Springer Series in Light Scattering

Alexander Kokhanovsky
Editor

Springer Series in Light Scattering

Volume 1: Multiple Light Scattering, Radiative Transfer and Remote Sensing

Springer

Springer Series in Light Scattering

The main purpose of new ***SPRINGER Series in Light Scattering*** is to present recent advances and progress in light scattering media optics. The topic is very broad and incorporates such diverse areas as atmospheric optics, ocean optics, optics of close-packed media, radiative transfer, light scattering, absorption, and scattering by single scatterers and also by systems of particles, biomedical optics, optical properties of cosmic dust, remote sensing of atmosphere and ocean, etc. The topic is of importance for material science, environmental science, climate change, and also for optical engineering. Although main developments in the solutions of radiative transfer and light scattering problems have been achieved in the 20th century by efforts of many scientists including V. Ambartsumian, S. Chandrasekhar, P. Debye, H. C. van de Hulst, G. Mie, and V. Sobolev, the light scattering media optics still have many puzzles to be solved such as radiative transfer in closely packed media, 3D radiative transfer as applied to the solution of inverse problems, optics of terrestrial and planetary surfaces, etc. Also it has a broad range of applications in many branches of modern science and technology such as biomedical optics, atmospheric and oceanic optics, and astrophysics, to name a few. It is planned that the Series will raise novel scientific questions, integrate data analysis, and offer new insights in optics of light scattering media.

More information about this series at http://www.springer.com/series/15365

Alexander Kokhanovsky
Editor

Springer Series in Light Scattering

Volume 1: Multiple Light Scattering, Radiative Transfer and Remote Sensing

Springer

Editor
Alexander Kokhanovsky
Vitrociset Belgium
Darmstadt
Germany

ISSN 2509-2790 ISSN 2509-2804 (electronic)
Springer Series in Light Scattering
ISBN 978-3-319-70795-2 ISBN 978-3-319-70796-9 (eBook)
https://doi.org/10.1007/978-3-319-70796-9

Library of Congress Control Number: 2017957680

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

Contributors

Adrian Doicu Institut für Methodik der Fernerkundung (IMF), Deutsches Zentrum für Luft- und Raumfahrt (DLR), Weßling, Oberpfaffenhofen, Germany

Dmitry S. Efremenko Institut für Methodik der Fernerkundung (IMF), Deutsches Zentrum für Luft- und Raumfahrt (DLR), Weßling, Oberpfaffenhofen, Germany

Otto P. Hasekamp SRON Netherlands Institute for Space Research, Utrecht, The Netherlands

Olli Jokinen Espoo, Finland

Konrad Kandler Institut fuer Angewandte Geowissenschaften, Technische Universität Darmstadt, Darmstadt, Germany

Osku Kemppinen Finnish Meteorological Institute, Helsinki, Finland

Hannakaisa Lindqvist Finnish Meteorological Institute, Helsinki, Finland

Valery A. Loiko Institute of Physics, National Academy of Sciences of Belarus, Minsk, Belarus

Diego Loyola Institut für Methodik der Fernerkundung (IMF), Deutsches Zentrum für Luft- und Raumfahrt (DLR), Weßling, Oberpfaffenhofen, Germany

Alexander A. Miskevich Institute of Physics, National Academy of Sciences of Belarus, Minsk, Belarus

Curtis D. Mobley Sequoia Scientific, Inc., Bellevue, WA, USA

Antonio Di Noia SRON Netherlands Institute for Space Research, Utrecht, The Netherlands

Timo Nousiainen Finnish Meteorological Institute, Helsinki, Finland

Thomas Trautmann Institut für Methodik der Fernerkundung (IMF), Deutsches Zentrum für Luft- und Raumfahrt (DLR), Weßling, Oberpfaffenhofen, Germany

Abbreviations

1D, 2D, 3D	One-dimensional, two-dimensional, three-dimensional
CPA	Coherent potential approximation
DDA	Discrete dipole approximation
EFA	Effective field approximation
EM	Electromagnetic
EMA	Effective medium approximation
fcc	Face-centered cubic
FDTD	Finite difference time domain
FOV	Field of view
IA	Interference approximation
KKR	Korringa–Kohn–Rostoker (method)
LC	Liquid crystal
LEED	Low-energy electron diffraction
LMS	Layer-multiple-scattering (method)
PBG	Photonic band gap
PC	Planar crystal
PCF	Pair correlation function
PDLC	Polymer-dispersed liquid crystal
PhC	Photonic crystal
QCA	Quasicrystalline approximation
RDF	Radial distribution function
RTE	Radiative transfer equation
SC	Solar cell
SNOM	Scanning near-field optical microscopy
SPR	Surface plasmon resonance
SSA	Single scattering approximation
TIR	Total internal reflection
TMM	Transfer-matrix method
TMSW	Theory of multiple scattering of waves
VSWF	Vector spherical wave function

Chapter 1
Invariant Imbedding Theory for the Vector Radiative Transfer Equation

Curtis D. Mobley

1.1 Introduction

This chapter develops invariant imbedding theory as needed to solve the time-independent vector (polarized) radiative transfer equation (VRTE) for a plane-parallel water body bounded by a wind-blown sea surface and a reflective bottom. The scalar (unpolarized) version of invariant imbedding theory applied to the same geometry is described in *Light and Water* (Mobley 1994) and is employed in the widely used HydroLight software (www.hydrolight.info). The development here parallels that work and extends the scalar equations to the vector level. However, there are additional complications in the vector theory because of loss of certain symmetries in going from the scalar to the vector theory.

The radiative transfer problem for this geometry requires solution of a linear integro-differential equation (the VRTE) subject to boundary conditions at the sea surface and bottom, which may be either finitely or infinitely deep. The VRTE constitutes a local formulation of the radiative transfer problem, which means that the equation involves spatial derivatives of the relevant variables (the radiance in the present case) and that the properties of the medium (the absorption and scattering properties) are described by their values at each point in space (Preisendorfer 1965).

The essence of invariant imbedding theory as applied to transport problems is that it transforms this local, linear, two-point boundary value problem to a pair of non-linear initial values problems in the form of ordinary differential equations for diffuse reflectances and transmittances (Preisendorfer 1958; Bellman et al. 1960). The solution in terms of reflectances and transmittances is a global formulation of the problem, which considers the response of the medium (e.g., the reflectance and transmittance of finitely thick layers of water or even of the ocean as a whole) to its inputs (e.g., the radiance incident onto the sea surface).

C. D. Mobley (✉)
Sequoia Scientific, Inc., 2700 Richards Road, Suite 107, Bellevue, WA, USA
e-mail: curtis.mobley@sequoiasci.com

A. Kokhanovsky (ed.), *Springer Series in Light Scattering*, Springer Series in Light Scattering, https://doi.org/10.1007/978-3-319-70796-9_1

The seminal idea of invariant imbedding theory in the geophysical setting traces back to Ambartsumian (1943), who was interested in computing the reflectance of light by planetary atmospheres. He realized that adding a layer of material to an optically infinitely thick medium leaves the reflectance unchanged (invariant). (Stokes (1862) used a similar idea to compute the reflectance of a stack of plates.) From this idea, Ambartsumian was able to derive an equation that could be solved for the reflectance of the entire atmosphere without first having to compute the radiance at each depth within the atmosphere. This was a pioneering global solution of a radiative transfer problem. Ambartsumian's idea was further developed by many others, most notably Chandrasekhar (1950), Sobolev (1956), Bellman and Kalaba (1956), and Preisendorfer (1958). The formulation used here follows that of Preisendorfer as presented for the scalar theory in his *Hydrologic Optics* treatise (Preisendorfer 1976).

The validity of radiative transfer theory rightly has been questioned (e.g., Preisendorfer 1965; Mishchenko 2013a, 2014) because of its phenomenological nature and heuristic (and sometimes physically indefensible) assumptions and derivations. R.W. Preisendorfer, a mathematician, worked to develop mathematically rigorous formulations of concepts such as radiance and invariance principles (Preisendorfer 1965, 1976). More recently, M.I. Mishchenko has expended enormous effort over many years to develop a physically and mathematically rigorous connection between Maxwell's equations and the VRTE, or as Preisendorfer (1965, p. 389) worded it, to construct "an analytical bridge between the mainland of physics and the island of radiative transfer theory." That connection is now rigorous (Mishchenko 2008a, 2016). Fortunately, after all of the physical and mathematical dust has settled, the VRTE still stands as a useful approximation to reality that gives results that are, for a wide range of situations, sufficiently accurate for many problems (Mishchenko 2013b). The detail and rigor that are lost in circumventing Maxwell's equations are often repaid by computational efficiency. The present paper therefore begins with a particular form of the VRTE and does not worry further about its foundations or interpretation.

The initial section formulates the problem in terms of continuous variables for depth, direction, and wavelength. The continuous variables are then discretized as needed for numerical solution of the equations. Boundary conditions at the air-water surface and ocean bottom are discussed in detail. Much of the material in these sections is well-known but is repeated here both for completeness and to emphasize certain points.

The subsequent sections comprise the core of the development: the formulation of invariant imbedding theory to solve the VRTE within the water body. The VRTE is first partitioned into separate sets of equations for upwelling and downwelling radiance, which is a key step for the application of invariant imbedding theory (Preisendorfer 1958; Bellman et al. 1960). These equations are then Fourier decomposed in the azimuthal direction. This leads first to local interaction equations, which govern how infinitesimally thin layers of water reflect and transmit light, and then to global interaction equations, which govern how finitely thick layers of water reflect and transmit light. Upward and downward sets of differential equations are developed for the operators occurring in the global interaction equations. Given the

absorbing and scattering properties of the water, solution of these equations gives a "bare slab" or "boundaryless" solution of the VRTE within the water column. Imbed and union rules are then developed, which enable the incorporation of the physical boundary conditions into the bare-slab solution for the interior of the water body. Those boundary conditions define specific environmental conditions such as the incident sun and sky radiance, surface wave state, and bottom reflectance. The solution at any depth within the water body is then obtained from the radiance incident onto the water body. Physical interpretation of the often complicated equations is emphasized and aided by graphical illustration of the physical processes. An associated technical report (Mobley 2014a) gives additional details and discusses issues such as Monte Carlo computation of the reflection and transmission properties of a wind-blown sea surface.

