mathbf{e}}_{p,2}^+(\mathbf{p})$ are then constructed according to Eqs. (6). Note that they lie on a sphere of radius $|\,\mathbf{k}\,| = n_2\,\omega/c$. The color associated to the vectors $\hat{\mathbf{e}}_{p,2}^+(\mathbf{p})$ and $\mathbf{k}_2^+(\mathbf{p})$ helps us to identify the $\mathbf{k}_2^+(\mathbf{p})$ associated to each $\hat{\mathbf{e}}_{p,2}^+(\mathbf{p})$ (they share the same color). (d) The wave vectors $\mathbf{k}_2^+(\mathbf{p})$ are projected on the sphere of radius $|\,\mathbf{k}\,| = n_1\,\omega/c$ following the x_3-direction, hence giving the wave vectors $\mathbf{k}_1^+(\mathbf{p})$ of zero $p \to p$ reflection. (e) The incoherent component of the MDRC is shown on the scattering sphere together with the set of wave vectors $\mathbf{k}_1^+(\mathbf{p})$ obtained in (d). (f) Projection of (e) in the $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2)$-plane. We verify in (e) and (f) that the constructed wave vectors indeed follow the curve of zero scattering of the incoherent component of the MDRC.

dipole radiation in free space discussed above in the case when the dipole oscillates in the $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2)$-plane. For normal incidence all the zero order waves and the incident wave have an oscillating electric field either along $\hat{\mathbf{e}}_1$ for p polarization or along $\hat{\mathbf{e}}_2$ for s polarization. Thus the dipoles in the media[41] oscillate along the direction of the incident field. For an s-polarized wave (field along $\hat{\mathbf{e}}_2$) we have seen that the dipole radiation in free space yields zero s-polarized emission in the $(\hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3)$-plane and an overall $|\sin(\phi_r - \pi/2)|$ intensity, which is consistent with what is observed in Fig. 9(f). Note that for a given polar angle of reflection θ_r, the variation along ϕ_r of the incoherent component of the MDRC to lowest non-zero order in the surface roughness for $s \to s$ polarized scattering is *exactly* proportional to $|\sin(\phi_r - \pi/2)|$ since $\rho_{ss}^{(1)}$ is proportional to $\hat{\mathbf{e}}_s(\mathbf{p}) \cdot \hat{\mathbf{e}}_s(\mathbf{p}_0)$, as can be seen from Eq. (30a), and this is the only ϕ_r dependence for normal incidence. This observation holds for all the polarization couplings up to a rotation by $\pi/2$ for cross-polarization. For example, for an s-polarized incident field and p-polarized reflected light the ϕ_r dependence is proportional to $|\sin\phi_r|$. For the transmitted light [Fig. 9(j-r)] the behaviour is similar, but in addition we now observe the Yoneda

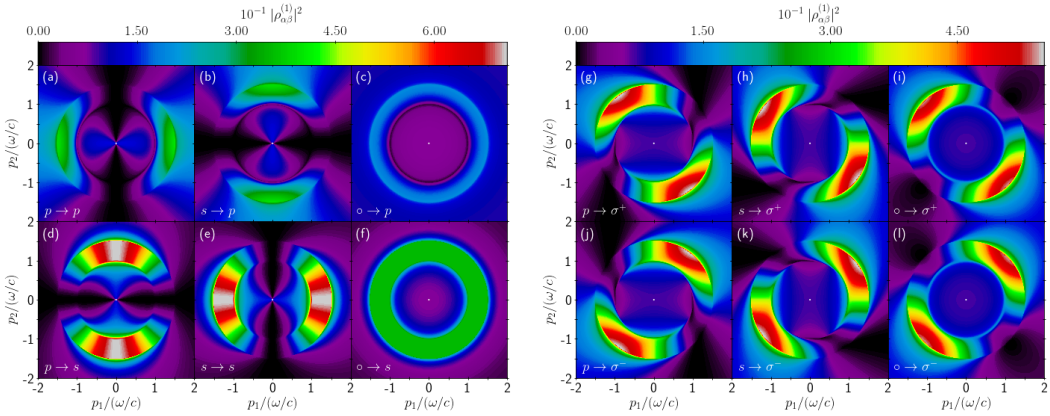


FIG. 12. The full angular distribution of $|\rho^{(1)}_{\alpha\beta}|^2$ for normal incidence, $\theta_0 = 0°$, $\epsilon_1 = 2.25$, $\epsilon_2 = 1$, for incident polarization $\beta \in \{p, s\}$ or unpolarized ($\circ$) and outgoing polarization $\alpha \in \{p, s, \sigma^+, \sigma^-\}$.

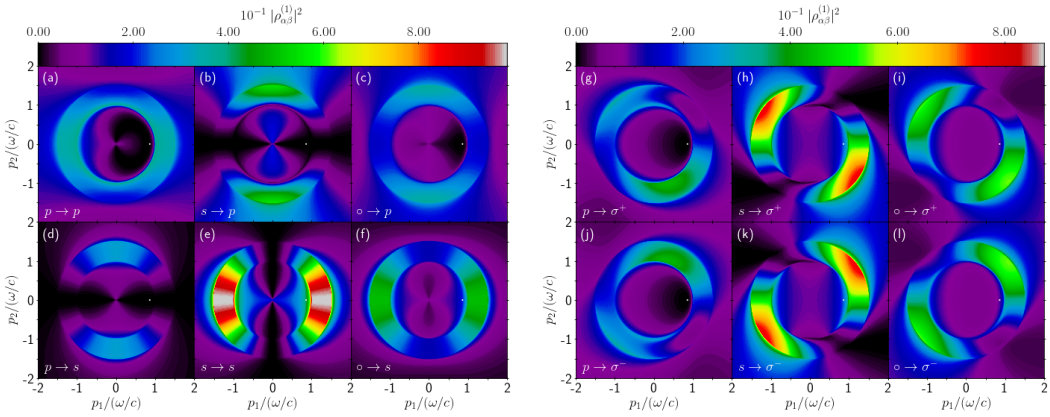


FIG. 13. Same as Fig. 12 but for the angle of incidence $\theta_0 = 35°$.

phenomenon. This is the enhancement of the diffuse contribution to the MDTC intensity above the critical lateral wave vector for the scattered light $|\,\mathbf{p}\,| > p_c$, as discussed extensively in Sec. IV C, which for normal incidence is directly observable for outgoing s-polarized light, especially in Fig. 9(r). For transmitted p-polarized light we observe a black ring of zero scattering intensity along the circle $|\,\mathbf{p}\,| = p_c$. This is the two-dimensional extension of our discussion in Sec. IV D for in-plane scattering, where we found that at normal incidence two Brewster waves with $\mathbf{p}_B = \pm p_c\ \hat{\mathbf{e}}_1$ could be found. Now we see that in two-dimensional $\mathbf{p}$-space the solution to Eq. (35) is in fact given by $|\,\mathbf{p}\,| = p_c$. In terms of dipole radiation in free space this corresponds to the vanishing radiation of p-polarized light in the equatorial plane for the case $\vartheta = 90°$ as illustrated in Fig. 8(c). A ring of zero intensity for the p-polarized *reflected* waves can also be found, but then in the evanescent regime as a two-dimensional extension of the corresponding discussion for in-plane scattering.

The similitude with the polarization of the radiation emitted by an oscillating dipole in free space is clear for normal incidence. Let us now consider a larger angle of incidence, $\theta_0 = 70°$, for which the diffuse contributions to the MDRC and MDTC for incidence in vacuum are shown in Fig. 10. First, we observe that for $p \to p$ reflection [Fig. 10(b)], there exists a closed curve of zero intensity in the forward scattering direction. Similarly, we observe a closed curve of zero intensity for $p \to p$ transmission [Fig. 10(k)] but in the backscattering region. These features are analogous to those observed in

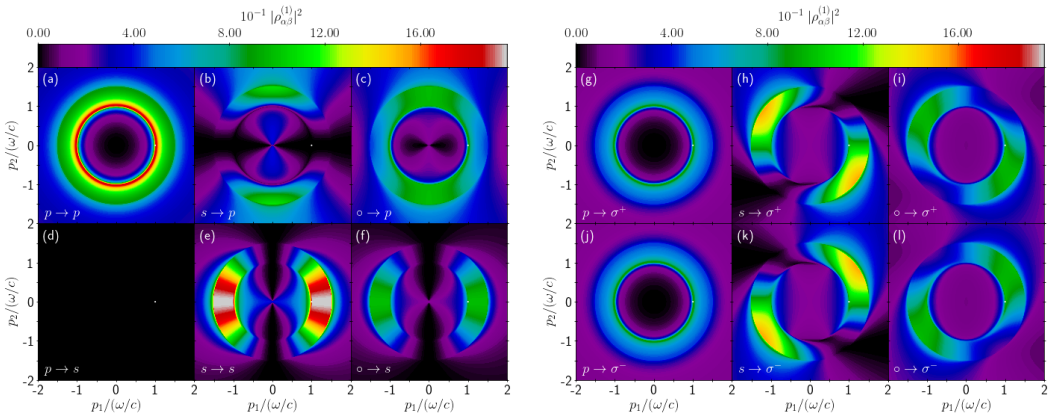


FIG. 14. Same as Fig. 12 but for the angle of incidence equal to the critical angle for total internal reflection, $\theta_0 = \theta_c = 41.81°$.

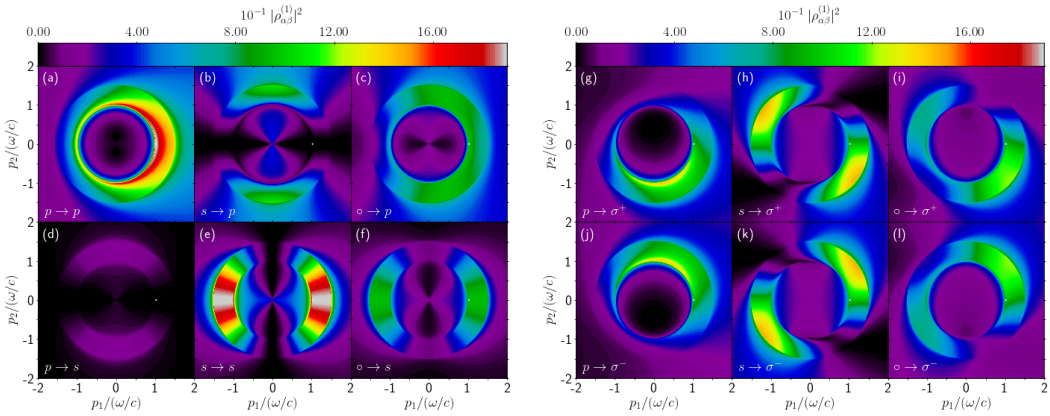


FIG. 15. Same as Fig. 12 but for the angle of incidence $\theta_0 = 43°$.

the case of the p polarization component of the dipole radiation in free space in the case where the dipole tilting angle is such that $0° < \vartheta < 90°$, e.g. as is illustrated in Fig. 8(b) for $\vartheta = 45°$. We can interpret the curves of zero intensity for $p \to p$ scattering in Fig. 10 as the signature of an overall dipole radiation whose dipole moment is tilted from the x_3-axis by some angle ϑ, where the polarization of the reflected light is derived from the northern hemisphere of the radiation polarization pattern while the polarization of the transmitted light is derived from the southern hemisphere of the radiation polarization pattern.

Let us now interpret Eq. (35) geometrically for $p \to p$ scattering for the case of reflection and $\theta_0 = 70°$. This construction is a generalization of the one made for scattering in the plane of incidence presented in Sec. IV D. Figure 11 provides illustrations of the main steps of the geometrical construction of the set of directions of zero $p \to p$ reflection in three dimensions. First, the wave vectors of incidence $\mathbf{k}_1^-(\mathbf{p}_0)$ and of the reflected zero order wave $\mathbf{k}_1^+(\mathbf{p}_0)$ are drawn [Fig. 11(a)]. Second, one determines the direction of the total zero order field amplitude $\mathbf{E}^{(0)}_{p,1}(\mathbf{p}_0)$ which is contained in the plane of incidence. The steps to geometrically construct the total zero order field have been treated in detail for s- and p-polarizations in Ref. 42, and thus we do not show these here for clarity. Once $\mathbf{E}^{(0)}_{p,1}(\mathbf{p}_0)$ is determined, we can construct the plane orthogonal to it: $\mathbf{E}^{(0)}_{p,1}(\mathbf{p}_0)^\perp$ [Fig. 11(b)]. Note that in general this plane does *not* contain the incident wave vector as made clear by the dashed line, showing the intersection of the plane of incidence with the plane $\mathbf{E}^{(0)}_{p,1}(\mathbf{p}_0)^\perp$. According to Eq. (35), all the polarization vectors $\hat{\mathbf{e}}^+_{p,2}(\mathbf{p})$ must be contained in the plane

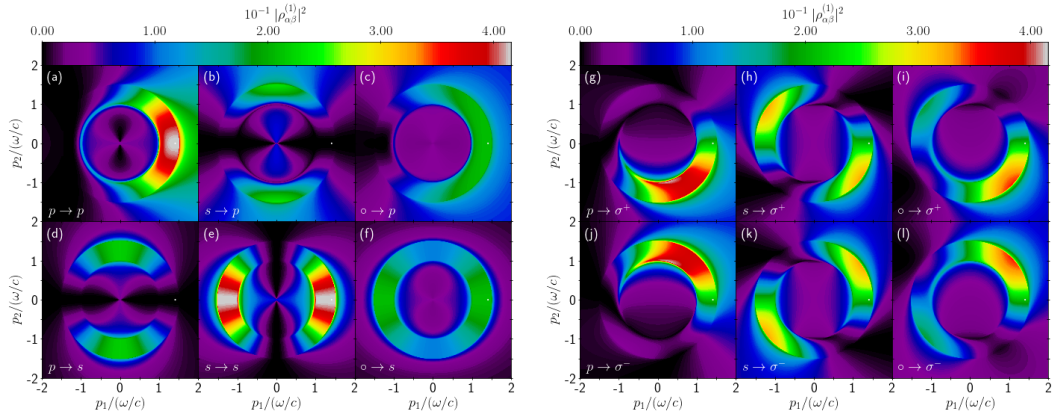


FIG. 16. Same as Fig. 12 but for the angle of incidence $\theta_0 = 70°$.

$\mathbf{E}^{(0)}_{p,1}(\mathbf{p}_0)^\perp$. Moreover, since the $\hat{\mathbf{e}}^+_{p,2}(\mathbf{p})$ vectors are normalized their end points are distributed on a circle of unit radius. The set of all $\hat{\mathbf{e}}^+_{p,2}(\mathbf{p})$ vectors satisfying Eq. (35) therefore spans a half circle in the plane $\mathbf{E}^{(0)}_{p,1}(\mathbf{p}_0)^\perp$ as shown on Fig. 11(c), where a sample of polarization vectors are represented. The fact that only the lower half circle is needed comes from the definition[43] of a polarization vector $\hat{\mathbf{e}}^+_{p,2}(\mathbf{p})$. For each polarization vector satisfying Eq. (35) we can construct its corresponding wave vector $\mathbf{k}^+_2(\mathbf{p})$, using for example that the direction is given by $\hat{\mathbf{e}}^+_{p,2}(\mathbf{p}) \times [\hat{\mathbf{e}}^+_{p,2}(\mathbf{p}) \times \hat{\mathbf{e}}_3]$ and that $\mathbf{k}^+_2(\mathbf{p})$ lies on the northern hemisphere of radius $k_2 = n_2\omega/c$. We thus obtain the set of all wave vectors $\mathbf{k}^+_2(\mathbf{p})$ whose corresponding p polarization vector satisfies Eq. (35). A sample of such vectors are represented for $\hat{\mathbf{e}}^+_{p,2}(\mathbf{p})$ and $\mathbf{k}^+_2(\mathbf{p})$ in Fig. 11(c). The last step consists in projecting the vectors $\mathbf{k}^+_2(\mathbf{p})$ along $\hat{\mathbf{e}}_3$ onto the sphere of radius $|\,\mathbf{k}\,| = n_1\,\omega/c$ to obtain the wave vectors $\mathbf{k}^+_1(\mathbf{p})$ of zero $p \to p$ reflection [Fig. 11(d)]. Figures 11(e) and 11(f) show the resulting sampled wave vectors $\mathbf{k}^+_1(\mathbf{p})$ together with the diffuse contribution to the the MDRC, mapped to the hemisphere and its projection in the $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2)$-plane respectively. We verify that the set of constructed wave vectors correspond to the observed curve of zero intensity for $p \to p$ reflection.

Figure 10(n) shows that the $s \to p$ transmitted light exhibits a circle of zero intensity, for $|\,\mathbf{p}\,| = n_1\,\omega/c$ similar to what was observed for normal incidence [Fig. 9(n)]. This feature is also present in reflection but in the evanescent region, and is observed by considering the complex amplitude instead of the MDRC. The reason for the invariance of the circle of zero intensity with the angle of incidence for the $s \to p$ scattering is simple to understand in terms of the dipole radiation in free space. For s-polarized incident light the dipoles in the media are all oriented along $\hat{\mathbf{e}}_2$, independent of the angle of incidence. Thus when measuring the p-polarized component of the radiated light we expect to obtain an underlying pattern of zero intensity consistent with that obtained in the case of the oscillating dipole in free space as illustrated in Fig. 8(c).

Circularly-polarized Brewster scattering — It is instructive to study the modulus square of the amplitudes rather than the MDRC and MDTC in order to appreciate the behavior of the amplitudes of the waves scattered in the evanescent region as well as the ones scattered in the propagating region. Furthermore, in order to illustrate, to our knowledge, a new effect which can be considered as a generalization of the Brewster scattering effect for light scattered from p-polarized to circularly-polarized light, we show in Figs. 12-16 $|\rho^{(1)}_{\alpha\beta}|^2$ in the $\mathbf{p}$-plane for different polar angles of incidence. We let $\beta \in \{p, s, \circ\}$ represent the polarization of the incident light, where $\circ$ indicates unpolarized light, and we let $\alpha \in \{p, s, \sigma^+, \sigma^-\}$ represent the polarization of the light scattered from the surface. The subscripts $\sigma^\pm$ denote respectively left and right circular polarization states and the corresponding reflection amplitudes are derived from the p and s polarization states by

$$\rho^{(1)}_{\sigma^\pm\beta} = \frac{1}{\sqrt{2}}\left[\rho^{(1)}_{p\beta} \pm i\rho^{(1)}_{s\beta}\right], \tag{49}$$

and similarly for the transmission amplitudes. We consider here only the case for which the medium of incidence is the denser one, as the circularly-polarized Brewster scattering effect only takes place when the reflected zero order wave undergoes total internal reflection. Note, however, that the effect can be observed both in the reflected *and* the transmitted scattered light. In Sec. IV D we have seen that the Brewster scattering effect exhibits a sudden transition when the reflected zero order wave undergoes total internal reflection. We have seen that, when restricted to scattering in the plane of incidence,

the direction of zero p-polarized reflected intensity goes towards the x_3-direction as the polar angle of incidence approaches the critical angle for total internal reflection. Then the zero direction suddenly disappears from the propagating region as the polar angle of incidence goes beyond the critical angle for total internal reflection. This sudden transition was argued to be attributed to a transition of the dipolar response of the media, going from an oscillating behavior to a rotating behavior due to the phase shift between the incident excitation and the reflected zero order wave. We are now studying this transition in the full **p**-plane with particular attention on the scattered *circularly polarized* light, as it was shown in Sec. IV E that the radiation emitted by a rotating dipole in free space exhibits characteristic signatures in the emitted circularly-polarized light out of the plane of incidence.

First, for polar angles of incidence smaller than the critical angle, $\theta_0 < \theta_c$, we have seen that both $\rho^{(1)}_{p\beta}$ and $\rho^{(1)}_{s\beta}$ are real for scattering angles smaller than the Yoneda threshold. In that case the right-hand side in Eq. (49) vanishes if and only if both $\rho^{(1)}_{p\beta}$ and $\rho^{(1)}_{s\beta}$ are zero simultaneously. For an incident p-polarized wave, $\beta = p$, this occurs only where the curve of zero $p \to p$ scattering (cf. previous subsection) intersects with the plane of incidence in which $p \to s$ scattering is identically zero. This is illustrated for normal incidence, $\theta_0 = 0°$, in Figs. 12(a) and (d) showing $|\rho^{(1)}_{pp}|^2$ and $|\rho^{(1)}_{sp}|^2$ in the **p**-plane, where we recognize the curves of zero scattering for the p- and s-polarized light discussed in previous sections. It is also illustrated in Figs. 12(g) and (j) showing $|\rho^{(1)}_{\sigma^\pm p}|^2$ where two directions of zero $p \to \sigma^\pm$ scattering are present at $\mathbf{p} = \pm p_c\,\hat{\mathbf{e}}_1$, although they are hard to spot on this figure. The effect is clearer for oblique incidence, as in Figs. 13(g) and (j), for which $\theta_0 = 35°$. Figures 13(g) and (j) show a clear unique direction of zero intensity in $p \to \sigma^\pm$ scattering in the plane of incidence.

As the angle of incidence reaches the critical angle of incidence, $\theta_0 = \theta_c = 41.81°$, the direction of zero intensity in $p \to \sigma^\pm$ scattering reaches the x_3-direction, as illustrated in Figs. 14(g) and (j). Note that the x_3-direction also implies zero $p \to p$ scattering intensity as already explained earlier, and that the distribution of $|\rho^{(1)}_{pp}|^2$ and $|\rho^{(1)}_{\sigma^\pm p}|^2$ are cylindrical symmetric as shown in Fig. 14(a), (g) and (j). The cylindrical symmetry can be understood based on the radiation of an oscillating dipole aligned with the x_3-axis. Indeed, we have seen in Section IV E that the p-polarized radiation from such a dipole is cylindrically symmetric with zero radiation at the poles of the unit sphere. The radiation from such a dipole is also purely p-polarized, which has two consequences: (i) the s-polarized scattered light vanishes identically for *all* **p** [Fig. 14(d)]; (ii) the radiation can be decomposed into σ^+ and σ^- components of *equal* intensity, as can be observed in Figs. 14(g) and (j). Even though we have now based our interpretation on the radiation of an oscillating dipole in free space for the sake of simplicity, it is straightforward to verify these assertions based on the expressions of the amplitudes given in Eq. (30). For example, it is clear that for $\theta_0 = \theta_c$, the total zero order field $\mathbf{E}^{(0)}_{1,p}(\mathbf{p}_0)$ is along $\hat{\mathbf{e}}_3$ and the dot product in Eq. (30c) vanishes for all **p**.

For $\theta_0 > \theta_c$, it is convenient to expand the right-hand side in Eq. (49). By inserting Eq. (30) into Eq. (49), the reduced first order reflection amplitude, $\hat{\rho}^{(1)}_{\sigma^\pm p}$, for $\sigma^\pm$-polarized light scattered from incident p-polarized light is then given by

$$\hat{\rho}^{(1)}_{\sigma^\pm p}(\mathbf{p}\,|\,\mathbf{p}_0) = \frac{1}{\sqrt{2}}\left[\gamma(\mathbf{p})\,\hat{\mathbf{e}}^+_{p,2}(\mathbf{p}) \pm i\,\hat{\mathbf{e}}_s(\mathbf{p})\right]\cdot\mathbf{E}^{(0)}_{1,p}(\mathbf{p}_0)\,. \tag{50}$$

Here we have used the short-hand notation $\gamma(\mathbf{p}) = (\hat{\mathbf{e}}^+_{p,2}(\mathbf{p})\cdot\hat{\mathbf{e}}^+_{p,1}(\mathbf{p}))^{-1}$. For $\theta_0 > \theta_c$, the total zero order field amplitude $\mathbf{E}^{(0)}_{1,p}(\mathbf{p}_0)$ is complex. Therefore neither $\rho^{(1)}_{pp}$ nor $\rho^{(1)}_{sp}$ can be zero for propagating waves. We have seen in Section IV D that a zero intensity $p \to p$ scattering point can be found in the evanescent region since $\hat{\mathbf{e}}^+_{p,2}(\mathbf{p})$ becomes complex. However, a zero point in $p \to \sigma^\pm$ scattering may be found in the propagating region. Indeed, the fact that the square bracket in Eq. (50) is complex even for purely real values of $\hat{\mathbf{e}}^+_{p,2}(\mathbf{p})$ and $\hat{\mathbf{e}}_s(\mathbf{p})$ may compensate for the fact that $\mathbf{E}^{(0)}_{1,p}(\mathbf{p}_0)$ is complex and make the dot product in Eq. (50) vanish. Note the similarity with the right-hand side in Eq. (44) for the case of the radiation emitted by a rotating dipole, with the important difference that the p polarization vector is that of the Snell-conjugate wave. Since $\mathbf{E}^{(0)}_{1,p}(\mathbf{p}_0)$ represents a state of polarization of the media in which the dipole rotates in the plane of incidence (cf. discussion in Section IV D), we expect to find a zero in the $\sigma^\pm$ scattering intensity on each side ($\phi_r = \pm\pi/2$) of the plane of incidence. This is indeed what we observe in Figs. 15(a) and (d) in the $|\rho^{(1)}_{\sigma^\pm p}|^2$ distribution of $p \to \sigma^\pm$ scattering.

Finally, let us comment on $s \to \sigma^\pm$ scattering. In Figs. 12-16 it can be observed that the distribution for $|\rho^{(1)}_{\sigma^\pm s}|^2$ stays identical, up to an overall factor, as the angle of incidence varies. This can be understood from the dipole picture. For s-polarized incident light, the incident and zero order waves are s-polarized, so the dipoles oscillate along the x_2-direction independently of the angle of incidence. For scattering in the plane of incidence the first order waves are purely s-polarized and the two $\sigma^\pm$ components have equal intensity. For scattering at $\phi_r = \pm\pi/2$, the first order waves are purely p-polarized and the two $\sigma^\pm$ components have again have equal intensity. We obtain the largest contrast between σ^+ and σ^- for ϕ_r being a multiple of 45° since then the p- and s-polarized components are of similar amplitudes.

V. CONCLUSION

Based on a perturbative solution of the reduced Rayleigh equations to first order in the surface profile function, we have achieved a detailed mathematical and physical analysis of the scattering of polarized light by a weakly rough interface between two dielectric media. The first order amplitudes are factorized as a product of a scalar component, mainly representing the relative phases of the different scattering paths, and a polarization component. The polarization component can be interpreted as the signature of the polarization state of the dipoles in the media induced by the incident and zero order fields.

We have seen that the Yoneda phenomenon can be explained simply based on a scalar wave, single scattering picture as an intensity enhancement induced by the evanescence of the component of a scattered couple mode existing in the lesser dense medium, while all the energy allocated to the couple mode is radiated away by the component existing in the denser medium. This mechanism clearly answers previous questions put forward in the literature: we conclude that the phenomenon results from a so-called single scattering mechanism, and is not associated with surface (eigen) modes. In particular, the Yoneda phenomenon is nothing else but the continuous analogue of a Rayleigh anomaly for periodic dielectric grating, in the sense that the diffuse light here plays the role of probing what the efficiency of a diffracted order would be if it were tracked as the period of the grating would vary. This claim is easily verified with straightforward numerical calculation and the exact same perturbation analysis we have exposed here but adapted to gratings.

By factorizing the scalar behavior from that specific to a polarized wave, we have identified the geometrical criterion for the Brewster scattering phenomenon for p-polarized excitation, and more generally, for predicting the zeros of scattered intensity and amplitude for any polarization state. Simply put, these zeros are not different from those found for the radiation from a tilted oscillating dipole in free space, when the polarization of the emitted radiation is adequately measured in a fixed frame of reference. To be more accurate one may say that the physical essence is that of oscillating dipoles, but one must include the fact that arrays of dipoles yield conjugate waves as was described by e.g. Ewald and Doyle [26, 29]. The directions of zero scattering (also for evanescent waves) can then be easily interpreted geometrically in terms of Snell-conjugate waves. Moreover, we have discovered an interesting phenomenon of circularly-polarized Brewster scattering in the reflected and transmitted light scattered out of the plane of incidence when the light is incident in the dense medium and the zero order wave undergoes total internal reflection. The physical mechanism responsible for this effect was explained based on the emission of dipoles rotating in the plane of incidence (and by Snell-conjugate waves), which are induced by the fact that the reflected and transmitted zero order waves are out of phase with the incident wave.

In the present work, particular attention is given to the average phase of the scattered waves compared to previously published works on the Yoneda and Brewster scattering phenomena. We have seen that the Brewster scattering phenomenon is associated with a phase jump, while the region of polar scattering angles beyond the Yoneda threshold is associated with a gradually changing phase. These considerations on the phase of the scattered waves can be of particular interest for testing the theory against experiment, e.g. the phase behavior could be tested by the use of interferometry techniques. A simple way to measure the phase behavior associated with the Yoneda and the Brewster scattering effects is to study the scattering of light by a thin dielectric film deposited on a dielectric substrate, as was recently suggested and observed numerically in Ref. 30. For such a system, Selényi rings, which are interference rings in the intensity of the diffusely scattered light, are expected to exhibit: (i) a reversal of angular positions of the maxima and minima of intensity of the rings for p-polarized light as the Brewster scattering angle is surpassed; (ii) a gradual shift of the angular positions of the rings with respect to those predicted by the simple path difference argument for light scattered at angles beyond the Yoneda threshold due to the additional gradual phase change associated with the Yoneda phenomenon. In addition, a scattering experiment such as the one achieved in Ref. 16, but where the outgoing circularly-polarized light is measured instead of the linearly polarized light, would be of particular interest to verify the existence of a circularly-polarized Brewster scattering phenomenon out of the plane of incidence as it would strengthen the rotating dipole interpretation from which it originates.

Finally, we emphasize that the results presented in this work are approximate and are expected to be valid only for weakly rough surfaces. Additional experimental and theoretical investigations are therefore welcome to assess the range of validity of the presented hypotheses.

ACKNOWLEDGMENTS

This research was supported in part by The Research Council of Norway Contract No. 216699. The research of I.S. was supported in part by the French National Research Agency (ANR-15-CHIN-0003). This research was supported in part by NTNU and the Norwegian metacenter for High Performance Computing (NOTUR) by the allocation of computer time.

Appendix A: Perturbative solution

This appendix is devoted to the derivation of the method known as small amplitude perturbation theory (SAPT) for obtaining approximate solutions of the reduced Rayleigh equations. The basic principle of the method is to expand the kernel factor $\mathcal{J}^{b,a}_{l,m}$ in a series of Fourier transforms of the power of the surface profile function ζ and to expand the unknown reflection and transmission amplitudes in a similar series and matching terms of the same order. The expansions can be expressed as follows

$$\begin{aligned}\mathcal{J}^{b,a}_{l,m}(\mathbf{p}\,|\,\mathbf{q}) &= [b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q})]^{-1} \int \exp[-i(\mathbf{p}-\mathbf{q})\cdot\mathbf{x}_\parallel]\, \exp[-i(b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q}))\,\zeta(\mathbf{x}_\parallel)]\,\mathrm{d}^2x_\parallel \\ &= \sum_{n=0}^{\infty} \frac{(-i)^n}{n!}\,[b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q})]^{n-1}\,\hat{\zeta}^{(n)}(\mathbf{p}-\mathbf{q}) \qquad \text{(A1a)}\\ \mathbf{R}(\mathbf{q}\,|\,\mathbf{p}_0) &= \sum_{j=0}^{\infty} \frac{(-i)^j}{j!}\,\mathbf{R}^{(j)}(\mathbf{q}\,|\,\mathbf{p}_0) \qquad \text{(A1b)}\\ \mathbf{T}(\mathbf{q}\,|\,\mathbf{p}_0) &= \sum_{j=0}^{\infty} \frac{(-i)^j}{j!}\,\mathbf{T}^{(j)}(\mathbf{q}\,|\,\mathbf{p}_0)\ . \qquad \text{(A1c)}\end{aligned}$$

In equation Eq. (A1a), we have defined the Fourier transform of the n^{th} power of ζ, which we will refer to as the n^{th} *Fourier moment of the surface profile*, as

$$\hat{\zeta}^{(n)}(\mathbf{q}) = \int \zeta^n(\mathbf{x}_\parallel)\,\exp[-i\,\mathbf{q}\cdot\mathbf{x}_\parallel]\,\mathrm{d}^2x_\parallel\ . \qquad \text{(A2)}$$

We are now ready to proceed with the perturbative method.