Dedication

Rudolph W. Preisendorfer (1927–1986) was one of the founding fathers of hydrologic optics. He was a pencil-and-paper mathematician whose driving interest was to put radiative transfer theory on a solid mathematical and physical foundation. The mathematics he developed for solving the scalar radiative transfer equation (SRTE) is the core of the HydroLight software, which I began developing with him during my postdoctoral and early career years (1977–1986). Invariant imbedding theory has proved to be robust, accurate, and extremely fast compared to many other techniques for solving the SRTE. Preisendorfer however said little about polarization and, to my knowledge, neither he nor anyone else has extended invariant imbedding theory to the VRTE in the way it is done here. If I have accomplished anything in the present work, it is because he showed the way. I humbly dedicate this extension of his work to his memory.

1.2 Formulation of the Vector Radiative Transfer Problem

This chapter assumes a basic knowledge of radiometry, optical oceanography, and polarization as given in *Light and Water* (Mobley 1994), Kattawar and Adams (1989), Kattawar (1994), and the introductory parts of Mishchenko et al. (2002). The mathematics developed here to solve the vector (or polarized) radiative transfer equation (VRTE) parallels that for the scalar (unpolarized) radiative transfer equation (SRTE). The development and solution of the SRTE as used in the unpolarized HydroLight software are given in *Light and Water*, which guides the development of the corresponding vector equations. Occasional reference will therefore be made to the scalar equations in *Light and Water*, with those scalar equations being prefaced here by an "L&W", e.g. Eq. (L&W 8.74). The mathematical development in this chapter is admittedly detailed and lengthy, but meticulous care is necessary to formulate a complete solution algorithm suitable for implementation in a computer code.

Invariant imbedding is mathematically complicated, but its computational efficiency and accuracy have been well established for the SRTE by over two decades of HydroLight usage. Invariant imbedding has several desirable features, the most important of which are

1. Radiances are computed in all directions simultaneously.
2. All orders of multiple scattering are included in the calculations.
3. Computer run time is linearly proportional to the optical depth to which the VRTE is solved.
4. The numerical calculations are in terms of initial value problems, which are well suited to numerical solution.
5. Inhomogeneous media are easily treated with no extra computational expense.

As will been seen, the invariant imbedding algorithm involves several steps:

1. The VRTE and associated boundary conditions are directionally discretized by integrating continuous functions of direction over finite solid angles. The numerically computed Stokes vectors are then exact values averaged over the finite solid angles.
2. The VRTE is rewritten as a pair of equations for upward and downward directions.
3. The VRTE is Fourier decomposed in azimuthal direction, which allows the solution in the interior of the water body to be computed as a sequence of independent "small" problems (one for each Fourier mode), rather than as one "large" problem (for all Fourier modes simultaneously).
4. These upward and downward equations are reformulated in terms of reflectance and transmittance functions for finitely thick layers of water. Those reflectance and transmittance functions depend only in the inherent optical properties of the water body and can be computed independently of the particular boundary conditions (i.e., incident lighting and bottom reflectance) applied to the water body.
5. The surface and bottom boundary conditions are then applied to the generic interior solution to obtain the solution of the VRTE for specific conditions of incident lighting, bottom reflectance, and internal sources.
6. Finally, the physical Stokes vectors are then reconstituted from the Fourier amplitudes.

1.2.1 Terminology and Notation

The state of polarization of a light field is specified by the four-component Stokes vector, whose elements are related to the complex amplitudes of the electric field vector $\boldsymbol{E}$ resolved into directions that are parallel ($E_\parallel$) and perpendicular ($E_\perp$) to a conveniently chosen reference plane. In the present development, this reference plane is the meridian plane, which is perpendicular to the mean sea surface. However, there are two versions of the Stokes vector seen in the literature, and these two versions have

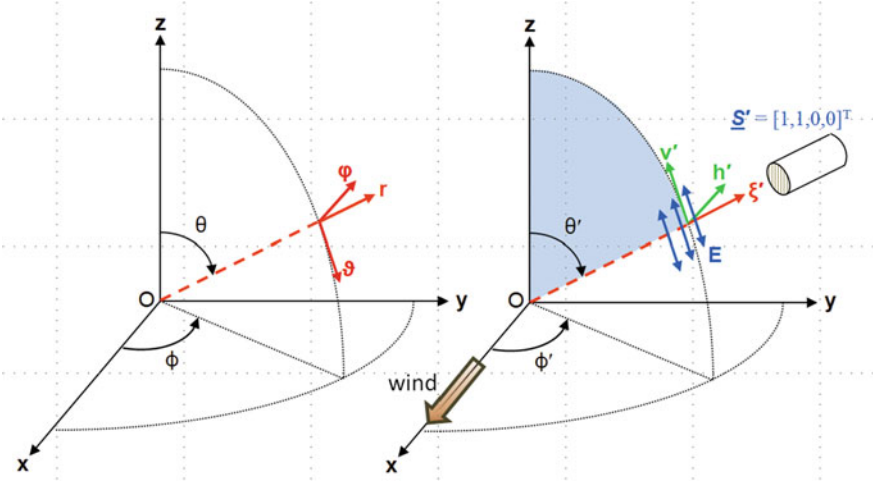

Fig. 1.1 Coordinate systems. The left panel shows the **x**-**y**-**z** Cartesian system, which is fixed in space and used to define the spherical (r, θ, ϕ) coordinate system. The right panel shows incident radiance propagating in direction $\boldsymbol{\xi}'$ whose Stokes vector is $\underline{S}'$. The $\mathbf{h}'$-$\mathbf{v}'$-$\boldsymbol{\xi}'$ system is a local system (changing with θ', ϕ') for defining direction and resolving Stokes vectors with respect to directions that are perpendicular and parallel to the incident meridian plane, part of which is shaded in blue. In the oceanographic setting, this figure is turned "upside down" as seen in Fig. 1.5, with the **x-y** plane being the mean sea surface, **x** pointing downwind, and **z** pointing downward into the water

different units and refer to different physical quantities. The *coherent* Stokes vector describes a quasi-monochromatic plane wave propagating in one exact direction, and the vector components have units of power per unit area (i.e., irradiance) on a surface perpendicular to the direction of propagation. With the choice of the meridian plane as the reference, $E_\parallel = E_\theta$ is the θ component of the electric field and $E_\perp = E_\phi$ is the ϕ component, as seen in Fig. 1.1. The coherent Stokes vector is defined as (Mishchenko et al. 2002)

$$\underline{S} = \begin{bmatrix} I \\ Q \\ U \\ V \end{bmatrix} = \frac{1}{2}\sqrt{\frac{\epsilon_m}{\mu_m}} \begin{bmatrix} E_\theta E_\theta^* + E_\phi E_\phi^* \\ E_\theta E_\theta^* - E_\phi E_\phi^* \\ -E_\theta E_\phi^* - E_\phi E_\theta^* \\ i(E_\phi E_\theta^* - E_\theta E_\phi^*) \end{bmatrix} . \tag{1.1}$$

Here ϵ_m is the electric permittivity of the medium (with units of Farad/m^2 or SI units of $\mathrm{A}^2\,\mathrm{s}^4\,\mathrm{kg}^{-1}\,m^{-3}$), and μ_m is the magnetic permeability of the medium (Henry/m or $\mathrm{kg\,m\,s}^{-2}\,\mathrm{A}^{-2}$). Electric fields have units of Newton/coulomb or $\mathrm{kg\,m\,s}^{-3}\,\mathrm{A}^{-1}$. Thus the elements of the coherent Stokes vector have units of $\mathrm{kg\,s}^{-3}$ or Watt/m^2, i.e. of irradiance. E^* denotes complex conjugate, hence the components of the Stokes vector are real numbers.

The *diffuse* Stokes vector is defined as in Eq. (1.1) but describes light propagating in a small set of directions surrounding a particular direction and has units of power

Table 1.1 Notation

Notation	Examples	Usage
Primed variables	$\theta', \phi', \boldsymbol{\xi}'$	Incident or unscattered directions
Unprimed variables	$\theta, \phi, \boldsymbol{\xi}$	Final or scattered directions
Bold face	$\mathbf{p}, \boldsymbol{\xi}$	Points in space or unit vectors defining directions in 3D space
Underline	$\underline{S}, \underline{P}$	4×1 Stokes vectors $\underline{S}$ or 4×4 matrices whose elements are numbers
Double underline	$\underline{\underline{R}}, \underline{\underline{\tau}}$	Composite matrices whose elements are themselves matrices or vectors
Tilde	$\underline{\tilde{M}}, \tilde{P}_{i,j}$	Matrices or their elements that have been "reduced" by factoring out the (1,1) element of the corresponding matrix, e.g. $\underline{P}$, so that $\tilde{P}_{i,j} = P_{i,j}/P_{1,1}$; hence $\tilde{P}_{1,1} = 1$
Caret or hat	$\underline{\hat{\tau}}, \hat{P}_{i,j}$	Fourier amplitudes of the respective (unhatted) physical quantity

per unit area per unit solid angle (i.e., radiance). It is the diffuse Stokes vector that appears in the radiative transfer equation as developed here. The differences in coherent and diffuse Stokes vectors are rigorously presented in Mishchenko et al. (2002).

Authors commonly omit the $\frac{1}{2}\sqrt{\epsilon/\mu_m}$ factor in Eq. (1.1) because they are interested only in relative values such as the degree of polarization, not absolute magnitudes, but this omission is both confusing and physically incorrect. Units and magnitudes matter! In particular, the different units of coherent and diffuse Stokes vectors, and whether or not the $\frac{1}{2}\sqrt{\epsilon/\mu_m}$ factor is included in the definition of the Stokes vector, have subtle but very important consequences regarding how light propagation across a dielectric interface such as the air-water surface is formulated, as will be seen below.

SI units are used here for all quantities except wavelength, which is in nanometers as is customary in optical oceanography. Thus spectral radiance has units of $\mathrm{W\,m^{-2}\,sr^{-1}\,nm^{-1}}$ and spectral irradiances have units of $\mathrm{W\,m^{-2}\,nm^{-1}}$. The mathematics needed to solve the VRTE using invariant imbedding theory becomes quite complicated, e.g., matrices whose elements are matrices, all of which eventually must be turned into a computer program. Precise terminology and notation are required to keep everything straight. Table 1.1 shows the notation used.

1.2.2 Coordinate Systems

Figure 1.1 shows the coordinate systems used to resolve Stokes vectors as needed for scattering calculations. Depth and direction are defined in a three-dimensional (3D)

Cartesian coordinate system with depth measured positive downward from 0 at the mean sea surface. Polar angle θ is defined from 0 in the $+\mathbf{z}$ or downwelling direction to π in the $-\mathbf{z}$ upwelling direction. $+\mathbf{x}$ points in the downwind direction; thus $\pm\mathbf{y}$ is the cross-wind direction. As will be seen in Sect. 1.6.1, using a wind-centered coordinate system makes it easier to model a random sea surface with different along-wind and cross-wind slope statistics. Azimuthal angle ϕ is measured counterclockwise from $+\mathbf{x}$ when looking in the $-\mathbf{z}$ direction. The sun is placed at an azimuthal angle of ϕ_{sun} relative $+\mathbf{x}$. If the sun is placed in the $+\mathbf{x}$ azimuthal direction at $\phi_{\text{sun}} = 0$, unscattered light from the sun is then traveling in the $-\mathbf{x}$ direction at $\phi = 180°$.