Reflection: We start by inserting Eqs. (A1a,A1b) into the reduced Rayleigh equation Eq. (17) in the case of reflection (see Eq. (18)). We obtain

$$\begin{aligned}&\sum_{n=0}^{\infty}\sum_{j=0}^{\infty} \frac{(-i)^{n+j}}{n!\,j!} \int [\alpha_2(\mathbf{p}) - \alpha_1(\mathbf{q})]^{n-1}\,\hat{\zeta}^{(n)}(\mathbf{p}-\mathbf{q})\,\mathbf{M}^{+,+}_{2,1}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{R}^{(j)}(\mathbf{q}\,|\,\mathbf{p}_0)\,\frac{\mathrm{d}^2q}{(2\pi)^2} \\ &= -\sum_{m=0}^{\infty} \frac{(-i)^m}{m!}\,[\alpha_2(\mathbf{p}) + \alpha_1(\mathbf{p}_0)]^{m-1}\,\hat{\zeta}^{(m)}(\mathbf{p}-\mathbf{p}_0)\,\mathbf{M}^{+,-}_{2,1}(\mathbf{p}\,|\,\mathbf{p}_0)\ . \qquad \text{(A3)}\end{aligned}$$

A summation over all $(n,j)\in\mathbb{N}^2$ is equivalent to a summation over subsets $\mathcal{S}_m = \{(n,j)\in\mathbb{N}^2\,|\,n+j=m\}$ of pairs of constant sum $m = n+j$, i.e. that we have $\sum_{n,j=0}^{\infty} \equiv \sum_{m=0}^{\infty}\sum_{(n,j)\in\mathcal{S}_m}$, therefore the previous equation can be recast as

$$\begin{aligned}&\sum_{m=0}^{\infty} \frac{(-i)^m}{m!} \sum_{n=0}^{m} \binom{m}{n} \int [\alpha_2(\mathbf{p}) - \alpha_1(\mathbf{q})]^{n-1}\,\hat{\zeta}^{(n)}(\mathbf{p}-\mathbf{q})\,\mathbf{M}^{+,+}_{2,1}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{R}^{(m-n)}(\mathbf{q}\,|\,\mathbf{p}_0)\,\frac{\mathrm{d}^2q}{(2\pi)^2} \\ &\qquad = -\sum_{m=0}^{\infty} \frac{(-i)^m}{m!}\,[\alpha_2(\mathbf{p}) + \alpha_1(\mathbf{p}_0)]^{m-1}\,\hat{\zeta}^{(m)}(\mathbf{p}-\mathbf{p}_0)\,\mathbf{M}^{+,-}_{2,1}(\mathbf{p}\,|\,\mathbf{p}_0)\ . \qquad \text{(A4)}\end{aligned}$$

Note that here we have used that $\frac{1}{n!\,(m-n)!} = \frac{1}{m!}\binom{m}{n}$ by definition of the binomial coefficients. The perturbation procedure consists in matching orders in both side of the equation. The order zero only consists of one term $n=m=0$ and gives

$$\begin{aligned}&\int [\alpha_2(\mathbf{p}) - \alpha_1(\mathbf{q})]^{-1}\,\hat{\zeta}^{(0)}(\mathbf{p}-\mathbf{q})\,\mathbf{M}^{+,+}_{2,1}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{R}^{(0)}(\mathbf{q}\,|\,\mathbf{p}_0)\,\frac{\mathrm{d}^2q}{(2\pi)^2} \\ &= -[\alpha_2(\mathbf{p}) + \alpha_1(\mathbf{p}_0)]^{-1}\,\hat{\zeta}^{(0)}(\mathbf{p}-\mathbf{p}_0)\,\mathbf{M}^{+,-}_{2,1}(\mathbf{p}\,|\,\mathbf{p}_0)\ . \qquad \text{(A5)}\end{aligned}$$

By using that $\hat{\zeta}^{(0)}(\mathbf{p}-\mathbf{q}) = (2\pi)^2\,\delta(\mathbf{p}-\mathbf{q})$, we finally obtain the zero order reflection amplitude

$$\mathbf{R}^{(0)}(\mathbf{p}\,|\,\mathbf{p}_0) = (2\pi)^2\,\delta(\mathbf{p}-\mathbf{p}_0)\,\frac{\alpha_1(\mathbf{p}_0)-\alpha_2(\mathbf{p}_0)}{\alpha_2(\mathbf{p}_0)+\alpha_1(\mathbf{p}_0)}\,\left[\mathbf{M}^{+,+}_{2,1}(\mathbf{p}_0\,|\,\mathbf{p}_0)\right]^{-1}\,\mathbf{M}^{+,-}_{2,1}(\mathbf{p}_0\,|\,\mathbf{p}_0) = (2\pi)^2\,\delta(\mathbf{p}-\mathbf{p}_0)\,\boldsymbol{\rho}^{(0)}(\mathbf{p}_0) \quad \text{(A6)}$$

We have just obtained that the zero order of the reflection amplitude corresponds exactly to the reflection amplitude for a planar surface and it is straightforward to show that $\boldsymbol{\rho}^{(0)}(\mathbf{p}_0)$ is a diagonal matrix containing the Fresnel amplitudes. This was to be expected in the sense that the zero order of the surface profile corresponds to its averaged plane. For orders $m \geq 1$, we have

$$\sum_{n=0}^{m} \binom{m}{n} \int [\alpha_2(\mathbf{p}) - \alpha_1(\mathbf{q})]^{n-1}\, \hat{\zeta}^{(n)}(\mathbf{p}-\mathbf{q})\; \mathbf{M}_{2,1}^{+,+}(\mathbf{p}\,|\,\mathbf{q})\; \mathbf{R}^{(m-n)}(\mathbf{q}\,|\,\mathbf{p}_0)\; \frac{\mathrm{d}^2 q}{(2\pi)^2}$$
$$= -[\alpha_2(\mathbf{p}) + \alpha_1(\mathbf{p}_0)]^{m-1}\, \hat{\zeta}^{(m)}(\mathbf{p}-\mathbf{p}_0)\; \mathbf{M}_{2,1}^{+,-}(\mathbf{p}\,|\,\mathbf{p}_0)\,, \tag{A7}$$

which by isolating the term of interest, $n = 0$ gives $\mathbf{R}^{(m)}$ as a function of $\mathbf{R}^{(m-1)} \cdots \mathbf{R}^{(0)}$, in other words we have a recursive relation for determining all orders,

$$\mathbf{R}^{(m)}(\mathbf{p}\,|\,\mathbf{p}_0) = [\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p})] \left[\mathbf{M}_{2,1}^{+,+}(\mathbf{p}\,|\,\mathbf{p})\right]^{-1} \Bigg[(\alpha_2(\mathbf{p}) + \alpha_1(\mathbf{p}_0))^{m-1}\, \hat{\zeta}^{(m)}(\mathbf{p}-\mathbf{p}_0)\; \mathbf{M}_{2,1}^{+,-}(\mathbf{p}\,|\,\mathbf{p}_0)$$
$$+ \sum_{n=1}^{m} \binom{m}{n} \int [\alpha_2(\mathbf{p}) - \alpha_1(\mathbf{q})]^{n-1}\, \hat{\zeta}^{(n)}(\mathbf{p}-\mathbf{q})\; \mathbf{M}_{2,1}^{+,+}(\mathbf{p}\,|\,\mathbf{q})\; \mathbf{R}^{(m-n)}(\mathbf{q}\,|\,\mathbf{p}_0)\; \frac{\mathrm{d}^2 q}{(2\pi)^2} \Bigg]. \tag{A8}$$

In general, the evaluation of high orders would require the evaluation of as many integrals as the order and can become costly. For the first order, only one such integral is to be evaluated and is straightforward thanks to the fact that $\mathbf{R}^{(0)}(\mathbf{q}\,|\,\mathbf{p}_0) \propto \delta(\mathbf{q}-\mathbf{p}_0)$. Applying the above equation for $m = 1$ gives

$$\mathbf{R}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_0) = [\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p})] \left[\mathbf{M}_{2,1}^{+,+}(\mathbf{p}\,|\,\mathbf{p})\right]^{-1} \Bigg[\hat{\zeta}^{(1)}(\mathbf{p}-\mathbf{p}_0)\; \mathbf{M}_{2,1}^{+,-}(\mathbf{p}\,|\,\mathbf{p}_0)$$
$$+ \int \hat{\zeta}^{(1)}(\mathbf{p}-\mathbf{q})\; \mathbf{M}_{2,1}^{+,+}(\mathbf{p}\,|\,\mathbf{q})\; \mathbf{R}^{(0)}(\mathbf{q}\,|\,\mathbf{p}_0)\; \frac{\mathrm{d}^2 q}{(2\pi)^2} \Bigg]$$
$$= [\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p})]\, \hat{\zeta}^{(1)}(\mathbf{p}-\mathbf{p}_0)\; \left[\mathbf{M}_{2,1}^{+,+}(\mathbf{p}\,|\,\mathbf{p})\right]^{-1} \left[\mathbf{M}_{2,1}^{+,-}(\mathbf{p}\,|\,\mathbf{p}_0) + \mathbf{M}_{2,1}^{+,+}(\mathbf{p}\,|\,\mathbf{p}_0)\; \boldsymbol{\rho}^{(0)}(\mathbf{p}_0) \right]$$
$$= [\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p})]\, \hat{\zeta}^{(1)}(\mathbf{p}-\mathbf{p}_0)\, \hat{\boldsymbol{\rho}}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_0) = \hat{\zeta}^{(1)}(\mathbf{p}-\mathbf{p}_0)\, \boldsymbol{\rho}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_0)\,. \tag{A9}$$

In Eq. (A9), we define the amplitude $\hat{\boldsymbol{\rho}}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_0)$ and $\boldsymbol{\rho}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_0) = (\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p}))\, \hat{\boldsymbol{\rho}}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_0)$. The reason for these two alternative expressions is that the first one gives a factorization which is more easily interpreted from a physical point of view while the latter factorization aims at separating what depends on the realization of the surface profile, which is just $\hat{\zeta}$ here, and the amplitude factor $\boldsymbol{\rho}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_0)$ which remains independent of the specific realization of the surface profile (see Section III).

Transmission: Repeating the reasoning for the transmission amplitudes, we start by inserting Eqs. (A1a,A1c) into Eq. (17) for transmission (see Eq. (18)) and get

$$\sum_{n=0}^{\infty} \sum_{j=0}^{\infty} \frac{(-i)^{n+j}}{n!\, j!} \int [-\alpha_1(\mathbf{p}) + \alpha_2(\mathbf{q})]^{n-1}\, \hat{\zeta}^{(n)}(\mathbf{p}-\mathbf{q})\; \mathbf{M}_{1,2}^{-,-}(\mathbf{p}\,|\,\mathbf{q})\; \mathbf{T}^{(j)}(\mathbf{q}\,|\,\mathbf{p}_0)\; \frac{\mathrm{d}^2 q}{(2\pi)^2}$$
$$= \frac{2\sqrt{\epsilon_1 \epsilon_2}\, \alpha_1(\mathbf{p}_0)}{\epsilon_2 - \epsilon_1} (2\pi)^2\, \delta(\mathbf{p}-\mathbf{p}_0)\, \mathbf{I}_2\,. \tag{A10}$$

By using the same re-summation argument as for reflection, the previous equation thus becomes

$$\sum_{m=0}^{\infty} \sum_{n=0}^{m} \frac{(-i)^m}{m!} \binom{m}{n} \int [-\alpha_1(\mathbf{p}) + \alpha_2(\mathbf{q})]^{n-1}\, \hat{\zeta}^{(n)}(\mathbf{p}-\mathbf{q})\; \mathbf{M}_{1,2}^{-,-}(\mathbf{p}\,|\,\mathbf{q})\; \mathbf{T}^{(m-n)}(\mathbf{q}\,|\,\mathbf{p}_0)\; \frac{\mathrm{d}^2 q}{(2\pi)^2}$$
$$= \frac{2\sqrt{\epsilon_1 \epsilon_2}\, \alpha_1(\mathbf{p}_0)}{\epsilon_2 - \epsilon_1} (2\pi)^2\, \delta(\mathbf{p}-\mathbf{p}_0)\, \mathbf{I}_2\,. \tag{A11}$$

Next we match the zero order to the right hand side and the other orders to zero. The zero order only consists of one term $n = m = 0$ and gives

$$\mathbf{T}^{(0)}(\mathbf{p}\,|\,\mathbf{p}_0) = \frac{2\sqrt{\epsilon_1 \epsilon_2}\alpha_1(\mathbf{p}_0)}{\epsilon_2 - \epsilon_1} (2\pi)^2\, \delta(\mathbf{p}-\mathbf{p}_0)\, [\alpha_2(\mathbf{p}_0) - \alpha_1(\mathbf{p}_0)]\; \left[\mathbf{M}_{1,2}^{-,-}(\mathbf{p}_0\,|\,\mathbf{p}_0)\right]^{-1} \tag{A12}$$
$$= (2\pi)^2\, \delta(\mathbf{p}-\mathbf{p}_0)\, \boldsymbol{\tau}^{(0)}(\mathbf{p}_0) \tag{A13}$$

Here we have used that $\hat{\zeta}^{(0)}(\mathbf{p}-\mathbf{q}) = (2\pi)^2\,\delta(\mathbf{p}-\mathbf{q})$. As observed for the reflection amplitudes, we have just obtained that the zero order of the transmission amplitudes corresponds exactly to the transmission amplitudes for a planar surface, i.e. that $\boldsymbol{\tau}^{(0)}(\mathbf{p}_0)$ is a diagonal matrix containing the Fresnel transmission amplitudes. For orders $m \geq 1$, we have

$$\sum_{n=0}^{m} \binom{m}{n} \int [\alpha_2(\mathbf{q}) - \alpha_1(\mathbf{p})]^{n-1}\, \hat{\zeta}^{(n)}(\mathbf{p}-\mathbf{q})\; \mathbf{M}_{1,2}^{-,-}(\mathbf{p}\,|\,\mathbf{q})\; \mathbf{T}^{(m-n)}(\mathbf{q}\,|\,\mathbf{p}_0)\; \frac{\mathrm{d}^2 q}{(2\pi)^2} = \mathbf{0}\,, \tag{A14}$$

which by isolating the term of interest, $n = 0$ gives $\mathbf{T}^{(m)}$ as a function of $\mathbf{T}^{(m-1)} \cdots \mathbf{T}^{(0)}$, in other words we have a recursive relation for determining all orders,

$$\begin{aligned}\mathbf{T}^{(m)}(\mathbf{p}\,|\,\mathbf{p}_0) =& [\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p})]\; \left[\mathbf{M}_{1,2}^{-,-}(\mathbf{p}\,|\,\mathbf{p})\right]^{-1} \\ & \sum_{n=1}^{m} \binom{m}{n} \int [\alpha_2(\mathbf{q}) - \alpha_1(\mathbf{p})]^{n-1}\, \hat{\zeta}^{(n)}(\mathbf{p}-\mathbf{q})\; \mathbf{M}_{1,2}^{-,-}(\mathbf{p}\,|\,\mathbf{q})\; \mathbf{T}^{(m-n)}(\mathbf{q}\,|\,\mathbf{p}_0)\; \frac{\mathrm{d}^2 q}{(2\pi)^2}\,. \end{aligned} \tag{A15}$$

Applying the above equation for $m = 1$ and using that $\mathbf{T}^{(0)}(\mathbf{q}\,|\,\mathbf{p}_0) \propto \delta(\mathbf{q}-\mathbf{p}_0)$ gives

$$\begin{aligned}\mathbf{T}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_0) =& [\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p})]\, \hat{\zeta}^{(1)}(\mathbf{p}-\mathbf{p}_0)\; \left[\mathbf{M}_{1,2}^{-,-}(\mathbf{p}\,|\,\mathbf{p})\right]^{-1}\; \mathbf{M}_{1,2}^{-,-}(\mathbf{p}\,|\,\mathbf{p}_0)\; \boldsymbol{\tau}^{(0)}(\mathbf{p}_0) \\ =& [\alpha_1(\mathbf{p}) - \alpha_2(\mathbf{p})]\, \hat{\zeta}^{(1)}(\mathbf{p}-\mathbf{p}_0)\, \hat{\boldsymbol{\tau}}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_0) = \hat{\zeta}^{(1)}(\mathbf{p}-\mathbf{p}_0)\, \boldsymbol{\tau}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_0)\,. \end{aligned} \tag{A16}$$

Appendix B: Differential reflection coefficient

Assuming we have obtained the reflection amplitudes $R_{\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_0)$ either by using the perturbative approach or by the purely numerical simulation, we can now proceed to express the differential reflection coefficient (DRC) defined as the time-averaged flux radiated around a given scattering direction (θ_r, ϕ) per unit solid angle and per unit incident flux and denoted $\partial R/\partial\Omega_r(\mathbf{p}\,|\,\mathbf{p}_0)$. Let a virtual hemisphere of radius $r \gg c/\omega$ lie on the plane $x_3 = 0$ on top of the scattering system. The support of this hemisphere is a disk of area $S = \pi r^2$. We consider the scattering from a *truncated* version of the scattering system in which the surface profiles are set to be flat outside the disk support. Consequently, the field amplitudes we will manipulate are not strictly speaking those of the full system of interest but will converge to them as $r \to \infty$. We will nevertheless keep the same notations as that from the full system introduced in Section III for simplicity. The time-averaged flux incident on this disk is given by

$$\begin{aligned} P_{\mathrm{inc}/S} &= -\mathrm{Re}\frac{c}{8\pi}\int_S \left[\mathbf{E}_0^*(\mathbf{p}_0) \times \left(\frac{c}{\omega}\,\mathbf{k}_1^-(\mathbf{p}_0) \times \mathbf{E}_0(\mathbf{p}_0)\right)\right] \cdot \hat{\mathbf{e}}_3\, \exp\left[-i(\mathbf{k}_1^{-*}(\mathbf{p}_0) - \mathbf{k}_1^-(\mathbf{p}_0)) \cdot \mathbf{x}\right] \mathrm{d}^2 x_\parallel \\ &= -\frac{c^2}{8\pi\omega}\,\mathrm{Re}\int_S \left[|\,\mathbf{E}_0(\mathbf{p}_0)|^2\;\mathbf{k}_1^-(\mathbf{p}_0) - \left(\mathbf{E}_0^*(\mathbf{p}_0)\cdot\mathbf{k}_1^-(\mathbf{p}_0)\right)\cdot\mathbf{E}_0(\mathbf{p}_0)\right] \cdot \hat{\mathbf{e}}_3 \mathrm{d}^2 x_\parallel \\ &= S\,\frac{c^2}{8\pi\omega}\,\alpha_1(\mathbf{p}_0)\,|\,\mathbf{E}_0(\mathbf{p}_0)|^2 \\ &= S\,\frac{c^2}{8\pi\omega}\,\alpha_1(\mathbf{p}_0)\,\left[\left|\mathcal{E}_{0,p}\right|^2 + \left|\mathcal{E}_{0,s}\right|^2\right]. \end{aligned} \tag{B1}$$

Here, the * denotes the complex conjugate, and incident field amplitude $\mathbf{E}_0(\mathbf{p}_0) = \mathcal{E}_{0,p}\;\hat{\mathbf{e}}_p^-(\mathbf{p}_0) + \mathcal{E}_{0,s}\;\hat{\mathbf{e}}_s(\mathbf{p}_0)$ as defined in Eq. (13), the vector identity $\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = (\mathbf{a}\cdot\mathbf{c})\mathbf{b} - (\mathbf{a}\cdot\mathbf{b})\mathbf{c}$ and the orthogonality between the field and the wave vector $\mathbf{E}_0^*(\mathbf{p}_0)\cdot\mathbf{k}_1^-(\mathbf{p}_0) = 0$ have been used. Note that the flux incident on the disk is proportional to the disk area. Let us now consider the outgoing flux crossing an elementary surface $\mathrm{d}\sigma = r^2\sin\theta_r \mathrm{d}\theta_r \mathrm{d}\phi = r^2 \mathrm{d}\Omega_r$ around a point $\mathbf{r} = r\,(\sin\theta_r\cos\phi\,\hat{\mathbf{e}}_1 + \sin\theta_r\sin\phi\,\hat{\mathbf{e}}_2 + \cos\theta_r\,\hat{\mathbf{e}}_3) = r\,\hat{\mathbf{n}}$. The flux crossing this elementary surface is given by

$$P_{\mathrm{d}\sigma} = \frac{c}{8\pi}\,\mathrm{Re}\left[\mathbf{E}_1^{+*}(\mathbf{r}) \times \mathbf{H}_1^+(\mathbf{r})\right] \cdot \hat{\mathbf{n}}\,\mathrm{d}\sigma. \tag{B2}$$

We then use the well-known asymptotic expansion of the field in the far-field given by (see Refs. 44 and 45)

$$\mathbf{E}_1^+(\mathbf{r}) \sim -i\,\epsilon_1^{1/2}\,\frac{\omega}{2\pi\,c}\cos\theta_r\,\frac{\exp(i\epsilon_1^{1/2}\frac{\omega}{c}r)}{r}\;\mathbf{E}_1^+(\mathbf{p}) \tag{B3a}$$

$$\mathbf{H}_1^+(\mathbf{r}) \sim -i\,\epsilon_1\,\frac{\omega}{2\pi\,c}\cos\theta_r\,\frac{\exp(i\epsilon_1^{1/2}\frac{\omega}{c}r)}{r}\,\hat{\mathbf{n}}\times\mathbf{E}_1^+(\mathbf{p}) \tag{B3b}$$

where $\mathbf{p} = \sqrt{\epsilon_1}\frac{\omega}{c}(\sin\theta_r\cos\phi\,\hat{\mathbf{e}}_1 + \sin\theta_r\sin\phi\,\hat{\mathbf{e}}_2)$. This asymptotic approximation will become more and more accurate as we let $r \to \infty$. Plugging Eq. (B3) into Eq. (B2) we obtain

$$P_{\mathrm{d}\sigma} = \epsilon_1^{3/2}\left(\frac{\omega}{2\pi\,c}\right)^2\cos^2\theta_r\,\frac{c}{8\pi}\,|\,\mathbf{E}_1^+(\mathbf{p})|^2\,\mathrm{d}\Omega_r = \epsilon_1^{3/2}\left(\frac{\omega}{2\pi\,c}\right)^2\cos^2\theta_r\,\frac{c}{8\pi}\,\left(|\,\mathcal{E}_{1,p}^+(\mathbf{p})|^2 + |\,\mathcal{E}_{1,s}^+(\mathbf{p})|^2\right)\,\mathrm{d}\Omega_r. \tag{B4}$$

The total differential reflection coefficient is then given by

$$\frac{\partial R}{\partial\Omega_r}(\mathbf{p}\,|\,\mathbf{p}_0) = \lim_{r\to\infty}\frac{P_{\mathrm{d}\sigma}}{P_{\mathrm{inc}/S}\,\mathrm{d}\Omega_r} = \lim_{r\to\infty}\frac{\epsilon_1}{S}\left(\frac{\omega}{2\pi\,c}\right)^2\frac{\cos^2\theta_r}{\cos\theta_0}\,\frac{|\,\mathcal{E}_{1,p}^+(\mathbf{p})|^2 + |\,\mathcal{E}_{1,s}^+(\mathbf{p})|^2}{|\,\mathcal{E}_{0,p}\,|^2 + |\,\mathcal{E}_{0,s}\,|^2}. \tag{B5}$$

From the total differential reflection coefficient given by Eq. (B5), we deduce the differential reflection coefficient when an incident plane wave of polarization β, with lateral wave vector $\mathbf{p}_0$ is reflected into a plane wave of polarization α with lateral wave vector $\mathbf{p}$ given as

$$\frac{\partial R_{\alpha\beta}}{\partial\Omega_r}(\mathbf{p}\,|\,\mathbf{p}_0) = \lim_{r\to\infty}\frac{\epsilon_1}{S}\left(\frac{\omega}{2\pi\,c}\right)^2\frac{\cos^2\theta_r}{\cos\theta_0}\,|R_{\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_0)|^2 = \lim_{r\to\infty}\frac{\partial R_{\alpha\beta}^{(S)}}{\partial\Omega_r}(\mathbf{p}\,|\,\mathbf{p}_0). \tag{B6}$$

As we are interested in averaging the optical response over realizations of the surface profiles, we consider the following ensemble average

$$\left\langle\frac{\partial R_{\alpha\beta}^{(S)}}{\partial\Omega_r}(\mathbf{p}\,|\,\mathbf{p}_0)\right\rangle = \frac{\epsilon_1}{S}\left(\frac{\omega}{2\pi\,c}\right)^2\frac{\cos^2\theta_r}{\cos\theta_0}\,\left\langle|R_{\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_0)|^2\right\rangle. \tag{B7}$$

A similar derivation for the differential transmitted coefficient yields

$$\left\langle\frac{\partial T_{\alpha\beta}^{(S)}}{\partial\Omega_t}(\mathbf{p}\,|\,\mathbf{p}_0)\right\rangle = \frac{\epsilon_2^{3/2}}{\epsilon_1^{1/2}S}\left(\frac{\omega}{2\pi\,c}\right)^2\frac{\cos^2\theta_t}{\cos\theta_0}\,\left\langle|T_{\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_0)|^2\right\rangle. \tag{B8}$$

By decomposing the reflection amplitudes as the sum of the mean and fluctuation (deviation from the mean)

$$R_{\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_0) = \langle R_{\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_0)\rangle + \left[R_{\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_0) - \langle R_{\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_0)\rangle\right], \tag{B9}$$

we can decompose the MDRC as the sum of a coherent term and an incoherent term

$$\left\langle\frac{\partial R_{\alpha\beta}^{(S)}}{\partial\Omega_r}(\mathbf{p}\,|\,\mathbf{p}_0)\right\rangle = \left\langle\frac{\partial R_{\alpha\beta}^{(S)}}{\partial\Omega_r}(\mathbf{p}\,|\,\mathbf{p}_0)\right\rangle_{\mathrm{coh}} + \left\langle\frac{\partial R_{\alpha\beta}^{(S)}}{\partial\Omega_r}(\mathbf{p}\,|\,\mathbf{p}_0)\right\rangle_{\mathrm{incoh}}, \tag{B10}$$

where

$$\left\langle\frac{\partial R_{\alpha\beta}^{(S)}}{\partial\Omega_r}(\mathbf{p}\,|\,\mathbf{p}_0)\right\rangle_{\mathrm{coh}} = \frac{\epsilon_1}{S}\left(\frac{\omega}{2\pi\,c}\right)^2\frac{\cos^2\theta_r}{\cos\theta_0}\,|\langle R_{\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_0)\rangle|^2 \tag{B11a}$$

$$\left\langle\frac{\partial R_{\alpha\beta}^{(S)}}{\partial\Omega_r}(\mathbf{p}\,|\,\mathbf{p}_0)\right\rangle_{\mathrm{incoh}} = \frac{\epsilon_1}{S}\left(\frac{\omega}{2\pi\,c}\right)^2\frac{\cos^2\theta_r}{\cos\theta_0}\,\left[\left\langle|R_{\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_0)|^2\right\rangle - |\langle R_{\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_0)\rangle|^2\right]. \tag{B11b}$$

If we now use the expression found in A for the reflection amplitudes to first order in the product of surface profiles,

$$\mathbf{R}\,(\mathbf{p}\,|\,\mathbf{p}_0) \approx \mathbf{R}^{(0)}\,(\mathbf{p}\,|\,\mathbf{p}_0) - i\,\mathbf{R}^{(1)}\,(\mathbf{p}\,|\,\mathbf{p}_0), \tag{B12}$$

where $\mathbf{R}^{(0)}\,(\mathbf{p}\,|\,\mathbf{p}_0)$ is the response from the corresponding system with planar interface, Eq. (A6), and $\mathbf{R}^{(1)}\,(\mathbf{p}\,|\,\mathbf{p}_0)$ is given in Eq. (A9), we obtain that the factor in the square bracket in Eq. (B11b) reads

$$\left\langle|R_{\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_0)|^2\right\rangle - |\langle R_{\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_0)\rangle|^2 = \left\langle\left|R_{\alpha\beta}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_0)\right|^2\right\rangle = \left\langle|\hat{\zeta}_S(\mathbf{p}-\mathbf{p}_0)|^2\right\rangle\,|\rho_{\alpha\beta}^{(1)}\,(\mathbf{p}\,|\,\mathbf{p}_0)\,|^2. \tag{B13}$$

Note here that we are still dealing with a scattering system whose surface profiles are flat outside the disk of radius r, hence the subscript S. For the statistical properties attributed to the surface profiles in Sec. II, we have

$$\begin{aligned}\left\langle\hat{\zeta}_S(\mathbf{q})\hat{\zeta}_S^*(\mathbf{q})\right\rangle &= \left\langle\int_S\int_S\zeta(\mathbf{x})\zeta(\mathbf{x}')\exp\left[i\,\mathbf{q}\cdot(\mathbf{x}-\mathbf{x}')\right]\,\mathrm{d}^2x\,\mathrm{d}^2x'\right\rangle\\ &= \int_S\int_S\langle\zeta(\mathbf{x})\zeta(\mathbf{x}')\rangle\,\exp\left[i\,\mathbf{q}\cdot(\mathbf{x}-\mathbf{x}')\right]\,\mathrm{d}^2x\,\mathrm{d}^2x'\\ &= \int_S\int_S\sigma^2\,W(\mathbf{x}-\mathbf{x}')\,\exp\left[i\,\mathbf{q}\cdot(\mathbf{x}-\mathbf{x}')\right]\,\mathrm{d}^2x\,\mathrm{d}^2x'.\end{aligned} \tag{B14}$$

Here we have used the definition of the Fourier transform, and the fact that ensemble average commutes with the integration of the surfaces and the definition of the correlation function. Via the change of variable $\mathbf{u} = \mathbf{x} - \mathbf{x}'$ we obtain

$$\left\langle \hat{\zeta}_S(\mathbf{q})\hat{\zeta}_S^*(\mathbf{q}) \right\rangle = S\,\sigma^2 \int_S W(\mathbf{u})\, \exp(i\,\mathbf{q}\cdot\mathbf{u})\, \mathrm{d}^2u = S\,\sigma^2\, g_S(\mathbf{q}). \tag{B15}$$

Thus

$$\left\langle |R_{\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_0)|^2 \right\rangle - |\langle R_{\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_0)\rangle|^2 = S\,\sigma^2\, g_S(\mathbf{p}-\mathbf{p}_0)\, \left|\rho^{(1)}_{\alpha\beta}\,(\mathbf{p}\,|\,\mathbf{p}_0)\right|^2. \tag{B16}$$

Finally, by plugging the above equation into Eq. (B11b), the surface area S cancels and letting $r \to \infty$, $g_S \to g$ (where we remind the reader that g is the power spectrum of the surface profiles) and we finally obtain the expression for the incoherent component of the MDRC for the entire (infinite) system under the first order approximation of the reflected amplitudes in product of the surface profiles

$$\left\langle \frac{\partial R_{\alpha\beta}}{\partial \Omega_r}(\mathbf{p}\,|\,\mathbf{p}_0) \right\rangle_{\text{incoh}} = \epsilon_1 \left(\frac{\omega}{2\pi c}\right)^2 \frac{\cos^2\theta_r}{\cos\theta_0}\, g(\mathbf{p}-\mathbf{p}_0)\, \sigma^2\, \left|\rho^{(1)}_{\alpha\beta}\,(\mathbf{p}\,|\,\mathbf{p}_0)\right|^2. \tag{B17}$$

Similarly, for the transmitted light we obtain

$$\left\langle \frac{\partial T_{\alpha\beta}}{\partial \Omega_t}(\mathbf{p}\,|\,\mathbf{p}_0) \right\rangle_{\text{incoh}} = \frac{\epsilon_2^{3/2}}{\epsilon_1^{1/2}} \left(\frac{\omega}{2\pi c}\right)^2 \frac{\cos^2\theta_t}{\cos\theta_0}\, g(\mathbf{p}-\mathbf{p}_0)\, \sigma^2\, \left|\tau^{(1)}_{\alpha\beta}\,(\mathbf{p}\,|\,\mathbf{p}_0)\right|^2. \tag{B18}$$

[1] Ø. S. Hetland, A. A. Maradudin, T. Nordam, and I. Simonsen, Phys. Rev. A **93**, 053819 (2016).
[2] Ø. S. Hetland, A. A. Maradudin, T. Nordam, P. A. Letnes, and I. Simonsen, Phys. Rev. A **95**, 043808 (2017).
[3] Y. Yoneda, Phys. Rev. **131**, 2010 (1963).
[4] G. H. Vineyard, Physical Review B **26**, 4146 (1982).
[5] S. K. Sinha, E. B. Sirota, S. Garoff, and H. B. Stanley, Phys. Rev. B **38**, 2297 (1988).
[6] E. E. Gorodnichev, S. L. Dudarev, D. B. Rogozkin, and M. I. Ryazanov, Sov. Phys. JETP Lett. **48**, 147 (1988).
[7] T. A. Leskova and A. A. Maradudin, Waves in Random Media **7**, 395 (1997).
[8] G. Renaud, R. Lazzari, and F. Leroy, Surf. Sci. Rep. **64**, 255 (2009).
[9] H. Dosch, Phys. Rev. B **35**, 2137 (1987).
[10] S. Stepanov, in *Exploration of Subsurface Phenomena by Particle Scattering*, edited by N. Lam, C. Melendres, and S. Sinha (International Advanced Studies Institute Press, North East, Maryland, 2000) pp. 119–137.
[11] A. Kitahara, K. Inoue, H. Kikkawa, K. Matsushita, and I. Takahashi, *X-ray Reflectivity study on polymeric surfaces near glass transition temperature*, Photon Factory Activity Report No. 20, Part B (unpublished) (2002) p. 83.
[12] J.-C. Gasse, D. Lützenkirchen-Hecht, R. Wagner, and R. Frahm, J. Phys.: Conf. Ser. **712**, 012028 (2016).
[13] T. Kawanishi, H. Ogura, and Z. L. Wang, Wave. Random Media **7**, 351 (1997).
[14] In an earlier numerical investigation of light scattering from one-dimensional dielectric rough surfaces, Nieto-Vesperinas and Sánchez-Gil [46] observed "sidelobes" in the angular intensity distributions. They did, however, not associate this observation with the Yoneda phenomenon.
[15] T. Tamir, in *Electromagnetic Surface Modes*, edited by A. D. Boardman (John Wiley & Sons, New York, 1982) Chap. 13.
[16] A. K. González-Alcalde, J.-P. Banon, Ø. S. Hetland, A. A. Maradudin, E. R. Méndez, T. Nordam, and I. Simonsen, Opt. Express **24**, 25995 (2016).
[17] D. Brewster, Philos. Trans. R. Soc. London **105**, 125 (1815).
[18] A. Lakhtakia, Opt. News **15**, 14 (1989).
[19] J. D. Jackson, *Classical Electrodynamics*, 3rd ed. (John Wiley & Sons, New York, 1999).
[20] E. Hecht, *Optics* (Addison-Wesley, 2002).
[21] A. Sommerfeld, J. Opt. Soc. Am. **7**, 501 (1923).
[22] J.-J. Greffet and A. Sentenac, in *Wave Propag. Scatt. Varied Media II*, Vol. 1558, edited by V. K. Varadan (1991) pp. 288–294.
[23] A. Sentenac and J.-J. Greffet, J. Opt. Soc. Am. A **15**, 528 (1998).
[24] O. Calvo-Perez, A. Sentenac, and J. J. Greffet, Radio Sci. **34**, 311 (1999).
[25] J. Lekner, *Theory of Reflection of Electromagnetic and Particle Waves*, Vol. 3 (Springer Science & Business Media, 1987).
[26] W. T. Doyle, Am. J. Phys. **53**, 463 (1985).
[27] J. Sein, Opt. Commun. **2**, 170 (1970).
[28] D. Pattanayak and E. Wolf, Opt. Commun. **6**, 217 (1972).
[29] P. P. Ewald, Annalen der Physik **354**, 1 (1916).
[30] J.-P. Banon, Ø.S. Hetland, and I. Simonsen, Annals of Physics **389**, 352 (2018).

[31] A. Soubret, G. Berginc, and C. Bourrely, Phys. Rev. B **63**, 245411 (2001).
[32] G. Brown, V. Celli, M. Haller, and A. Marvin, Surf. Sci. **136**, 381 (1984).
[33] A. Soubret, G. Berginc, and C. Bourrely, J. Opt. Soc. Am. A **18**, 2778 (2001).
[34] A. A. Maradudin and E. R. Méndez, Appl. Opt. **32**, 3335 (1993).
[35] A. Maradudin, A. R. McGurn, and E. R. Méndez, J. Opt. Soc. Am. A **12**, 2500 (1995).
[36] A. R. McGurn and A. A. Maradudin, Waves in Random Media **6**, 251 (1996), http://dx.doi.org/10.1088/0959-7174/6/3/006.
[37] I. Simonsen, Eur. Phys. J.-Spec. Top. **181**, 1 (2010).
[38] Assuming of course $\epsilon_1 \neq \epsilon_2$, otherwise the problem is trivial.
[39] U. Fano, J. Opt. Soc. Am. **31**, 213 (1941).
[40] The dot product here must be taken as the Hermitian inner product for complex vectors $\mathbf{a} \cdot \mathbf{b} = \sum_j a_j^* b_j$.
[41] Here limited to one medium, but in general both media sharing the interface could be dielectrics.
[42] W. T. Doyle, Am. J. Phys. **48**, 643 (1980).
[43] One may extend the construction to all vectors on the circle defined as the intersection of the unit sphere and the plane $\mathbf{E}^{(0)}_{p,1}(\mathbf{p}_0)^\perp$, but it would result in constructing twice the same set of wave vectors of zero scattering.
[44] G. S. Agarwal, Phys. Rev. B **15**, 2371 (1977).
[45] K. Miyamoto and E. Wolf, J. Opt. Soc. Am. **52**, 615 (1962).
[46] M. Nieto-Vesperinas and J. A. Sánchez-Gil, J. Opt. Soc. Am. A **9**, 424 (1992).