The unit vectors in the directions of increasing (r, θ, ϕ) are

$$\begin{aligned}\mathbf{r} &= \sin\theta\cos\phi\,\mathbf{x} + \sin\theta\sin\phi\,\mathbf{y} + \cos\theta\,\mathbf{z}\,,\\ \boldsymbol{\vartheta} &= \cos\theta\cos\phi\,\mathbf{x} + \cos\theta\sin\phi\,\mathbf{y} - \sin\theta\,\mathbf{z}\,,\\ \boldsymbol{\varphi} &= -\sin\phi\,\mathbf{x} + \cos\phi\,\mathbf{y}\,.\end{aligned}$$

Let $\boldsymbol{\xi}$ denote a unit vector pointing in the direction of light propagation, as given by angles (θ, ϕ). (In Fig. 1.1, $\boldsymbol{\xi}' = \mathbf{r}$.) The components of $\boldsymbol{\xi}$ are given by

$$\boldsymbol{\xi} = \xi_x\mathbf{x} + \xi_y\mathbf{y} + \xi_z\mathbf{z} = (\xi_x, \xi_y, \xi_y) \tag{1.2}$$

where

$$\xi_x = \sin\theta\cos\phi\,, \quad \xi_y = \sin\theta\sin\phi\,, \quad \xi_z = \cos\theta\,.$$

Direction (θ, ϕ) always refers to the direction of propagation. Thus light traveling straight down is traveling in the $+\mathbf{z}$ or $\theta = 0$ direction. To detect radiance in the (θ, ϕ) direction, an instrument is pointed in the viewing direction $(\theta_v, \phi_v) = (\theta - \pi, \phi + \pi)$, which is sometimes more convenient for plotting. A prime generally denotes an incident or unscattered direction, e.g., $\boldsymbol{\xi}'$ or (θ', ϕ'). Unprimed variables denote final or scattered directions, e.g. $\boldsymbol{\xi}$ or (θ, ϕ). However, in the development of the sea-surface reflectance and transmittance equations in Sect. 1.4, it will be more convenient to use subscripts i, r, and t for incident, reflected, and transmitted (refracted) directions, respectively, e.g. $\boldsymbol{\xi}_i$, $\boldsymbol{\xi}_r$, and $\boldsymbol{\xi}_t$.

There will be frequent need to define unit vectors in directions perpendicular and parallel to a given plane (which can be a meridian plane, a scattering plane in the water volume, or the plane of incident, reflected, and transmitted light for light incident onto an air-water or bottom surface, all of which are defined below). The following conventions define these unit vectors. Vector $\mathbf{p}$ denotes a vector parallel to a plane, and $\mathbf{s}$ denotes a vector perpendicular to a plane. The perpendicular vector $\mathbf{s}$ is chosen to be in the direction given by the vector cross product of the incident direction crossed with the final direction. The parallel vector $\mathbf{p}$ is then defined as the direction of propagation cross the perpendicular direction. Thus the perpendicular cross parallel directions give the direction of light propagation: $\mathbf{s} \times \mathbf{p} = \boldsymbol{\xi}$, where $\times$ denotes the vector cross product.

For an incident direction $\boldsymbol{\xi}'$ and the associated Stokes vector $\underline{S}'$ specified in the incident meridian plane, the first vector is taken to be $\mathbf{z}$ and the second is the direction of propagation. Thus $\mathbf{h}' \equiv \mathbf{z} \times \boldsymbol{\xi}' / |\mathbf{z} \times \boldsymbol{\xi}'|$ as seen in the right panel of Fig. 1.1. In this case $|\mathbf{z} \times \boldsymbol{\xi}'| = \sin\theta'$, and $\mathbf{h}' = \boldsymbol{\varphi}'$. Similarly, $\mathbf{v}' \equiv \boldsymbol{\xi}' \times \mathbf{h}' = -\boldsymbol{\vartheta}'$, and $\mathbf{h}' \times \mathbf{v}' = \boldsymbol{\xi}'$. Meridian planes are perpendicular to the mean sea surface. The $\mathbf{h}'$ vector as just defined is therefore parallel to the mean sea surface and therefore is often referred to as the "horizontal" direction; $\mathbf{v}'$ lies in a vertical plane and is correspondingly called the "vertical" direction. For a final direction $\boldsymbol{\xi}$ and its Stokes vector $\underline{S}$ in the final meridian plane, the first vector is the direction $\boldsymbol{\xi}$ and the second is $\mathbf{z}$. This vector cross product algorithm for specifying perpendicular and parallel directions is convenient for sea surface reflectance and transmission calculations in which light can propagate from one tilted wave facet to another without reference to meridian planes, except for the incident and final directions when light enters or leaves the region of the sea surface (see Mobley 2015).

The Q and U components of a Stokes vector describe linear polarization with the plane of polarization specified relative to a particular coordinate system. The I component is the total radiance, and V describes circular polarization; these quantities do not depend on the coordinate system and are invariant under a rotation of the coordinate system. The blue arrows in the right panel of Fig. 1.1 represent the plane of oscillation of the electric field vector $\boldsymbol{E}$ parallel to the meridian plane, i.e. for vertical plane polarization. Thus $E_\theta \neq 0$ and $E_\phi = 0$. According to Eq. (1.1), the I and Q components are non-zero and equal, and the U and V components are zero. The components of $\underline{S}' = [1, 1, 0, 0]^T$ (superscript T denotes transpose) as shown thus represent radiance of magnitude 1 $\mathrm{W\,m^{-2}\,sr^{-1}\,nm^{-1}}$ that is 100% vertically plane polarized. If the blue arrows of Fig. 1.1 are perpendicular to the meridian plane, Eq. (1.1) gives the Stokes vector for horizontal polarization (parallel to the mean sea surface) proportional to $\underline{S}' = [1, -1, 0, 0]^T$.

For the development of invariant imbedding theory for the VRTE, it will be convenient to divide the entire water body into the sea surface, the water column, and the bottom, as shown in Fig. 1.2. Depth $z = a = 0$ refers to a location in the air just above the mean sea surface; depth $z = w = 0$ is in the water just below the mean sea surface. The sea surface is thus a "layer" of zero thickness, which physically represents a discontinuity in the index of refraction between the air above and the water below. This layer will be denoted $[a, w]$. The water column extends from depth $z = w = 0$ to some maximum depth of interest at $z = m$. This depth m is the maximum depth to which the VRTE will be solved. The water body can be divided into sublayers, e.g., $[w, z]$ is the water from the surface to any depth z, and $[z, m]$ is the water from depth z to the maximum depth of interest. The absorption and scattering properties within the water body are generally functions of depth; the water-column layer notation $[w, z]$ and $[z, m]$ does not imply that the layers are homogeneous. A bottom boundary condition is applied at depth m. This bottom boundary condition can describe either an opaque reflecting bottom physically located at depth $z = m = b$, or it can describe the reflectance properties of an infinitely deep, homogeneous, source-free

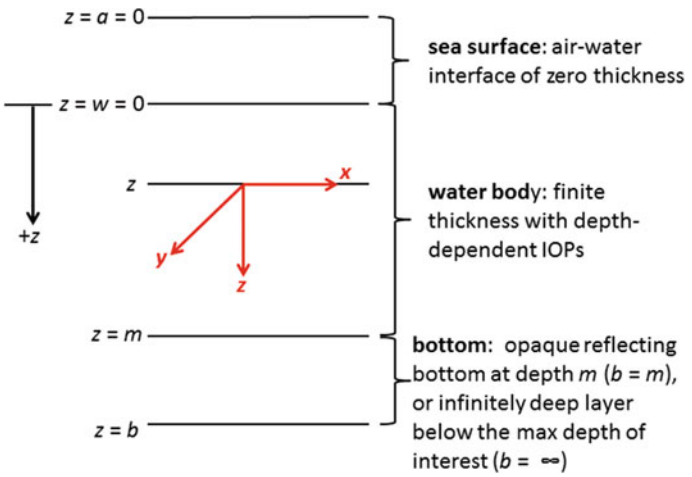

Fig. 1.2 Partition of the entire water body into surface, water column, and bottom layers

water body below depth m, in which case $b = \infty$ and the bottom layer of water is denoted $[m, \infty]$. The entire water column is the union of these surface, water, and bottom layers: $[a, b] = [a, w] \cup [w, m] \cup [m, b]$.

1.2.3 The Vector Radiative Transfer Equation

The time-independent, source-free VRTE for particles in an absorbing medium has the form (Mishchenko 2008a, Mishchenko 2008b)

$$\boldsymbol{\xi} \cdot \nabla \underline{S}(\mathbf{r}, \boldsymbol{\xi}) = \left[-2k'' - n_o \langle \underline{K}(\boldsymbol{\xi}) \rangle\right] \underline{S}(\mathbf{r}, \boldsymbol{\xi}) + n_o \int_{4\pi} \langle \underline{Z}(\boldsymbol{\xi}', \boldsymbol{\xi}) \rangle \, \underline{S}(\boldsymbol{\xi}') \, d\Omega(\boldsymbol{\xi}') \,. \tag{1.3}$$

Here k'' is the complex part of the wavenumber of the host medium. $\langle \underline{K} \rangle$ and $\langle \underline{Z} \rangle$ are, respectively, the 4×4 Stokes extinction and phase matrices, averaged over all particle states. n_o is the particle density in units of particles per cubic meter. $d\Omega(\boldsymbol{\xi}')$ denotes an element of solid angle centered on direction $\boldsymbol{\xi}'$, and the integration is over all 4π steradians of direction. Mishchenko (2007) gives $\underline{K}$ and $\underline{Z}$ as functions of the amplitude scattering matrix of electromagnetic theory.

Invariant imbedding theory is applicable to one-dimensional (1D) geometries, i.e. to physical situations for which the optical properties of the medium and the radiance vary in only one spatial dimension. Consider, therefore, a plane-parallel ocean in which the absorbing and scattering properties of the water can vary arbitrarily with depth but do not vary with horizontal position. These properties, collectively

called the inherent optical properties (IOPs), also vary with wavelength λ. The same holds true of the surface and bottom boundary conditions; they are the same for all horizontal positions. The 3D VRTE of Eq. (1.3) is therefore restricted to depth dependence only.