Paper 3:

J.-P. Banon, Ø. S. Hetland, and I. Simonsen. Selective enhancement of Selényi rings induced by the cross-correlation between the interfaces of a two-dimensional randomly rough dielectric film. *Ann. Phys.*, 389:352–382, 2018

Annals of Physics 389 (2018) 352–382

Contents lists available at ScienceDirect

Annals of Physics

journal homepage: www.elsevier.com/locate/aop

Selective enhancement of Selényi rings induced by the cross-correlation between the interfaces of a two-dimensional randomly rough dielectric film

J.-P. Banon [a,*], Ø.S. Hetland [a], I. Simonsen [a,b]

[a] Department of Physics, NTNU – Norwegian University of Science and Technology, NO-7491 Trondheim, Norway

[b] Surface du Verre et Interfaces, UMR 125 CNRS/Saint-Gobain, F-93303 Aubervilliers, France

ARTICLE INFO

Article history:
Received 6 September 2017
Accepted 9 December 2017
Available online 19 December 2017

Keywords:
Light scattering
Interference
Randomly rough surfaces
Numerical simulation

ABSTRACT

By the use of both perturbative and non-perturbative solutions of the reduced Rayleigh equation, we present a detailed study of the scattering of light from two-dimensional weakly rough dielectric films. It is shown that for several rough film configurations, Selényi interference rings exist in the diffusely scattered light. For film systems supported by dielectric substrates where only one of the two interfaces of the film is weakly rough and the other planar, Selényi interference rings are observed at angular positions that can be determined from simple phase arguments. For such single-rough-interface films, we find and explain by a single scattering model that the contrast in the interference patterns is better when the top interface of the film (the interface facing the incident light) is rough than when the bottom interface is rough. When both film interfaces are rough, Selényi interference rings exist but a potential cross-correlation of the two rough interfaces of the film can be used to selectively enhance some of the interference rings while others are attenuated and might even disappear. This feature may in principle be used in determining the correlation properties of interfaces of films that otherwise would be difficult to access.

* Corresponding author.
E-mail address: jean-philippe.banon@ntnu.no (J.-P. Banon).

https://doi.org/10.1016/j.aop.2017.12.003
0003-4916/

1. Introduction

Interference effects in the diffuse light scattered by thin and rough dielectric films can look both stunning and unexpected, and they have fascinated their observers for centuries. First formally described in modern times as colorful rings in the diffusely scattered light originating from a dusty back-silvered mirror by Newton [1], what is today known as Quételet- and Selényi-rings have been thoroughly analyzed theoretically [2–6] and experimentally [7,8]. An example of a non-laboratory situation where one may observe this phenomenon is in light reflections from bodies of water if appropriate algae are present on the water surface. This phenomenon, modeled as a thin layer of spherical scatterers suspended on a reflecting planar surface, was investigated by Suhr and Schlichting [6].

In a theoretical study of the scattering from one-dimensional randomly rough surfaces ruled on dielectric films on perfectly conducting substrates, Lu et al. [4] concluded that the degree of surface roughness had the biggest impact on which interference phenomena could be observed. For films with a thickness on the order of several wavelengths they were able to explain the periodic fringes they observed in the mean differential reflection coefficient through simple phase arguments. The patterns in the diffusely scattered light were shown to undergo a transition, with increasing surface roughness, from an intensity pattern exhibiting fringes whose angular positions are independent of the angle of incidence (Selényi rings [9]) to one with fringes whose angular positions depend on the angle of incidence (Quételet rings [7]) and eventually into a fringeless pattern with a backscattering peak, which is a signature of multiple scattering [10]. Although the Selényi rings are centered around the mean surface normal, with their position being independent of the angle of incidence, their amplitude, however, is modulated by the angle of incidence. According to the current understanding of the Selényi rings, their main origin is due to the interference between light scattered back directly from the top scattering layer and light reflected by the film after being scattered within it. In this paper we seek to complete this interpretation of the interference phenomena within a single scattering approximation, enabling a sound interpretation of the Selényi rings for the previously unexplored case when the rough surface is shifted to the non-incident face of the film.

A similar system to the one studied by Lu et al. was also thoroughly studied perturbatively and experimentally by Kaganovskii et al. [8]. They concluded that the long-range (smooth) component of the surface roughness, whenever present, can have a deciding effect on the interference pattern observed in the diffusely scattered light.

However, most of the relevant studies conducted on the topic so far have been restricted to investigations of scattering from a single rough interface. Allowing for more than one rough interface significantly increases the complexity of the problem both analytically and computationally, but it also opens a door to a richer set of scattering phenomena. Such stacked, multi-layered systems will in many cases better represent the real-world scattering systems we are attempting to model [2]. Two or more of these randomly rough interfaces in the stack will also often be correlated, either naturally occurring, by design or by method of production [11,12]. Since both Quételet- and Selényi-rings may enable a practical way of remote sensing and surface characterization for certain geometries and layer thicknesses, it is important also to model the impact of such roughness cross-correlation.

In this paper we investigate interference effects in the light scattered diffusely from an optical system composed of two semi-infinite media separated by a single thin dielectric film where both interfaces may be rough [Fig. 1(a)]. After describing the statistical properties of the interfaces in Section 2, we derive, in Section 3, a set of reduced Rayleigh equations (RREs) for the case of electromagnetic scattering from a system with two rough interfaces, inspired by the work of Soubret et al. [13]. Although only the case of reflected light will be analyzed in detail, the RREs for both the reflection and the transmission amplitudes are given for completeness; furthermore, this also serves to show that the presented framework can easily be generalized to an arbitrary number of rough interfaces. A perturbative method and a purely numerical method for solving the RREs are described in Section 4. Since solving the RREs for a set of two, or more, two-dimensional randomly rough surfaces by purely numerical means is a highly computationally intensive task, the perturbative method will be our main investigation tool for simulating and interpreting interference effects in such geometries. In Section 5.1 we discuss rough film geometries where either the top interface or bottom interface of the film is allowed to be randomly rough and the other interface is planar. For

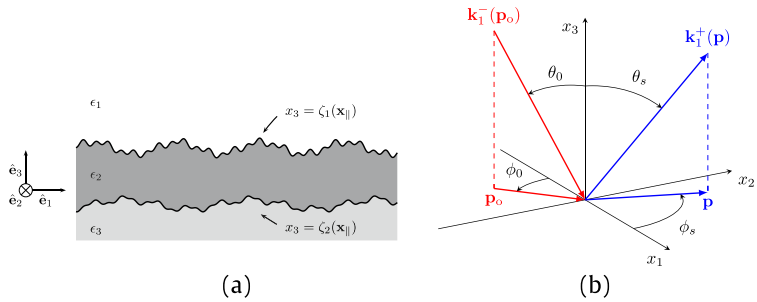


Fig. 1. (a) Layered system with two rough interfaces. (b) Definitions of the angles of incidence and scattering and wave vectors.

such geometries, we compare the predictions for the scattered intensities obtained on the basis of the perturbative and non-perturbative methods. After having established the apparent validity of the perturbative method for the level of roughness assumed, we continue to investigate rough film geometries where both interfaces of the film are randomly rough and have a varying cross-correlation [Section 5.2]. Section 5.3 gives a brief discussion concerning additional effects one expects to observe in transmission. Finally, Section 6 presents the conclusions that we have drawn from this study.

2. Scattering systems

An overview of a typical system geometry is provided in Fig. 1. We consider the case where both interfaces of the film may be randomly rough and possess non-trivial auto- and cross-correlation. Furthermore, we will be interested in scattering systems for which the mean thickness of the film is several wavelengths so that interference fringes can be observed in the diffusely reflected or transmitted intensities. The definition of the geometry is set in the three-dimensional space endowed with a Cartesian coordinate system $(O, \hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3)$, with the vector plane $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2)$ parallel to the mean plane of the interfaces [Fig. 1(b)]. The origin, O, can be arbitrarily chosen, only affecting the complex reflection and transmission amplitudes by an overall phase factor which plays no role in the intensity of the scattered light. The scattering system splits space into a slab of three domains, or layers, that will be denoted by the indices $j \in \{1, 2, 3\}$. The mean thickness of the film will be denoted $d > 0$, and the jth interface separating media j and $j + 1$ can be described by the equation

$$x_3 = \zeta_j(\mathbf{x}_\parallel) = d_j + h_j(\mathbf{x}_\parallel)\,, \tag{1}$$

for $j \in \{1, 2\}$, where $\mathbf{x}_\parallel = x_1\,\hat{\mathbf{e}}_1 + x_2\,\hat{\mathbf{e}}_2$, $d_j = \langle \zeta_j \rangle$ denotes the average of the jth profile (and we have $d_1 - d_2 = d$), and the term h_j will be assumed to be a continuous, differentiable, single-valued, stationary, isotropic, Gaussian random process with zero mean and given auto-correlation. More specifically, the surface profile functions are assumed to satisfy the following properties

$$\big\langle h_j(\mathbf{x}_\parallel) \big\rangle = 0 \tag{2a}$$

$$\big\langle h_j(\mathbf{x}_\parallel) h_j(\mathbf{x}'_\parallel) \big\rangle = \sigma_j^2\, W(\mathbf{x}_\parallel - \mathbf{x}'_\parallel). \tag{2b}$$

Here and in the following, the angle brackets denote an average over an ensemble of realizations of the stochastic process, σ_j denotes the rms roughness of interface j and $W(\mathbf{x}_\parallel)$ represents the height auto-correlation function normalized so that $W(\mathbf{0}) = 1$. For reasons of simplicity we here restrict ourselves to the situation where both interfaces are characterized by the same form of the correlation function. In particular, we will here assume a Gaussian form of the auto-correlation function that is defined by

$$W(\mathbf{x}_\parallel) = \exp\left(-\frac{|\mathbf{x}_\parallel|^2}{a^2}\right)\,, \tag{3}$$

where a is the correlation length. The corresponding power spectrum (defined as the Fourier transform of W) is then

$$g(\mathbf{p}) = \pi a^2 \exp\left(-\frac{|\mathbf{p}|^2 a^2}{4}\right) , \tag{4}$$

with $\mathbf{p} = p_1\,\hat{\mathbf{e}}_1 + p_2\,\hat{\mathbf{e}}_2$. In addition, the two interfaces will be assumed to be cross-correlated in the following way

$$\left\langle h_1(\mathbf{x}_\parallel) h_2(\mathbf{x}'_\parallel)\right\rangle = \gamma\ \sigma_1\sigma_2\ W(\mathbf{x}_\parallel - \mathbf{x}'_\parallel) , \tag{5}$$

where $\gamma \in [-1, 1]$ is a dimensionless cross-correlation coupling variable. When $\gamma = 0$ the two interfaces are uncorrelated, and the extreme cases $\gamma = \pm 1$ and $\sigma_1 = \sigma_2$ can be viewed respectively as the second interface being a shifted copy of the first one by a vector $-d\,\hat{\mathbf{e}}_3$, or as the second interface being a symmetric copy of the first one with respect to the plane $x_3 = (d_1 + d_2)/2$. We can summarize the correlations expressed by Eqs. (2b) and (5) by the following relation

$$\left\langle h_i(\mathbf{x}_\parallel) h_j(\mathbf{x}'_\parallel)\right\rangle = [\delta_{ij} + \gamma(1 - \delta_{ij})]\ \sigma_i\sigma_j\ W(\mathbf{x}_\parallel - \mathbf{x}'_\parallel) , \tag{6}$$

where δ_{ij} denotes the Kronecker delta.

3. Formulation of the problem

The theoretical approach used in this work to study the scattering of light from the systems of interest is based on the so-called reduced Rayleigh equations. A reduced Rayleigh equation is an integral equation in which the integral kernel encodes the materials and geometry of the scattering system and the unknowns are the reflection or transmission amplitudes for each polarization. In the following, in order to establish the notation and highlight the main assumptions of the method, we will briefly recall the key ideas of the derivation of the reduced Rayleigh equations for a system composed of three media separated by two disjoint rough interfaces. We will use, to our knowledge, the most general form of the reduced Rayleigh equations for a single interface derived by Soubret et al. in Ref. [13] and used by these authors in Refs. [13,14] in the case of a single interface system and a film geometry. Once the general framework is established, we will apply it to the specific geometries of interest.

3.1. The reduced Rayleigh equations

All physical quantities introduced hereafter will be indexed with respect to the medium (domain) they belong to. The electromagnetic response of the media is modeled by non-magnetic, homogeneous, isotropic, linear constitutive relations in the frequency domain, i.e. that *a priori* each medium is characterized by frequency dependent scalar complex dielectric functions, $\epsilon_j(\omega)$, where ω denotes the frequency of the electromagnetic wave excitation. We consider the presence of an electromagnetic field $(\mathbf{E}, \mathbf{H})$ in the whole space. The fields will be denoted by a subscript j depending on their containing medium. As an example, the electric field evaluated at a point $\mathbf{x}$ in medium 1 at time t is denoted $\mathbf{E}_1(\mathbf{x}, t) = \mathbf{E}_1(\mathbf{x}, \omega)\exp(-i\omega t)$. The source free Maxwell equations, together with homogeneous, linear and isotropic constitutive relations in the frequency domain, result in the electric and magnetic fields satisfying the Helmholtz equation in each region. Namely, for all $j \in \{1, 2, 3\}$,

$$\nabla^2\mathbf{E}_j(\mathbf{x}, \omega) + \epsilon_j(\omega)\left(\frac{\omega}{c}\right)^2 \mathbf{E}_j(\mathbf{x}, \omega) = \mathbf{0} , \tag{7}$$

and a similar equation satisfied for $\mathbf{H}$. Here, ∇^2 denotes the Laplace operator and c represents the speed of light in vacuum. In the following, we will drop the time, or frequency, dependence, since we assume a stationary regime where time contributes only by an overall phase factor $\exp(-i\omega t)$. It is known that a solution to the Helmholtz equation can be written as a linear combination of plane waves, thus the representation of the electric field in each region can be written as

$$\mathbf{E}_j(\mathbf{x}) = \sum_{a=\pm} \int_{\mathbb{R}^2} \left[\mathcal{E}^a_{j,p}(\mathbf{q})\ \hat{\mathbf{e}}^a_{p,j}(\mathbf{q}) + \mathcal{E}^a_{j,s}(\mathbf{q})\ \hat{\mathbf{e}}_s(\mathbf{q})\right] \exp\left(i\,\mathbf{k}^a_j(\mathbf{q})\ \cdot \mathbf{x}\right) \frac{\mathrm{d}^2 q}{(2\pi)^2} , \tag{8}$$

where

$$\alpha_j(\mathbf{q}) = \sqrt{\epsilon_j\Big(\frac{\omega}{c}\Big)^2 - \mathbf{q}^2}, \qquad \mathrm{Re}\,(\alpha_j), \mathrm{Im}\,(\alpha_j) \geq 0\,, \tag{9a}$$

$$\mathbf{k}_j^{\pm}(\mathbf{q}) = \mathbf{q} \pm \alpha_j(\mathbf{q})\,\hat{\mathbf{e}}_3\,, \tag{9b}$$

$$\hat{\mathbf{e}}_s(\mathbf{q}) = \hat{\mathbf{e}}_3 \times \hat{\mathbf{q}}\,, \tag{9c}$$

$$\hat{\mathbf{e}}_{p,j}^{\pm}(\mathbf{q}) = \frac{c}{\sqrt{\epsilon_j}\omega}\left(\pm\alpha_j(\mathbf{q})\,\hat{\mathbf{q}} - |\mathbf{q}|\,\hat{\mathbf{e}}_3\right). \tag{9d}$$

Here a caret over a vector indicates that the vector is a unit vector. Note that the wave vector $\mathbf{k}_j^{\pm}(\mathbf{q})$ of an elementary plane wave is decomposed into its projection $\mathbf{q}$ in the lateral vector plane $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2)$ and the component $\pm\alpha_j(\mathbf{q})$ along $\hat{\mathbf{e}}_3$. The sum for $a = \pm$ takes into account both upwards and downwards propagating and evanescent (and possibly growing) waves. The field amplitude is decomposed in the *local polarization basis* $(\hat{\mathbf{e}}_{p,j}^a(\mathbf{q})\,, \hat{\mathbf{e}}_s(\mathbf{q}))$, so that $\mathcal{E}_{j,\alpha}^a(\mathbf{q})$ denotes the component of the field amplitude in the polarization state α of the mode characterized by a and $\mathbf{q}$. In this basis, the directions given by $\hat{\mathbf{e}}_{p,j}^{\pm}(\mathbf{q})$, and $\hat{\mathbf{e}}_s(\mathbf{q})$ are respectively the directions of the p- and s-polarization of the electric field amplitude. Furthermore, the electromagnetic fields have to satisfy the boundary conditions $(j \in \{1, 2\})$

$$\mathbf{n}_j(\mathbf{x}_\parallel) \times \Big[\mathbf{E}_{j+1}(\mathbf{s}_j(\mathbf{x}_\parallel)) - \mathbf{E}_j(\mathbf{s}_j(\mathbf{x}_\parallel))\Big] = \mathbf{0} \tag{10a}$$

$$\mathbf{n}_j(\mathbf{x}_\parallel) \times \Big[\mathbf{H}_{j+1}(\mathbf{s}_j(\mathbf{x}_\parallel)) - \mathbf{H}_j(\mathbf{s}_j(\mathbf{x}_\parallel))\Big] = \mathbf{0}\,, \tag{10b}$$

where $\mathbf{n}_j(\mathbf{x}_\parallel)$ is a vector that is normal to surface j at the surface point $\mathbf{s}_j(\mathbf{x}_\parallel) = \mathbf{x}_\parallel + \zeta_j(\mathbf{x}_\parallel)\hat{\mathbf{e}}_3$, and given by

$$\mathbf{n}_j(\mathbf{x}_\parallel) = \hat{\mathbf{e}}_3 - \frac{\partial\zeta_j}{\partial x_1}(\mathbf{x}_\parallel)\,\hat{\mathbf{e}}_1 - \frac{\partial\zeta_j}{\partial x_2}(\mathbf{x}_\parallel)\,\hat{\mathbf{e}}_2\,. \tag{11}$$

Here, $\partial/\partial x_k$ denotes the partial derivative along the direction $\hat{\mathbf{e}}_k$. Following Soubret et al. [13], for a given surface indexed by j, by substituting the field expansion Eq. (8) into Eq. (10) and by a clever linear integral combination of the boundary conditions, one can show that the upward or downward field amplitudes in medium $j+1$ can be linked to the upward and downward field amplitudes in medium j via the following integral equation defined for $a_{j+1} = \pm$, $j \in \{1, 2\}$, and $\mathbf{p}$ in the vector plane $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2)$:

$$\sum_{a_j=\pm}\int \mathcal{J}_{j+1,j}^{a_{j+1},a_j}(\mathbf{p}\,|\,\mathbf{q})\;\mathbf{M}_{j+1,j}^{a_{j+1},a_j}(\mathbf{p}\,|\,\mathbf{q})\;\boldsymbol{\mathcal{E}}_j^{a_j}(\mathbf{q})\,\frac{\mathrm{d}^2q}{(2\pi)^2} = \frac{2\,a_{j+1}\,\sqrt{\epsilon_j\epsilon_{j+1}}\,\alpha_{j+1}(\mathbf{p})}{\epsilon_{j+1}-\epsilon_j}\,\boldsymbol{\mathcal{E}}_{j+1}^{a_{j+1}}(\mathbf{p})\,. \tag{12}$$

Here $\boldsymbol{\mathcal{E}}_j^a(\mathbf{q}) = (\mathcal{E}_{j,p}^a(\mathbf{q})\,, \mathcal{E}_{j,s}^a(\mathbf{q}))^{\mathrm{T}}$ denotes a column vector of the polarization components of the field amplitude in medium j. Moreover, $\mathbf{M}_{l,m}^{b,a}(\mathbf{p}\,|\,\mathbf{q})$ is a 2×2 matrix which originates from a change of coordinate system between the local polarization basis $(\hat{\mathbf{e}}_{p,l}^b(\mathbf{p})\,, \hat{\mathbf{e}}_s(\mathbf{p}))$ and $(\hat{\mathbf{e}}_{p,m}^a(\mathbf{q})\,, \hat{\mathbf{e}}_s(\mathbf{q}))$, defined for $a = \pm$, $b = \pm$, and $l, m \in \{j, j+1\}$ such that $l \neq m$ as

$$\mathbf{M}_{l,m}^{b,a}(\mathbf{p}\,|\,\mathbf{q}) = \begin{pmatrix} |\mathbf{p}||\mathbf{q}| + ab\,\alpha_l(\mathbf{p})\,\alpha_m(\mathbf{q})\,\hat{\mathbf{p}}\cdot\hat{\mathbf{q}} & -b\,\sqrt{\epsilon_m}\,\dfrac{\omega}{c}\,\alpha_l(\mathbf{p})\,[\hat{\mathbf{p}}\times\hat{\mathbf{q}}]\cdot\hat{\mathbf{e}}_3 \\ a\,\sqrt{\epsilon_l}\,\dfrac{\omega}{c}\,\alpha_m(\mathbf{q})\,[\hat{\mathbf{p}}\times\hat{\mathbf{q}}]\cdot\hat{\mathbf{e}}_3 & \sqrt{\epsilon_l\epsilon_m}\,\dfrac{\omega^2}{c^2}\,\hat{\mathbf{p}}\cdot\hat{\mathbf{q}} \end{pmatrix}. \tag{13}$$

The kernel scalar factor $\mathcal{J}_{l,m}^{b,a}(\mathbf{p}\,|\,\mathbf{q})$ encodes the surface geometry and is defined as

$$\mathcal{J}_{l,m}^{b,a}(\mathbf{p}\,|\,\mathbf{q}) = (b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q}))^{-1}\int \exp\left[-i(\mathbf{k}_l^b(\mathbf{p}) - \mathbf{k}_m^a(\mathbf{q}))\cdot(\mathbf{x}_\parallel + \zeta_j(\mathbf{x}_\parallel)\,\hat{\mathbf{e}}_3)\right]\mathrm{d}^2x_\parallel. \tag{14}$$

Notice that, as already pointed out in Ref. [13], due to the symmetry of the boundary conditions, one may also show in the same way that

$$\sum_{a_{j+1}=\pm} \int \mathcal{T}_{j,j+1}^{a_j,a_{j+1}}(\mathbf{p}\,|\,\mathbf{q})\ \mathbf{M}_{j,j+1}^{a_j,a_{j+1}}(\mathbf{p}\,|\,\mathbf{q})\ \boldsymbol{\mathcal{E}}_{j+1}^{a_{j+1}}(\mathbf{q})\,\frac{\mathrm{d}^2q}{(2\pi)^2} = \frac{2\,a_j\,\sqrt{\epsilon_j\epsilon_{j+1}}\,\alpha_j(\mathbf{p})}{\epsilon_j-\epsilon_{j+1}}\,\boldsymbol{\mathcal{E}}_j^{a_j}(\mathbf{p})\ , \tag{15}$$

which can be obtained from Eq. (12) by interchanging j and $j+1$. Typically, Eq. (12) is appropriate to solve the problem of reflection whereas Eq. (15) is appropriate to solve the problem of transmission, as we will see later. In the following, it will be convenient to define

$$\boldsymbol{\Theta}_{j+1,j}^{a_{j+1},a_j}(\mathbf{p}\,|\,\mathbf{q}) = \alpha_{j+1}^{-1}(\mathbf{p})\mathcal{T}_{j+1,j}^{a_{j+1},a_j}(\mathbf{p}\,|\,\mathbf{q})\ \mathbf{M}_{j+1,j}^{a_{j+1},a_j}(\mathbf{p}\,|\,\mathbf{q}) \tag{16}$$

and

$$\boldsymbol{\Theta}_{j,j+1}^{a_j,a_{j+1}}(\mathbf{p}\,|\,\mathbf{q}) = \alpha_j^{-1}(\mathbf{p})\mathcal{T}_{j,j+1}^{a_j,a_{j+1}}(\mathbf{p}\,|\,\mathbf{q})\ \mathbf{M}_{j,j+1}^{a_j,a_{j+1}}(\mathbf{p}\,|\,\mathbf{q}) \tag{17}$$

which we will refer to as the forward and backward *single interface transfer kernels* between media j and $j+1$, respectively. Our aim is to study reflection from and transmission through the whole system, i.e. we need to relate the field amplitudes in regions 1 and 3 without having to explicitly consider the field amplitudes in region 2. To this end, we have to combine Eq. (12) for $j=1$ and $j=2$ in order to eliminate $\boldsymbol{\mathcal{E}}_2^{\pm}$. A systematic way of doing this, and which can be generalized to an arbitrary number of layers, is presented below. The key observation lies in the fact that one can choose the sign a_{j+1} in Eq. (12) and therefore Eq. (12) contains two vector equations for a given j. For reasons that will soon become clear, the variable $\mathbf{p}$ that appears in Eq. (12) is renamed $\mathbf{p}_2$. By left-multiplying both sides of Eq. (12) taken at $j=1$ by $a_2\,\boldsymbol{\Theta}_{3,2}^{a_3,a_2}(\mathbf{p}\,|\,\mathbf{p}_2)$, where $a_3=\pm$ can be arbitrarily chosen, we obtain

$$\sum_{a_1=\pm} a_2 \int \boldsymbol{\Theta}_{3,2}^{a_3,a_2}(\mathbf{p}\,|\,\mathbf{p}_2)\,\boldsymbol{\Theta}_{2,1}^{a_2,a_1}(\mathbf{p}_2\,|\,\mathbf{q})\,\boldsymbol{\mathcal{E}}_1^{a_1}(\mathbf{q})\,\frac{\mathrm{d}^2q}{(2\pi)^2} = \frac{2\sqrt{\epsilon_1\epsilon_2}}{\epsilon_2-\epsilon_1}\,\boldsymbol{\Theta}_{3,2}^{a_3,a_2}(\mathbf{p}\,|\,\mathbf{p}_2)\,\boldsymbol{\mathcal{E}}_2^{a_2}(\mathbf{p}_2)\,.$$

By integrating this equation over $\mathbf{p}_2$ divided by $(2\pi)^2$ and summing over $a_2=\pm$, one obtains that the right-hand-side of the resulting equation is, up to a constant factor, equal to the left-hand-side of Eq. (12) evaluated for $j=2$. In this way we obtain

$$\sum_{a_1=\pm} \int \boldsymbol{\Theta}_{3,1}^{a_3,a_1}(\mathbf{p}\,|\,\mathbf{q})\,\boldsymbol{\mathcal{E}}_1^{a_1}(\mathbf{q})\,\frac{\mathrm{d}^2q}{(2\pi)^2} = a_3\,\frac{4\sqrt{\epsilon_1\epsilon_2^2\epsilon_3}}{(\epsilon_3-\epsilon_2)(\epsilon_2-\epsilon_1)}\,\boldsymbol{\mathcal{E}}_3^{a_3}(\mathbf{p})\,, \tag{18}$$

where the *forward two-interface transfer kernel* $\boldsymbol{\Theta}_{3,1}^{a_3,a_1}(\mathbf{p}|\mathbf{q})$ is defined by the composition rule

$$\boldsymbol{\Theta}_{3,1}^{a_3,a_1}(\mathbf{p}\,|\,\mathbf{q}) = \sum_{a_2=\pm} a_2 \int \boldsymbol{\Theta}_{3,2}^{a_3,a_2}(\mathbf{p}\,|\,\mathbf{p}_2)\,\boldsymbol{\Theta}_{2,1}^{a_2,a_1}(\mathbf{p}_2\,|\,\mathbf{q})\,\frac{\mathrm{d}^2p_2}{(2\pi)^2}. \tag{19}$$

By a similar method and by the use of Eq. (15), we obtain the backward relation

$$\sum_{a_3=\pm} \int \boldsymbol{\Theta}_{1,3}^{a_1,a_3}(\mathbf{p}\,|\,\mathbf{q})\,\boldsymbol{\mathcal{E}}_3^{a_3}(\mathbf{q})\,\frac{\mathrm{d}^2q}{(2\pi)^2} = a_1\,\frac{4\sqrt{\epsilon_1\epsilon_2^2\epsilon_3}}{(\epsilon_1-\epsilon_2)(\epsilon_2-\epsilon_3)}\,\boldsymbol{\mathcal{E}}_1^{a_1}(\mathbf{p})\,, \tag{20}$$

where the *backward two-interface transfer kernel* $\boldsymbol{\Theta}_{1,3}^{a_1,a_3}(\mathbf{p}|\mathbf{q})$ is defined as

$$\boldsymbol{\Theta}_{1,3}^{a_1,a_3}(\mathbf{p}\,|\,\mathbf{q}) = \sum_{a_2=\pm} a_2 \int \boldsymbol{\Theta}_{1,2}^{a_1,a_2}(\mathbf{p}\,|\,\mathbf{p}_2)\,\boldsymbol{\Theta}_{2,3}^{a_2,a_3}(\mathbf{p}_2\,|\,\mathbf{q})\,\frac{\mathrm{d}^2p_2}{(2\pi)^2}. \tag{21}$$

Let us now make a few remarks on Eqs. (18) and (19). Eq. (18) is an integral equation of the same form as Eq. (12) but it only relates the field amplitudes in medium 1 and 3. Our aim of eliminating the field amplitudes in the intermediary medium is therefore achieved. However, this comes at a cost since the new transfer kernel $\boldsymbol{\Theta}_{3,1}^{a_3,a_1}(\mathbf{p}\,|\,\mathbf{q})$ is defined as an integral of the product of two single interface kernels

as can be seen in Eq. (19). We will see that this pays off in the case where one of the interfaces is flat, but that the cost can be significant in terms of computational load when both interfaces are rough.

So far, we have stayed general and simply assumed the presence of an electromagnetic field decomposed in propagating and non-propagating waves in each region. Therefore, there is no uniqueness in the solutions to the transfer equations, Eqs. (18) and (20). To ensure a unique solution, one needs to impose some constraints on the field. First, we need to introduce an incident field to our model. This will split the field expansion into a sum of an incident field, which is given by our model of the problem, and a scattered field. Note that within this framework, the incident field may be chosen to be in either medium, or to be a combination of excitations incident from different media. Second, we need to impose the Sommerfeld radiation condition at infinity. This implies that the non-propagating waves are indeed only evanescent waves in the media unbounded in the $\hat{\mathbf{e}}_3$-direction and that the propagating ones are directed outwards.

In our case, the incident field will be taken as a plane wave incident from medium 1 and defined as

$$\mathbf{E}_0(\mathbf{x}) = \left[\mathcal{E}_{0,p}\,\hat{\mathbf{e}}^-_{p,1}(\mathbf{p}_0) + \mathcal{E}_{0,s}\,\hat{\mathbf{e}}_s(\mathbf{p}_0)\right]\exp\left(i\mathbf{k}^-_1(\mathbf{p}_0)\cdot\mathbf{x}\right), \tag{22}$$

where $\mathbf{p}_0$ is the projection of the incident wave's wave vector in the $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2)$ plane, with the property $|\mathbf{p}_0| \leq \sqrt{\epsilon_1}\,\omega/c$, i.e. we consider an incident propagating wave. The fact that this is the only incident wave considered, together with the Sommerfeld radiation condition at infinity, gives, apart from the incident field, that the only elementary waves allowed in the scattered field are those with wave vectors of the form $\mathbf{k}^+_1(\mathbf{p})$ and $\mathbf{k}^-_3(\mathbf{p})$ in medium 1 and 3, respectively. This property can be expressed by defining the field amplitudes

$$\boldsymbol{\mathcal{E}}^-_1(\mathbf{q}) = (2\pi)^2\,\delta(\mathbf{q}-\mathbf{p}_0)\,\boldsymbol{\mathcal{E}}_0\,, \tag{23a}$$

$$\boldsymbol{\mathcal{E}}^+_3(\mathbf{q}) = \mathbf{0}\,, \tag{23b}$$

where $\boldsymbol{\mathcal{E}}_0 = (\mathcal{E}_{0,p}, \mathcal{E}_{0,s})^{\mathrm{T}}$. Next, we assume that the scattered field amplitudes are linearly related to the incident field amplitude $\boldsymbol{\mathcal{E}}_0$ via the reflection and transmission amplitudes, $\mathbf{R}(\mathbf{q}\,|\,\mathbf{p}_0)$ and $\mathbf{T}(\mathbf{q}\,|\,\mathbf{p}_0)$, defined as

$$\boldsymbol{\mathcal{E}}^+_1(\mathbf{q}) = \mathbf{R}(\mathbf{q}|\mathbf{p}_0)\boldsymbol{\mathcal{E}}_0, \tag{24a}$$

$$\boldsymbol{\mathcal{E}}^-_3(\mathbf{q}) = \mathbf{T}(\mathbf{q}|\mathbf{p}_0)\boldsymbol{\mathcal{E}}_0. \tag{24b}$$

The reflection and transmission amplitudes are therefore described by 2×2 matrices, i.e. for $\mathbf{X} = \mathbf{R}$ or $\mathbf{T}$

$$\mathbf{X} = \begin{pmatrix} X_{pp} & X_{ps} \\ X_{sp} & X_{ss} \end{pmatrix}. \tag{25}$$

From a physical point of view, the coefficient $R_{\alpha\beta}(\mathbf{q}|\mathbf{p}_0)$ (resp. $T_{\alpha\beta}(\mathbf{q}|\mathbf{p}_0)$) for $\alpha, \beta \in \{p, s\}$ is the field amplitude for the reflected (resp. transmitted) light with lateral wave vector $\mathbf{q}$ in the polarization state α from a unit incident field with lateral wave vector $\mathbf{p}_0$ in the polarization state β. The reflection and transmission amplitudes are then the unknowns in our scattering problem. The equations we need to solve are deduced from the general equations Eqs. (18) and (20) by applying them respectively at $a_3 = +$ and $a_1 = -$ and by using Eqs. (23) and (24) for the model of the field expansion. This yields the following two decoupled integral equations for the reflection or transmission amplitudes, the so-called reduced Rayleigh equations, that can be written in the following general form, for $\mathbf{X} = \mathbf{R}$ or $\mathbf{T}$ [15]

$$\int \mathbf{M}_{\mathbf{X}}(\mathbf{p}|\mathbf{q})\,\mathbf{X}(\mathbf{q}|\mathbf{p}_0)\frac{\mathrm{d}^2q}{(2\pi)^2} = -\mathbf{N}_{\mathbf{X}}(\mathbf{p}|\mathbf{p}_0)\,, \tag{26}$$

where the matrices $\mathbf{M}_{\mathbf{X}}$ and $\mathbf{N}_{\mathbf{X}}$ are given by

$$\mathbf{M}_{\mathbf{R}}(\mathbf{p}|\mathbf{q}) = \boldsymbol{\Theta}^{+,+}_{3,1}(\mathbf{p}|\mathbf{q}) \tag{27a}$$

$$\mathbf{M}_{\mathbf{T}}(\mathbf{p}|\mathbf{q}) = \boldsymbol{\Theta}^{-,-}_{1,3}(\mathbf{p}|\mathbf{q}) \tag{27b}$$

$$\mathbf{N_R}(\mathbf{p}|\mathbf{q}) = \Theta_{3,1}^{+,-}(\mathbf{p}|\mathbf{q}) \tag{27c}$$

$$\mathbf{N_T}(\mathbf{p}|\mathbf{q}) = \frac{4\sqrt{\epsilon_1\epsilon_2^2\epsilon_3}}{(\epsilon_1-\epsilon_2)(\epsilon_2-\epsilon_3)}\,(2\pi)^2\,\delta(\mathbf{p}-\mathbf{q})\,\mathbf{I}_2, \tag{27d}$$

with $\mathbf{I}_2$ denoting the 2×2 identity matrix. In the cases where only one interface is rough and the other interface is planar, the complexity associated with the transfer kernels is equivalent to that of a single rough interface separating two media. For instance, if the second interface is planar and the first interface is rough, we can choose the origin of the coordinate system such that $\zeta_2(\mathbf{x}_\parallel) = d_2 = 0$, and Eq. (14) yields, for $l, m \in \{2, 3\}$ and $l \neq m$,

$$\mathcal{J}_{l,m}^{b,a}(\mathbf{p}\,|\,\mathbf{q}) = \frac{(2\pi)^2\,\delta(\mathbf{p}-\mathbf{q})}{b\alpha_l(\mathbf{p}) - a\alpha_m(\mathbf{q})}. \tag{28}$$

The Dirac distribution then simplifies the wave vector integration present in the two-interface transfer kernels and one gets

$$\Theta_{3,1}^{a_3,a_1}(\mathbf{p}\,|\,\mathbf{q}) = \sum_{a_2=\pm} a_2\,\frac{\mathbf{M}_{3,2}^{a_3,a_2}(\mathbf{p}\,|\,\mathbf{p})\,\Theta_{2,1}^{a_2,a_1}(\mathbf{p}\,|\,\mathbf{q})}{\alpha_3(\mathbf{p})\,[a_3\alpha_3(\mathbf{p}) - a_2\alpha_2(\mathbf{p})]}, \tag{29a}$$

and

$$\Theta_{1,3}^{a_1,a_3}(\mathbf{p}\,|\,\mathbf{q}) = \sum_{a_2=\pm} a_2\,\frac{\Theta_{1,2}^{a_1,a_2}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{M}_{2,3}^{a_2,a_3}(\mathbf{q}\,|\,\mathbf{q})}{\alpha_2(\mathbf{q})\,[a_2\alpha_2(\mathbf{q}) - a_3\alpha_3(\mathbf{q})]}. \tag{29b}$$

If the first interface is planar and the second interface rough, we can choose the origin of the coordinate system such that $\zeta_1(\mathbf{x}_\parallel) = d_1 = 0$, and Eq. (28) holds for $l, m \in \{1, 2\}$ and $l \neq m$, and the two-interface transfer kernels read

$$\Theta_{3,1}^{a_3,a_1}(\mathbf{p}\,|\,\mathbf{q}) = \sum_{a_2=\pm} a_2\,\frac{\Theta_{3,2}^{a_3,a_2}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{M}_{2,1}^{a_2,a_1}(\mathbf{q}\,|\,\mathbf{q})}{\alpha_2(\mathbf{q})\,[a_2\alpha_2(\mathbf{q}) - a_1\alpha_1(\mathbf{q})]}, \tag{30a}$$

and

$$\Theta_{1,3}^{a_1,a_3}(\mathbf{p}\,|\,\mathbf{q}) = \sum_{a_2=\pm} a_2\,\frac{\mathbf{M}_{1,2}^{a_1,a_2}(\mathbf{p}\,|\,\mathbf{p})\,\Theta_{2,3}^{a_2,a_3}(\mathbf{p}\,|\,\mathbf{q})}{\alpha_1(\mathbf{p})\,[a_1\alpha_1(\mathbf{p}) - a_2\alpha_2(\mathbf{p})]}. \tag{30b}$$

3.2. Observables

The observable of interest in this study is the so-called incoherent (or diffuse) component of the *mean differential reflection coefficient* (DRC) that we denote $\langle \partial R_{\alpha\beta}(\mathbf{p}|\mathbf{p}_0)/\partial\Omega_s\rangle_{\text{incoh}}$. It is defined as the ensemble average over realizations of the surface profile function of the incoherent component of the radiated reflected flux of an α-polarized wave around direction $\hat{\mathbf{k}}_1^+(\mathbf{p})$, per unit incident flux of a β-polarized plane wave of wave vector $\mathbf{k}_1^-(\mathbf{p}_0)$, and per unit solid angle. The precise mathematical definition and the derivation of the expression for the mean DRC as a function of the reflection amplitudes is given in Appendix B.