In addition, it is sufficient for the present purposes to consider the case of a dilute mixture of randomly oriented, mirror-symmetric particles. (Here "dilute" means that the individual particles are separated by many times the wavelength of the light, so that each particle is in the scattering far field of the others.) For such particles, the extinction matrix becomes diagonal and each of the diagonal elements $\underline{K}_{i,i}$ is identical. In addition, the phase matrix then depends on the included angle ψ between $\boldsymbol{\xi}'$ and $\boldsymbol{\xi}$ and not on the directions themselves. Using Eq. (1.2) to express the incident direction $\boldsymbol{\xi}'$ and scattered direction $\boldsymbol{\xi}$ in terms of the incident and scattered polar and azimuthal angles gives the scattering angle ψ:

$$\cos\psi = \boldsymbol{\xi}' \cdot \boldsymbol{\xi} = \cos\theta' \cos\theta + \sin\theta' \sin\theta \cos(\phi - \phi') \,, \tag{1.4}$$

where the dot in $\boldsymbol{\xi}' \cdot \boldsymbol{\xi}$ denotes the vector dot or inner product. $\underline{Z}$ also becomes block diagonal for randomly oriented, mirror-symmetric particles.

To conform with the notation used in optical oceanography, the particle number density n_o is combined with the effective extinction matrix to give the total beam attenuation coefficient

$$c = 2k'' + n_o \langle \underline{K}_{1,1} \rangle$$

In Eq. (1.3), $\langle \underline{K}(\boldsymbol{\xi}) \rangle$ defines the average extinction cross section per particle and has units of area. Multiplying by the particle number density n_o gives units of inverse meters, as is customary for the beam attenuation coefficient. Similarly, the phase matrix is reformulated as

$$\underline{P}(\boldsymbol{\xi}' \cdot \boldsymbol{\xi}) = n_o \langle \underline{Z}(\boldsymbol{\xi}' \cdot \boldsymbol{\xi}) \rangle \,.$$

The resulting VRTE then becomes

$$\begin{aligned} \cos\theta \frac{d\underline{S}(z, \boldsymbol{\xi}, \lambda)}{dz} &= -c(z, \lambda)\underline{S}(z, \boldsymbol{\xi}, \lambda) \\ &\quad + \int_{4\pi} \underline{P}(z, \boldsymbol{\xi}' \cdot \boldsymbol{\xi}, \lambda)\, \underline{S}(z, \boldsymbol{\xi}', \lambda)\, d\Omega(\boldsymbol{\xi}') \\ &\quad + \underline{\sigma}(z, \boldsymbol{\xi}, \lambda) \,. \end{aligned} \tag{1.5}$$

For generality, an internal source term $\underline{\sigma}$ has been added to the VRTE. This term represents radiance created per unit path length at wavelength λ; $\underline{\sigma}$ has units of radiance per meter or W m^{-3} sr^{-1} nm^{-1}. If the source is bioluminescence, the radiant energy is created from chemical energy. If the source represents Raman scatter or fluorescence, the emitted energy at wavelength λ comes from inelastic scatter from other wavelengths. The absorbing and scattering properties of the water body are

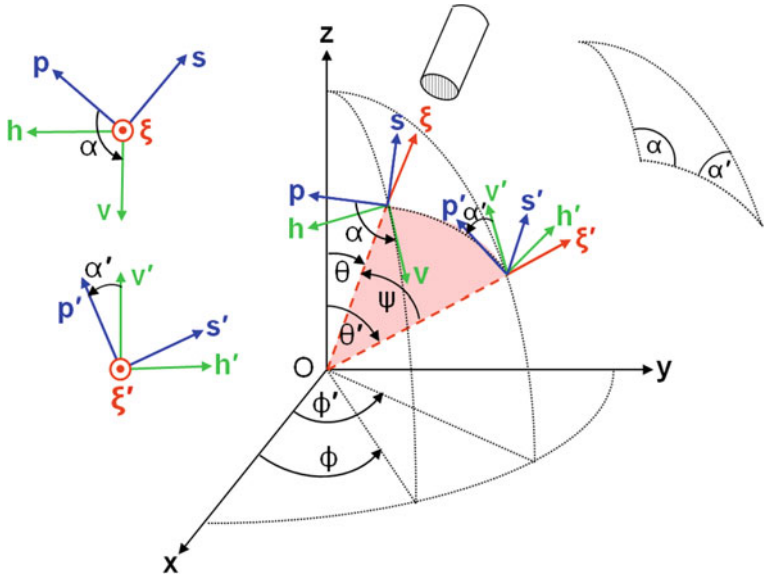

Fig. 1.3 The scattering plane and the incident and final meridian planes showing the coordinate systems and rotations used to specify scattering of polarized light. The scattering plane is partly shaded in red

now specified by c and $\underline{P}$, which are taken to be functions of depth z and wavelength λ. These inherent optical properties (IOPs) fully specify the optical properties of the water column. In practice, these quantities are often obtained as sums of the contributions by pure water, particles (phytoplankton, minerals, detritus), and dissolved substances for each particular water body. An example of the construction of these IOPs can be seen in Chowdhary et al. (2006). Although the contributions of the various water-column constituents to the total IOPs are of interest in understanding the ocean as an ecosystem, it is only the total IOPs that appear in the VRTE.

1.2.3.1 Scattering Geometry

The reduction of the directional dependence of $\underline{P}$ to $\boldsymbol{\xi}' \cdot \boldsymbol{\xi}$ gives an additional important simplification of the VRTE. The elements of input and output Stokes vectors are defined relative to meridian planes, as described above. However, scattering from an incident direction $\boldsymbol{\xi}'$ to a final direction $\boldsymbol{\xi}$ is defined in terms of the included scattering angle ψ and the scattering plane, as illustrated in Fig. 1.3.

Consider an incident beam of light propagating in direction $\boldsymbol{\xi}' = \mathbf{r}$ as in Fig. 1.1. Direction $\boldsymbol{\xi}'$ is specified by polar and azimuthal directions (θ', ϕ'). The $\mathbf{z}$ axis and the direction of light propagation $\boldsymbol{\xi}'$ define the incident meridian plane, part of which is shaded in blue in the right panel of Fig. 1.1. The 4×1 (diffuse) Stokes vector

$\underline{S}' = [I', Q', U', V']^T$ for this beam of light is described with reference to "horizontal" and "vertical" directions, $+\mathbf{h}'$ and $+\mathbf{v}'$ respectively, which were defined above. Note that the horizontal unit vector $+\mathbf{h}'$ is perpendicular to the meridian plane, and the vertical vector $+\mathbf{v}'$ is parallel to the meridian plane.

To compute how an incident Stokes vector $\underline{S}'$ is scattered to a final vector $\underline{S}$, the horizontal and vertical components of $\underline{S}'$ in the incident meridian plane must first be transformed ("rotated") into components parallel and perpendicular to the scattering plane. The coordinate system after rotation of $\mathbf{v}'$ and $\mathbf{h}'$ about the $\boldsymbol{\xi}'$ axis is labeled $\mathbf{p}'$ (parallel to the scattering plane) and $\mathbf{s}'$ (perpendicular to the scattering plane). Note that $\mathbf{s}' \times \mathbf{p}'$ still gives the direction of propagation $\boldsymbol{\xi}'$. As shown in Fig. 1.3, rotation angle α' takes $\mathbf{v}'$ into $\mathbf{p}'$ (and $\mathbf{h}'$ into $\mathbf{s}'$). In this chapter, rotation angles are defined as positive for counterclockwise rotations when looking "into the beam," e.g. in the $-\boldsymbol{\xi}'$ direction. This is similar to rotations about the $\mathbf{z}$ axis of Fig. 1.1 having positive angles ϕ for counterclockwise rotations when looking in the $-\mathbf{z}$ direction.

When computing single scattering with both $\boldsymbol{\xi}'$ and $\boldsymbol{\xi}$ being expressed in their respective meridian planes, the rotation angles can be obtained from spherical trigonometry applied to the triangle defined by $\mathbf{z}$, $\boldsymbol{\xi}'$, and $\boldsymbol{\xi}$, which is shown in the inset in Fig. 1.3. Given θ', ϕ', θ, ϕ, spherical trigonometry gives the rotation angles α' and α as (e.g., van de Hulst 1980, Vol. 2, p. 499; or Mishchenko et al. 2002, p. 90)

$$\cos\alpha' = (\cos\theta - \cos\theta'\cos\psi)/(\sin\psi\sin\theta') \tag{1.6}$$

or

$$\sin\alpha' = -\sin\theta\sin(\phi - \phi')/\sin\psi\,. \tag{1.7}$$

and

$$\cos\alpha = (\cos\theta' - \cos\theta\cos\psi)/(\sin\psi\sin\theta) \tag{1.8}$$

or

$$\sin\alpha = -\sin\theta'\sin(\phi - \phi')/\sin\psi \tag{1.9}$$

for $\psi \neq 0$ or π and for $0 \leq \phi - \phi' \leq \pi$. If $\pi < \phi - \phi' < 2\pi$, then α' and α are given by the negatives of these equations. The scattering angle ψ is given by Eq. (1.4). The rotation and scattering angles depend only on the difference in azimuthal angles via $\phi - \phi'$. Special cases are required when θ', θ, or ψ are zero. For $\psi = 0$, set $\alpha' = \alpha = 0$ since $\phi' = \phi \implies \sin(\phi - \phi') = 0$. For $\psi = \pi$, set $\alpha' = \alpha = 0$ since $\phi = \phi' + \pi \implies \sin(\phi - \phi') = 0$. If $\sin\theta' = 0$, replace Eqs. (1.6) and (1.8) with (Hu et al. 2001)

$$\cos\alpha' = -\cos\theta'\cos(\phi - \phi') \tag{1.10}$$

$$\cos\alpha = \cos\theta'\,. \tag{1.11}$$

If $\sin\theta = 0$, replace Eqs. (1.6) and (1.8) with

$$\cos\alpha' = \cos\theta \tag{1.12}$$

$$\cos\alpha = -\cos\theta\cos(\phi - \phi')\,. \tag{1.13}$$

When doing calculations of multiple scattering between sea surface wave facets, a light ray can reflect from one wave facet to another several times before the incident ray finally leaves the surface region and needs to be rotated into the final meridian plane. In this case, it is more convenient to obtain the rotation angles from the perpendicular (or parallel) axes as determined for the incident ray direction onto a facet and the normal to the tilted wave facet. The details of these calculations are given in Mobley (2014b, 2015) and will not be repeated here. However, it can be seen from Fig. 1.3 that the rotation angles can be obtained from $\cos^{-1}(\mathbf{v}' \cdot \mathbf{p}')$ and $\cos^{-1}(\mathbf{p} \cdot \mathbf{v})$.