4. Numerical methods

Solutions of the reduced Rayleigh equation, Eq. (26), are obtained via both a perturbative and a non-perturbative numerical approach. In this work we investigate systems with two interfaces; For the case when one of these interfaces is planar we are able to employ both approaches, but when both interfaces are rough we will exclusively use the perturbative approach due to the high computational cost of the non-perturbative approach.

4.1. Perturbative method

The approximated solution of Eq. (26) for the reflection amplitudes, and to first order in product of surface profiles, obtained by small amplitude perturbation theory (SAPT) is derived in Appendix A and given by

$$\mathbf{R}(\mathbf{p} \mid \mathbf{p}_0) \approx \mathbf{R}^{(0)}(\mathbf{p} \mid \mathbf{p}_0) - i\mathbf{R}^{(1)}(\mathbf{p} \mid \mathbf{p}_0)\,, \tag{31a}$$

$$\mathbf{R}^{(1)}(\mathbf{p} \mid \mathbf{p}_0) = \hat{h}_1(\mathbf{p} - \mathbf{p}_0)\boldsymbol{\rho}_1(\mathbf{p} \mid \mathbf{p}_0) + \hat{h}_2(\mathbf{p} - \mathbf{p}_0)\boldsymbol{\rho}_2(\mathbf{p} \mid \mathbf{p}_0). \tag{31b}$$

Here $\mathbf{R}^{(0)}(\mathbf{p} \mid \mathbf{p}_0)$ is the response from the corresponding system with planar interfaces (i.e. that of a Fabry–Perot interferometer), $\hat{h}_j$ are the Fourier transforms of the stochastic component of the surface profiles and $\boldsymbol{\rho}_j(\mathbf{p} \mid \mathbf{p}_0)$ are matrix-valued amplitudes depending *only* on the mean film thickness, the dielectric constants of all media and the wave vectors of incidence and scattering. The explicit expressions for these matrices are given in Appendix A (see Eq. (A.20)). The corresponding expression for the incoherent component of the mean differential reflection coefficient reads Appendices A and B

$$\left\langle \frac{\partial R_{\alpha\beta}(\mathbf{p}|\mathbf{p}_0)}{\partial \Omega_s} \right\rangle_{\text{incoh}} = \epsilon_1 \left(\frac{\omega}{2\pi c}\right)^2 \frac{\cos^2\theta_s}{\cos\theta_0}\, g(\mathbf{p} - \mathbf{p}_0)\Big[\sigma_1^2\, |\rho_{1,\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0)|^2 + \sigma_2^2\, |\rho_{2,\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0)|^2 + 2\gamma\sigma_1\sigma_2\, \mathrm{Re}\left\{\rho_{1,\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0)\rho^*_{2,\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0)\right\}\Big]\,, \tag{32}$$

where the wave vectors

$$\mathbf{p} = \sqrt{\epsilon_1}\frac{\omega}{c}\sin\theta_s(\cos\phi_s\,\hat{\mathbf{e}}_1 + \sin\phi_s\,\hat{\mathbf{e}}_2) \tag{33a}$$

and

$$\mathbf{p}_0 = \sqrt{\epsilon_1}\frac{\omega}{c}\sin\theta_0(\cos\phi_0\,\hat{\mathbf{e}}_1 + \sin\phi_0\,\hat{\mathbf{e}}_2) \tag{33b}$$

are defined in terms of the angles of scattering (θ_s, ϕ_s) and incidence (θ_0, ϕ_0), respectively [see Fig. 1]. The three terms present in the angular brackets of Eq. (32) can be interpreted as follows. The term containing $\sigma_1^2\, |\rho_{1,\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0)|^2$ (resp. $\sigma_2^2\, |\rho_{2,\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0)|^2$) corresponds to the contribution to the diffuse intensity of the associated system for which the first (resp. second) interface would be rough and the other planar. Indeed, this would be the only remaining term if we were to set $\sigma_2 = 0$ (resp. $\sigma_1 = 0$) in Eq. (32). The sum of the two first terms would correspond to the sum of intensity of the aforementioned associated systems, which would be the expected overall response if the two interface *were not* correlated, i.e. if $\gamma = 0$. The last term in Eq. (32), which does not vanish for $\gamma \neq 0$, can be interpreted physically as taking into account the interference between paths resulting from single scattering events on the top interface and those resulting from single scattering events on the bottom interface. Note that this last term, in contrast to the two first, may take positive *and* negative values as the incident and scattering wave vectors are varied, and hence may result in cross-correlation induced constructive and destructive interference. It is clear from the derivation, however, that the overall incoherent component of the mean differential coefficient remains non-negative, as is required for any intensity.

4.2. Nonperturbative method

Solutions of Eq. (26) were also obtained in a rigorous, purely numerical, nonperturbative manner according to the method described in detail in Ref. [16]; only a brief summary of the method is presented here. This method has previously been used for the investigations of the two-dimensional rough surface scattering of light from metallic or perfectly conducting surfaces [16–18]; from and through single dielectric interfaces [17,19,20] and film geometries [21–23]. In this method, an ensemble of realizations of the surface profile function $\zeta_j(\mathbf{x}_\parallel)$ is generated by the use of the Fourier filtering method [24] on a square grid of $N_x \times N_x$ surface points, covering an area of $S = L^2$ in the $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2)$-plane. The integral equation, Eq. (26), is solved numerically with finite limits $\pm Q$ and

discretization $\Delta q = 2\pi/L$ with $N_q \times N_q$ points in wave vector space according to the Nyquist sampling theorem given the spatial discretization of the surface. On evaluating the kernel scalar factors $\mathcal{J}^{b,a}_{l,m}(\mathbf{p}\,|\,\mathbf{q})$, defined in Eq. (14), we first expand the integrand in powers of $\zeta_j(\mathbf{x}_\parallel)$, truncate this expansion after 20 terms, and integrate the resulting sum term-by-term. The Fourier integral of $\zeta^n_j(\mathbf{x}_\parallel)$ that remains now only depends on the surface profile function and the difference in lateral wave vectors $\mathbf{p}-\mathbf{q}$, and not on $\alpha_l(\mathbf{p})$ and $\alpha_m(\mathbf{q})$. These Fourier integrals are therefore calculated only once, on a $\mathbf{p}-\mathbf{q}$ grid, for every surface realization by the use of the fast Fourier transform. The resulting matrix equations are then solved by LU factorization and back substitution, using the ScaLAPACK library [25]. This process is repeated for a large number N_p of realizations of the surface profile function, enabling the calculation of the ensemble averaged observables of interest; like the mean DRC.

It remains to mention that Eqs. (29) and (30), giving the transfer kernels in the case where only one of the interfaces is rough and the other planar, have been written in a rather compact form. Numerically, these expressions tend to lead to instabilities due to factors of the form $\exp(-i\alpha_2(\mathbf{q})d)$ or $\exp(-i\alpha_2(\mathbf{p})d)$ which grow for evanescent waves inside the film. This technical issue is resolved by using the following two ideas: (i) expanding the two terms in the kernels (i.e. for $a_2 = \pm$) and factorizing out the troublesome exponential factor and canceling it on both sides of the reduced Rayleigh equation (if the exponential factor is a function of the variable $\mathbf{p}$) or (ii) making a change of variable such that the troublesome exponential factor is absorbed into the reflection or transmission amplitudes (if the exponential factor is a function of the variable $\mathbf{q}$). One may also shift the x_3-axis in order to facilitate the aforementioned steps. We chose here not to give more details on the explicit implementation, as these modifications are to be done in a case by case basis depending on which surface is planar and whether the reflected or transmitted light is considered.

5. Results and discussion

5.1. Single rough interface

As a direct comparison between results obtained by the perturbative and nonperturbative solutions of Eq. (26), Fig. 2 shows the angular distributions of the co-polarized ($\alpha = \beta$) incoherent contribution to the mean DRC for light incident from vacuum ($\epsilon_1 = 1$) that is reflected diffusively into the plane of incidence (i.e. $|\hat{\mathbf{p}} \cdot \hat{\mathbf{p}}_0| = 1$) from a randomly rough dielectric film (photoresist, $\epsilon_2 = 2.69$) deposited on a silicon substrate ($\epsilon_3 = 15.08 + 0.15i$) for the cases where only one of the interfaces is rough and the other planar. Results for the case where only the top interface (the interface facing the medium of incidence) is rough ($\sigma_2 = 0$) and where only the bottom interface is rough ($\sigma_1 = 0$) are shown in Figs. 2(a)–(b) and (c)–(d), respectively. Light was incident on the dielectric film from the vacuum side in the form of a plane wave of wavelength $\lambda = 632.8$ nm with angles of incidence $(\theta_0, \phi_0) = (16.8^\circ, 0^\circ)$. The two interfaces were characterized by rms-roughness $\sigma_1 = \lambda/30, \sigma_2 = 0$ [Figs. 2(a)–(b)] or $\sigma_1 = 0, \sigma_2 = \lambda/30$ [Figs. 2(c)–(d)], correlation length $a = \lambda/3$, and the film thickness was assumed to be $d = 8\lambda \approx 5\ \mu$m. The scattering system was chosen in order to highlight the interference phenomena and to purposely deviate from the more historically typical scattering system of a dielectric film on a perfect electric conductor. The dashed curves in Fig. 2 display the results of computations of the perturbative solution of the RRE, Eq. (32), to leading order, while the solid curves in Fig. 2 show the non-perturbative solutions of the RRE, Eq. (26). In obtaining these latter results the following parameters, defined in Section 4.2, were used: $N_x = 449$, $L = 45\lambda$, $N_q = 225$ and $N_p = 325$, implying integration limits in wavevector space $Q = \pm 2.5\omega/c$. Since these non-perturbative results for the mean DRC are obtained through an ensemble average over a finite number of surface realizations, they are less smooth than their perturbative counterparts, for which the averaging is performed analytically. Using a larger number of surface realizations in obtaining the ensemble average would have produced smoother results, but we have chosen not to do so here due to the high associated computational cost.

Figs. 2(a)–(b) show excellent agreement between the results for the mean DRC obtained by the analytical perturbative method and the corresponding results obtained by a full solution of the RRE for the chosen parameters for the case where only the upper interface is rough. In particular, the fringes observed in these figures are consistently predicted by both calculation methods for the set

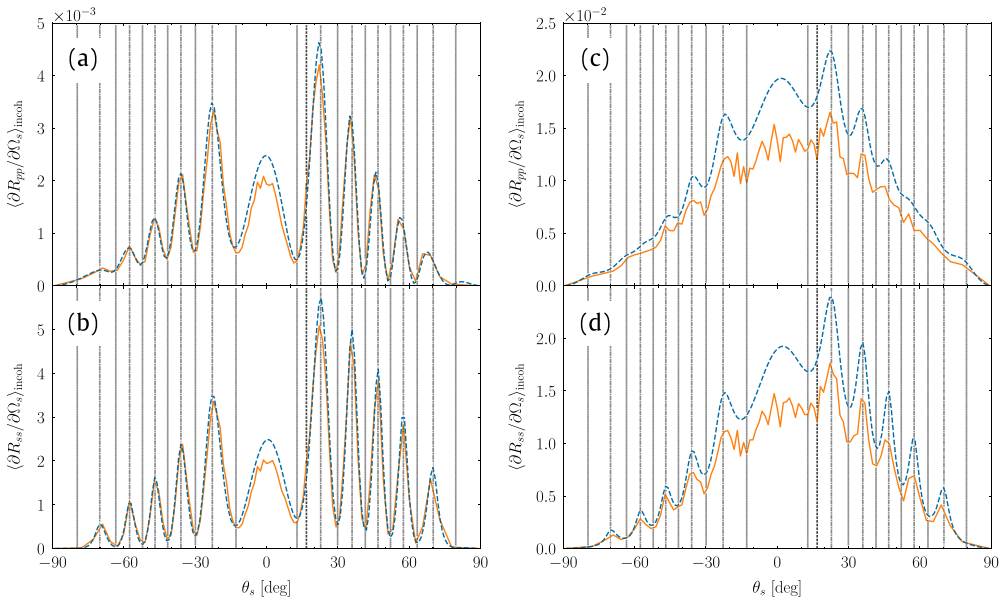


Fig. 2. Incoherent components of the mean DRCs for in-plane co-polarized scattering as functions of the polar angle of scattering, θ_s (note the convention $\theta_s < 0$ for $\phi_s = \phi_0 + 180°$). The light of wavelength $\lambda = 632.8$ nm was incident from vacuum on the rough photoresist film supported by a silicon substrate [$\epsilon_1 = 1.0, \epsilon_2 = 2.69, \epsilon_3 = 15.08 + 0.15i$]. The surface-height correlation length of the rough Gaussian correlated surface was $a = \lambda/3$, the mean film thickness was $d = 8\lambda$, and the angles of incidence were $(\theta_0, \phi_0) = (16.8°, 0°)$ in all cases. Panels (a) and (b) correspond to cases where only the top interface was rough, while panels (c) and (d) present the results for a film where only the bottom interface of the film is rough. In both cases, the rms-roughness of the rough interface was set to $\sigma = \lambda/30$. The results obtained on the basis of the non-perturbative method are shown as solid lines while those obtained with the perturbative method, Eq. (32), are shown as dashed lines. The position of the specular direction in reflection is indicated by the vertical dashed lines. The vertical dash-dotted and dotted lines indicate the angular positions of the maxima and minima predicted by Eq. (36), respectively.

of parameters assumed and their angular positions agree well with the expected angular positions (dashed–dotted vertical lines in Figs. 2(a)–(b)). When the lower surface is rough, the results presented in Figs. 2(c)–(d) show that the agreement between the two calculation methods is still satisfactory, but a larger discrepancy between them is now observed relative to what was found when the upper surface was rough. This larger discrepancy might be due to the fact that the error between the perturbative solution and the exact solution grows with the ratio of the dielectric constants of the media that are separated by the rough interface. Since the dielectric contrast between the silicon substrate and the photoresist film is larger than that between the photoresist film and vacuum, the corresponding error is also larger. Since the perturbative method is employed only to leading order, these agreements overall indicate that the physical phenomena that give rise to the scattered intensity distributions are well approximated as single scattering phenomena, at least for weakly rough surfaces.

We identify the interference fringes in Fig. 2 as in-plane scattering distributions of Selényi rings [9]. These rings are known to be centered around the mean surface normal, with their *angular position* being independent of the angle of incidence. Their *amplitude*, however, is modulated by the angle of incidence. This can indeed be observed if we vary the angle of incidence and record the resulting in-plane co-polarized angular scattering distributions, presented as contour plots in the first two columns of Fig. 3. Figs. 3(a)–(b) present, for *p*-polarized light, contour plots of the (θ_0, θ_s) dependence of the in-plane co-polarized incoherent component of the mean DRC when the top or bottom interface of the film is rough, respectively. Similar results but for *s*-polarized light are presented in Figs. 3(f)–(g). For both configurations, the co-polarized incoherent component of the mean DRC exhibits maxima that occur on a regular grid of (θ_0, θ_s)-points for *s*-polarized light [Figs. 3(f)–(g)]. A similar pattern

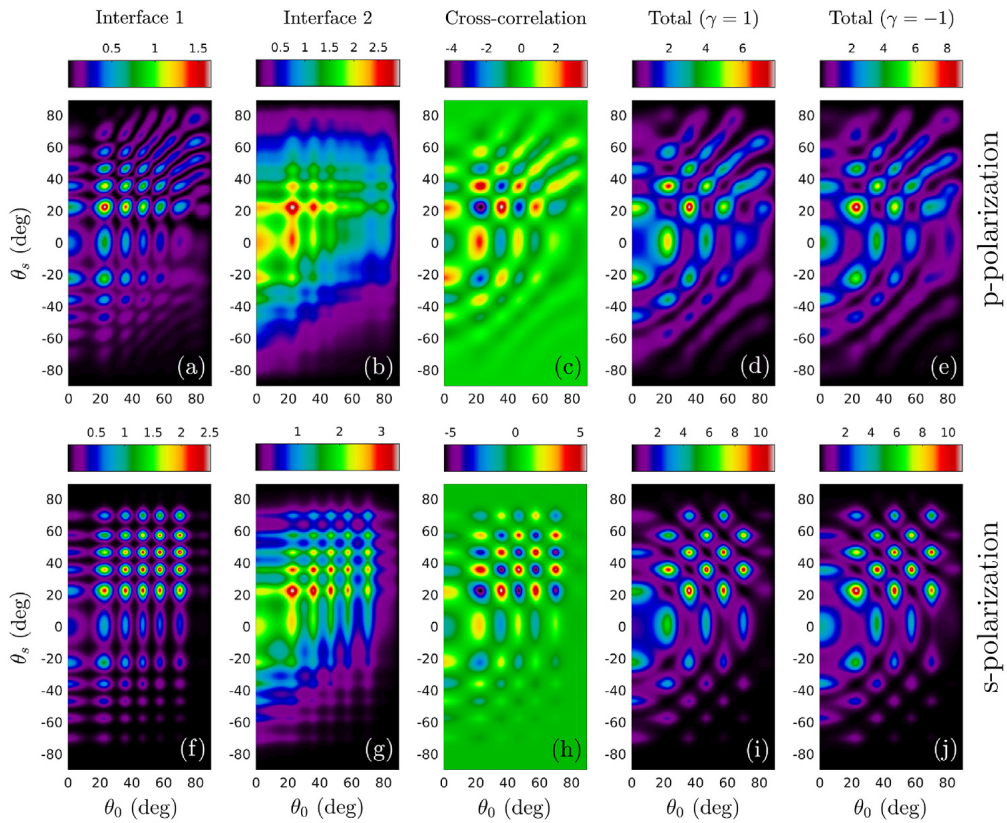


Fig. 3. Scaled incoherent component of the mean DRCs for in-plane co-polarized scattering, $100 \times \langle \partial R_{\alpha\alpha}/\partial \Omega_s \rangle_{\text{incoh}}$, as functions of the polar angle of incidence θ_0 and the polar angle of scattering θ_s obtained on the basis of Eq. (32). The first row of sub-figures [Figs. 3(a)–(e)] corresponds to *p*-polarized light (as marked in the figure), while the second row [Figs. 3(f)–(j)] corresponds to *s*-polarized light. These results were obtained under the assumption that the wavelength in vacuum was $\lambda = 632.8$ nm, the mean film thickness was $d = 8\lambda$, and the dielectric constants of the media were $\epsilon_1 = 1.0$, $\epsilon_2 = 2.69$, $\epsilon_3 = 15.08 + 0.15i$. The rms-roughness of the rough interfaces of the film were assumed to be $\sigma_1 = \sigma_2 = \lambda/30$, and the Gaussian correlation functions were characterized by the correlation length $a = \lambda/3$. The first column of sub-figures presents contour plots of the mean DRCs for a film geometry where only the top interface of the film is rough and the bottom interface planar. The second column shows similar results when the top film interface is planar and the bottom film interface is rough. In the third column, contour plots of *only* the cross-correlation term in Eq. (32) – that is, the contribution to the mean DRC produced by the last term in the square brackets of this equation – are depicted assuming a perfect correlation [$\gamma = 1$] between the rough top and rough bottom interface of the film. Finally, in the fourth and fifth column, contour plots of the total mean DRCs obtained on the basis of Eq. (32) are presented for two-rough-interface film geometries characterized by $\gamma = 1$ and $\gamma = -1$, respectively.

is observed for *p*-polarized light in Figs. 3(a)–(b), although the grid of maxima appears to lose some of its regularity for the larger polar angles of incidence and scattering [Figs. 3(a)–(b)]. We speculate that this is due to a Brewster effect, both in its traditional sense and through the Brewster scattering angles [19,20,26], but we will not delve further on this behavior here. In addition, by comparing the results presented in Figs. 2, 3(a)–(b), and 3(f)–(g), we note that the contrast in the interference pattern is better for the configurations where the top interface is rough than for those where the bottom interface is rough. In the following we will explain these observations in terms of a single scattering model which is an extension of the model previously proposed by Lu and co-workers [4].

Lu et al. suggested that, for sufficiently small roughness, the main effect of the rough interface is to produce scattered waves that cover a wide range of scattering angles both inside and outside the film, and the film may then be considered to approximate a planar waveguide for subsequent reflections and refractions within the film. This claim is supported by the observed agreement between the

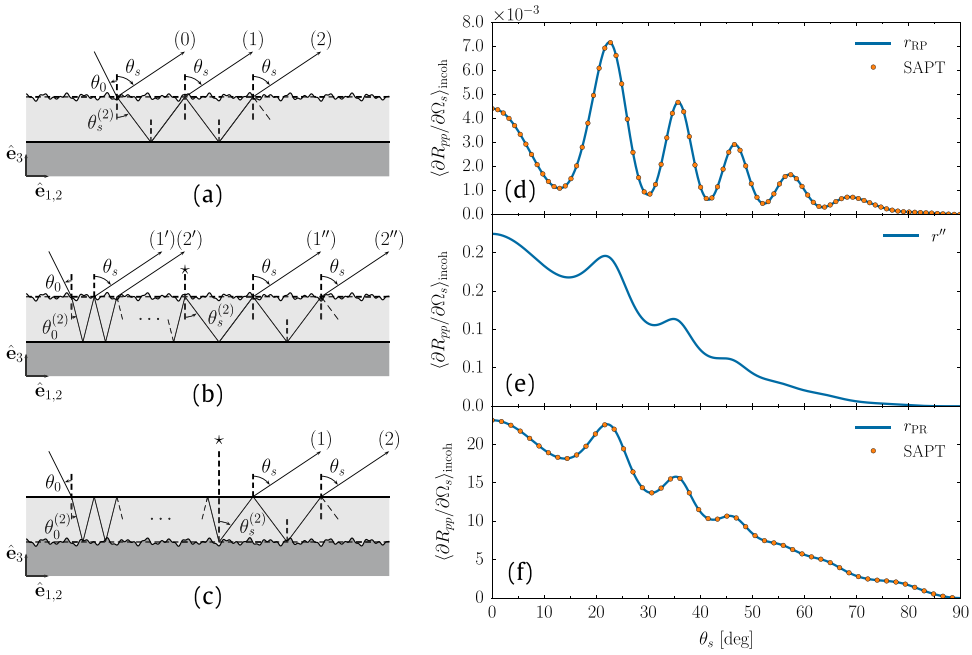


Fig. 4. Sketch of the optical paths involved in the single scattering model in the case of scattering from the top surface (a) and (b), or from the bottom interface (c). Incoherent component of the mean differential reflection coefficient for in-plane co-polarized scattering as a function of the polar angle of scattering for normal incidence for *p*-polarization (d) to (f). Apart from the angle of incidence the remaining parameters are the same as those from Fig. 2. In panels (d) and (f), the results were obtained from SAPT (circles), and from the single scattering model Eqs. (40)(d) and (41)(e) (solid line) respectively for the cases illustrated in (a–b) and (c). In panel (e), only the contribution of r'' (Eq. (39)) to the incoherent component of the mean DRC is shown.

mean DRC distributions obtained through the perturbative solution to leading order, whose physical interpretation is to take only single scattering events into account, and the full solutions of the RRE in Fig. 2, since the latter method allows for the full range of multiple scattering events. As the incident light interacts with the rough interface, whether it is located at the top or bottom interface, multiple wave components are generated in the film. These waves then undergo multiple specular reflections within the film while also being partially refracted back into the medium of incidence. Since Lu et al. only investigated the case where the rough interface is on top, their results were adequately explained under the assumption that the incident light was scattered by the rough interface during its first encounter with the interface. However, a more detailed analysis of the possible optical paths in the system is necessary in order to fully understand the case where the rough interface is at the bottom of the film, as illustrated by the more complete depiction of optical paths in Figs. 4(a)–(c). We will now analyze the different optical paths involving a single scattering event in the two configurations in more detail, and also construct a model for the resulting reflection amplitudes. Let $r_{ji}(\mathbf{p} \mid \mathbf{p}_0)$ and $t_{ji}(\mathbf{p} \mid \mathbf{p}_0)$ denote the reflection and transmission amplitudes obtained by small amplitude perturbation theory to first order in the surface profile separating *two* media with dielectric constants ϵ_i and ϵ_j (with the incident wave in medium i). Note that these amplitudes are different from those obtained for the full system considered in this paper. The expressions for these reflection amplitudes can be found e.g. in Refs. [19,20]. Moreover, let $r_{ji}^{(\mathrm{F})}(\mathbf{p})$ and $t_{ji}^{(\mathrm{F})}(\mathbf{p})$ represent the corresponding Fresnel amplitudes. All the amplitudes considered here may represent either *p*-polarization or *s*-polarization as we treat in-plane co-polarized scattering for simplicity.

In the case where only the top interface is rough the scattering event may occur on the first intersection between the path and the top interface, yielding a reflected scattered path denoted (0) in Fig. 4(a). Alternatively, on the first intersection the scattering event may yield a refracted (and

scattered) wave in the film. Since the single scattering event allowed in our analysis has then occurred, subsequent reflections within the film and refractions through the top interface are treated according to Snell's law of reflection and refraction, resulting in the paths denoted (1) and (2) (and so on) in Fig. 4(a). With each such non-scattering interaction with an interface, the reflection/transmission amplitude associated with the path is given by the Fresnel amplitude. Following the different paths depicted in Fig. 4(a) and summing the corresponding (partial) reflection amplitudes we obtain the following reflection amplitude:

$$\begin{aligned} r(\mathbf{p}|\mathbf{p}_0) &= r_{21}(\mathbf{p}|\mathbf{p}_0) + t_{12}^{(\mathrm{F})}(\mathbf{p})\, r_{32}^{(\mathrm{F})}(\mathbf{p})\, t_{21}(\mathbf{p}|\mathbf{p}_0) \exp(2i\varphi_s) \sum_{n=0}^{\infty} \left[r_{12}^{(\mathrm{F})}(\mathbf{p})\, r_{32}^{(\mathrm{F})}(\mathbf{p})\, \exp(2i\varphi_s) \right]^n \\ &= r_{21}(\mathbf{p}|\mathbf{p}_0) + \frac{t_{12}^{(\mathrm{F})}(\mathbf{p})\, r_{32}^{(\mathrm{F})}(\mathbf{p})\, t_{21}(\mathbf{p}|\mathbf{p}_0)\, \exp(2i\varphi_s)}{1 - r_{12}^{(\mathrm{F})}(\mathbf{p})\, r_{32}^{(\mathrm{F})}(\mathbf{p}) \exp(2i\varphi_s)}\,, \end{aligned} \tag{34}$$

where $\varphi_s = 2\pi\sqrt{\epsilon_2}d\cos\theta_s^{(2)}/\lambda$. The positions of the maxima in the resulting angular intensity distribution $|r(\mathbf{p}|\mathbf{p}_0)|^2$ are consistent with the predictions given by Lu et al. [4]. The difference in optical path length between path (0) and (1), and between (1) and (2), and more generally between two such consecutive paths, can be expressed as

$$\Delta = 2\sqrt{\epsilon_2}d\cos\theta_s^{(2)}, \tag{35}$$

where θ_s in the vacuum is related to $\theta_s^{(2)}$ in the film by $\sqrt{\epsilon_2}\sin\theta_s^{(2)} = \sqrt{\epsilon_1}\sin\theta_s$ according to Snell's law. The polar angles of scattering for which the diffusely scattered intensity has local maxima are given by

$$\frac{2\pi\sqrt{\epsilon_2}d}{\lambda}\cos\theta_s^{(2)} = \frac{2\pi d}{\lambda}\left(\epsilon_2 - \epsilon_1\sin^2\theta_s\right)^{1/2} = (\nu + 1/2)\pi\,, \tag{36a}$$

while the positions of the minima are determined from the relation

$$\frac{2\pi\sqrt{\epsilon_2}d}{\lambda}\cos\theta_s^{(2)} = \frac{2\pi d}{\lambda}\left(\epsilon_2 - \epsilon_1\sin^2\theta_s\right)^{1/2} = \nu\pi\,, \tag{36b}$$

where $\nu \in \mathbb{Z}$. The angular positions of the maxima and minima predicted by Eq. (36) are indicated by vertical dash-dotted and dotted vertical lines, respectively, in Fig. 2, and these predictions agree well with the maxima and minima that can be observed in the in-plane co-polarized mean DRC. Equation (36) does not depend on the polar angle of incidence θ_0, which supports the observation that the positions of the maxima and minima of the incoherent components of the mean DRC do not move with angle of incidence for weakly rough films. However, the modulation of the fringes with the angle of incidence cannot be explained if we consider solely the paths depicted in Fig. 4(a). Indeed, additional paths involving a single scattering event may be drawn as illustrated in Fig. 4(b). It is possible for the incident path not to experience a scattering event when it encounters the top interface for the first time, and it may also bounce within the film an arbitrary number of times before it experiences a scattering event while finally being refracted into the vacuum. Such paths are denoted (1′) and (2′) in Fig. 4(b). The resulting (partial) reflection amplitude corresponding to the "single-primed" paths in Fig. 4(b) reads

$$\begin{aligned} r'(\mathbf{p}|\mathbf{p}_0) &= t_{12}(\mathbf{p}|\mathbf{p}_0)\, r_{32}^{(\mathrm{F})}(\mathbf{p}_0)\, t_{21}^{(\mathrm{F})}(\mathbf{p}_0) \exp(2i\varphi_0) \sum_{n=0}^{\infty} \left[r_{12}^{(\mathrm{F})}(\mathbf{p}_0)\, r_{32}^{(\mathrm{F})}(\mathbf{p}_0)\, \exp(2i\varphi_0) \right]^n \\ &= \frac{t_{12}(\mathbf{p}|\mathbf{p}_0)\, r_{32}^{(\mathrm{F})}(\mathbf{p}_0)\, t_{21}^{(\mathrm{F})}(\mathbf{p}_0) \exp(2i\varphi_0)}{1 - r_{12}^{(\mathrm{F})}(\mathbf{p}_0)\, r_{32}^{(\mathrm{F})}(\mathbf{p}_0)\, \exp(2i\varphi_0)}, \end{aligned} \tag{37}$$

where $\varphi_0 = 2\pi\sqrt{\epsilon_2}d\cos\theta_0^{(2)}/\lambda$. The difference in optical path length between path (1′) and (2′) is given by

$$\Delta = 2\sqrt{\epsilon_2}d\cos\theta_0^{(2)}, \tag{38}$$

where $\sqrt{\epsilon_2}\sin\theta_0^{(2)} = \sqrt{\epsilon_1}\sin\theta_0$ according to Snell's law. Hence, we again obtain a series of maxima and minima in the mean DRC if we replace $\theta_s^{(2)}$ by $\theta_0^{(2)}$ in Eq. (36), but this time the positions of the maxima and minima are indeed a function of the polar angle of incidence θ_0. This interference phenomenon serves to modulate the intensity of the Selényi interference patterns. The static fringe pattern and the modulation introduced by the angle of incidence is clearly observed in the in-plane scattered intensities displayed in Fig. 3(a) and (f). However, we still have more optical paths to take into account. Indeed, paths yielding outgoing paths of type (1′) and (2′) may experience a scattering event while being reflected on the top surface instead of being refracted into the vacuum. Such a scattering event is indicated by the star in Fig. 4(b), and thereon the path may be reflected within the film an arbitrary number of times before being refracted into the vacuum as depicted by the paths denoted (1″) and (2″) in Fig. 4(b). In order to obtain the reflection amplitudes corresponding to all such paths, it suffices to multiply the overall reflection amplitude for *all* paths bouncing *any* arbitrary number of times with an angle $\theta_0^{(2)}$ within the film before the scattering event, with the overall reflection amplitude of *all* paths starting from the scattering event and bouncing *any* number of times within the film before being refracted into the vacuum. In this way we obtain the reflection amplitude

$$\begin{aligned} r''(\mathbf{p}|\mathbf{p}_0) &= t_{21}^{(\mathrm{F})}(\mathbf{p}_0)\, r_{32}^{(\mathrm{F})}(\mathbf{p}_0)\, \exp(2i\varphi_0) \sum_{n=0}^{\infty} \left[r_{12}^{(\mathrm{F})}(\mathbf{p}_0)\, r_{32}^{(\mathrm{F})}(\mathbf{p}_0)\, \exp(2i\varphi_0) \right]^n \\ &\quad \times t_{12}^{(\mathrm{F})}(\mathbf{p})\, r_{32}^{(\mathrm{F})}(\mathbf{p})\, r_{12}(\mathbf{p}|\mathbf{p}_0)\, \exp(2i\varphi_s) \sum_{n'=0}^{\infty} \left[r_{12}^{(\mathrm{F})}(\mathbf{p})\, r_{32}^{(\mathrm{F})}(\mathbf{p})\, \exp(2i\varphi_s) \right]^{n'} \\ &= \frac{t_{12}^{(\mathrm{F})}(\mathbf{p})\, r_{32}^{(\mathrm{F})}(\mathbf{p})\, r_{12}(\mathbf{p}|\mathbf{p}_0)\, r_{32}^{(\mathrm{F})}(\mathbf{p}_0)\, t_{21}^{(\mathrm{F})}(\mathbf{p}_0)\, \exp(2i(\varphi_0+\varphi_s))}{\left[1 - r_{12}^{(\mathrm{F})}(\mathbf{p})\, r_{32}^{(\mathrm{F})}(\mathbf{p})\, \exp(2i\varphi_s)\right] \left[1 - r_{12}^{(\mathrm{F})}(\mathbf{p}_0)\, r_{32}^{(\mathrm{F})}(\mathbf{p}_0)\, \exp(2i\varphi_0)\right]}. \end{aligned} \tag{39}$$

Note that the paths (1″) and (2″) are somewhat ill-defined in Fig. 4(b). Indeed, each path represents a family of paths with different history prior to the scattering event. For a given path, the path prior to the scattering event consists of a number of specular reflections within the film for which amplitudes dependent on the angle of incidence θ_0, as seen previously for the paths represented by r', while the path that follows after the scattering event consists of a number of specular reflections within the film which are dependent on the angle of scattering θ_s. Therefore, the phase difference between any two such paths will, in general, contain an integer combination of φ_0 and φ_s depending on the number of bounces prior to and after the scattering event. Eq. (39) hence contains both φ_0 and φ_s. The total reflection amplitude for all possible paths involving a single scattering event for the rough-planar (RP) film [Figs. 4(a) and (b)] is obtained by summing the amplitudes obtained from all the previously analyzed diagrams, namely

$$r_{\mathrm{RP}}(\mathbf{p}|\mathbf{p}_0) = r(\mathbf{p}|\mathbf{p}_0) + r'(\mathbf{p}|\mathbf{p}_0) + r''(\mathbf{p}|\mathbf{p}_0). \tag{40}$$

The intensity distribution corresponding to Eq. (40) is shown in Fig. 4(d) for normal incidence and *p*-polarized light, and is compared to results based on small amplitude perturbation theory to leading order, Eq. (32), in the case where only the top interface is rough. The two results are literally indistinguishable. Similar results were also found in the case of *s*-polarized light, but the results are not shown (in order to keep the figure simple). These findings strongly suggest that the two methods are equivalent. In particular, this means that the perturbative solution to leading order derived in Appendix A can indeed be interpreted as a sum of all paths involving a single scattering event, although this was not obvious from the derivation itself. The model presented here thus justifies this physical picture. Fig. 4(e) shows the incoherent contribution to the in-plane co-polarized mean DRC one would obtain if *only* paths of type (1″), (2″), and so on were present, in other words the intensity distribution resulting from Eq. (39). The relative contribution from r'' [Fig. 4(e)] to r_{RP} [Fig. 4(d)] is so small that it to some approximation may be ignored, as it was in Ref. [4], but we will soon see that this path type is crucial in the case of a system with the rough interface shifted to the bottom of the film.