Once the incident Stokes vector is specified in the scattering plane, a scattering matrix $\underline{M}(\psi)$ is applied to obtain the final Stokes vector, which is then expressed in the **s**-**p**-$\boldsymbol{\xi}$ scattering plane coordinate system defined for the final direction: $\mathbf{s} \times \mathbf{p} = \boldsymbol{\xi}$. Finally, the parallel and perpendicular components of the final Stokes vector must be expressed as horizontal and vertical components in the final meridian plane as specified by the **h**-**v**-$\boldsymbol{\xi}$ system. As illustrated in Fig. 1.3, this requires a counterclockwise rotation through an angle of α, where α is the "interior" angle of the spherical triangle illustrated in the figure. If $\underline{R}(\gamma)$ represents a counterclockwise (positive) rotation through angle γ and $\underline{M}(\psi)$ represents scattering through scattering angle ψ, then this scattering process is symbolically represented by

$$\underline{S} = \underline{R}(\alpha)\underline{M}(\psi)\underline{R}(\alpha')\underline{S}'\,. \tag{1.14}$$

In other words, the phase matrix can be decomposed as

$$\underline{P}(\zeta; \theta', \phi', \rightarrow \theta, \phi; \lambda) = \underline{R}(\alpha)\,\underline{M}(\psi)\,\underline{R}(\alpha')\,. \tag{1.15}$$

The phase matrix transforms $\underline{S}'$ into $\underline{S}$, with both vectors expressed in their respective meridian planes. This corresponds to how Stokes vectors are measured in the field. The scattering matrix corresponds to how scattering is measured as a function of scattering angle in an instrument.

For the choice of a positive rotation being counterclockwise when looking into the beam, the Stokes vector rotation matrix is (e.g., Mishchenko et al. 2002, p. 25)

$$\underline{R}(\gamma) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\gamma & -\sin 2\gamma & 0 \\ 0 & \sin 2\gamma & \cos 2\gamma & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}\,. \tag{1.16}$$

The choice of coordinate systems and rotation angles is not unique. Kattawar and Adams (1989), Kattawar (1994), Zhai et al. (2012), and Mishchenko et al. (2002) all choose the reference plane to the be meridian plane. (These authors use somewhat

different notation; our **h** and **v** are Kattawar's **r** and **l**, respectively. Mishchenko et al. (2002, p. 16) on the other hand use the spherical coordinate system unit vectors $\boldsymbol{\vartheta}$ and $\boldsymbol{\varphi}$ to specify the vertical and horizontal directions.) Thus in the ocean setting these authors all regard the "horizontal" direction (parallel to the mean sea surface) as being the "perpendicular" direction (relative to the meridian plane), and "vertical" to the mean sea surface as being the "parallel" direction. However, Bohren and Huffman (1983) and Hecht (1989) choose their "parallel" direction to be parallel to a horizontal plane, such as a laboratory bench top or the mean sea surface, and their vertical direction is perpendicular to the bench top or mean sea surface. Thus their Stokes vector for horizontal polarization (parallel to the mean sea surface) is proportional to $\underline{S} = [1, 1, 0, 0]^T$. This different choice arises from the viewpoint of describing polarization in a convenient way for a laboratory experiment with reference to a table top, versus modeling light incident onto the sea surface with reference to meridian planes.

Similar confusion is found in the choice of rotation angles. Kattawar (and his students in their papers) and Bohren and Huffman (1983) define a positive rotation as being *clockwise* when looking into the beam. Since a clockwise rotation through angle γ is the same as a counterclockwise rotation through $-\gamma$, Kattawar's rotation matrix is the transpose of the one in Eq. (1.16). Thus Kattawar (e.g., in Kattawar and Adams 1989, Eq. 10) writes Eq. (1.14) as $\underline{S} = \underline{R}(-\alpha)\underline{M}(\psi)\underline{R}(-\alpha')\underline{S}'$ (again, with minor differences in notation; α' here is Kattawar's Φ, etc.). Others often write a rotation as $\underline{R}(\pi - \alpha)$ rather than $\underline{R}(-\alpha)$; these are equivalent because Eq. (1.16) shows that $\underline{R}(\pi - \alpha) = \underline{R}(-\alpha)$. Chandrasekhar (1950) also defines a positive rotation as being clockwise when looking into the beam. However, he uses a different definition for the Stokes vector for which only the fourth component is independent of coordinate system, so his rotation matrix is more complicated. Mishchenko defines a positive rotation as being clockwise when looking in the direction of propagation. This is equivalent to counterclockwise when looking into the beam as used here; thus his rotation matrix is the same as that in Eq. (1.16). van de Hulst (1980) also uses the same rotation convention as is used here. All of this is considered well known, so the details are often omitted in publications, with confusing apparent differences among papers being the price of brevity. Fortunately, the only real requirement for Stokes vectors, coordinate systems, and rotations is consistency in usage once a choice has been made.

With the simplification of Eq. (1.14), the water-column IOPs are fully specified by the attenuation coefficient $c(z, \lambda)$ and the 4×4 scattering matrix $\underline{M}(z, \psi, \lambda)$. The (1,1) element of $\underline{M}$ is the volume scattering function (VSF) β, which describes scattering of unpolarized radiance into unpolarized radiance. The units of β and $\underline{M}$ are $\mathrm{m}^{-1}\ \mathrm{sr}^{-1}$. (The term scattering matrix is commonly used when the scattering is caused by particles or molecules, as in the ocean. In laboratory optics the transformation from $\underline{S}'$ to $\underline{S}$ is often caused by lenses, filters, mirrors, etc., and the term Mueller matrix is commonly used. However, scattering matrix and Mueller matrix are synonymous in that they both transform Stokes vectors in the scattering plane.)

The integral of the VSF over all scattering directions gives the scattering coefficient $b(z, \lambda)$:

$$b(z,\lambda) = \int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi} \beta(z,\theta',\phi' \to \theta,\phi,\lambda))d\Omega(\theta,\phi) \quad \text{for any } (\theta',\phi')$$
$$= 2\pi \int_0^{\pi} \beta(z,\psi,\lambda)\sin\psi d\psi\,.$$

The second form holds for isotropic media, in which case scattering of unpolarized light depends only on the polar scattering angle ψ and not on azimuthal scattering angle. That is the situation assumed here. The scattering phase function is defined by $\tilde{\beta}(\psi) = \beta(\psi)/b$. In general, the phase function gives the probability per unit solid angle Ω centered on (θ,ϕ) that radiance traveling in any direction (μ',ϕ') will be scattered into any other direction (μ,ϕ). The phase function therefore satisfies the normalization condition

$$1 = \int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi} \tilde{\beta}(\theta',\phi' \to \theta,\phi))d\Omega(\theta,\phi) \quad \text{for any } (\theta',\phi')$$
$$= 2\pi \int_0^{\pi} \tilde{\beta}(\psi)\sin\psi d\psi\,. \tag{1.17}$$

That is, the probability is 1 that the radiance will be scattered into some direction. As above, the second equation holds for isotropic media.

The attenuation coefficient $c(z,\lambda)$ is the sum of the absorption and scattering coefficients:

$$c(z,\lambda) = a(z,\lambda) + b(z,\lambda)\,.$$

Finally, the nondimensional albedo of single scattering is the ratio of the scattering coefficient to the attenuation coefficient:

$$\omega_o(z,\lambda) = \frac{b(z,\lambda)}{c(z,\lambda)}\,.$$

In the oceanographic setting, depth z is expressed in meters, positive downward from the mean sea surface at $z = 0$. However, it is the non-dimensional optical depth ζ that governs light propagation. The optical depth is defined by

$$d\zeta(\lambda) \equiv c(z,\lambda)dz\,. \tag{1.18}$$

This equation can be integrated to convert geometric depth to optical depth, or vice versa:

$$\zeta(z,\lambda) = \int_0^z c(z',\lambda)dz' \quad \text{or} \quad z(\zeta,\lambda) = \int_0^{\zeta} \frac{d\zeta'}{c(\zeta',\lambda)}\,.$$

Note that the optical depth corresponding to a given physical depth z depends on wavelength via $c(z,\lambda)$.

Furthermore, the scattering coefficient $b(z,\lambda)$ can be factored out of the phase matrix to obtain the normalized phase matrix $\underline{\tilde{P}} \equiv \underline{P}/b(\zeta,\lambda)$. This is similar to

factoring the scattering coefficient $b(z, \lambda)$ out of the VSF to obtain the scattering phase function. Both the phase function $\tilde{\beta}$ and the normalized phase matrix $\underline{\tilde{P}}$ have units of sr^{-1}.

Dividing Eq. (1.5) by c and using Eq. (1.18) gives the VRTE in terms of optical depth. It is common to use $\mu = \cos\theta$ as the polar angle variable. An element of solid angle is then $d\Omega = \sin\theta\, d\theta\, d\phi = d\mu\, d\phi$. Writing the phase matrix as $\underline{P} = b\underline{\tilde{P}}$ and taking $b(z, \lambda)$ outside the integral over incident directions allows the IOPs to be written in terms of the albedo of single scattering $\omega_o = b/c$. These steps allow Eq. (1.5) to be rewritten as

$$\begin{aligned} \mu \frac{d\underline{S}(\zeta, \mu, \phi, \lambda)}{d\zeta} &= -\underline{S}(\zeta, \mu, \phi, \lambda) \\ &\quad + \omega_o(\zeta, \lambda) \int_0^{2\pi} \int_0^{\pi} \underline{\tilde{P}}(\zeta; \mu', \phi' \rightarrow \mu, \phi; \lambda)\, \underline{S}(\zeta, \mu', \phi', \lambda)' d\mu' d\phi' \\ &\quad + \underline{\Sigma}(\zeta, \mu, \phi, \lambda)\,. \end{aligned} \tag{1.19}$$

The $1/c$ factor multiplying the source function $\underline{\sigma}$ in Eq. (1.5) has for notational convenience been incorporated into the definition of the vector source function $\underline{\Sigma} = \underline{\sigma}/c$, which therefore has units of radiance. The IOPs are now defined via the albedo of single scattering $\omega_o(\zeta, \lambda)$ and the normalized phase matrix $\underline{\tilde{P}}(\zeta, \psi, \lambda)$; $c(\zeta, \lambda)$ is needed only if there are internal sources.

This matrix equation represents 4 scalar equations for the 4 components of the Stokes vector. The integral term connects these four components and shows how scattering transforms various states of polarization (components of the Stokes vector) for light traveling in all incident directions (μ', ϕ') into other states of polarization of light traveling in the (μ, ϕ) direction of interest. In particular, the (i, j) element of $\underline{\tilde{P}}$ shows how scattering converts incident light in polarization state j (the jth element of $\underline{S}'$) into polarization state i of the final Stokes vector $\underline{S}$.

Equation (1.19) is the form of the VRTE that will be solved by invariant imbedding. Indeed, the remainder of this chapter can be viewed as a pure mathematics problem: solve Eq. (1.19) subject to the boundary conditions specified in Sects. 1.2.5 and 1.2.6.

1.2.4 Phase Matrices

The terminology relating to phase matrices and related quantities is confusing, and there is no uniformity of notation in the literature. Further comment is therefore warranted. The phase matrix for randomly oriented, mirror-symmetric particles can be written is several ways:

$$\underline{P}(\zeta;\mu',\phi'\rightarrow\mu,\phi;\lambda)=\underline{P}(\zeta;\mu',\mu,\phi-\phi';\lambda)=\underline{P}(\zeta;\psi;\lambda) \tag{1.20}$$

$$=\underline{R}(\alpha)\,\underline{M}(\psi)\,\underline{R}(\alpha') \tag{1.21}$$

$$=\beta(\psi)\,\underline{R}(\alpha)\,\underline{\tilde{M}}(\psi)\,\underline{R}(\alpha') \tag{1.22}$$

$$=b(\zeta,\lambda)\,\tilde{\beta}(\psi)\,\underline{R}(\alpha)\,\underline{\tilde{M}}(\psi)\,\underline{R}(\alpha') \tag{1.23}$$

$$=b(\zeta,\lambda)\,\tilde{\beta}(\psi)\,\underline{\tilde{\tilde{P}}} \tag{1.24}$$

$$=b(\zeta,\lambda)\,\underline{\tilde{P}}\,. \tag{1.25}$$

The quantities seen in these equations are described as follows:

- The *phase matrix* $\underline{P}$ is a 4×4 matrix that transforms the incident or unscattered Stokes vector $\underline{S}'$ into the final or scattered vector $\underline{S}$, with both vectors expressed in meridian planes. The directional arguments seen in Eq. (1.20) highlight the dependence on ψ as seen in Eq. (1.4).
- The scattering matrix $\underline{M}$ is a 4×4 matrix that transforms $\underline{S}'$ into $\underline{S}$, with both vectors expressed in the scattering plane.
- $\underline{R}(\alpha')$ and $\underline{R}(\alpha)$ are the *rotation matrices* that carry Stokes vectors into and out of the scattering plane. The rotation matrices are non-dimensional; it is the scattering matrix that carries the units of $\mathrm{m}^{-1}\,\mathrm{sr}^{-1}$, and the phase matrix therefore has the same units.
- The (1,1) element of the scattering matrix, $M_{1,1}$, transforms the total incident radiance, the I' element of $\underline{S}'$, into total scattered radiance, the I element of $\underline{S}$. In different terminology, $M_{1,1}(\psi)$ is thus the volume scattering function $\beta(\psi)$ as used in scalar radiative transfer theory.
- The volume scattering function $\beta(\psi)=M_{1,1}(\psi)$ can be factored out of each element of $\underline{M}$ to obtain a dimensionless *reduced scattering matrix* $\underline{\tilde{M}}$ whose elements are $\tilde{M}_{i,j}(\psi)=M_{i,j}(\psi)/M_{1,1}(\psi)$.
- $\underline{\tilde{\tilde{P}}}\equiv\underline{R}(\alpha)\,\underline{\tilde{M}}(\psi)\,\underline{R}(\alpha')$ is called the *reduced phase matrix* because it is constructed from the reduced scattering matrix $\underline{\tilde{M}}$; it is likewise dimensionless.
- $\underline{\tilde{P}}$ as seen in in Eqs. (1.19) and (1.25) is called the *normalized phase matrix*. Note that the (1,1) element of the normalized phase matrix $\underline{\tilde{P}}$ is just the scattering phase function $\tilde{\beta}$. $\underline{\tilde{P}}$ therefore has units of sr^{-1}.

As has been noted, for a random collection of mirror-symmetric particles, the scattering matrix becomes block diagonal. The reduced scattering matrix then has the form (Mishchenko et al. 2002)

$$\underline{\tilde{M}}=\begin{bmatrix}1 & \tilde{M}_{12}(\psi) & 0 & 0)\\ \tilde{M}_{12}(\psi) & \tilde{M}_{22}(\psi) & 0 & 0\\ 0 & 0 & \tilde{M}_{33}(\psi) & \tilde{M}_{34}(\psi)\\ 0 & 0) & -\tilde{M}_{34}(\psi) & \tilde{M}_{44}(\psi)\end{bmatrix}. \tag{1.26}$$

The sixth independent element of the scattering matrix $\underline{M}$ is the volume scattering function $\underline{M}_{1,1}(\psi)$, which has been factored out in creating $\underline{\tilde{M}}$.

The normalized phase matrix $\underline{\tilde{P}}$ corresponding to a reduced scattering matrix of the form (1.26), as used in Eq. (1.19), can be written in full as

$$\begin{aligned}\underline{\tilde{P}} &= \tilde{\beta}\,\underline{R}(\alpha)\,\underline{\tilde{M}}(\psi)\,\underline{R}(\alpha') \\ &= \tilde{\beta}(\psi)\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & \cos 2\alpha & -\sin 2\alpha & 0\\ 0 & \sin 2\alpha & \cos 2\alpha & 0\\ 0 & 0 & 0 & 1\end{bmatrix} \\ &\times \begin{bmatrix}1 & \tilde{M}_{12}(\psi) & 0 & 0\\ \tilde{M}_{12}(\psi) & \tilde{M}_{22}(\psi) & 0 & 0\\ 0 & 0 & \tilde{M}_{33}(\psi) & \tilde{M}_{34}(\psi)\\ 0 & 0 & -\tilde{M}_{34}(\psi) & \tilde{M}_{44}(\psi)\end{bmatrix}\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & \cos 2\alpha' & -\sin 2\alpha' & 0\\ 0 & \sin 2\alpha' & \cos 2\alpha' & 0\\ 0 & 0 & 0 & 1\end{bmatrix}. \end{aligned} \tag{1.27}$$

Letting $c = \cos 2\alpha$, $s' = \sin 2\alpha'$, etc., and $m_{ij} = \tilde{M}_{i,j}(\psi)$, the normalized phase matrix becomes

$$\underline{\tilde{P}} = \tilde{\beta}(\psi)\begin{bmatrix}1 & c'\,m_{12} & -s'\,m_{12} & 0\\ c\,m_{12} & c'c\,m_{22} - s's\,m_{33} & -s'c\,m_{22} - c's\,m_{33}) & -s\,m_{34}\\ s\,m_{12} & c's\,m_{22} + s'c\,m_{33} & -s's\,m_{22} + c'c\,m_{33}) & c\,m_{34}\\ 0 & -s'\,m_{34} & -c'\,m_{34} & m_{44}\end{bmatrix}. \tag{1.28}$$

Note that although the scattering matrix elements $\tilde{M}_{ij}(\psi)$ and $\tilde{M}_{ji}(\psi)$ have the symmetries seen in Eq. (1.26), those symmetries are lost in the phase matrix because of the rotations.

1.2.4.1 Dependence of the Phase Matrix on Azimuthal Angle

The phase function $\tilde{\beta}(\psi)$ in the SRTE depends on the scattering angle ψ, computed via Eq. (1.4). This equation shows that the phase function depends on the incident and scattered azimuthal angles via cosine of the difference in the azimuthal angles. Because $\tilde{\beta}(\mu', \phi' \rightarrow \mu, \phi)$ depends only on the difference $\phi - \phi'$, ϕ' can be set to 0 without loss of generality; this choice merely anchors the difference $\phi - \phi'$ to $\phi' = 0$ for computational purposes. The dependence on $\cos(\phi - \phi')$ also means that the phase function can be expanded as a Fourier cosine series in the azimuthal angle ϕ.

Unfortunately this $\cos(\phi - \phi')$ dependence is lost in the phase matrix because of the rotation matrices (van de Hulst 1980). The phase matrix still depends on $\phi - \phi'$, but some phase matrix elements depend on $\cos(\phi - \phi')$ and some depend on $\sin(\phi - \phi')$. Figure 1.4 illustrates this for the simple case of the normalized Rayleigh phase matrix defined by Eq. (1.28) and the Rayleigh scattering matrix

$$\underline{M}_{\text{Ray}} = \beta_{\text{Ray}}\underline{\tilde{M}}_{\text{Ray}} = b_{\text{Ray}}\tilde{\beta}_{\text{Ray}}\underline{\tilde{M}}_{\text{Ray}}$$
$$= b_{\text{Ray}}\frac{3}{16\pi}(1+\cos^2\psi)\begin{bmatrix} 1 & -\frac{\sin^2\psi}{1+\cos^2\psi} & 0 & 0 \\ -\frac{\sin^2\psi}{1+\cos^2\psi} & 1 & 0 & 0 \\ 0 & 0 & \frac{2\cos\psi}{1+\cos^2\psi} & 0 \\ 0 & 0 & 0 & \frac{2\cos\psi}{1+\cos^2\psi} \end{bmatrix}. \quad (1.29)$$

(Note that the Rayleigh phase function $\tilde{\beta}_{\text{Ray}} = \frac{3}{16\pi}(1+\cos^2\psi)$ satisfies the normalization Eq. 1.17.)

The non-zero elements of the upper left and lower right 2×2 blocks of matrix elements in Fig. 1.4 are non-zero at $\phi - \phi' = 0$ and 180°. These 8 matrix elements can be expanded as Fourier cosine series, just as for the scalar phase function. However, the non-zero elements of the lower left and upper right 2×2 blocks of elements are 0 at $\phi - \phi' = 0$ and 180°. These matrix elements therefore must be expanded as Fourier sine series. This result holds true for any scattering matrix. This mixed cosine and sine and dependence of the phase matrix elements on azimuthal angle considerably complicates the solution of the VRTE compared to that of the SRTE, as will be seen in Sect. 1.5.