Let us now analyze the case where only the bottom interface is rough, as illustrated in Fig. 4(c). If we follow paths (1) and (2) in Fig. 4(c), it becomes evident that a path must first undergo a Snell refraction

from vacuum into the film before it may interact with the rough interface. Following this refraction into the film a given path may undergo an arbitrary number of Snell reflections within the film, now at a polar angle $\theta_0^{(2)}$ with the normal to the mean film interfaces, before it is scattered by the rough interface as indicated by the star in Fig. 4(c). The path then performs an arbitrary number of Snell reflections within the film, now at a polar angle of scattering $\theta_s^{(2)}$ with the normal to the mean film interfaces, before it exits into the vacuum. All possible paths involving a single scattering event are for the present configuration depicted in Fig. 4(c), and it is now immediately evident that these paths bear close resemblance to those shown in Fig. 4(b) which correspond to the amplitude r''. Consequently the resulting intensity pattern associated with the paths in Fig. 4(c) will exhibit, by construction of the paths, dependencies on *both* the polar angles of incidence and scattering as given by Eqs. (38) and (35). This is supported both by the resulting reflection amplitude [Eq. (41)] and the angular positions of the maxima and minima of the in-plane co-polarized mean DRC displayed in Figs. 2(c) and (d), indicated as vertical dashed–dotted and dotted lines, respectively. Similar to what was done for the paths of type (1″) and (2″) in the configuration depicted in Fig. 4(b), the resulting reflection amplitude for the paths shown in Fig. 4(c) can be expressed as the product of the partial reflection amplitude resulting from all possible paths *prior* to the scattering event and the partial reflection amplitude resulting from all possible paths that may follow *after* the scattering event. The resulting reflection amplitude for the planar-rough (PR) film [Fig. 4(c)] obtained in this way reads

$$\begin{aligned} r_{\mathrm{PR}}(\mathbf{p}|\mathbf{p}_0) &= t_{21}^{(\mathrm{F})}(\mathbf{p}_0)\,\exp(i\varphi_0)\sum_{n=0}^{\infty}\left[r_{12}^{(\mathrm{F})}(\mathbf{p}_0)\,r_{32}^{(\mathrm{F})}(\mathbf{p}_0)\,\exp(2i\varphi_0)\right]^n \\ &\quad\times t_{12}^{(\mathrm{F})}(\mathbf{p})\,r_{32}(\mathbf{p}|\mathbf{p}_0)\,\exp(i\varphi_s)\sum_{n'=0}^{\infty}\left[r_{12}^{(\mathrm{F})}(\mathbf{p})\,r_{32}^{(\mathrm{F})}(\mathbf{p})\,\exp(2i\varphi_s)\right]^{n'} \\ &= \frac{t_{12}^{(\mathrm{F})}(\mathbf{p})\,r_{32}(\mathbf{p}|\mathbf{p}_0)\,t_{21}^{(\mathrm{F})}(\mathbf{p}_0)\,\exp(i(\varphi_0+\varphi_s))}{\left[1-r_{12}^{(\mathrm{F})}(\mathbf{p})\,r_{32}^{(\mathrm{F})}(\mathbf{p})\,\exp(2i\varphi_s)\right]\left[1-r_{12}^{(\mathrm{F})}(\mathbf{p}_0)\,r_{32}^{(\mathrm{F})}(\mathbf{p}_0)\,\exp(2i\varphi_0)\right]}. \end{aligned} \tag{41}$$

The intensity pattern predicted by Eq. (41) is presented as a solid line in Fig. 4(f) for normal incident *p*-polarized light; in the same figure, the filled circles represent the prediction from Eq. (32). As was the case when only the top interface was rough, we find an excellent agreement between the two approaches also when only the bottom interface is rough. A similar agreement was also found when the incident light was *s*-polarized (results not shown). These findings support our single scattering interpretation of the perturbative solution to leading order. We have now explained the angular positions of the Selényi rings and their amplitude modulation with the angle of incidence based on optical path analysis.

It remains to explain the difference in contrast observed in the interference patterns corresponding to the geometries where the rough surface is either located on the top of the film or at the bottom of the film (with the other film interface planar). In providing such an explanation, the expressions given by Eqs. (40) and (41) will prove to be useful alternative representations of the perturbative solutions of the RRE to leading order. Indeed, we can now investigate the relative contribution from each type of path by artificially removing terms. In our analysis of the type of paths in the two configurations, we have identified that paths of type (1″) and (2″), in the configuration where the top interface is rough, are similar to paths (1) and (2) for the configuration where the bottom interface is rough. As was mentioned previously, Fig. 4(e) shows the (diffuse) in-plane mean DRC we would obtain if *only* paths of type (1″), (2″), *etc.* were present; in other words the scattering intensities originating in Eq. (39). We observe that the curve in Fig. 4(e) exhibits poor contrast, and is very similar to the scattering intensities observed in the case where the bottom interface is rough [Fig. 4(f)]. This clearly hints towards the idea that the poor contrast observed when the bottom film interface is rough is intrinsically linked to the nature of the paths. Moreover, we have seen that ignoring the contribution from r'' in Eq. (40) gives a result similar to when all terms of the same equation are included. This indicates that the contribution from r'' can be neglected relative to the other two terms in Eq. (40). However, since paths similar to (1″), (2″), *etc.* are the *only* paths allowed for the configuration where the bottom interface is rough, the contrast is poor by default. In both cases, and as we have seen, a

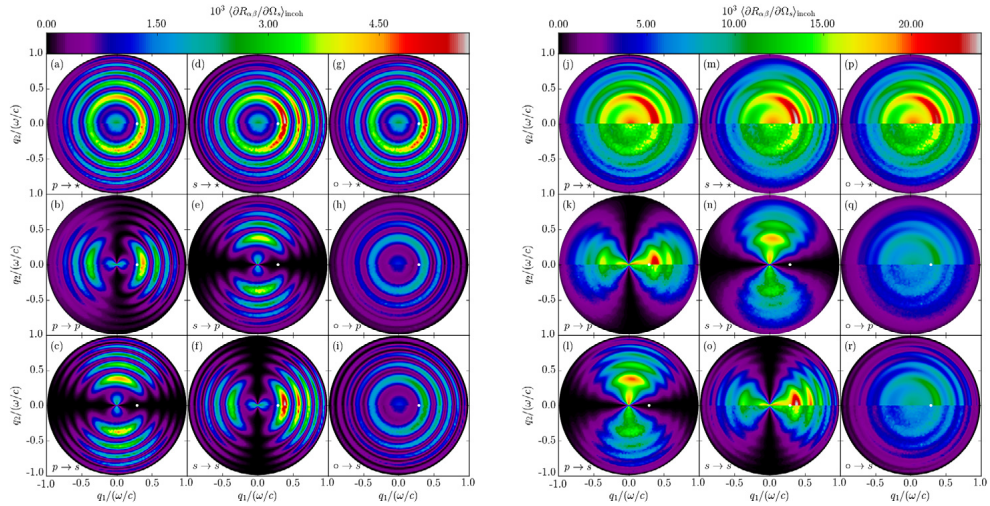


Fig. 5. The full angular distribution of the incoherent component of the mean DRC, $\left\langle \partial R_{\alpha\beta}/\partial\Omega_s \right\rangle_{\text{incoh}}$, as function of the lateral wave vector $\mathbf{q}$ of the light that is scattered from a rough film where either the top interface is rough [Figs. 5(a)–(i)] or the bottom interface is rough [Figs. 5(j)–(r)] and the other interface of the film is planar. The light of wavelength $\lambda = 632.8$ nm was incident from vacuum on the rough photoresist film supported by a silicon substrate [$\epsilon_1 = 1.0, \epsilon_2 = 2.69, \epsilon_3 = 15.08 + 0.15i$]. The rms-roughness of the rough film interface was $\sigma_1 = \lambda/30$, $\sigma_2 = 0$ [Figs. 5(a)–(i)] and $\sigma_1 = 0$, $\sigma_2 = \lambda/30$ [Figs. 5(j)–(r)]. The surface-height correlation length was $a = 211\text{nm} = \lambda/3$, the film thickness was $d = 5062.4\text{nm} = 8\lambda$ and the angles of incidence were $(\theta_0, \phi_0) = (16.8^\circ, 0^\circ)$ for all panels. The positions of the specular directions in reflection are indicated by white dots. The remaining parameters assumed for the scattering geometry and used in performing the numerical simulations had values that are identical to those assumed in obtaining the results of Fig. 2. The upper halves of all panels are results from the small amplitude perturbation method to leading order, while the lower halves show results obtained through the non-perturbative solutions of the RRE. The sub-figures in Figs. 5(a)–(i) and (j)–(r) are both organized in the same manner and show how incident β-polarized light is scattered by the one-rough-interface film geometry into α-polarized light [with $\alpha = p, s$ and $\beta = p, s$] and denoted $\beta \to \alpha$. Moreover, the notation $\circ \to \star$ is taken to mean that the incident light was unpolarized while the polarization of the scattered light was not recorded. For instance, this means that the data shown in Fig. 5(a) are obtained by adding the data sets presented in Figs. 5(b)–(c); similarly, the data shown in Fig. 5(g) result from the addition and division by a factor two of the data sets presented in Figs. 5(a) and (d); *etc.*. Finally, the in-plane intensity variations from Figs. 5(b, f) and (k, o) are the curves depicted in Figs. 2(a)–(b) and (c)–(d), respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

typical path must undergo a number of non-scattering reflections within the film both before *and* after the scattering event occurs. Consequently, the phase difference between any two such paths will in general involve integer combinations of φ_0 and φ_s, as can be seen from Eqs. (39) and (41). This phase mixing is the fundamental reason for the difference in contrast found in the contributions to the total intensity r_{RP} from the three components of Eq. (40). The difference in contrast can also be investigated mathematically by estimating the contrast directly, as explained in Appendix C.

We now turn to the full angular distributions for the mean DRC. Figs. 5(a)–(i) and (j)–(r) show the full angular distributions of the incoherent contribution to the mean DRC, for simulation parameters corresponding to those assumed in obtaining the results of Figs. 2(a)–(b) and (c)–(d), respectively. In fact, the non-perturbative results presented in Figs. 2(a)–(b) and (c)–(d) correspond to in-plane cuts along the q_1 axis from Figs. 5(b, f, k, o). The results of Fig. 5 show that, in addition to the interference phenomena already mentioned, the distributions of the incoherent contributions to the mean DRC are also weighted by the shifted power spectrum of the rough interface. In the current work this is a Gaussian envelope centered at the angle of specular reflection, where the width of the envelope is directly influenced by the surface-height correlation length a. This is shown explicitly in the case of small amplitude perturbation theory to leading order as the term $g(\mathbf{p} - \mathbf{p}_0)$ in Eq. (32), and its impact on the scattering distributions should not be confused with the interference phenomena.

The reader may verify that the maxima and minima are located at the same positions as predicted for Fig. 2, as is predicted by Eq. (36). However, for Figs. 5(j)–(r) the contrast in the oscillations of the incoherent contribution to the mean DRC is now less pronounced, as explained for in-plane scattering.

The lower left 2×2 panels in each of the panel collections in Fig. 5 display overall dipole-like patterns oriented along the plane of incidence for co-polarized scattering and perpendicular to it for cross-polarized scattering. These features are consequences of the definition used for the polarization vectors of our system. They are similar to the scattered intensity patterns obtained in recent studies of light scattering from single two-dimensional randomly rough surfaces [16,19,20,27–29].

5.2. Two rough interfaces

We will now turn to the discussion of the geometry where *both* the top and bottom interfaces of the film are rough. In the following it will be assumed that these rough interfaces are characterized by Eq. (6), and for simplicity it will be assumed that their rms-roughness are the same and equal to $\sigma_1 = \sigma_2 = \lambda/30$. The cross-correlation between these two interfaces is characterized by the parameter γ which is allowed to take values in the interval from -1 to 1. All the remaining experimental parameters are identical to those assumed in the preceding sections of this paper.

For the case where only one of the two interfaces of the film was rough, we demonstrated that good agreement exists between the results obtained by a purely numerical solution of the RRE and those obtained on the basis of a perturbative solution of the same equation [SAPT]. A purely numerical solution of the RRE associated with a film geometry where more than one of the interfaces are rough is a challenging task that requires extensive computational resources to obtain, and to the best of our knowledge such a purely numerical solution has not yet been reported. Therefore, for film geometries where both interfaces are rough we will only solve the corresponding RRE through SAPT to obtain the incoherent component of the mean DRC to second order in products of the surface profile functions, for which the relevant expression is given by Eq. (32). In the following it will be assumed that for the level of surface roughness that we consider here, which provided accurate results for the corresponding one-rough-interface film geometry considered in the preceding subsection, such a perturbative solution method is sufficiently accurate to adequately describe the physics of the problem under investigation.

The first set of scattering results for a film bounded by two rough interfaces is presented in Fig. 6. In particular, Figs. 6(a)–(c) present the incoherent component of the mean DRC for in-plane co-polarized scattering (*i.e.* $|\hat{\mathbf{p}} \cdot \hat{\mathbf{p}}_0| = 1$ and $\alpha = \beta$) as a function of the polar angle of scattering θ_s, for given polar angle of incidence equal to $\theta_0 = 0^\circ$, and for three extreme values of the cross-correlation parameter $\gamma \in \{0, 1, -1\}$. These three values of γ physically correspond to the situations of uncorrelated film interfaces; perfectly positively correlated interfaces so that the film thickness measured along any vertical line segment will be constant and equal to d; and perfectly negatively correlated or anti-correlated interfaces, respectively. From Fig. 6(a) one observes that for uncorrelated interfaces of the film [$\gamma = 0$], the number of interference fringes and their angular positions remain unchanged as compared to what was found when only one of the two interfaces of the film was rough. This is found to be the case for both p- and s-polarized incident light. Such behavior can easily be understood in terms of the expression in Eq. (32); when $\gamma = 0$ only the first two terms in the square brackets on the right-hand-side of this equation contribute. These two terms are the only non-zero contributions to the incoherent component of the mean DRC (to second order) for a film system bounded by two *uncorrelated* rough surfaces. Moreover, these two contributions are, respectively, identical to the incoherent component of the mean DRC obtained for film geometries where either the top or the bottom interface of the film is rough and the other planar. Summing these two contributions will hence result in summing two similar interference intensity patterns. Consequently, the resulting interference pattern maintains the same number of fringes at the same positions as the pattern obtained from scattering from the corresponding one-rough interface film geometry. However, by gradually introducing more cross-correlation between the two rough interfaces of the film [$\gamma \neq 0$], one observes that half of the fringes observed for the system for which $\gamma = 0$ are significantly attenuated whereas the other half are enhanced [Figs. 6(b) and (c)]. Furthermore, it is observed from the results in Figs. 6(a)–(c) that the fringes that are enhanced (attenuated) for the case when $\gamma = 1$

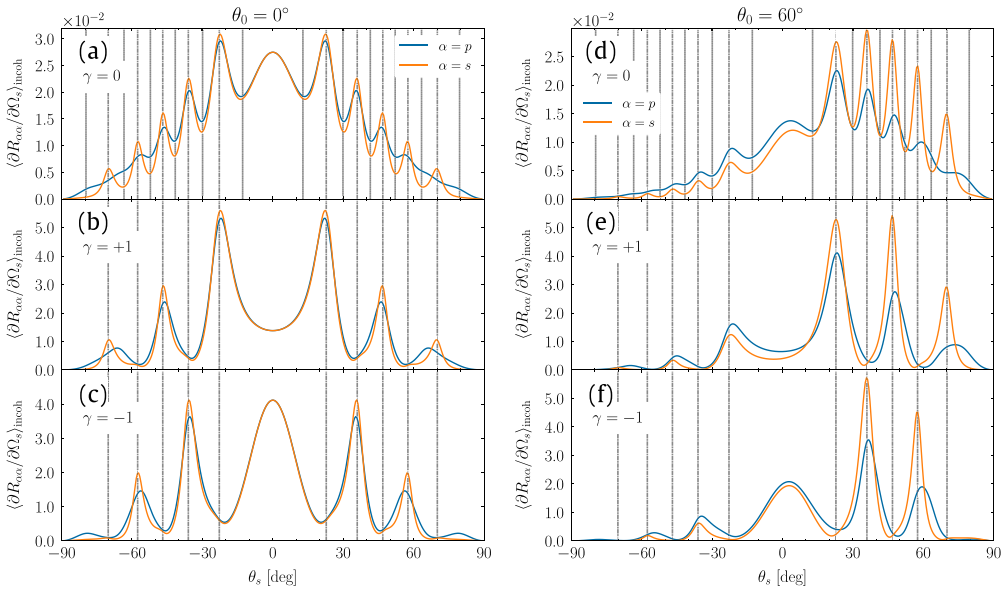


Fig. 6. Incoherent components of the mean differential reflection coefficients $\langle \partial R_{\alpha\alpha}/\partial\Omega_s \rangle_{\text{incoh}}$ for in-plane co-polarized scattering from a two-rough-interface film geometry for the polar angle of incidence $\theta_0 = 0°$ [Figs. 6(a)–(c)] and $\theta_0 = 60°$ [Figs. 6(d)–(e)]. The wavelength of the incident light was $\lambda = 632$ nm, the mean thickness of the film $d = 8\lambda$, and the dielectric constants of the media were $\epsilon_1 = 1.0, \epsilon_2 = 2.69, \epsilon_3 = 15.08 + 0.15i$. The rms-roughness of the interfaces were $\sigma_1 = \sigma_2 = \lambda/30$, and the Gaussian correlation functions of each of the surfaces were characterized by the correlation length $a = \lambda/3$. The cross-correlation function between the rough top and rough bottom interface of the film had the form (5) and was characterized by the parameter γ with values as indicated in each of the panels. The vertical dash-dotted and dotted lines indicate the expected angular positions of the maxima and minima of the scattered intensity as predicted by Eq. (36b), respectively. For reasons of clarity only the expected positions of the minima of the in-plane mean DRCs are indicated in Figs. 6(a) and (d).

are the fringes being attenuated (enhanced) for the case when $\gamma = -1$. This phenomenon can be attributed to the last term in the square brackets in Eq. (32) which is linear in γ and can take both positive and negative values and hence increase or decrease the value of the intensity pattern resulting from the superposition of the scattering amplitudes obtained for the two independent aforementioned one-rough-interface film geometries.

The last term in the square brackets of Eq. (32) is an interference term. Physically it can be interpreted as the interference between a path formed by a single scattering event occurring on the *top* interface of the film such as one depicted in Figs. 4(a–b), and a path consisting of a single scattering event taking place on the *bottom* interface as depicted in Fig. 4(c). When the two interfaces are uncorrelated, the phase difference between these two optical paths will form an uncorrelated random variable so that the ensemble average of the term where it appears in Eq. (32) will be zero and the mean DRC will equal the sum of the intensities of the two corresponding one-rough-interface geometries, i.e. it will be given by the two first terms of Eq. (32). However, when the two interfaces of the film are completely or partially correlated, $|\gamma| > 0$, the phase difference of these two paths becomes a correlated random variable so that the interference term – the last term in (32) – does not average to zero; this results in an optical interference effect. Consequently, the observed interference pattern for $|\gamma| > 0$ will obtain a non-zero contribution from the last term in the square brackets of Eq. (32), which thus will make it different from the pattern obtained for an uncorrelated film geometry that corresponds to $\gamma = 0$.

Figs. 6(d)–(f) present for polar angle of incidence $\theta_0 = 60°$ similar results to those presented in Figs. 6(a)–(c) for normal incidence. Except for the increased intensity of the light scattered into the forward direction defined by $\theta_s > 0°$ relative to what is scattered into angles $\theta_s < 0°$, and the increased contrast of the fringes observed for *s*-polarized light in the forward direction, the behavior of the mean DRC curves is rather similar for the two angles of incidence. In particular, for the same value

of γ, fringes are observed at the same angular positions for the two angles of incidence. Moreover, which of the fringes that are enhanced or attenuated by the introduction of (positive or negative) cross-correlation between the two rough interfaces of the film are also the same for the two angles of incidence. Such behavior is as expected for Selényi fringes.

A close inspection of the perturbative results presented in Fig. 6 reveals that for both $\theta_0 = 0°$ and $\theta_0 = 60°$ the angular positions of the maxima of the in-plane, co-polarized mean DRC curves are more accurately predicted by Eq. (36) for *s*-polarized light than for *p*-polarized light; this seems in particular to be the case for the larger values of $|\theta_s|$. We speculate that such behavior is related to a phase change associated with the Brewster scattering phenomenon [19,20,26] that exists in the case of *p*-polarized light, reminiscent of the well known phase change associated with the Brewster angle found for planar interfaces.

So far in our analysis of the two-rough-interface film geometry, we have observed that the enhancement or attenuation of the diffusely scattered co-polarized intensity are localized to regions around the polar angles determined by Eq. (36a). In order to make this observation more apparent, Figs. 3(a)–(e) present various terms, or combinations of terms, from Eq. (32) when the incident and scattered light is *p*-polarized; Figs. 3(f)–(j) depict similar results for *s*-polarized incident and scattered light. The three first columns of sub-figures that are present in Fig. 3 – labeled "Interface 1", "Interface 2", and "Cross-correlation" – represent the terms in Eq. (32) that contain the factors σ_1^2, σ_2^2, and $\sigma_1\sigma_2$, respectively. The cross-correlation terms, Figs. 3(c) and (h), where obtained from the last term of Eq. (32) with $\gamma = 1$. Furthermore, a contour plot that appears in the 4th column of Fig. 3 [labeled "Total ($\gamma = 1$)"], displays the sum of the data used to produce the three first mean DRC contour plots appearing in the same row. In other words, the results depicted in Figs. 3(d) and (i) are the contour plots of the incoherent component of the mean DRC for a film geometry bounded by two perfectly correlated rough interfaces and therefore given by the expression in Eq. (32) with $\gamma = 1$. Similarly, the incoherent component of the mean DRCs for a geometry where the two rough interfaces of the film are perfectly anti-correlated are displayed in the last column of Fig. 3 [Figs. 3(e) and (j)] and labeled "Total ($\gamma = -1$)". These latter results correspond to Eq. (32) with $\gamma = -1$, and can be obtained by summing the results of the two first columns and subtracting the result of the third column of Fig. 3.

The contour plots of the cross-correlation terms presented in Figs. 3(c) and (h), which are obtained under the assumption that $\gamma = 1$, display extrema localized on the same grid of points in the (θ_0, θ_s)-plane as the extrema of the incoherent component of the mean DRC obtained when only one of the film interfaces is rough [Figs. 3(a)–(b) and (f)–(g)]. An important observation should be made from these results. The minima of the former (the cross-correlation terms) are negative while the latter are always non-negative. Hence, the incoherent component of the mean DRC for $\gamma = 1$, which according to Eq. (32) corresponds to the addition of the results used to produce the three first columns of each row of Fig. 3, will cause fringes localized at the minima of the cross-correlation terms to be *attenuated* (or disappear) and those localized at the maxima of the cross-correlation terms to be *enhanced* [see Figs. 3(d) and (i)].

The preceding discussion stays valid when considering the full angular distribution of the incoherent component of the mean DRC. Fig. 7 presents the full angular distribution of the incoherent component of the mean DRC, obtained on the basis of Eq. (32), for the two polar angles of incidence $\theta_0 = 0°$ [Figs. 7(a)–(f)] and $\theta_0 = 60°$ [Figs. 7(g)–(l)]. In this figure, each column formed by the sub-plots corresponds to either *p*- or *s*-polarized incident light, and in all cases the polarization of the scattered light was not recorded. Moreover, each of the three rows of sub-figures that are present in Fig. 7 corresponds to different values for the cross-correlation parameter $\gamma \in \{0, 1, -1\}$ as indicated in the figure. From the results presented in Fig. 7 it should be apparent that what appear as fringes in the in-plane angular dependence of the mean DRCs indeed are expressed as interference rings in the full-angular distribution of the same quantity; this is particularly apparent for normal incidence where the intensity of the (Selényi) interference rings is independent of the azimuthal angle of scattering ϕ_s (due to the rotational invariance of the system and the source). The angular distributions in Figs. 7(a)–(f) also demonstrate very clearly how the possible interference rings present for uncorrelated interfaces of the film [$\gamma = 0$] are enhanced or attenuated when $|\gamma| \neq 0$, *i.e.* when cross-correlation exists between the two rough interfaces of the film.

Figs. 7(g)–(l) show that interference rings are also present for non-normal incidence and that they are present for the same polar scattering angles θ_s as was found for normal incidence. However, for

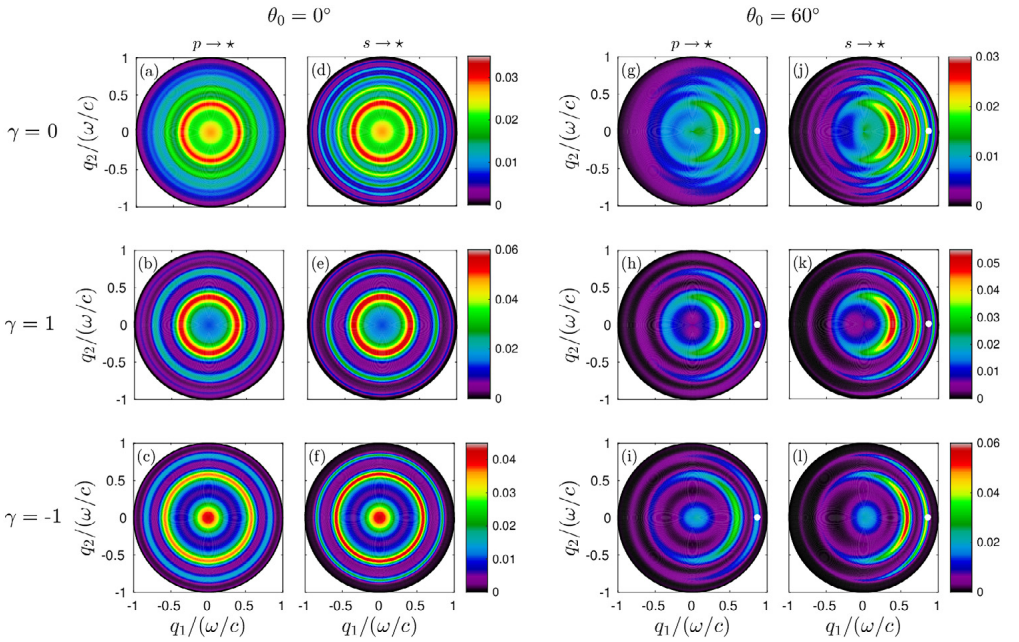


Fig. 7. The full angular distribution of the incoherent component of the mean DRC, $\langle \partial R_{\alpha\beta}/\partial\Omega_s \rangle_{\text{incoh}}$, for incident β-polarized light that is scattered by a two-rough-interface film geometry into α-polarized light [with $\alpha = p, s$ and $\beta = p, s$]. When the polarization of the scattered light is not observed, the relevant mean DRC quantity is $\sum_{\alpha=p,s} \langle \partial R_{\alpha\beta}/\partial\Omega_s \rangle_{\text{incoh}}$ and this situation is labeled as $\beta \to \star$. The reported results were obtained on the basis of SAPT, Eq. (32), and the polar angles of incidence were $\theta_0 = 0°$ [Figs. 7(a)–(f)] and $\theta_0 = 60°$ [Figs. 7(g)–(l)]. The incident in-plane wave vector is indicated by the white dot for non-normal incidence [Figs. 7(g)–(l)]. The cross-correlation function between the rough top and rough bottom interface of the film had the form (5) and was characterized by the parameter γ as indicated in the figure (and constant for each row of sub-figure). The remaining roughness parameters are identical to those assumed in producing the results presented in Fig. 6. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

non-normal incidence the intensity of the rings does depend on the azimuthal angle of scattering. It is found that the intensity of the interference rings are concentrated to the forward scattering plane [$|\phi_s - \phi_0| < 90°$].

For normal incidence Fig. 8 presents, for completeness, the full angular distribution of $\langle \partial R_{\alpha\beta}/\partial\Omega_s \rangle_{\text{incoh}}$ for all possible linear polarization couplings, *i.e.* from incident β-polarized light to scattered α-polarized light. The values assumed for the cross-correlation parameter in obtaining these results were $\gamma \in \{0, 1, -1\}$. It should be observed from the results of Fig. 8 that interference structures are observed but they are *not* ring structures of a constant amplitude as was seen in Figs. 7(a)–(f). The reason for this difference is that in the results presented in Fig. 8 only scattered light of a given linear polarization was observed; this contrasts with the situation assumed in producing Fig. 7 where all scattered light was observed and not only scattered light of a given linear polarization.

We have here only shown the extreme cases of cross-correlation, but one may also consider intermediate values for the cross-correlation parameter γ. The effect found for $\gamma = \pm 1$ remains also for $0 < |\gamma| < 1$ but with less pronounced enhancement and attenuation of the rings. The reader is invited to take a look at the animations in the Supplementary Materials, where the contour plots of the incoherent component of the mean DRCs are featured for smoothly varying cross-correlation parameter γ over the interval from -1 to 1, for both normal incidence and for $\theta_0 = 60°$ incidence.