The phase function for the SRTE can be written as a function of either $\cos(\phi - \phi')$ or $\cos(\phi' - \phi)$. However, reversing the order of ϕ and ϕ' in the phase matrix of the VRTE introduces a sign change in the elements that depend on $\sin(\phi - \phi')$. For consistency, the development below always writes azimuthal angles in the order of scattered direction minus incident direction, i.e. $\phi - \phi'$ (or $\phi_v - \phi_s$ after discretization).

Hovenier (1969) derived several symmetry relations that must be obeyed by phase matrices. These symmetries result from considerations of time-reversal invariance of the scattering process and from the geometric symmetries of the scattering particles. These equations need not be repeated here, but they do provide important checks on on the correctness of the discretized phase matrices, which will be computed as described in Sect. 1.3.2.

1.2.5 Boundary Conditions at the Sea Surface

The VRTE in the previous section describes how light is absorbed and scattered within the water. It is also necessary to determine how light enters and leaves the water, and how it is reflected by the bottom if the water has a finite depth. Thus, in order to solve the VRTE, boundary conditions for the Stokes vector must be specified at the sea-surface and for an opaque but reflecting surface at the sea bottom.

Just as for "volume scattering" within the water, the "surface scattering" processes of reflection and refraction are described by 4×4 matrices. There are four of these *sea-surface radiance transfer functions*. $\underline{r}_{aw}(\theta', \phi' \to \theta, \phi)$ specifies how polarized radiance incident onto the sea surface in downwelling direction (θ', ϕ') is reflected

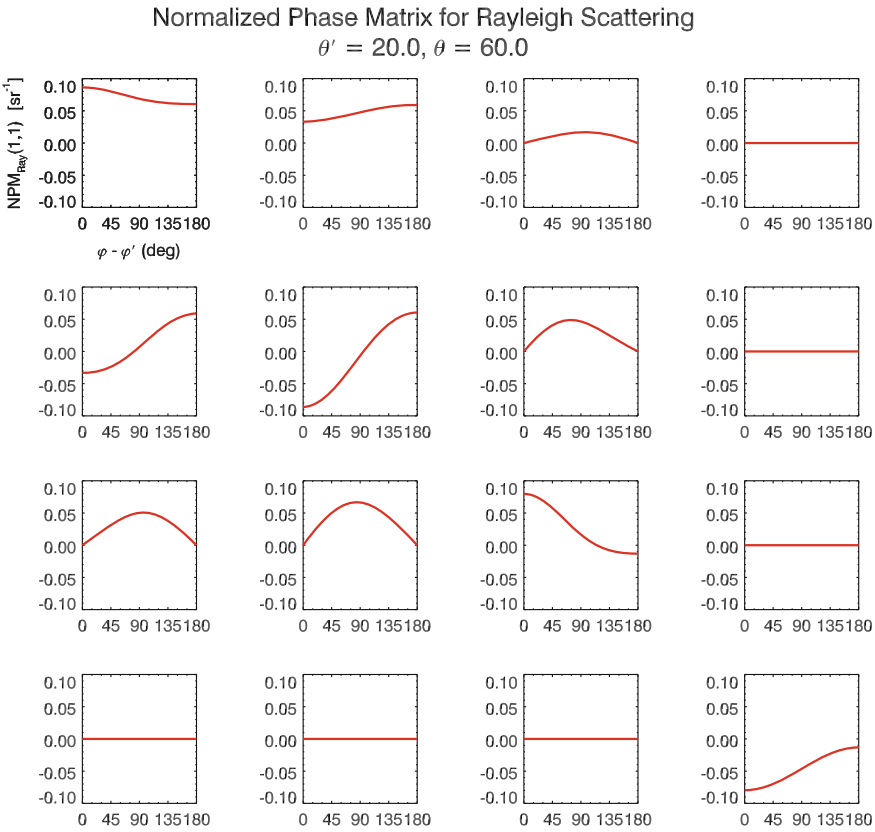

Fig. 1.4 The normalized phase matrix for Rayleigh scattering, $\underline{\tilde{P}}_{\text{Ray}}$, plotted as a function of $\phi - \phi'$ for fixed polar angles $\theta' = 20°$ and $\theta = 60°$. Note the block structure as regards expansion of elements in either cosine or sine series

back upward into direction (θ, ϕ) by the water surface, per unit of solid angle. Similarly, $\underline{t}_{aw}$ describes how air-incident light is transmitted through the surface into the water, $\underline{r}_{wa}$ reflects water-incident light back to the water, and $\underline{t}_{wa}$ transmits light from the water into the air. These transfer functions have units of sr^{-1}. The boundary conditions at the sea surface for the VRTE are given by

$$\begin{aligned}\underline{S}(a,\theta,\phi) = &\iint_{2\pi_d} \underline{r}_{aw}(\theta',\phi' \to \theta,\phi)\, \underline{S}(a,\theta',\phi')\, d\Omega(\theta',\phi') \\ &+ \iint_{2\pi_u} \underline{t}_{wa}(\theta',\phi' \to \theta,\phi)\, \underline{S}(w,\theta',\phi')\, d\Omega(\theta',\phi') \quad \text{for } (\theta,\phi) \in 2\pi_u\end{aligned} \tag{1.30}$$

and

$$\underline{S}(w,\theta,\phi) = \iint_{2\pi_u} \underline{r}_{wa}(\theta',\phi' \to \theta,\phi)\, \underline{S}(w,\theta',\phi')\, d\Omega(\theta',\phi') + \iint_{2\pi_d} \underline{t}_{aw}(\theta',\phi' \to \theta,\phi)\, \underline{S}(a,\theta',\phi')\, d\Omega(\theta',\phi') \quad \text{for } (\theta,\phi) \in 2\pi_d \tag{1.31}$$

The notation $\iint_{2\pi_d}$ ($\iint_{2\pi_u}$) denotes integration over all θ', ϕ' in the downward (upward) hemisphere of directions (2π steradians). Equation (1.30) thus shows that the upwelling radiance in the air (depth argument a) just above the sea surface comes from downwelling radiance in the air reflected back upward by the sea surface, and from upwelling radiance incident onto the sea surface from below (depth w) and transmitted through the surface. A similar interpretation holds for Eq. (1.31). When solving the VRTE, the sky radiance $\underline{S}(a,\theta',\phi')$ incident onto the sea surface is considered known. Equations (1.30) and (1.31) are the vector equivalents of the scalar equations seen in (L&W 4.3) and (L&W 4.4).

The four sea-surface transfer functions are conceptually like the phase matrix seen in the VRTE of Eq. (1.19). That is, they describe reflection or transmission (i.e., scattering) from incident direction (θ',ϕ') to final direction (θ,ϕ), with incident and final Stokes vectors expressed in their meridian planes. For a level sea surface, the incident and final meridian planes are the same. For a wind-blown sea surface, the reflected and transmitted light for a single incident ray will generally lie in a different meridian plane than the incident light. Indeed, for a time- or space-averaged collection of random sea surfaces, a given incident direction yields reflected and transmitted light in all directions.

1.2.6 Bottom Boundary Conditions

For complete generality, the bottom boundary condition can be formulated with a pair of equations like those at the sea surface. However, for problems of oceanographic interest, it is sufficient to consider a bottom that reflects light but does not allow light to enter the water from below the bottom. In that case, there is only radiance incident onto the bottom being reflected back upward into the water column. There is no radiance being transmitted upward through the bottom, nor are we interested in radiance being transmitted downward below the bottom. Thus the bottom boundary condition reduces to just

$$\underline{S}(m,\theta,\phi) = \iint_{2\pi_d} \underline{r}_{mb}(\theta',\phi' \to \theta,\phi)\, \underline{S}(m,\theta',\phi')\, d\Omega(\theta',\phi') \quad \text{for } (\theta,\phi) \in 2\pi_u\,,$$

where depth argument m indicates the depth where the bottom boundary condition is applied. Recall from Fig. 1.2 that this is the maximum depth to which the VRTE is to be solved. Depth m can be either a finite depth with a physical bottom, or a depth in the water column below which the water in infinitely deep, homogeneous, and source-free. In either case, $\underline{r}_{mb}$ is a 4×4 matrix that describes the reflectance properties of the bottom layer $[m, b]$. It does not matter whether $\underline{r}_{mb}$ describes a finite-depth, opaque, physical bottom (i.e., a layer $[m, b] = [m, m]$) or an infinitely deep water column (i.e., a layer $[m, b] = [m, \infty]$), so long as $\underline{r}_{mb}$ describes how downwelling light at depth m is reflected back upward by whatever is at depth m and below.

1.2.6.1 Finite-Depth Bottoms

In the case of an opaque bottom at depth $z = m = b$, the bottom boundary condition is commonly written in terms of the vector bidirectional reflectance distribution function (VBRDF):

$$\underline{S}(m, \theta, \phi) = \iint_{2\pi_d} \underline{VBRDF}(\theta', \phi' \rightarrow \theta, \phi) \cos\theta' \, \underline{S}(m, \theta', \phi') \, d\Omega(\theta', \phi') \, .$$

Each (i, j) element of the VBRDF has the functional form of a scalar BRDF that reflects light from incident direction (θ', ϕ') into direction (θ, ϕ), and from polarization state j to i. The (1,1) element of the VBRDF is the scalar BRDF for unpolarized radiance. In general, these matrix elements will depend on wavelength. The $\cos\theta'$ factor appears in this equation because the radiance reflectance function $\underline{r}_{mb}$ as used here is defined for surfaces normal to the directions of light propagation, whereas $\underline{VBRDF}$ is defined for light incident onto a horizontal surface. Just as for the sea surface, $\underline{r}_{mb}$ and $\underline{VBRDF}$ have units of sr^{-1} and include the effects of rotations between meridian planes and the reflection plane.

Although some models have been developed for the VBRDF of terrestrial vegetation and soils (e.g., Schott 1999), there are no such measurements or models for typical ocean bottom materials such as sand, mud, sea grass, or coral. Such materials are typically highly scattering rough surfaces that diffusely reflect light into all directions. For most oceanographic applications, it will be sufficient to consider VBRDFs that depend only on the difference in reflected and incident azimuthal angles, i.e.,

$$\underline{VBRDF}(\theta', \phi' \rightarrow \theta, \phi) = \underline{VBRDF}(\theta', \theta, \phi - \phi') \, .$$

Such VBRDFs should describe most bottom materials, and the algorithm developed here is constructed to handle any VBRDF of this form. However, such VBRDFs would not describe a sandy bottom with a linear ripple structure, which would reflect light differently for "along-ripple" vs "cross-ripple" directions. Such bottoms can be modeled in the same way as the sea surface reflectance, with corresponding modifications to the equations in Sect. 1.5.5.2 below.