5.3. Transmitted light

Finally, we would like to briefly comment on what would be observed in transmission if a non-absorbing medium was chosen, such as silica. No results will be presented here, but we have observed

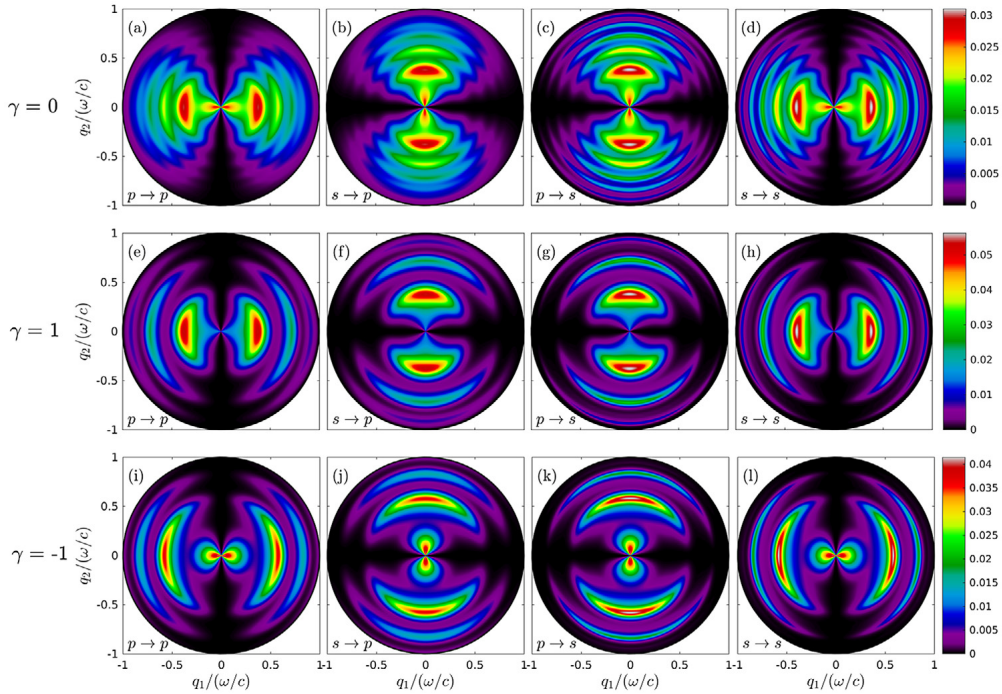


Fig. 8. The full angular distribution of the incoherent component of the mean DRC, $\langle \partial R_{\alpha\beta}/\partial\Omega_s \rangle_{\text{incoh}}$, for incident β-polarized light of polar angle $\theta_0 = 0^\circ$ that is scattered by a two-rough-interface film geometry into α-polarized light [with $\alpha = p, s$ and $\beta = p, s$] and labeled $\beta \to \alpha$ in the sub-figures. The cross-correlation function between the rough top and rough bottom interface of the film had the form (5) and was characterized by the parameter γ as marked in the figure. The reported results were obtained on the basis of SAPT, Eq. (32). The remaining experimental and roughness parameters are identical to those assumed in producing the results presented in Figs. 6 and 7.

that interference rings are also observed in the diffusely transmitted light and that the effect of enhancement and attenuation of the rings induced by the interface cross-correlation still holds. Furthermore, additional features attributed to the so-called Brewster scattering angles and Yoneda effects in the diffusely transmitted light would then be present. As presented in Ref. [20] for scattering systems of comparable surface roughness and materials, the diffusely transmitted intensity as a function of angle of transmission will be modulated by a typical Yoneda intensity pattern. At normal incidence this pattern exhibits a peak at some critical angle of scattering for *s*-polarized light and a vanishing intensity for *p*-polarized light (see Ref. [20] for details). However, we observed that not only did the overall intensity distribution undergo such modulation: the angular positions of the fringes were also affected compared to the predictions provided by naive optical path arguments, analogous to what was presented in this paper for reflection. The angular positions of the fringes predicted by optical path arguments leading to equations similar to Eq. (36) still hold for scattering angles below the Yoneda critical angle, but must be corrected for scattering angles larger than the Yoneda critical angle. We speculate that this is due to a gradual phase shift that occurs above the critical angle, and that it is associated with the Yoneda phenomenon. Note that this phenomenon is also observed in the diffusely reflected light if the medium of incidence has a higher refractive index than that of the substrate (i.e. $\epsilon_1 > \epsilon_3$) [19,22]. Moreover, we have also observed that when scattered to larger polar angles than the Brewster scattering angle the *p*-polarized transmitted light exhibits an additional phase shift, as compared to *s*-polarized transmitted light, resulting in a switch in the positions for the maxima and minima. These and other features of the interference rings in the diffusely transmitted light will be discussed in more detail in a dedicated paper.

6. Conclusion

Based on both non-perturbative and perturbative solutions of the reduced Rayleigh equation, we have in this paper demonstrated that for systems composed of two-dimensional weakly rough dielectric films, Selényi rings can be observed in the diffusely scattered light. These rings make up a static interference pattern that is modulated by the polar angle of incidence. We have illustrated that the interference mechanism at play can be explained by simple optical path arguments, leading to a simple model capable of predicting both the angular positions of the rings and the expected difference in contrast of the rings for film geometries where either the top or the bottom interface of the film is rough (but not both interfaces).

Furthermore, by investigating the influence of the cross-correlation between the film interfaces when both interfaces are rough, we have shown that a selective enhancement or attenuation of the interference rings in the diffusely scattered light can be observed. This suggests that the positions and the amplitudes of Selényi rings can, when combined with reflectivity and/or ellipsometry measurements, in principle enable the determination of the mean film thickness, the dielectric constant of the film material and the statistical properties of the interfaces. In particular, numerical experiments show that the cross-correlation between interfaces can be assessed. Alternatively, film geometries consisting of cross-correlated interfaces can be designed to control the intensity pattern of the diffusely scattered light that they give rise to. Sensors can also be designed in such a way that the interference rings observed for a clean system with known cross-correlated interfaces will be modified by the adsorption of a substance or nano-particles onto the first interface, hence partially destroying the effective cross-correlation between the interfaces. These possibilities are, however, likely to be limited by the ordering of length scales $d > \lambda > \sigma$, which expresses the fact that the film thickness must be on the order of a few wavelengths to observe interference rings in the diffusely scattered light and that the rms-roughness of the interfaces should be small compared to the wavelength. Such a length scale ordering combined with controlled interface cross-correlation may be challenging to achieve experimentally.

While the main results presented in this paper considered the diffusely scattered light, the theoretical framework that it presents also allows for the investigation into the light transmitted diffusely through transparent film structures with one or several rough interfaces. The developed theoretical framework is readily generalized to the case of an arbitrary number of correlated layers and allows, for example, for the study of the effect of gradually changing cross-correlations over many interfaces.

We hope that the results presented in this paper can motivate experimental investigations into the scattering of light from rough film systems so that the predictions that are reported here based on theoretical grounds can be confirmed experimentally.

Acknowledgments

This research was supported in part by The Research Council of Norway Contract No. 216699. The research of I.S. was supported in part by the French National Research Agency (ANR-15-CHIN-0003). This research was supported in part by NTNU and the Norwegian metacenter for High Performance Computing (NOTUR) by the allocation of computer time.

Appendix A. Perturbative solution

We present here a method known as small amplitude perturbation theory that we apply to find an approximate solution of the reduced Rayleigh equations. We will illustrate the method considering a system made of a stack of three media separated by two randomly rough interfaces, like the one depicted in Fig. 1. Using the notation introduced in Section 3, we know that the reduced Rayleigh equations for the reflection amplitude is given by Eqs. (26) and (27)

$$\int \boldsymbol{\Theta}_{3,1}^{+,+}(\mathbf{p} \mid \mathbf{q})\, \mathbf{R}(\mathbf{q} \mid \mathbf{p}_0) \frac{\mathrm{d}^2 q}{(2\pi)^2} = -\boldsymbol{\Theta}_{3,1}^{+,-}(\mathbf{p} \mid \mathbf{p}_0)\,, \tag{A.1}$$

where we recall that the *forward two-interface transfer kernel* is defined as

$$\Theta_{3,1}^{a_3,a_1}(\mathbf{p}_3 \,|\, \mathbf{p}_1) = \sum_{a_2=\pm} a_2 \int \Theta_{3,2}^{a_3,a_2}(\mathbf{p}_3 \,|\, \mathbf{p}_2)\, \Theta_{2,1}^{a_2,a_1}(\mathbf{p}_2 \,|\, \mathbf{p}_1)\, \frac{\mathrm{d}^2 p_2}{(2\pi)^2}\,, \tag{A.2}$$

with the *single-interface kernels* $\Theta_{l,m}^{b,a}$ defined for successive media, i.e. $l, m \in \{1, 3\}$ such that $|l-m| = 1$, $a, b \in \{\pm\}$, as

$$\Theta_{l,m}^{b,a}(\mathbf{p} \,|\, \mathbf{q}) = \alpha_l^{-1}(\mathbf{p}) \mathcal{J}_{l,m}^{b,a}(\mathbf{p} \,|\, \mathbf{q})\, \mathbf{M}_{l,m}^{b,a}(\mathbf{p} \,|\, \mathbf{q}). \tag{A.3}$$

The perturbative method consists in expanding each single-interface kernel in a series of Fourier moments. In order to avoid unnecessary lengthy expansion, we first introduce some notations that will allow us to keep a compact derivation and proved to be useful for generalizing to an arbitrary number of layers and for numerical implementation. We define

$$\begin{aligned}\tilde{\Theta}_{3,1}^{a_3,a_1,(\mathbf{m})}(\mathbf{p}_3 \,|\, \mathbf{p}_2 \,|\, \mathbf{p}_1) = \sum_{a_2=\pm} a_2\, \alpha_3^{-1}(\mathbf{p}_3) \big[a_3\alpha_3(\mathbf{p}_3) - a_2\alpha_2(\mathbf{p}_2)\big]^{m_2-1} \\ \times\ \exp\big[-i\left\{a_3\alpha_3(\mathbf{p}_3) - a_2\alpha_2(\mathbf{p}_2)\right\} d_2\big] \\ \times\ \alpha_2^{-1}(\mathbf{p}_2) \big[a_2\alpha_2(\mathbf{p}_2) - a_1\alpha_1(\mathbf{p}_1)\big]^{m_1-1} \\ \times\ \exp\big[-i\left\{a_2\alpha_2(\mathbf{p}_2) - a_1\alpha_1(\mathbf{p}_1)\right\} d_1\big] \\ \times\ \mathbf{M}_{3,2}^{a_3,a_2}(\mathbf{p}_3 \,|\, \mathbf{p}_2)\, \mathbf{M}_{2,1}^{a_2,a_1}(\mathbf{p}_2 \,|\, \mathbf{p}_1)\,,\end{aligned} \tag{A.4}$$

where $\mathbf{m} = (m_1, m_2) \in \mathbb{N}^2$ is a pair-index (i.e. a two component multi-index). Here, we have made the choice of factorizing the phase factor $e^{-i(a_{j+1}\alpha_{j+1}(\mathbf{p}_{j+1}) - a_j\alpha_j(\mathbf{p}_j))d_j}$, with $d_j = \langle \zeta_j \rangle$ being the offset height of the jth interface, from each factor $\mathcal{J}_{j+1,j}^{a_{j+1},a_j}(\mathbf{p}_{j+1} \,|\, \mathbf{p}_j)$ for later convenience. Given this definition, an expansion of the two-interface kernel in Fourier moments is given by

$$\begin{aligned}\Theta_{3,1}^{a_3,a_1}(\mathbf{p}_3 \,|\, \mathbf{p}_1) &= \sum_{\mathbf{m}=0}^{\infty} \frac{(-i)^{|\mathbf{m}|}}{\mathbf{m}!} \int \hat{h}_2^{(m_2)}(\mathbf{p}_3 - \mathbf{p}_2)\, \hat{h}_1^{(m_1)}(\mathbf{p}_2 - \mathbf{p}_1) \tilde{\Theta}_{3,1}^{a_3,a_1,(\mathbf{m})}(\mathbf{p}_3 \,|\, \mathbf{p}_2 \,|\, \mathbf{p}_1) \frac{\mathrm{d}^2 p_2}{(2\pi)^2} \\ &= \sum_{\mathbf{m}=0}^{\infty} \frac{(-i)^{|\mathbf{m}|}}{\mathbf{m}!}\, \mathbf{Z}_{3,1}^{a_3,a_1,(\mathbf{m})}(\mathbf{p}_3 \,|\, \mathbf{p}_1)\,,\end{aligned} \tag{A.5}$$

where $\sum_{\mathbf{m}=0}^{\infty} \equiv \sum_{m_1=0}^{\infty} \sum_{m_2=0}^{\infty}$, $|\mathbf{m}| = m_1 + m_2$ is the *length* of the pair-index, and $\mathbf{m}! = m_1!\, m_2!$, and for all $j \in \{1, 2\}$,

$$\hat{h}_j^{(m_j)}(\mathbf{q}) = \int \exp\big[-i\mathbf{q}\cdot\mathbf{x}\big]\ \big[\zeta_j(\mathbf{x}) - d_j\big]^{m_j} \mathrm{d}^2 x\,, \tag{A.6}$$

is the Fourier moment of $h_j = \zeta_j - d_j$ of order m_j. It is then clear that $\mathbf{Z}_{3,1}^{a_3,a_1,(\mathbf{m})}(\mathbf{p}_3 \,|\, \mathbf{p}_1)$ is a term of order $|\mathbf{m}|$ in product of surface profiles. The reflection amplitude can be expanded as

$$\mathbf{R}(\mathbf{q} \,|\, \mathbf{p}_0) = \sum_{j=0}^{\infty} \frac{(-i)^j}{j!} \mathbf{R}^{(j)}(\mathbf{q} \,|\, \mathbf{p}_0)\,, \tag{A.7}$$

where the term $\mathbf{R}^{(j)}(\mathbf{q} \,|\, \mathbf{p}_0)$ is of order j in product of surface profiles. We are now ready to start the derivation of the perturbative expansion. By plugging Eqs. (A.5) and (A.7) into Eq. (A.1) we obtain

$$\sum_{\substack{\mathbf{m}'=0 \\ j=0}}^{\infty} \frac{(-i)^{|\mathbf{m}'|+j}}{\mathbf{m}'!\, j!} \int \mathbf{Z}_{3,1}^{+,+,(\mathbf{m}')}(\mathbf{p} \,|\, \mathbf{q})\, \mathbf{R}^{(j)}(\mathbf{q} \,|\, \mathbf{p}_0) \frac{\mathrm{d}^2 q}{(2\pi)^2} = -\sum_{\mathbf{m}=0}^{\infty} \frac{(-i)^{|\mathbf{m}|}}{\mathbf{m}!} \mathbf{Z}_{3,1}^{+,-,(\mathbf{m})}(\mathbf{p} \,|\, \mathbf{p}_0). \tag{A.8}$$

Summing over all multi-index $\mathbf{m}$ is equivalent to summing over subsets $\mathcal{S}_m = \{\mathbf{m} \in \mathbb{N}^2 \,||\mathbf{m}| = m\}$ of multi-index of constant length m, i.e. that we have $\sum_{\mathbf{m}=0}^{\infty} \equiv \sum_{m=0}^{\infty} \sum_{\mathbf{m}\in\mathcal{S}_m}$, therefore the previous

equation can be re-written as

$$\sum_{\substack{m'=0\\ j=0}}^{\infty} \sum_{\mathbf{m}'\in\mathcal{S}_{m'}} \frac{(-i)^{m'+j}}{\mathbf{m}'!\,j!} \int \mathbf{Z}_{3,1}^{+,+,(\mathbf{m}')}(\mathbf{p}\mid\mathbf{q})\,\mathbf{R}^{(j)}(\mathbf{q}\mid\mathbf{p}_0)\frac{\mathrm{d}^2q}{(2\pi)^2} = -\sum_{m=0}^{\infty}\sum_{\mathbf{m}\in\mathcal{S}_m}\frac{(-i)^m}{\mathbf{m}!}\mathbf{Z}_{3,1}^{+,-,(\mathbf{m})}(\mathbf{p}\mid\mathbf{p}_0). \tag{A.9}$$

We then use the definition of the multinomial coefficient in multi-index form as $|\mathbf{m}|!/\mathbf{m}! = \binom{|\mathbf{m}|}{\mathbf{m}}$ to obtain

$$\sum_{\substack{m'=0\\ j=0}}^{\infty} \frac{(-i)^{m'+j}}{m'!\,j!} \sum_{\mathbf{m}'\in\mathcal{S}_{m'}} \binom{m'}{\mathbf{m}'} \int \mathbf{Z}_{3,1}^{+,+,(\mathbf{m}')}(\mathbf{p}\mid\mathbf{q})\,\mathbf{R}^{(j)}(\mathbf{q}\mid\mathbf{p}_0)\frac{\mathrm{d}^2q}{(2\pi)^2} = -\sum_{m=0}^{\infty}\frac{(-i)^m}{m!}\sum_{\mathbf{m}\in\mathcal{S}_m}\binom{m}{\mathbf{m}}\mathbf{Z}_{3,1}^{+,-,(\mathbf{m})}(\mathbf{p}\mid\mathbf{p}_0). \tag{A.10}$$

We now make a change of summation index $j \leftrightarrow m - m'$ on the left hand side of the above equation. This makes clearly appear terms of order m in product of surface profiles. We obtain

$$\sum_{m=0}^{\infty}\sum_{m'=0}^{m} \frac{(-i)^m}{m'!\,(m-m')!} \sum_{\mathbf{m}'\in\mathcal{S}_{m'}} \binom{m'}{\mathbf{m}'} \int \mathbf{Z}_{3,1}^{+,+,(\mathbf{m}')}(\mathbf{p}\mid\mathbf{q})\,\mathbf{R}^{(m-m')}(\mathbf{q}\mid\mathbf{p}_0)\frac{\mathrm{d}^2q}{(2\pi)^2} = -\sum_{m=0}^{\infty}\frac{(-i)^m}{m!}\sum_{\mathbf{m}\in\mathcal{S}_m}\binom{m}{\mathbf{m}}\mathbf{Z}_{3,1}^{+,-,(\mathbf{m})}(\mathbf{p}\mid\mathbf{p}_0)\,, \tag{A.11}$$

which can be re-written with the use of the definition of the binomial coefficient $\binom{m}{m'}$ as

$$\sum_{m=0}^{\infty}\frac{(-i)^m}{m!}\sum_{m'=0}^{m}\binom{m}{m'} \sum_{\mathbf{m}'\in\mathcal{S}_{m'}} \binom{m'}{\mathbf{m}'} \int \mathbf{Z}_{3,1}^{+,+,(\mathbf{m}')}(\mathbf{p}\mid\mathbf{q})\,\mathbf{R}^{(m-m')}(\mathbf{q}\mid\mathbf{p}_0)\frac{\mathrm{d}^2q}{(2\pi)^2} = -\sum_{m=0}^{\infty}\frac{(-i)^m}{m!}\sum_{\mathbf{m}\in\mathcal{S}_m}\binom{m}{\mathbf{m}}\mathbf{Z}_{3,1}^{+,-,(\mathbf{m})}(\mathbf{p}\mid\mathbf{p}_0).$$

It is now time to identify terms of same orders in the left and right hand sides. For $m = 0$, only the term for $\mathbf{m}' = (0, 0)$ remains in the left hand side, only the term $\mathbf{m} = (0, 0)$ remains in the right hand side and we have

$$\int \mathbf{Z}_{3,1}^{+,+,(\mathbf{0})}(\mathbf{p}\mid\mathbf{q})\,\mathbf{R}^{(0)}(\mathbf{q}\mid\mathbf{p}_0)\frac{\mathrm{d}^2q}{(2\pi)^2} = -\mathbf{Z}_{3,1}^{+,-,(\mathbf{0})}(\mathbf{p}\mid\mathbf{p}_0). \tag{A.12}$$

which, when expanded, reads

$$\iint \hat{h}_2^{(0)}(\mathbf{p}-\mathbf{p}_2)\,\hat{h}_1^{(0)}(\mathbf{p}_2-\mathbf{q})\tilde{\Theta}_{3,1}^{+,+,(\mathbf{0})}(\mathbf{p}\mid\mathbf{p}_2\mid\mathbf{q})\frac{\mathrm{d}^2p_2}{(2\pi)^2}\,\mathbf{R}^{(0)}(\mathbf{q}\mid\mathbf{p}_0)\frac{\mathrm{d}^2q}{(2\pi)^2} = -\int \hat{h}_2^{(0)}(\mathbf{p}-\mathbf{p}_2)\,\hat{h}_1^{(0)}(\mathbf{p}_2-\mathbf{p}_0)\tilde{\Theta}_{3,1}^{+,-,(\mathbf{0})}(\mathbf{p}\mid\mathbf{p}_2\mid\mathbf{p}_0)\frac{\mathrm{d}^2p_2}{(2\pi)^2}. \tag{A.13}$$

From the definition of the zero order Fourier moment, we have $\hat{h}_j^{(0)}(\mathbf{q}) = (2\pi)^2\,\delta(\mathbf{q})$, which yields

$$\tilde{\Theta}_{3,1}^{+,+,(\mathbf{0})}(\mathbf{p}\mid\mathbf{p}\mid\mathbf{p})\,\mathbf{R}^{(0)}(\mathbf{p}\mid\mathbf{p}_0) = -(2\pi)^2\tilde{\Theta}_{3,1}^{+,-,(\mathbf{0})}(\mathbf{p}_0\mid\mathbf{p}_0\mid\mathbf{p}_0)\,\delta(\mathbf{p}-\mathbf{p}_0). \tag{A.14}$$

Here, the reader may recognize the solution of the reflection problem for a stack of layers with flat interfaces, i.e. Fresnel amplitudes

$$\mathbf{R}^{(0)}(\mathbf{p}|\mathbf{p}_0) = -\left[\bar{\Theta}_{3,1}^{+,+}(\mathbf{p}_0)\right]^{-1} \bar{\Theta}_{3,1}^{+,-}(\mathbf{p}_0)\,(2\pi)^2\delta(\mathbf{p}-\mathbf{p}_0) = -\rho_0(\mathbf{p}_0)\,(2\pi)^2\delta(\mathbf{p}-\mathbf{p}_0)\,, \tag{A.15}$$

where $\bar{\Theta}_{3,1}^{+,+}(\mathbf{p}_0) \equiv \tilde{\Theta}_{3,1}^{+,+,(\mathbf{0})}(\mathbf{p}_0 \,|\, \mathbf{p}_0 \,|\, \mathbf{p}_0)$ and $\bar{\Theta}_{3,1}^{+,-}(\mathbf{p}_0) \equiv \tilde{\Theta}_{3,1}^{+,-,(\mathbf{0})}(\mathbf{p}_0 \,|\, \mathbf{p}_0 \,|\, \mathbf{p}_0)$. Thus, the order zero of the perturbative expansion corresponds to the Fresnel coefficients for the corresponding system with flat interfaces. For orders $m \geq 1$, we have

$$\begin{aligned}\sum_{m'=0}^{m} \binom{m}{m'} \sum_{\mathbf{m}'\in\mathcal{S}_{m'}} \binom{m'}{\mathbf{m}'} \int \mathbf{Z}_{3,1}^{+,+,(\mathbf{m}')}(\mathbf{p}\,|\,\mathbf{q})\,\mathbf{R}^{(m-m')}(\mathbf{q}\,|\,\mathbf{p}_0)\frac{\mathrm{d}^2q}{(2\pi)^2} \\ = -\sum_{\mathbf{m}\in\mathcal{S}_m} \binom{m}{\mathbf{m}} \mathbf{Z}_{3,1}^{+,-,(\mathbf{m})}(\mathbf{p}\,|\,\mathbf{p}_0).\end{aligned}$$

By isolating the term corresponding to $m' = 0$, hence $\mathbf{m}' = (0, 0)$ and using that for all $j \in \{1, 2\}$ we have $\hat{h}_j^{(0)}(\mathbf{q}) = (2\pi)^2\,\delta(\mathbf{q})$, we obtain

$$\begin{aligned}\mathbf{R}^{(m)}(\mathbf{p}\,|\,\mathbf{p}_0) = -\left[\bar{\Theta}_{3,1}^{+,+}(\mathbf{p})\right]^{-1} \Bigg[&\sum_{\mathbf{m}\in\mathcal{S}_m} \binom{m}{\mathbf{m}} \mathbf{Z}_{3,1}^{+,-,(\mathbf{m})}(\mathbf{p}\,|\,\mathbf{p}_0) \\ &+ \sum_{m'=1}^{m} \binom{m}{m'} \int \sum_{\mathbf{m}'\in\mathcal{S}_{m'}} \binom{m'}{\mathbf{m}'} \mathbf{Z}_{3,1}^{+,+,(\mathbf{m}')}(\mathbf{p}\,|\,\mathbf{q})\mathbf{R}^{(m-m')}(\mathbf{q}\,|\,\mathbf{p}_0)\frac{\mathrm{d}^2q}{(2\pi)^2} \Bigg].\end{aligned} \tag{A.16}$$

We have finally obtained a recursive expression giving the mth order term in the reflection amplitude expansion as a function of lower order terms. For weakly rough surfaces, an approximation based on a truncation of the expansion of the reflection amplitude Eq. (A.7) to the first non-trivial order often yields accurate physical insights. For $m = 1$, we obtain that

$$\begin{aligned}\mathbf{R}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_0) &= -\left[\bar{\Theta}_{3,1}^{+,+}(\mathbf{p})\right]^{-1} \Bigg[\mathbf{Z}_{3,1}^{+,-,(1,0)}(\mathbf{p}\,|\,\mathbf{p}_0) + \mathbf{Z}_{3,1}^{+,-,(0,1)}(\mathbf{p}\,|\,\mathbf{p}_0) \\ &\quad + \int \left(\mathbf{Z}_{3,1}^{+,+,(1,0)}(\mathbf{p}\,|\,\mathbf{q}) + \mathbf{Z}_{3,1}^{+,+,(0,1)}(\mathbf{p}\,|\,\mathbf{q})\right)\mathbf{R}^{(0)}(\mathbf{q}\,|\,\mathbf{p}_0)\frac{\mathrm{d}^2q}{(2\pi)^2}\Bigg] \\ &= -\left[\bar{\Theta}_{3,1}^{+,+}(\mathbf{p})\right]^{-1} \Bigg[\mathbf{Z}_{3,1}^{+,-,(1,0)}(\mathbf{p}\,|\,\mathbf{p}_0) + \mathbf{Z}_{3,1}^{+,-,(0,1)}(\mathbf{p}\,|\,\mathbf{p}_0) \\ &\quad - \left(\mathbf{Z}_{3,1}^{+,+,(1,0)}(\mathbf{p}\,|\,\mathbf{p}_0) + \mathbf{Z}_{3,1}^{+,+,(0,1)}(\mathbf{p}\,|\,\mathbf{p}_0)\right)\left[\bar{\Theta}_{3,1}^{+,+}(\mathbf{p}_0)\right]^{-1}\bar{\Theta}_{3,1}^{+,-}(\mathbf{p}_0)\Bigg],\end{aligned} \tag{A.17}$$

where we have used the previously obtained expression for $\mathbf{R}^{(0)}(\mathbf{q}\,|\,\mathbf{p}_0)$ in Eq. (A.15), and in particular the fundamental property of the Dirac delta. From the definition of $\mathbf{Z}_{3,1}^{a_3,a_1,(\mathbf{m})}$ [Eq. (A.5)] it is clear that for $\mathbf{m} = (1, 0)$ or $(0, 1)$ the integration reduces to

$$\mathbf{Z}_{3,1}^{a_3,a_1,(1,0)}(\mathbf{p}\,|\,\mathbf{p}_0) = \hat{h}_1^{(1)}(\mathbf{p}-\mathbf{p}_0)\tilde{\Theta}_{3,1}^{a_3,a_1,(1,0)}(\mathbf{p}\,|\,\mathbf{p}\,|\,\mathbf{p}_0) \tag{A.18a}$$

$$\mathbf{Z}_{3,1}^{a_3,a_1,(0,1)}(\mathbf{p}\,|\,\mathbf{p}_0) = \hat{h}_2^{(1)}(\mathbf{p}-\mathbf{p}_0)\tilde{\Theta}_{3,1}^{a_3,a_1,(0,1)}(\mathbf{p}\,|\,\mathbf{p}_0\,|\,\mathbf{p}_0). \tag{A.18b}$$

It is convenient to group terms with common factor $\hat{h}_j \equiv \hat{h}_j^{(1)}$ in Eq. (A.17), which leads to

$$\mathbf{R}^{(1)}(\mathbf{p}\,|\,\mathbf{p}_0) = \hat{h}_1(\mathbf{p}-\mathbf{p}_0)\rho_1(\mathbf{p}\,|\,\mathbf{p}_0) + \hat{h}_2(\mathbf{p}-\mathbf{p}_0)\rho_2(\mathbf{p}\,|\,\mathbf{p}_0)\,, \tag{A.19}$$

with

$$\rho_1(\mathbf{p}\,|\,\mathbf{p}_0) = \left[\bar{\Theta}_{3,1}^{+,+}(\mathbf{p})\right]^{-1} \left[\tilde{\Theta}_{3,1}^{+,+,(1,0)}(\mathbf{p}\,|\,\mathbf{p}\,|\,\mathbf{p}_0)\rho_0(\mathbf{p}_0) - \tilde{\Theta}_{3,1}^{+,-,(1,0)}(\mathbf{p}\,|\,\mathbf{p}\,|\,\mathbf{p}_0)\right] \tag{A.20a}$$

$$\rho_2(\mathbf{p}\,|\,\mathbf{p}_0) = \left[\bar{\Theta}_{3,1}^{+,+}(\mathbf{p})\right]^{-1} \left[\tilde{\Theta}_{3,1}^{+,+,(0,1)}(\mathbf{p}\,|\,\mathbf{p}_0\,|\,\mathbf{p}_0)\rho_0(\mathbf{p}_0) - \tilde{\Theta}_{3,1}^{+,-,(0,1)}(\mathbf{p}\,|\,\mathbf{p}_0\,|\,\mathbf{p}_0)\right]. \tag{A.20b}$$

We have treated the case of reflection as a representative example, but the same method applies for transmission.

Appendix B. Differential reflection coefficient

Assuming we have obtained the reflection amplitudes $R_{\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0)$ either by using the perturbative approach or by the purely numerical simulation, we can now proceed to express the differential reflection coefficient (DRC) defined as the time-averaged flux radiated around a given scattering direction (θ_s, ϕ_s) per unit solid angle and per unit incident flux and denoted $\partial R/\partial\Omega_s(\mathbf{p} \mid \mathbf{p}_0)$. Let a virtual hemisphere of radius $r \gg c/\omega$ lie on the plane $x_3 = 0$ on top of the scattering system. The support of this hemisphere is a disk of area $S = \pi r^2$. We consider the scattering from a *truncated* version of the scattering system in which the surface profiles are set to be flat outside the disk support. Consequently, the field amplitudes we will manipulate are not strictly speaking those of the full system of interest but will converge to them as $r \to \infty$. We will nevertheless keep the same notations as that from the full system introduced in Section 3 for simplicity. The time-averaged flux incident on this disk is given by

$$\begin{aligned} P_{\mathrm{inc}/S} &= -\mathrm{Re}\frac{c}{8\pi}\int_S \left[\mathbf{E}_0^*(\mathbf{p}_0) \times \left(\frac{c}{\omega}\,\mathbf{k}_1^-(\mathbf{p}_0) \times \mathbf{E}_0(\mathbf{p}_0)\right)\right] \cdot \hat{\mathbf{e}}_3 \;\exp\left[-i(\mathbf{k}_1^{-*}(\mathbf{p}_0) - \mathbf{k}_1^-(\mathbf{p}_0)) \cdot \mathbf{x}\right] \mathrm{d}^2x_\parallel \\ &= -\frac{c^2}{8\pi\omega}\,\mathrm{Re}\int_S \left[|\mathbf{E}_0(\mathbf{p}_0)|^2\,\mathbf{k}_1^-(\mathbf{p}_0) - \left(\mathbf{E}_0^*(\mathbf{p}_0) \cdot \mathbf{k}_1^-(\mathbf{p}_0)\right) \cdot \mathbf{E}_0(\mathbf{p}_0)\right] \cdot \hat{\mathbf{e}}_3 \mathrm{d}^2x_\parallel \\ &= S\,\frac{c^2}{8\pi\omega}\,\alpha_1(\mathbf{p}_0)\,|\mathbf{E}_0(\mathbf{p}_0)|^2 \\ &= S\,\frac{c^2}{8\pi\omega}\,\alpha_1(\mathbf{p}_0)\left[\left|\mathcal{E}_{0,p}\right|^2 + \left|\mathcal{E}_{0,s}\right|^2\right]. \end{aligned} \tag{B.1}$$

Here, the * denotes the complex conjugate, and incident field amplitude $\mathbf{E}_0(\mathbf{p}_0) = \mathcal{E}_{0,p}\,\hat{\mathbf{e}}_{p,1}^-(\mathbf{p}_0) + \mathcal{E}_{0,s}\,\hat{\mathbf{e}}_s(\mathbf{p}_0)$ as defined in Eq. (22), the vector identity $\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c}$ and the orthogonality between the field and the wave vector $\mathbf{E}_0^*(\mathbf{p}_0) \cdot \mathbf{k}_1^-(\mathbf{p}_0) = 0$ have been used. Note that the flux incident on the disk is proportional to the disk area. Let us now consider the outgoing flux crossing an elementary surface $\mathrm{d}\sigma = r^2 \sin\theta_s \mathrm{d}\theta_s \mathrm{d}\phi_s = r^2 \mathrm{d}\Omega_s$ around a point $\mathbf{r} = r\,(\sin\theta_s \cos\phi_s\,\hat{\mathbf{e}}_1 + \sin\theta_s \sin\phi_s\,\hat{\mathbf{e}}_2 + \cos\theta_s\,\hat{\mathbf{e}}_3) = r\,\hat{\mathbf{n}}$. The flux crossing this elementary surface is given by

$$P_{\mathrm{d}\sigma} = \frac{c}{8\pi}\,\mathrm{Re}\left[\mathbf{E}_1^{+*}(\mathbf{r}) \times \mathbf{H}_1^+(\mathbf{r})\right] \cdot \hat{\mathbf{n}}\,\mathrm{d}\sigma. \tag{B.2}$$

We then use the well-known asymptotic expansion of the field in the far-field given by (see Refs. [30,31])

$$\mathbf{E}_1^+(\mathbf{r}) \sim -i\,\epsilon_1^{1/2}\,\frac{\omega}{2\pi\,c}\,\cos\theta_s\,\frac{\exp(i\epsilon_1^{1/2}\frac{\omega}{c}r)}{r}\,\mathbf{E}_1^+(\mathbf{p}) \tag{B.3a}$$

$$\mathbf{H}_1^+(\mathbf{r}) \sim -i\,\epsilon_1\,\frac{\omega}{2\pi\,c}\,\cos\theta_s\,\frac{\exp(i\epsilon_1^{1/2}\frac{\omega}{c}r)}{r}\,\hat{\mathbf{n}} \times \mathbf{E}_1^+(\mathbf{p}) \tag{B.3b}$$

where $\mathbf{p} = \sqrt{\epsilon_1}\frac{\omega}{c}(\sin\theta_s \cos\phi_s\,\hat{\mathbf{e}}_1 + \sin\theta_s \sin\phi_s\,\hat{\mathbf{e}}_2)$. This asymptotic approximation will become more and more accurate as we let $r \to \infty$. Plugging Eq. (B.3) into Eq. (B.2) we obtain

$$\begin{aligned} P_{\mathrm{d}\sigma} &= \epsilon_1^{3/2}\left(\frac{\omega}{2\pi\,c}\right)^2 \cos^2\theta_s\,\frac{c}{8\pi}\,|\mathbf{E}_1^+(\mathbf{p})|^2\,\mathrm{d}\Omega_s \\ &= \epsilon_1^{3/2}\left(\frac{\omega}{2\pi\,c}\right)^2 \cos^2\theta_s\,\frac{c}{8\pi}\,\left(|\mathcal{E}_{1,p}^+(\mathbf{p})|^2 + |\mathcal{E}_{1,s}^+(\mathbf{p})|^2\right)\mathrm{d}\Omega_s. \end{aligned} \tag{B.4}$$

The total differential reflection coefficient is then given by

$$\frac{\partial R}{\partial\Omega_s}(\mathbf{p} \mid \mathbf{p}_0) = \lim_{r\to\infty}\frac{P_{\mathrm{d}\sigma}}{P_{\mathrm{inc}/S}\,\mathrm{d}\Omega_s} = \lim_{r\to\infty}\frac{\epsilon_1}{S}\left(\frac{\omega}{2\pi\,c}\right)^2\frac{\cos^2\theta_s}{\cos\theta_0}\,\frac{|\mathcal{E}_{1,p}^+(\mathbf{p})|^2 + |\mathcal{E}_{1,s}^+(\mathbf{p})|^2}{|\mathcal{E}_{0,p}|^2 + |\mathcal{E}_{0,s}|^2}. \tag{B.5}$$

From the total differential reflection coefficient given by Eq. (B.5), we deduce the differential reflection coefficient when an incident plane wave of polarization β, with in-plane wave vector $\mathbf{p}_0$ is reflected into a plane wave of polarization α with in-plane wave vector $\mathbf{p}$ given as

$$\frac{\partial R_{\alpha\beta}}{\partial \Omega_s}(\mathbf{p} \mid \mathbf{p}_0) = \lim_{r\to\infty} \frac{\epsilon_1}{S}\left(\frac{\omega}{2\pi\, c}\right)^2 \frac{\cos^2\theta_s}{\cos\theta_0} \left|R_{\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0)\right|^2 = \lim_{r\to\infty} \frac{\partial R^{(S)}_{\alpha\beta}}{\partial \Omega_s}(\mathbf{p} \mid \mathbf{p}_0). \tag{B.6}$$

As we are interested in averaging the optical response over realizations of the surface profiles, we consider the following ensemble average

$$\left\langle \frac{\partial R^{(S)}_{\alpha\beta}}{\partial \Omega_s}(\mathbf{p} \mid \mathbf{p}_0) \right\rangle = \frac{\epsilon_1}{S}\left(\frac{\omega}{2\pi\, c}\right)^2 \frac{\cos^2\theta_s}{\cos\theta_0} \left\langle \left|R_{\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0)\right|^2 \right\rangle. \tag{B.7}$$

By decomposing the reflection amplitudes as the sum of the mean and fluctuation (deviation from the mean)

$$R_{\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0) = \left\langle R_{\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0) \right\rangle + \left[R_{\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0) - \left\langle R_{\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0) \right\rangle \right], \tag{B.8}$$

we can decompose the mean differential reflection coefficient as the sum of a coherent term and an incoherent term

$$\left\langle \frac{\partial R^{(S)}_{\alpha\beta}}{\partial \Omega_s}(\mathbf{p} \mid \mathbf{p}_0) \right\rangle = \left\langle \frac{\partial R^{(S)}_{\alpha\beta}}{\partial \Omega_s}(\mathbf{p} \mid \mathbf{p}_0) \right\rangle_{\text{coh}} + \left\langle \frac{\partial R^{(S)}_{\alpha\beta}}{\partial \Omega_s}(\mathbf{p} \mid \mathbf{p}_0) \right\rangle_{\text{incoh}}, \tag{B.9}$$

where

$$\left\langle \frac{\partial R^{(S)}_{\alpha\beta}}{\partial \Omega_s}(\mathbf{p} \mid \mathbf{p}_0) \right\rangle_{\text{coh}} = \frac{\epsilon_1}{S}\left(\frac{\omega}{2\pi\, c}\right)^2 \frac{\cos^2\theta_s}{\cos\theta_0} \left|\left\langle R_{\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0)\right\rangle\right|^2 \tag{B.10a}$$

$$\left\langle \frac{\partial R^{(S)}_{\alpha\beta}}{\partial \Omega_s}(\mathbf{p}|\mathbf{p}_0) \right\rangle_{\text{incoh}} = \frac{\epsilon_1}{S}\left(\frac{\omega}{2\pi\, c}\right)^2 \frac{\cos^2\theta_s}{\cos\theta_0} \left[\left\langle \left|R_{\alpha\beta}(\mathbf{p}|\mathbf{p}_0)\right|^2 \right\rangle - \left|\left\langle R_{\alpha\beta}(\mathbf{p}|\mathbf{p}_0)\right\rangle\right|^2\right]. \tag{B.10b}$$

If we now use the expression found in Appendix A for the reflection amplitudes to first order in the product of surface profiles,

$$\mathbf{R}(\mathbf{p} \mid \mathbf{p}_0) \approx \mathbf{R}^{(0)}(\mathbf{p} \mid \mathbf{p}_0) - i\mathbf{R}^{(1)}(\mathbf{p} \mid \mathbf{p}_0), \tag{B.11}$$

where $\mathbf{R}^{(0)}(\mathbf{p} \mid \mathbf{p}_0)$ is the response from the corresponding system with flat interfaces (i.e. that of a Fabry–Perot interferometer), Eq. (A.15), and $\mathbf{R}^{(1)}(\mathbf{p} \mid \mathbf{p}_0)$ is given in Eq. (A.19), we obtain that the factor in the square bracket in Eq. (B.10b) reads

$$\begin{aligned}
\left\langle \left|R_{\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0)\right|^2 \right\rangle - \left|\left\langle R_{\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0)\right\rangle\right|^2 &= \left\langle \left|R^{(1)}_{\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0)\right|^2 \right\rangle \\
&= \left\langle \left|\hat{h}_{1,S}(\mathbf{p}-\mathbf{p}_0)\right|^2 \right\rangle \left|\rho_{1,\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0)\right|^2 \\
&\quad + \left\langle \left|\hat{h}_{2,S}(\mathbf{p}-\mathbf{p}_0)\right|^2 \right\rangle \left|\rho_{2,\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0)\right|^2 \\
&\quad + 2\,\mathrm{Re}\left\langle \hat{h}_{1,S}(\mathbf{p}-\mathbf{p}_0)\hat{h}^*_{2,S}(\mathbf{p}-\mathbf{p}_0) \right\rangle \\
&\quad \times \left(\rho_{1,\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0)\rho^*_{2,\alpha\beta}(\mathbf{p} \mid \mathbf{p}_0)\right).
\end{aligned} \tag{B.12}$$

Note here that we are still dealing with a scattering system whose surface profiles are flat outside the disk of radius r, hence the subscript S. For the statistical properties attributed to the surface profiles in Section 2, we have

$$\begin{aligned}
\left\langle \hat{h}_{i,S}(\mathbf{q})\hat{h}^*_{j,S}(\mathbf{q}) \right\rangle &= \left\langle \int_S \int_S h_i(\mathbf{x}) h_j(\mathbf{x}') \exp\left[i\mathbf{q}\cdot(\mathbf{x}-\mathbf{x}')\right] \mathrm{d}^2x\, \mathrm{d}^2x' \right\rangle \\
&= \int_S \int_S \left\langle h_i(\mathbf{x}) h_j(\mathbf{x}') \right\rangle \exp\left[i\mathbf{q}\cdot(\mathbf{x}-\mathbf{x}')\right] \mathrm{d}^2x\, \mathrm{d}^2x'
\end{aligned}$$

$$= \int_S \int_S \gamma_{ij}\, W(\mathbf{x}-\mathbf{x}')\, \exp\left[i\mathbf{q}\cdot(\mathbf{x}-\mathbf{x}')\right]\, \mathrm{d}^2x\, \mathrm{d}^2x'. \tag{B.13}$$

Here we have used the definition of the Fourier transform, and the fact that ensemble average commutes with the integration of the surfaces and the definition of the correlation function. We have also introduced the shorthand $\gamma_{ij} = \left[\delta_{ij} + \gamma(1-\delta_{ij})\right]\, \sigma_i\, \sigma_j$. Via the change of variable $\mathbf{u} = \mathbf{x} - \mathbf{x}'$ we obtain

$$\left\langle \hat{h}_{i,S}(\mathbf{q}) \hat{h}^*_{j,S}(\mathbf{q}) \right\rangle = S\, \gamma_{ij} \int_S W(\mathbf{u})\, \exp(i\mathbf{q}\cdot\mathbf{u})\, \mathrm{d}^2u = S\, \gamma_{ij}\, g_S(\mathbf{q}). \tag{B.14}$$

Thus

$$\begin{aligned}\left\langle |R_{\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_0)|^2 \right\rangle - \left|\left\langle R_{\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_0)\right\rangle\right|^2 = S\, g_S(\mathbf{p}-\mathbf{p}_0) \Big[&\sigma_1^2 \left|\rho_{1,\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_0)\right|^2 + \sigma_2^2 \left|\rho_{2,\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_0)\right|^2 \\ &+ 2\gamma\sigma_1\sigma_2\, \mathrm{Re}\left\{\rho_{1,\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_0)\rho^*_{2,\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_0)\right\}\Big].\end{aligned} \tag{B.15}$$

Finally, by plugging the above equation into Eq. (B.10b), the surface area S cancels and letting $r \to \infty$, $g_S \to g$ (where we remind the reader that g is the power spectrum of the surface profiles) and we finally obtain the expression for the incoherent component of the mean differential reflection coefficient for the entire (infinite) system under the first order approximation of the reflected amplitudes in product of the surface profiles

$$\begin{aligned}\left\langle \frac{\partial R_{\alpha\beta}(\mathbf{p}|\mathbf{p}_0)}{\partial\Omega_s} \right\rangle_{\text{incoh}} = \epsilon_1 \left(\frac{\omega}{2\pi c}\right)^2 \frac{\cos^2\theta_s}{\cos\theta_0}\, g(\mathbf{p}-\mathbf{p}_0) \Big[&\sigma_1^2 \left|\rho_{1,\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_0)\right|^2 + \sigma_2^2 \left|\rho_{2,\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_0)\right|^2 \\ &+ 2\gamma\sigma_1\sigma_2\, \mathrm{Re}\left\{\rho_{1,\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_0)\rho^*_{2,\alpha\beta}(\mathbf{p}\,|\,\mathbf{p}_0)\right\}\Big].\end{aligned} \tag{B.16}$$

Appendix C. Contrast estimates

We propose here to motivate mathematically that the phase mixing in paths of type (1''), (2'') *etc.*, from Fig. 4(b) and those from Fig. 4(c) intrinsically leads to poorer contrast in the interference pattern found in the incoherent contribution to the mean DRC than, for example, paths of type (1), (2) in Fig. 4(a), where no phase mixing is allowed. As a prototypical reflection amplitude for a sum of paths that involves phase mixing and a sum of paths that does not (and will serve as reference), let us have respectively

$$r_{\text{mix}\varphi} = \frac{\tilde{r}}{[1 - r_0 \exp(2i\varphi_0)]\, [1 - r_s \exp(2i\varphi_s)]} \tag{C.1a}$$

$$r_{\text{ref}} = \frac{\tilde{r}}{1 - r_s \exp(2i\varphi_s)}. \tag{C.1b}$$

These reflection amplitudes mimic the structure from Eqs. (41) and Eq. (34) respectively, but we will see that the precise expressions for the numerators do not matter for the contrast, and are hence denoted by the same symbol $\tilde{r}$. Note that all the reflection amplitudes in Eq. (C.1) depend on angles of incidence and scattering, but for clarity we drop these arguments. Our first step consists in taking the square modulus of Eq. (C.1)

$$I_{\text{mix}\varphi} = \frac{|\tilde{r}|^2}{|1 - r_0 \exp(2i\varphi_0)|^2\, |1 - r_s \exp(2i\varphi_s)|^2} \tag{C.2a}$$

$$I_{\text{ref}} = \frac{|\tilde{r}|^2}{|1 - r_s \exp(2i\varphi_s)|^2}\ , \tag{C.2b}$$

and in bounding the intensity by using the triangular inequality

$$\frac{|\tilde{r}|^2}{(1+|r_0|)^2\,(1+|r_s|)^2} \quad \leq \quad I_{\text{mix}\varphi} \quad \leq \quad \frac{|\tilde{r}|^2}{(1-|r_0|)^2\,(1-|r_s|)^2} \tag{C.3a}$$

$$\frac{|\tilde{r}|^2}{(1+|r_s|)^2} \quad \leq \quad I_{\text{ref}} \quad \leq \quad \frac{|\tilde{r}|^2}{(1-|r_s|)^2}. \tag{C.3b}$$

It is clear from Eq. (C.3) that the intensity lies between two bounding curves. A fair estimate for the trend, i.e. the intensity without the oscillations would be given by $|\tilde{r}|^2$, and we thus estimate, or rather define, the *inverse contrast* as

$$\eta_{\text{mix}\varphi}^{-1} = (1+|r_0|)^2\,(1+|r_s|)^2 - (1-|r_0|)^2\,(1-|r_s|)^2 \tag{C.4a}$$

$$\eta_{\text{ref}}^{-1} = (1+|r_s|)^2 - (1-|r_s|)^2. \tag{C.4b}$$

This may not be the most *natural* definition for the contrast, but we choose this one since it is easier to work with and will not change the conclusion. By re-writing Eq. (C.4) by using straightforward algebra, we obtain

$$\eta_{\text{mix}\varphi}^{-1} = 4|r_s| + 4|r_0| + 4|r_0||r_s| + 4|r_0|^2|r_s| \tag{C.5a}$$

$$\eta_{\text{ref}}^{-1} = 4|r_s|. \tag{C.5b}$$

This shows that the inverse contrast for phase mixing is larger than that of the reference, i.e. that the *contrast* in the case of phase mixing is smaller than that of the reference. Indeed, the two last terms in Eq. (C.5a) are cross-terms resulting directly from the phase mixing nature of the initial reflection amplitude. Note that the choice for the reference was arbitrary and one could choose to study paths of type (1'), (2'), *etc.*, in Fig. 4(b), and hence replace $r_s \exp(2i\varphi_s)$ in Eq. (C.1) by $r_0 \exp(2i\varphi_0)$, and the conclusion would still hold.

Appendix D. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.aop.2017.12.003.

References

[1] I. Newton, Opticks, Or a Treatise of the Reflections, Refractions, Inflections and Colours of Light, William Innys, 1730.
[2] A.J. de Witte, Amer. J. Phys. 35 (1967) 301–313.
[3] V. Freilikher, V.I. Tatarskii, M. Pustilnik, I. Yurkevich, Opt. Lett. 19 (1994) 1382.
[4] J.Q. Lu, J.A. Sánchez-Gil, E.R. Méndez, Z.-H. Gu, A.A. Maradudin, J. Opt. Soc. Am. A 15 (1998) 185.
[5] O. Calvo-Perez, A. Sentenac, J.J. Greffet, Radio Sci. 34 (1999) 311–335.
[6] W. Suhr, H.J. Schlichting, Appl. Opt. 48 (2009) 4978–4984.
[7] C. Raman, G.L. Datta, Phil. Mag. 42 (1921) 826–840.
[8] Y.S. Kaganovskii, V.D. Freilikher, E. Kanzieper, Y. Nafcha, M. Rosenbluh, I.M. Fuks, J. Opt. Soc. Am. A 16 (1999) 331.
[9] P. Selényi, Ann. Phys. 340 (1911) 444–460.
[10] I. Simonsen, Eur. Phys. J.-Spec. Top. 181 (2010) 1–103.
[11] C. Amra, J.H. Apfel, E. Pelletier, Appl. Opt. 31 (1992) 3134.
[12] Y. Pan, Z. Wu, L. Hang, Appl. Surf. Sci. 256 (2010) 3503–3507.
[13] A. Soubret, G. Berginc, C. Bourrely, Phys. Rev. B 63 (2001) 245411.
[14] A. Soubret, G. Berginc, C. Bourrely, J. Opt. Soc. Am. A 18 (2001) 2778–2788.
[15] G.C. Brown, V. Celli, M. Haller, A. Marvin, Surf. Sci. 136 (1984) 381–397.
[16] T. Nordam, P.A. Letnes, I. Simonsen, Front. Phys. 1 (2013) 1–15.
[17] T.A. Leskova, P.A. Letnes, A.A. Maradudin, T. Nordam, I. Simonsen, Proc. SPIE 8172 (2011) 817209–817209–20.
[18] T. Nordam, P.A. Letnes, I. Simonsen, A.A. Maradudin, J. Opt. Soc. Am. A 31 (2014) 1126–1134.
[19] Ø.S. Hetland, A.A. Maradudin, T. Nordam, I. Simonsen, Phys. Rev. A 93 (2016) 053819.
[20] Ø.S. Hetland, A.A. Maradudin, T. Nordam, P.A. Letnes, I. Simonsen, Phys. Rev. A 95 (2017) 043808.
[21] T. Nordam, P.A. Letnes, I. Simonsen, A.A. Maradudin, Opt. Express 20 (2012) 11336–11350.
[22] A.K. González-Alcalde, J.-P. Banon, Ø.S. Hetland, A.A. Maradudin, E.R. Méndez, T. Nordam, I. Simonsen, Opt. Express 24 (2016) 25995–26005.
[23] J.-P. Banon, T. Nesse, Z. Ghadyani, M. Kildemo, I. Simonsen, Opt. Lett. 42 (2017) 2631–2634.
[24] A. Maradudin, T. Michel, A. McGurn, E. Mendez, Ann. Phys., NY 203 (1990) 255–307.
[25] L.S. Blackford, J. Choi, A. Cleary, E. D'Azevedo, J. Demmel, I. Dhillon, J. Dongarra, S. Hammarling, G. Henry, A. Petitet, K. Stanley, D. Walker, R.C. Whaley, ScaLAPACK Users' Guide, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1997.

[26] T. Kawanishi, H. Ogura, Z.L. Wang, Wave. Random Media 7 (1997) 351–384.
[27] I. Simonsen, A.A. Maradudin, T.A. Leskova, Phys. Rev. A 81 (2010) 013806.
[28] I. Simonsen, J.B. Kryvi, A.A. Maradudin, T.A. Leskova, Comput. Phys. Comm. 182 (2011) 1904–1908.
[29] I. Simonsen, A.A. Maradudin, T. Leskova, Phys. Rev. Lett. 104 (2010) 223904.
[30] G.S. Agarwal, Phys. Rev. B 15 (1977) 2371–2383.
[31] K. Miyamoto, E. Wolf, J. Opt. Soc. Amer. 52 (1962) 615–625.

Paper 4:

J.-P. Banon, Ø. S. Hetland, and I. Simonsen. A reduction of variance technique applied to the simulation of light scattering from randomly rough surfaces. *Manuscript in preparation*, 2018

Is not included due to copyright

Paper 5:

M. Kildemo, J.-P. Banon, A. Baron, B. B. Svendsen, T. Brakstad, and I. Simonsen. Optical response of gold hemispheroidal lattices on transparent substrates. *App. Surf. Sci.*, 421:593 – 600, 2017

Applied Surface Science 421 (2017) 593–600

Contents lists available at ScienceDirect

Applied Surface Science

journal homepage: www.elsevier.com/locate/apsusc

Optical response of gold hemispheroidal lattices on transparent substrates

Morten Kildemo[a,*], Jean-Philippe Banon[a], Alexandre Baron[b], Brage B. Svendsen[a], Thomas Brakstad[a], Ingve Simonsen[a,c]

[a] Department of Physics, NTNU Norwegian University of Science and Technology, NO-7491 Trondheim, Norway
[b] CNRS, Centre de Recherche Paul Pascal, UPR 8641 115 Avenue Schweitzer, 33600 Pessac, France
[c] Surface du Verre et Interfaces, UMR 125 CNRS/Saint-Gobain, F-93303 Aubervilliers, France

ARTICLE INFO

Article history:
Received 30 July 2016
Received in revised form 2 February 2017
Accepted 2 February 2017
Available online 20 February 2017

Keywords:
Mueller Matrix Ellipsometry
Plasmonics
Optical properties
Polarimetry

ABSTRACT

Square arrays of gold (Au) hemispheroids deposited on a UV-transparent glass substrate reveal a rich optical response when investigated by spectroscopic Mueller Matrix Ellipsometry. Two samples were studied; the first consisted of hemispheroids of parallel radius of 58 nm and height 30 nm with lattice constant 210 nm; the corresponding parameters for the second sample were 38 nm, 20 nm and 125 nm, respectively. By a full azimuthal rotation of the samples, we observe all the Rayleigh anomalies corresponding to grazing diffracted waves, with strong resonances for co-polarization scattered light near the high symmetry points and cross-polarization scattered light around the Localized Surface Plasmon Resonance. Polarization-conversion becomes particularly important at grazing incidence, and the cross-polarization follows the Rayleigh lines. The optical response (neglecting polarization conversion) is modelled in the quasi-static approximation using the so-called Bedeaux-Vlieger formalism, and the Finite Element Method using COMSOL. The direct inversion of the effective (substrate dependent) dielectric function is discussed.

1. Introduction

Mueller matrix ellipsometry is expected to play an important role in the characterization of plasmonic based devices, such as plasmonic metamaterials/metasurfaces [1,2], or composite nanoplasmonic devices with applications ranging from optoelectronics to biomedicine [3–5]. The fascinating properties of metamaterials can be exemplified by e.g. the negative refractive index [6] and its application to perfect lensing and subdiffraction imaging [7].

The metamaterials approach consists in determining effective electromagnetic parameters for an inhomogeneous periodic artificial material. As such, the metamaterials approach bridges the gap between low-frequency effective medium theory and the high frequency diffractive regime (therein photonic crystals) [8]. However, it is necessarily the combined response that will be observed by spectroscopy across a larger spectral range.

Here we study an apparently simple model system consisting of hemispheroidal Au particles organized in a square lattice on a UV-transparent flat glass substrate. The particle dimensions are such that the quasistatic approximation should be valid, at least for the longest wavelengths. This system was recently shown to have a rich optical response including polarization coupling around the Localized Surface Plasmon Resonance (LSPR), and around so-called Rayleigh anomalies (or Rayleigh lines) related to grazing diffracted waves just at the onset of diffracted orders [2].

Indeed, the importance of the polarization coupling can be directly observed in Mueller matrix spectroscopic studies with complete azimuthal sample rotation of inherently anisotropic systems, such as self-assembled Ag or Au particles along the ripples of a nanopatterned substrate [9–11], meta-surfaces of U-shaped particles [12], slanted metallic pillars [13], and chiral structures [14].

In this paper, the regular lattices of idealized metallic hemispheroids supported by a flat dielectric substrate (here SiO_2) are modelled by the Bedeaux-Vlieger theory [15]. Since this theory does not account for cross-polarized scattered light, only a block-diagonal Mueller matrix can be obtained within this approach. In order to account for cross-polarized scattered light, and therefore obtaining non-vanishing off-diagonal elements of the Mueller matrix, more accurate and time consuming modelling approaches have to be used. One possibility is to perform the modelling on

* Corresponding author.
E-mail address: Morten.Kildemo@ntnu.no (M. Kildemo).

http://dx.doi.org/10.1016/j.apsusc.2017.02.008
0169-4332/

the basis of the reduced Rayleigh equation [16,17] which recently was shown to produce reliable results for the Mueller matrix elements [18]. Other approaches, often used in the metamaterials community, are the full wave solutions obtained by the Finite Element Method (FEM) and Finite Difference Time Domain (FDTD) simulations. In this work we will present results obtained by FEM simulations.

2. Experimental

The samples were produced by evaporating a thin film of Au onto a clean (and flat) UV-grade fused silica surface using an e-beam evaporator (Pfeiffer Vacuum Classic 500). The deposited film thicknesses were 40 nm (Sample A) and 20 nm (Sample B). The resulting films were smooth but polycrystalline. The Au nano-structures on glass were produced by Focused Ion Beam (FIB)-milling using Ga ions (FEI Helios Dual-beam FIB). The two samples reported here were manufactured to make up Au hemispheroids distributed in a square pattern on a glass-surface, see Fig. 1(a); the lattice constants were $a = 210\,nm$ (Sample A) and $a = 125\,nm$ (Sample B). After the milling, the particles were found to be hemispheroids of lateral radius $R_{xy} = (58 \pm 4)$ nm (Sample A), and (38 ± 4) nm (Sample B), see Fig. 1(c). The heights of the particles (the perpendicular radii) were difficult to estimate accurately from the combination of AFM and SEM images. However, rough estimates are $R_z = 30\,nm$ for Sample A (<40 nm) and $R_z = 25\,nm$ for Sample B (with an expected uncertainty of several nanometers). Unfortunately, an over-etching into the substrate was observed for both samples, i.e. the Au particles are probably on top of a dielectric mound, as roughly sketched in Fig. 1(d). Both samples had an over-milling of at least 10 nm into the substrate.

The surface coverage for a square lattice of hemispheroidal particles is defined as $\Theta = \pi R_{xy}^2/a^2$, giving $\Theta_A = 0.20$ and $\Theta_B = 0.18$ for the coverage of Sample A and Sample B, respectively.

For the optical characterization of the samples, a variable angle multichannel dual rotating compensator Mueller matrix ellipsometer (RC2) from JA Woollam Company was used. Our instrument has a collimated 150 W Xe source and operates in the spectral range from 210 nm (5.9 eV) to 1700 nm (0.73 eV). As the FIB-milling is a time consuming process, the total milling area opened by FIB was limited to 240 μm × 240 μm. Focusing and collection lenses with a focal length of 20 mm and a Numerical Aperture of approximately 0.15, were applied, allowing a normal incidence spot size of smaller 100 μm. This spot size allowed us to study the full azimuthal rotation of the sample while ensuring that the spot-size remained within the 240 μm × 240 μm area. The spectroscopic Mueller matrix was measured for the polar incidence angles (with respect to surface normal) $\theta_0 = 45°$, 55° and 65°. Full azimuthal rotation of the sample around the sample normal (360°) in steps of 5° was performed for each polar angle of incidence in order to fully map any anisotropy in the optical response of the sample (see Brakstad et al. [2] for further details).

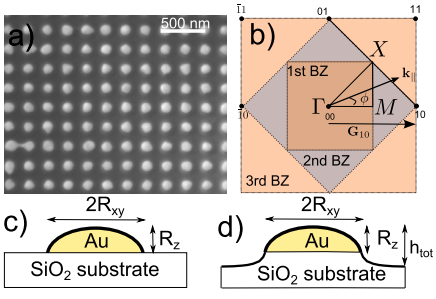


Fig. 1. (a) Real space Scanning Electron Microscopy (SEM) image of the particle array. (b) Schematic drawing of the 2-dimensional reciprocal lattice defining ϕ_0. (c) The ideal model consisting of a hemispheroidal Au particle on uv-transparent glass. (d) The presumed non-ideal model with the particles on a mound.

3. Results and discussion

3.1. Mueller matrix: LSPR, Rayleigh lines and polarization coupling

Fig. 2 presents an overview of the normalized Mueller matrix optical response, **m**, measured for Sample A as a function of the photon energy ($E = \hbar\omega$), for the specular direction $[(\theta_0, \phi_0) = (\theta_s, \phi_s)]$. A series of measurements were performed for a fixed polar angle of incidence $\theta_0 = 55°$, but different values of the azimuthal angle of incidence ϕ_0. The coordinate system is defined so that the value $\phi_0 = 0°$ represents a direction that coincides with one of the main axes of the square lattice. In Fig. 2 the (2×2) block diagonal elements of **m** are stacked as functions of the azimuthal angle of incidence (except m_{22}). The corresponding SEM image of Sample A and a schematic diagram of the reciprocal lattice are shown in Fig. 1(a) and (b), respectively, see also Figs. 1–3 in Brakstad et al. [2]. We observe from the diagonal elements of the Mueller matrix that a LSPR exists for a photon energy around 2.1 eV, which is here found to be slightly dispersive with azimuthal rotation angle ϕ_0, see also Ψ_{pp} and $\langle\varepsilon\rangle_{pp}$ in Ref. [2]. Furthermore, we observe polarization conversion around the LSPR that dependends on ϕ_0; this is seen from the off-block-diagonal elements (which are not stacked) in Fig. 2. We further observe that there are prominent features in both the block-diagonal elements (dips and peaks) and the off-block-diagonal elements (strong polarization conversion); these features are attributed to the Rayleigh anomalies [19] as previously described in detail in Brakstad et al. [2] and references therein.

In recent works [2,9,20,21] we have found it useful to present Mueller matrix data in terms of contour plots in polar coordinates as they clearly show their dependence on the rotation of the sample. In such plots, the polar coordinate corresponds to the azimuthal rotation angle (ϕ_0) and the radial coordinate represents the photon energy. Fig. 3 presents in this fashion the normalized Mueller matrix data of which the data-sets from Fig. 2 are subsets; the inner and outer circles in Fig. 3 correspond to 0.73 eV and 5.9 eV, respectively. Notice that for given azimuthal rotation angle (ϕ_0), a cut along the radial direction of the data-sets presented in Fig. 3 will result in curves for the same value of ϕ_0 that are similar to those in Fig. 2. Moreover, the data in Fig. 3 are organized such that $\phi_0 = \angle(\mathbf{k}_\parallel, \mathbf{G}_\parallel^{(10)})$, where the component of the incident wave vector parallel to the surface of the substrate is ($|\mathbf{k}| = k = n_0\omega/c$)

$$\mathbf{k}_\parallel = k \sin\theta_0(\cos\phi_0, \sin\phi_0, 0), \tag{1}$$

and the reciprocal lattice vector is defined as

$$\mathbf{G}_\parallel^{(\mathbf{l})} = l_1\mathbf{b}_1 + l_2\mathbf{b}_2 = G_\parallel^{(\mathbf{l})}(\cos\phi_\mathbf{l}, \sin\phi_\mathbf{l}, 0), \tag{2}$$

where l_1 and l_2 are integers and $G_\parallel^{(\mathbf{l})} = |\mathbf{G}_\parallel^{(\mathbf{l})}|$. In writing Eq. (2) we have introduced the primitive translation vectors of the reciprocal lattice $\mathbf{b}_1$ and $\mathbf{b}_2$ ($|\mathbf{b}_{1,2}| = 2\pi/a$), and defined the vector $\mathbf{l} = (l_1, l_2)$ (see Ref. [2] for further details).

It is observed from Fig. 3 that the Mueller matrix of Sample A is nearly block-diagonal for photon energies up till about 3 eV, and the same was found also for Sample B (results not shown). The LSPR is now observed as the nearly circular features around 2.1 eV in all the block diagonal elements presented in Fig. 3. The Rayleigh lines observed in e.g. the m_{12} element, now make up a set of features that resembles the first and second Brillouin zone (BZ), see Fig. 1(b).

Brakstad et al. [2] described thoroughly the main features of these Rayleigh lines [19]. In particular, in this work, a compact

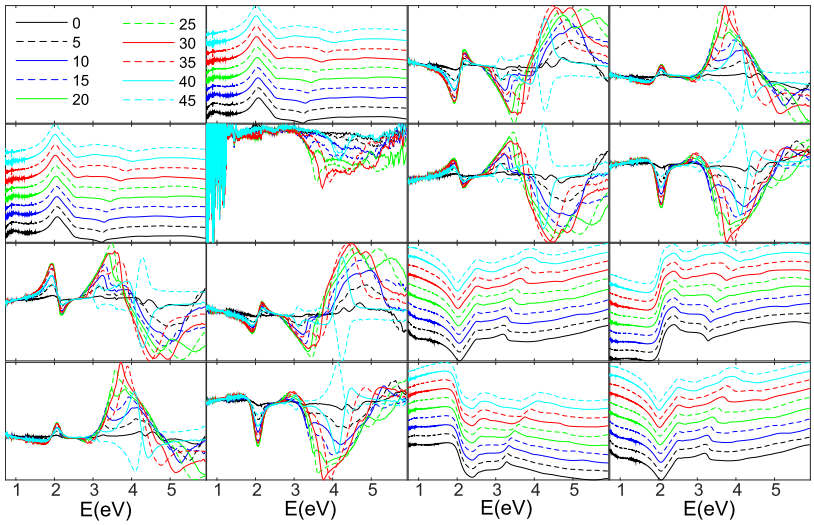


Fig. 2. Overview of the experimental normalized Mueller matrix elements for Sample A as functions of photon energy. The polar angle of incidence was $\theta_0 = 55^\circ$ and the azimuthal angle of incidence ϕ_0 was varied from 0° to 45° in steps of 5°. The measurements were done in the specular direction, i.e. $(\theta_s, \phi_s) = (\theta_0, \phi_0)$. The (2×2) block diagonals elements are stacked (except the m_{22} element), that is, a constant offset was added to each data set for reasons of clarity. The vertical limits for the off-diagonal elements are $m_{13} = [-0.103, 0.117]$, $m_{14} = [-0.079, 0.116]$, $m_{23} = [-0.097, 0.067]$, $m_{24} = [-0.148, 0.082]$. Similarly, $m_{31} = [-0.122, 0.102]$, $m_{32} = [-0.062, 0.098]$, $m_{41} = [-0.0696, 0.116]$, $m_{42} = [-0.149, 0.081]$, while $m_{22} = [0.96, 1]$. See the colorbars in Fig. 3 for the vertical limits for all elements and all azimuthal orientations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

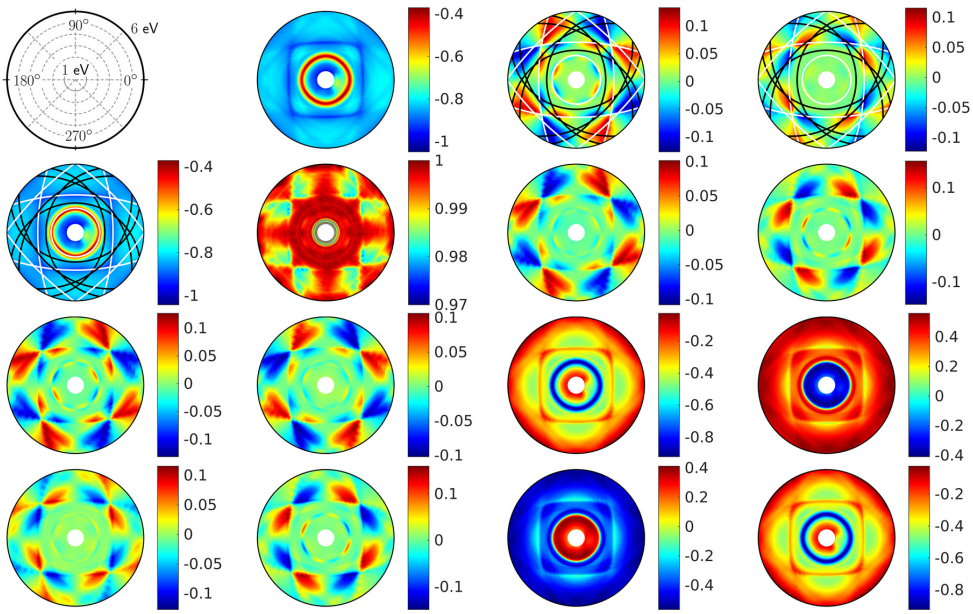


Fig. 3. Contour plots of the elements of the experimental normalized Mueller matrix **m** for the Sample A, measured for $\theta_0 = \theta_s = 55^\circ$ and $\phi_0 = \phi_s$ (where θ_s and ϕ_s denote the polar and azimuthal angles of scattering). The photon energy and the azimuthal rotation angle (ϕ_0) of the incident light represent the radius and the angle in these polar plots, respectively. The inner circles in the plots correspond to the photon energy 0.73 eV, while the outer correspond to 5.90 eV. The Rayleigh-lines for the first BZ (upright semi-square), the 2nd BZ (tilted semi-square) in air (white lines), and in the glass substrate (black lines) have been superimposed on the m_{21} element. In the m_{13} and m_{14} elements, the extended Rayleigh-lines for air (white lines) and glass (black lines) that were calculated for a 90° symmetry are superimposed, in addition to the LSPR resonance at 2.1 eV (white circles) as estimated from the quasi-static approximation. The circles in the schematic inset, replacing the m_{11} element (that is trivially one), correspond to the photon energy of the incident light from 1 eV (thick inner line) to 6 eV (thick outer line) in steps of 1 eV, while 0, 45, 90, etc. denote the azimuthal rotation angle (ϕ_0) in degrees. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

expression for the Rayleigh line condition was rederived, and it reads [2]

$$k^2 - \frac{2k\sin\theta_0 G_{\parallel}^{(\mathbf{l})}\cos(\phi_{\mathbf{l}} - \phi_0)}{n_i^2 - \sin^2\theta_0} - \frac{\left(G_{\parallel}^{(\mathbf{l})}\right)^2}{n_i^2 - \sin^2\theta_0} = 0, \tag{3}$$

where n_i denotes the refractive index of either the ambient (air) or the substrate (SiO_2). The form in Eq. (3) is extremely useful, as it allows to either solve for the Rayleigh lines, the lattice constants or even the refractive index of the substrate.

The Rayleigh lines calculated from Eq. (3) are shown in Fig. 3 as black lines for the substrate and white lines for air, and they describe the main features of the block-diagonal elements of the Mueller matrix. The polarization conversion is now clearly observed from Fig. 3 as regular dots around the LSPR (with maxima for incidence midway between $\Gamma - X$ and $\Gamma - M$). It is also clearly observed from the results presented in Fig. 3 that the polarization conversion is found around the Rayleigh lines [2], i.e. around the black (substrate) and white (air) Rayleigh lines drawn in the m_{13} and m_{14} elements.

We speculate that the polarization coupling around the LSPR is a result of spatial dispersion [22,12], while the polarization coupling around the Rayleigh lines is a result of a modification of the effective polarizability tensor as a result of the grazing diffracted waves. In the remaining part of this paper we will focus on modelling the approximately near block diagonal response below the Rayleigh lines, in particular using the Bedeaux-Vlieger formalism, and full wave COMSOL simulations.

3.2. Bedeaux-Vlieger model

Over several decades, starting in the 1970s, Bedeaux and Vlieger developed an approach to the calculation of the optical properties of thin island films (or rough surfaces) that is based on the use of effective boundary conditions [15,23–25,1,26]. The Bedeaux-Vlieger (BV) model introduces so-called surface suceptibilities, related to particle polarizabilities, that modifies the well-known Fresnel amplitudes of a flat surface to account for the presence of the island film (or surface roughness) [see Eqs. (4) and (7)]. When no particles (islands) are present at the flat surface of the substrate, the surface suceptibilities vanish and the modified Fresnel amplitudes reduce to the classic and well-known Fresnel amplitudes; in this sense the surface suceptibilities encode the effects of the presence of the island film. Within the BV model, the polarizability of a particle is calculated within the quasi-static approximation by means of adapting a multipole expansion of the scalar electric potential [15,23,24,27]. This is achieved by first calculating, to a high multipole order, the interaction between a single particle and the substrate by the use method of images [27]. Next, the single-particle polarizability is corrected for particle-particle interactions by assuming that this can be done adequately by only including dipolar or quadrupolar interactions. That is, the particle-particle interaction is included only to a low multipolar order which is expected to be a good approximation in the low particle coverage limit. In this way, the particle polarizabilities are calculated within the BV model, and from them, the surface suceptibilities and, therefore, the optical response can be calculated. For a more detailed discussion of the BV model the interested reader is referred to Refs. [15,23,24]. It should be mentioned that the BV formalism is implemented in the (open source) software GranFilm developed by Simonsen and Lazzari [25]. The advantage of the BV formalism is that it results in a fast calculation (fraction of seconds) of the full spectrum.

The BV model introduces two types of surface susceptibilities that are either *parallel* or *perpendicular* to the surface of the substrate; in accordance with Ref. [15], we will in the following denote them $\gamma(\omega)$ and $\beta(\omega)$, respectively.

When light of s polarization is incident from an ambient medium of refractive index n_0, onto an island film supported by the flat surface of a substrate that has refractive index n_2, the BV model predicts that the reflection amplitude should equal [15,25]:

$$r_{012s}(\omega) = \frac{n_0\cos\theta_0 - n_2\cos\theta_2 + i\frac{\omega}{c}\gamma(\omega)}{n_0\cos\theta_0 + n_2\cos\theta_2 - i\frac{\omega}{c}\gamma(\omega)}. \tag{4}$$

Here, θ_0 and θ_2 are the polar angles of incidence and transmission, respectively, and they are measured positive from the normal to the mean surface. It is convenient to rewrite Eq. (4) in the alternative form

$$r_{012s}(\omega) = \frac{r_{02s}(\omega) + i\frac{\omega}{c}\frac{\gamma(\omega)}{\eta_s}}{1 - i\frac{\omega}{c}\frac{\gamma(\omega)}{\eta_s}}, \tag{5}$$

with $\eta_s = n_0\cos\theta_0 + n_2\cos\theta_2$, so that the Fresnel reflection amplitude for the ambient-substrate system, $r_{02s}(\omega)$, enters explicitly; this latter quantity is defined by the expression obtained by putting $\gamma = 0$ in Eq. (4).

Similarly, for p-polarized incident light the reflection amplitude of an island film is expressed in the BV model as [15,25]:

$$r_{012p}(\omega) = \frac{\kappa_-(\omega) - i\frac{\omega}{c}\left[\gamma\cos\theta_0\cos\theta_2 - n_0 n_2\varepsilon_0\beta\sin^2\theta_0\right]}{\kappa_+(\omega) - i\frac{\omega}{c}\left[\gamma\cos\theta_0\cos\theta_2 + n_0 n_2\varepsilon_0\beta\sin^2\theta_0\right]} \tag{6a}$$

where

$$\kappa_\pm(\omega) = (n_2\cos\theta_0 \pm n_0\cos\theta_2)\left(1 - \frac{1}{4}\frac{\omega^2}{c^2}\gamma\beta\varepsilon_0\sin^2\theta_0\right). \tag{6b}$$

For reasons of a more compact presentation, we have in Eq. (6) not indicated explicitly the frequency dependence of the surface susceptibilities. Neglecting terms of second or higher order in the surface susceptibilities allow us to write Eq. (6a) approximately as

$$r_{012p}(\omega) \approx \frac{r_{02p}(\omega) - i\frac{\omega}{c}\frac{1}{\eta_p}\left[\gamma\cos\theta_0\cos\theta_2 - n_0 n_2\varepsilon_0\beta\sin^2\theta_0\right]}{1 - i\frac{\omega}{c}\frac{1}{\eta_p}\left[\gamma\cos\theta_0\cos\theta_2 + n_0 n_2\varepsilon_0\beta\sin^2\theta_0\right]}. \tag{7}$$

In writing Eq. (7) we have defined $\eta_p = n_2\cos\theta_0 + n_0\cos\theta_2$ and $r_{02p}(\omega)$ denotes the Fresnel reflection amplitude of the flat ambient-substrate surface for p-polarized illumination [27,28] and can be obtained by putting $\gamma = \beta = 0$ in Eq. (6).

Several parameters are needed to calculate the surface susceptibilities, $\gamma(\omega)$ and $\beta(\omega)$, in the approach of BV. In addition to the wavelength, angles of incidence (θ_0, ϕ_0), and refractive indices of the media involved, they are the *morphological parameters* of the hemispheroidal island film; the radii of the hemispherodal particles that are parallel (R_{xy}) and perpendicular (R_z) to the surface of the substrate, and the lattice constant, *a*. These morphological parameters are also those that one typically tries to extract during the inversion of experimental data sets using the BV model (more about this later). It should be noted that the BV model does not convey polarization coupling, and does not describe Rayleigh modes correctly; thus it is only an approximate solution for regular lattices. Indeed, the BV model was originally developed to treat the general problem of disordered particles on surfaces and the weak cross-polarized scattered signal was neglected. Moreover, the particle dimensions we will be concerned with, allow us to neglect both finite size and retardation effects in each particle (as is done in the BV model). We will in the following also neglect the plausible SiO_2 mound underneath each particle and treat the surface of the substrate as flat (i.e. as in Fig. 1(c)).

The open symbols in Figs. 4 and 5 represent the experimental normalized Mueller matrix elements m_{12} (*N*), m_{33} (*C*) and m_{34} (*S*) for Samples A and B, respectively, as functions of photon energy for

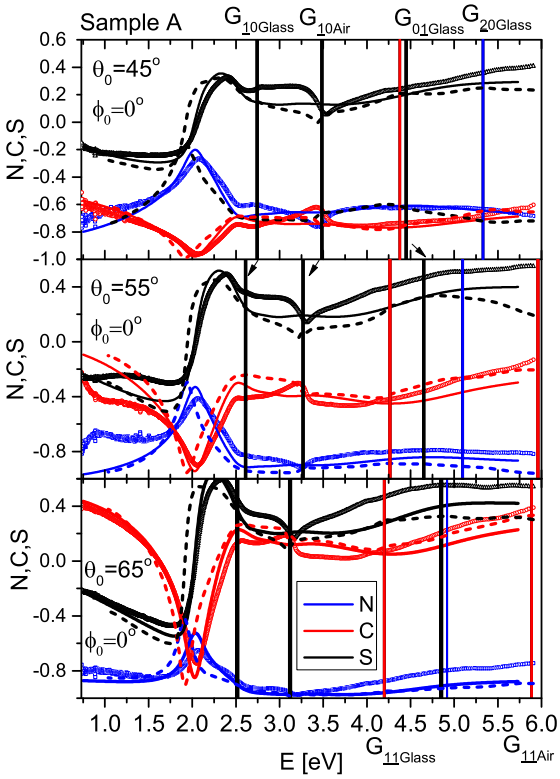


Fig. 4. The measured standard ellipsometric quantities $N = m_{21}$ (blue □); $C = m_{33}$ (red ○); and $S = m_{34}$ (black △) for Sample A as functions of photon energy for the three polar angles of incidence $\theta_0 = 45°$, 55° and 65° as indicated in each of the panels. The scattered signal was observed in the specular direction. The experimental data are shown as open symbols, and are selected for the azimuthal angle $\phi_0 = 0°$. The vertical lines indicate the photon energies of the corresponding Rayleigh lines. The results of the GranFilm simulations (the BV formalism) are shown as full lines, while the COMSOL simulations are given as dashed lines. The morphological parameters used in these simulations are given in Table 1; the GranFilm simulations assumed the parameter set corresponding to the smallest value of χ^2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

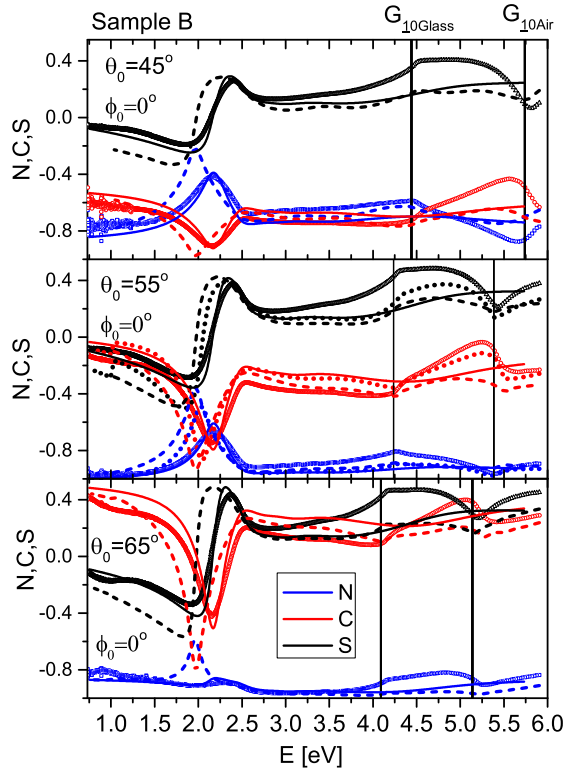


Fig. 5. Same as Fig. 4 but for Sample B. The COMSOL simulations used $(R_{xy}, R_z, a) = (38, 20, 125)$ nm (dashed lines), see Table 1. The GranFilm simulations assumed the parameter set corresponding to the smallest value of χ^2. The dotted lines shown in the middle panel for $\theta = 55°$ are COMSOL simulations with an added mound of height 20 nm, see Fig. 1(d).

the azimuthal angle of incidence $\phi_0 = 0°$ and three polar angles of incidence $\theta_0 = 45°$, 55° and 65°. The experimental data sets in Figs. 4 and 5 were then fitted with respect to the morphological parameters of the spheroidal lattice (R_{xy}, R_z and a) using the BV model as implemented in GranFilm [25]. The corresponding Mueller matrix elements obtained in this way are presented as full lines in Figs. 4 and 5 and the parameter sets that were obtained by this procedure are presented in Table 1. As can be observed from the values presented in Table 1, the morphological parameter obtained from SEM and AFM images (that are marked in boldface in Table 1) do not result in the smallest figure of merit, denoted χ^2 in the Table 1. For Sample A [Fig. 4] the lateral diameter of the particles $2R_{xy} = 116\,nm$ is at the limit of breakdown of the quasistatic approximation and some of the Rayleigh lines coincide with the position of the LSPR resulting in a perturbed signal around this photon energy. On the other hand, for Sample B a better agreement is found between the experimental morphological parameters and those resulting from a fitting procedure using the BV formalism, see Fig. 5; in this case the Rayleigh lines are shifted to higher energies and away from the LSPR, and the lateral particle size is smaller (than for Sample

Table 1
The morphology parameter sets (R_{xy}, R_z, a) (in nanometers) used in the numerical simulations performed using either GranFilm or COMSOL. The values that appear in boldface coincide with the values obtained from analyzing SEM and AFM images of the samples. It is noted that the values for R_z that were obtained from such images had considerable uncertainty associated with them. The GranFilm software was used for the purpose of reconstructing the morphology parameters of the samples by fitting the experimental data for the standard ellipsometric quantities presented in Figs. 4 and 5. This was done by defining a cost function that was minimized over parameter space using a mean square error estimator; the χ^2 values obtained in this way are reported. The star superscript (⋆) indicates that the (COMSOL) simulation result corresponds to a geometry where the particles were on top of mounds of height 20 nm.

	GranFilm				COMSOL			
	R_{xy}	R_z	a	χ^2	R_{xy}	R_z	a	χ^2
Sample A	**58**	**30**	**210**	3.7	**58**	**30**	**210**	5.5
	60	29	198	2.6	–	–	–	–
Sample B	**34**	**20**	**125**	3.5	**34**	**20**	**125**	6.4
	34	29	114	2.1	**34**	**20⋆**	**125**	3.0

A) so that the quasistatic approximation should be more accurate. However, as mentioned in Section 2, the manufactured samples did turn out to have some non-ideal characteristics, which probably will affect both the quality of the fit and the extracted morphological parameters that can be obtain from the BV approach. These aspects will be discussed in greater detail below.

It is worth noting that the BV model predicts that the optical response of a system consisting of a hemispheroidal island film supported by a planar substrate is *distinctly different* from that of an identical system but made with untruncated spheroidal particle (of the same volume) nearly touching the substrate [24,29,30]. Recently this latter system was studied by Mendoza-Galván et al. [26] who adapted the simplifying dipole approximation (known as the (modified) Yamaguchi model, see Refs. [15,31,32]). The advantage of considering untruncated particles, instead of truncated ones, in combination with the dipole approximation, is that one can easily obtain closed form expressions for the uniaxial effective medium dielectric functions [26].

3.3. Full wave simulations

The two structures, Sample A and Sample B, were also modelled using the Finite Element Method based commercial software package COMSOL Multiphysics, operating in the frequency-domain. The input to the simulations were the frequency dependent dielectric functions from the ellipsometric analysis of the deposited Au film, and SiO_2 data from the Woolam software database, in addition to the geometric parameters estimated from SEM and AFM, see Table 1. It is particularly interesting to use such a full wave tool in conjunction with real ellipsometric data, since the Finite Element Method has been regularly used for simulating the optical response of metamaterials, with consequent retrieval of effective optical parameters [8]. It is noted that this method can also be used to study modes, important in order to both interpret the optical response, but also in order to deduce bi-anisotropy and polarization coupling.

The simulations were performed using 4 ports with periodic Floquet boundary conditions to simulate the periodicity. The incident wave on port 1 was assumed either plane TE (s-polarized) or plane TM (p-polarized), along the $\mathbf{G}_{\parallel}^{(10)}$ direction. In the current work, the azimuthal angle was not changed.

The dashed lines in Figs. 4 and 5 present the standard ellipsometric quantities *N*, *C* and *S* calculated using COMSOL for both Samples A and B.

It is clear that the simulations to a large extent reproduce the block diagonal Mueller matrix, including the Rayleigh lines. This is evidently an important step in order to validate any EM simulator.

The simulations using the current model do not properly reproduce the experimental LSPR. As this problem occurs both in the BV and the COMSOL simulations, the models must be refined to include the SiO_2 mound as sketched in Fig. 1(d), which probably resulted from an over-etching during the FIB milling in between the Au particles. The dotted lines shown in the middle panel of Fig. 5 ($\theta_0 = 55^\circ$), was calculated using an additional 20 nm SiO_2 mound, as shown in Fig. 1(d). Indeed, the results show that the mound causes a blue shift of the LSPR, in addition to an improvement of the simulated response in the ultraviolet part of the spectrum. The simulation results that can be obtained using such refined models will be the topic for future work, among a full description of the polarization conversion.

In the simulations, both polarizations (TE and TM input) were run in parallel on a computer equipped with 32 GB RAM and an Intel i7-3930K processor operating at 3.2 GHz and running 64 bits Windows operating system. Typically, with an appropriate meshing, COMSOL used 1 minute per wavelength for Sample A and 13 seconds per wavelength for Sample B, with increasing computational time with added complexity (here the mound). Using a smaller number for the maximum allowed mesh size, considerably increases the computational time. It appears that for small unit cells and an optimized mesh size, fitting of ellipsometric data using finite element methods can in principle be performed on typical desk-top systems used in ellipsometric analysis. However, the required computational cost of such an approach will be orders of magnitude higher than when using GranFilm.

3.4. Direct Inversion (continuous film approximation)

It is common in the field ellipsometry to extract the effective optical properties of a thin film containing plasmonic nanoparticles. Therefore, we will below perform this analysis with the intention of comparing the results from such an approach to what can be obtained by other methods. A natural question then follows, what does an extracted dielectric function for an ultrathin layer represent? For such an ultrathin film (particles on the surface), and in view of the BV model, we learn that the resulting dielectric function is dependent on the substrate and thus not unique for the film. This behaviour is a direct consequence of the interaction between the particle and the substrate. A fitted effective dielectric function for the ultrathin effective film does not reveal the quantitative morphological parameters such as particle size/shape, density (lattice constant) and material, which is often sought in metrology. However, the effective dielectric function of this ultrathin layer may play an important role in terms of design and characterization of optical meta-surfaces, in particular with respect to controlling the propagation of surface bound waves, such as e.g. Surface Plasmon Polaritons on hyperbolic meta-surfaces [33].

The dielectric function of the effective film on the substrate was determined through a direct inversion of the Mueller matrix elements (principally sensitive to the *N*, *C*, and *S* quantities in Figs. 4 and 5). Fig. 6 shows the result of the inversion, where the dielectric tensor has been posed as an effective thin biaxial film. The film is as expected mainly uniaxial, but a small in-plane anisotropy was allowed for in order to reproduce some of the features observed in the off-diagonal elements of the Mueller matrix. The in-plane components were described by Bsplines where the initial guess was given by the fit of a Maxwell Garnett effective medium [1,34], while the out of plane component was simply described by a Bruggeman effective medium [1,34]. The thickness of the effective medium film was fixed to a "reasonable apriori-initial value" and not further fitted. The effective medium model is found to reproduce the data with high accuracy, and the fitted dielectric function gives indeed the position of the LSPR, and evidently it conveys non-quantitative information about particle-size and distribution (the distribution is evidently narrow in this idealized case), in addition to information about lattice interactions.

It is both interesting and instructive to establish a relationship between the BV formalism and the extracted dielectric function for the continuous effective film. To this end we start by neglecting the polarization coupling (in-plane anisotropy) and assume that the system has a uniaxial response. Under these assumptions, the relationship between the BV formalism and the effective medium film approach can be obtained by considering the first order Taylor expansion of the exponential function in the thin film formulas for a uniaxial film with the extraordinary axis perpendicular to the surface normal [28]. Let d be the thickness of the continuous thin film, and assume $d/\lambda \ll 1$, then

$$r_{012\nu} = \frac{r_{01\nu} + r_{12\nu}e^{-2iX_\nu}}{1 + r_{01\nu}r_{12\nu}e^{-2iX_\nu}} \approx \frac{r_{02\nu} - 2iX_\nu \frac{r_{12\nu}}{1+r_{01\nu}r_{01\nu}}}{1 - 2iX_\nu \frac{r_{01\nu}r_{12\nu}}{1+r_{01\nu}r_{12\nu}}}, \tag{8}$$

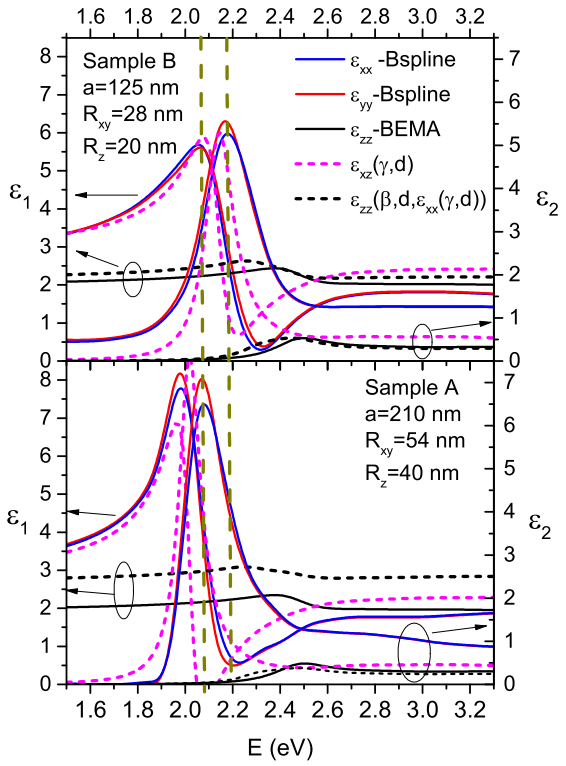


Fig. 6. Effective biaxial dielectric functions of Sample A (bottom Figure) and Sample B (top Figure). The biaxial Bspline inverted data are shown as full lines. The dielectric function calculated from γ and β (obtained from GranFilm) with the use of Eqs. (9) and (10) are shown as dashed lines. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

where $\nu = s, p$, and the expression for the Fresnel reflection amplitudes $r_{mn\nu}$ are given by Eqs. (4) and (6) after setting to zero the surface suceptibilities that appear in them; $\gamma = \beta = 0$ (see also Refs. [27,28]). In writing Eq. (8) we introduced the phase thickness (eigenmode) functions, X_ν, given in Azzam and Bashara [28].

To relate the reflection amplitudes in Eq. (8) to the reflection amplitudes obtained in the BV model, *e.g.*, Eqs. (5) and (7), we start by observing that the BV amplitude for s polarization depends only on the parallel susceptibility (γ), while the BV amplitude for p polarization depends both on the parallel and perpendicular susceptibilities (γ and β). Therefore, it is to be expected that the effective dielectric function $\varepsilon_{xx} = \varepsilon_{yy}$ will depend on γ but not β, in the same way as ε_{zz} will depend on β but not γ. In order to see if this assertion is true, we start by considering the case of s polarization. Equating the right hand side of Eq. (8) for $\nu = s$, with the corresponding BV formula in Eq. (5) gives the relationship:

$$\varepsilon_{xx} = \varepsilon_2 + \frac{\gamma}{d}, \tag{9}$$

where the thickness of the continuous thin film (d) must be chosen independently. Substituting the above expression for ε_{xx} into the right hand side of Eq. (8) for $\nu = p$, and equating the resulting expression with the corresponding BV formula, Eq. (7), results in

$$\frac{1}{\varepsilon_{zz}} = \frac{1}{\varepsilon_2} - \frac{\beta}{d}. \tag{10}$$

It is interesting to note that the effective medium dielectric functions add as disks on a substrate; additively parallel to the planar surface and capacitively perpendicular to it.

The fitted uniaxial dielectric function extracted from the ellipsometric data can now in principle be related to morphological parameters through the BV formalism as is the case for an effective medium theory. Using the calculated γ and β from the BV model implemented in GranFilm, and the relationship in Eqs. (9) and (10), we finally obtain the uniaxial dielectric function plotted (dashed lines) in Fig. 6. In both cases, the parameter d is adjusted to make a reasonable fit (we used $d = 1.0\,nm$ for Sample B and $d = 1.9\,nm$ for Sample A, but it is emphasized that these values depend on the initial choice for the thickness in the direct inversion).

The dielectric function calculated based on Eqs. (9) and (10) turns out to be in reasonable agreement with the one obtained by inversion (see Fig. 6). The difference is in particular, a small red-shift of the one calculated from the BV formalism, and the more asymmetric shape towards higher energies of the Bspline inverted one. The correspondence appears better for Sample B, as expected. This red-shift was also observed from the COMSOL simulations, and was probably a result of reduced screening by the substrate, again as a result of the over-etching of the substrate (i.e. the particles are sitting on a small SiO_2 mound). The real part of the calculated ε_z component is observed to be off-set from the inverted one. Indeed, it is difficult to invert for the z-component accurately using ellipsometry, and it is strongly correlated to the chosen film thickness. As a result, in future work, it appears better to replace the initial guess by the dielectric function produced by GranFilm.

4. Conclusions

Mueller Matrix Ellipsometry reveals a complex optical response from a simple square array of Au-nanoparticles supported by a glass substrate. In particular, polarization conversion is found around the Localized Surface Plasmon Resonance (LSPR), and along Rayleigh lines. The Rayleigh lines can be used to find the orientation of the sample, and through a simple second order equation be used to directly estimate lattice constants.

It is found that for a regular 2D lattice, the Bedaux-Vlieger formalism can extract reasonable parameters related to particle dimensions, as long as the Rayleigh anomalies are well above the LSPR. However, the weak polarization conversion around the plasmon resonance and the small dispersion of the plasmon resonance with respect to the azimuthal rotation of the substrate cannot be modelled within the current formulation of the latter formalism.

The finite element method appears, when using periodic boundary conditions, to reasonably well model the ellipsometric spectra including the Rayleigh lines.

Both methods show that imperfections in the system, such as a dielectric mound, must be included in the model in order to reproduce well the recorded data. The Bedaux-Vlieger formalism must be reworked in order to include such an over-etching into the substrate, while it can be easily added to the Finite Element model.

The GranFilm implementation of the Bedeaux-Vlieger model is computationally fast (spectra calculated in fraction of seconds), in addition to being highly instructive. The Finite Element Model appears much more versatile and seems accurate, but is computationally expensive (spectra calculated in 30 minutes to several hours, depending on number of spectral points, meshing and unit cell size).

The substrate dependent effective uniaxial dielectric function for an equivalent continuous ultra-thin film was extracted by direct inversion of the ellipsometric data. It is possibly to give a meaning to this dielectric function by e.g. correlating it to the one directly estimated in the continuous film limit of the Bedaux-Vlieger

formalism. It is envisaged that the latter dielectric function may be useful in the description of meta-surfaces, therein e.g. to estimate the propagation of surface waves.

Acknowledgements

The research of I.S. was supported in part by the Research Council of Norway, Contract No. 216699, and The French National Research Agency (ANR) under grant ANR-15-CHIN-0003-01.

References

[1] T. Oates, H. Wormeester, H. Arwin, Characterization of plasmonic effects in thin films and metamaterials using spectroscopic ellipsometry, Prog. Surf. Sci. 86 (2011) 328–376.
[2] T. Brakstad, M. Kildemo, Z. Ghadyani, I. Simonsen, Dispersion of polarization coupling, localized and collective plasmon modes in a metallic photonic crystal mapped by Mueller matrix ellipsometry, Opt. Express 23 (2015) 22800–22815.
[3] F. Monticone, A. Alù, Metamaterials and plasmonics: from nanoparticles to nanoantenna arrays, metasurfaces, and metamaterials, Chin. Phys. B 23 (2014) 047809.
[4] S. Maier, Plasmonics: Fundamentals and Applications, Springer, New York, 2007.
[5] N.G. Khlebtsov, L.A. Dykman, Optical properties and biomedical applications of plasmonic nanoparticles, J. Quant. Spectrosc. Radiat. Transf. 111 (2010) 1–35.
[6] R.A. Shelby, D.R. Smith, S. Schultz, Experimental verification of a negative index of refraction, Science 292 (2001) 77–79.
[7] N. Fang, H. Lee, C. Sun, X. Zhang, Subdiffraction-limited optical imaging with a silver superlens, Science 308 (2005) 534–537.
[8] D.R. Smith, S. Schultz, P. Markoš, C.M. Soukoulis, Determination of effective permittivity and permeability of metamaterials from reflection and transmission coefficients, Phys. Rev. B 65 (2002) 195104.
[9] L.M.S. Aas, M. Kildemo, C. Martella, M.C. Giordano, D. Chiappe, F.B. de Mongeot, Optical properties of biaxial nanopatterned gold plasmonic nanowired gridpolarizer, Opt. Express 21 (2013) 30918–30931.
[10] T.W.H. Oates, M. Ranjan, S. Facsko, H. Arwin, Highly anisotropic effective dielectric functions of silver nanoparticle arrays., Opt. Express 19 (2011) 2014–2028.
[11] M. Ranjan, T.W.H. Oates, S. Facsko, W. Möller, Optical properties of silver nanowire arrays with 35 nm periodicity, Opt. Lett. 35 (2010) 2576–2578.
[12] N. Guth, B. Gallas, J. Rivory, J. Grand, A. Ourir, G. Guida, R. Abdeddaim, C. Jouvaud, J. de Rosny, Optical properties of metamaterials: influence of electric multipoles, magnetoelectric coupling, and spatial dispersion, Phys. Rev. B 85 (2012) 115138.
[13] D. Schmidt, B. Booso, T. Hofmann, E. Schubert, A. Sarangan, M. Schubert, Generalized ellipsometry for monoclinic absorbing materials: determination of optical constants of cr columnar thin films, Opt. Lett. 34 (2009) 992–994.
[14] G.K. Larsen, Y. Zhao, Extracting the anisotropic optical parameters of chiral plasmonic nanostructured thin films using generalized ellipsometry, Appl. Phys. Lett. 105 (2014) 071109.
[15] D. Bedeaux, J. Vlieger, Optical Properties Of Surfaces, 2nd Ed., Imperial College Press, London, 2004.
[16] I. Simonsen, Optics of surface disordered systems: a random walk through rough surface scattering phenomena, Eur. Phys. J. Spec. Top. 181 (2010) 1–103.
[17] T. Nordam, P. Letnes, I. Simonsen, Numerical simulations of scattering of light from two-dimensional surfaces using the reduced Rayleigh equation, Front. Phys. 1 (2013) 8.
[18] J.-P. Banon, M. Kildemo, I. Simonsen, Accurate Metrology of Arrays of Rounded Plasmonic Particles Using Reduced Rayliegh Equations and Mueller Matrix Ellipsometry, 2017, unpublished.
[19] Lord Rayleigh, On the dynamical theory of gratings, Proc. R. Soc. (Lond.) Ser. A 79 (1907) 399–416.
[20] L.M.S. Aas, M. Kildemo, Y. Cohin, E. Søndergård, Determination of small tilt angles of short gasb nanopillars using UV–visible Mueller matrix ellipsometry, Thin Solid Films 541 (2012) 97–101.
[21] Z. Ghadyani, M. Kildemo, L.M.S. Aas, Y. Cohin, E. Søndergård, Anisotropic plasmonic Cu nanoparticles in sol–gel oxide nanopillars studied by spectroscopic Mueller matrix ellipsometry, Opt. Express 21 (2013) 30796–30811.
[22] C. Menzel, T. Paul, C. Rockstuhl, T. Pertsch, S. Tretyakov, F. Lederer, Validity of effective material parameters for optical fishnet metamaterials, Phys. Rev. B 81 (2010) 035320.
[23] I. Simonsen, R. Lazzari, J. Jupille, S. Roux, Numerical modeling of the optical response of supported metallic particles, Phys. Rev. B 61 (2000) 7722–7733.
[24] R. Lazzari, I. Simonsen, D. Bedeaux, J. Vlieger, J. Jupille, Polarizability of truncated spheroidal particles supported by a substrate: model and applications, Eur. Phys. J. B 24 (2001) 267–284.
[25] R. Lazzari, I. Simonsen, GranFilm: a software for calculating thin-layer dielectric properties and Fresnel coefficients, Thin Solid Films 419 (2002) 124–136.
[26] A. Mendoza-Galván, K. Järrendahl, A. Dmitriev, T. Pakizeh, M. Käll, H. Arwin, Optical response of supported gold nanodisks, Opt. Express 19 (2011) 12093–12107.
[27] J. Jackson, Classical Electrodynamics, 3rd Ed., John Wiley & Sons, New York, 2007.
[28] R.M.A. Azzam, N.M. Bashara, Ellipsometry and Polarized Light, North-Holland Publishing Company, Amsterdam, 1977.
[29] R. Ruppin, Surface modes and optical absorption of a small sphere above a substrate, Surf. Sci. 127 (1983) 108–118.
[30] R. Lazzari, S. Roux, I. Simonsen, J. Jupille, B. Bedeaux, V. Vlieger, Multipolar plasmon resonances in supported silver particles: the case of Ag/α-Al_2O_3(0001), Phys. Rev. B 65 (2002) 235424.
[31] T. Yamaguchi, S. Yoshida, A. Kinbara, Optical effect of the substrate on the anomalous absorption of aggregated silver films, Thin Solid Films 21 (1974) 173–187.
[32] V.A. Fedotov, V.I. Emel'yanov, K.F. MacDonald, N.I. Zheludev, Optical properties of closely packed nanoparticle films: spheroids and nanoshells, J. Opt. A, Pure Appl. Opt. 6 (2004) 155–160.
[33] J.S. Gomez-Diaz, M. Tymchenko, A. Alù, Hyperbolic metasurfaces: surface plasmons, light-matter interactions, and physical implementation using graphene strips, Opt. Mater. Express 5 (2015) 2313–2329.
[34] G.A. Niklasson, C.G. Granqvist, O. Hunderi, Effective medium models for the optical properties of inhomogeneous materials, Appl. Opt. 20 (1981) 26–30.

Paper 6:

J.-P. Banon, T. Nesse, Z. Ghadyani, M. Kildemo, and I. Simonsen. Critical dimension metrology of a plasmonic photonic crystal based on Mueller matrix ellipsometry and the reduced Rayleigh equation. *Opt. Lett.*, 42(13):2631–2634, 2017

Is not included due to copyright