Multiple scattering tends to depolarize light. Many bottom materials have rough surfaces that multiply scatter incident light in the process of eventually reflecting it away from the surface. Therefore, pending actual measurements, it is expedient, and perhaps even reasonable, to model sediments and aquatic vegetation as Lambertian, depolarizing surfaces. $\underline{r}_{mb}$ then has the simple form

$$\underline{r}_{mb} \equiv \underline{r}_{\text{Lamb}} = \frac{R(\lambda)\cos\theta'}{\pi} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} . \tag{1.32}$$

Here $R(\lambda) = E_u(\lambda)/E_d(\lambda)$ is the wavelength-dependent irradiance reflectance of the bottom for unpolarized light; π carries the units of sr^{-1}. Because only the (1,1) matrix element is nonzero, the rotation matrices, which do not affect the (1,1) matrix element, are not included in Eq. (1.32). Matrices of this form have been used by Zhai et al. (2008) in studies of underwater imaging.

1.2.6.2 Infinitely Deep Water

The vector reflectance properties of an infinitely deep layer of homogeneous, source-free water below depth $z = m$ are much more complicated than for an opaque bottom and have not yet been worked out. The development would presumably follow that for the scalar case seen in *Light and Water* Sect. 9.5. Meanwhile, infinitely deep bottoms can be simulated simply by placing an opaque bottom at finite depth m sufficiently deep that the bottom reflectance does not significantly affect the radiance at the greatest depth of interest. The only penalty for this is increased run time for solving the VRTE to a great depth.

1.3 Discretization of the VRTE

The VRTE and its boundary conditions represent a continuum of depths, directions, and wavelengths. In order to solve these equations on a computer, they must be discretized to obtain a finite number of output variables, namely Stokes vectors for a finite number of depths, directions, and wavelengths. There are various ways in which to do this. For example, a function of direction can be expanded as an infinite series and then truncated after a finite number of terms. The result would then be an approximate Stokes vector for an exact direction. The directional discretization used here (and in HydroLight for the SRTE), on the other hand, averages functions of direction over finite solid angles. The resulting Stokes vectors are exact values averaged over finite solid angles.

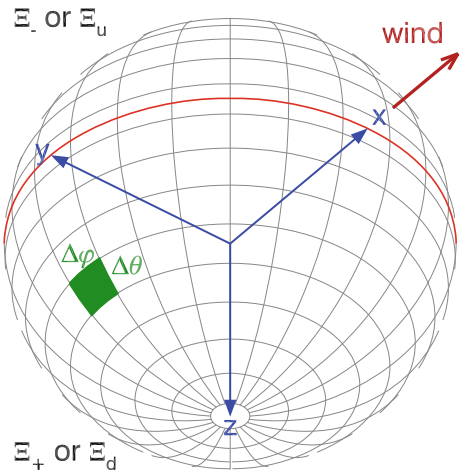

Fig. 1.5 The partition of the unit sphere of directions into a finite number of solid angles. The coordinate system is drawn with $+\mathbf{z}$ pointing downward, as in HydroPol. The $+\mathbf{x}$ direction is downwind. The perspective view of this figure can be thought of as being in the water looking obliquely upward

1.3.1 Discretization of Stokes Vectors

Figure 1.5 shows (θ, ϕ) directional bins that are the same as those used in the standard version of HydroLight. Let Ξ represent the set of all directions $\boldsymbol{\xi} = (\theta, \phi)$. In this figure, Ξ is divided into rectangular regions, called quads, of size 10° in polar angle by 15° in azimuthal angle, plus two polar caps of 5° half angle. The quad boundaries are analogous to lines of constant latitude and longitude on the Earth. This unit sphere of directions can be thought of as centered at each point of space in the **x**-**y**-**z** coordinate system, which is shown in blue. The red line at the "equator" divides the set of all directions into hemispheres of upward (Ξ_- or Ξ_u) and downward (Ξ_+ or Ξ_d) directions.

As will be seen in Sect. 1.5, invariant imbedding splits the VRTE into sets of "upward" and "downward" equations. Polar angle θ is normally measured from 0 in the $+\mathbf{z}$ direction to 90° at the equator. However, rather than continuing with the measurement of θ to π or 180° in the $-\mathbf{z}$ direction, it will be convenient to think of the upward and downward hemispheres as each having θ run from 0 at the pole to 90° at the equator. The θ quad boundaries as drawn thus run from 5° for the polar cap boundary, through 15, 25,…, 85, and 90°, for a total of $M = 10$ $\Delta\theta$ or $\Delta\mu$ bins in each hemisphere. These quads are numbered from $u = 1$ next to the equator to $u = M = 10$ for the polar caps. Because of the partitioning into upward and downward hemispheres, the equator must always be a quad boundary, hence the row of quads next to the equator has $\Delta\theta = 5°$. The two 5° $\Delta\theta$ quads next to the equator can, if desired, be combined during post-processing of the solution radiances to obtain a horizontal quad of size $\Delta\theta = 10°$.

As previously noted, azimuthal angle ϕ is measured from 0 in the $+\mathbf{x}$ direction, and the first ϕ quad is likewise centered on $+\mathbf{x}$. For reasons that will be seen in Sect. 1.6.1, the coordinate system is "wind centered," with $+\mathbf{x}$ being in the downwind direction.

Thus $\mathbf{x}$ is always downwind, $\mathbf{z}$ is downward into the water, and $\mathbf{y} = \mathbf{z} \times \mathbf{x}$ is the cross-wind direction. The sun can be placed at any azimuthal direction relative to the wind.

The invariant imbedding solution algorithm uses a Fourier decomposition in the azimuthal direction, which requires an even number of $\Delta\phi$ bins with each $\Delta\phi$ bin being the same size. The total number of azimuthal quads is therefore written as $2N$. For 15 deg $\Delta\phi$ quads, there are $2N = 24$ ϕ quads, numbered $v = 1$ to $v = 2N = 24$, with quad 1 being the one centered on $+\mathbf{x}$. In addition, selecting the number of quads to be a multiple of 4, as here, is convenient for calculation of radiances at azimuthal directions of 0, 90, 180, and 270° because those directions are then at the centers of quads.

The total number of quads and polar caps in Fig. 1.5 is thus $2(M-1)2N + 2 = 434$. The quad with θ band index u and ϕ band index v is denoted $Q(u, v)$ or Q_{uv}, or $Q(M)$ or Q_M for a polar cap. With these conventions, it can be seen that the shaded green quad in Fig. 1.5 corresponds to quad $Q(u, v) = Q(5, 7)$ in the downward hemisphere (recall that $+\mathbf{z}$ is downward as emphasized in Fig. 1.5, so the green quad as drawn is in the Ξ_d or Ξ_+ hemisphere). The (θ, ϕ) range of this quad is $\theta = 45$ to 55° and $\phi = 82.5$ to 97.5°. This quad would contain radiance traveling downward at a nominal angle (quad center) of 50° from the nadir direction and perpendicular to the wind direction. Polar cap quads have no ϕ dependence and must always be treated as a special case.

The solid angle of $Q(u, v)$ is $\Omega_{uv} = \Delta\mu_u \Delta\phi_v$, or $\Omega_M = 2\pi[1 - \cos(\theta_{(M-1)})] = 2\pi[1 - \mu_{(M-1)}]$ for a polar cap, where $\theta_{(M-1)}$ is the boundary of the polar cap.

To convert the variables in the VRTE from continuous functions of depth, direction, and wavelength, to a finite number of values for discrete depths, directions, and wavelengths, (1) integrate the VRTE over quads of finite solid angle Ω_{uv}; (2) integrate the VRTE over wavelength bands of finite bandwidth $\Delta\lambda_j$, $j = 1, ..., J$; and (3) save the solution at only a finite number of depths z_k, $k = 1, ..., K$. Thus the computed Stokes vectors are

$$\underline{S}(k, u, v, j) = \frac{1}{\Delta\lambda_j} \frac{1}{\Omega_{uv}} \int_{\Delta\lambda_j} \int_{\Delta\mu_u} \int_{\Delta\phi_v} \underline{S}(z_k, \mu, \phi, \lambda)\, d\mu\, d\phi\, d\lambda\,. \tag{1.33}$$

These integrations are applied to each of the four $[I, Q, U, V]^T$ Stokes components. The wavelength bands $\Delta\lambda_j$ do not need to be equal in size. The interpretation of $\underline{S}(k, u, v, j)$ is that it is the exact Stokes vector $\underline{S}(z, \mu, \phi, \lambda)$ averaged over direction within quad $Q(u, v)$ and over wavelength within band $\Delta\lambda_j$, at a particular depth z_k.

As will be seen, the depths where output is saved refer to exact depths in the water column, not to finite-depth bins over which variables are depth averaged. Indeed, the IOPs must be supplied as continuous functions of depth, and the solution algorithm solves the VRTE with arbitrarily fine depth resolution, which is determined by the accuracy parameters of the numerical differential equation solver. The solution is simply saved at the finite set of user-requested output depths z_k as the VRTE is solved as a continuous function of depth.

The chosen quad angular resolution is a balance between the conflicting needs of having sufficiently high angular resolution in the radiance distribution and keeping the run times acceptably small. Numerical studies with the scalar HydroLight code show that a finer angular resolution than that seen in Fig. 1.5 does not change computed quantities such as irradiances and reflectances by more than roughly one percent, whereas accuracy starts to degrade for a coarser resolution. However, the run time is proportional the number of quads squared because scattering must be computed from every quad into every other quad. Thus finer angular resolution comes at a high computational cost with almost no improvement in numerical accuracy for the quantities of interest to most users.

The wavelength bands are usually chosen to be comparable to those of oceanographic sensors, typically $\Delta\lambda = 5$ or 10 nm. Run time is roughly proportional to the number of wave bands. However, the optical depth is wavelength dependent, so wavelengths greater than 700 nm, where water absorption becomes very large, can take much longer than shorter wavelengths.

In HydroLight, neither the number of depths K where output is requested nor their spacing significantly affects run time. This is likely to remain true in the polarized case. Requesting additional output depths does not change the values at the computed depths, it only increase the size of the output files. The size of the output files is directly proportional to the number of output depths, quads, and wavelength bands.

1.3.2 Quad Averaging

For a function of direction $F(\theta, \phi) = F(\mu, \phi)$, the average over all directions in quad $Q(u, v)$, termed the quad average $F(u, v)$, is defined as (L&W 4.52)

$$\begin{aligned} F(u, v) &= \frac{1}{\Omega_{uv}} \int_{\Delta\theta_u} \int_{\Delta\phi_v} F(\theta, \phi)\, \sin\theta\, d\theta\, d\phi \\ &= \frac{1}{\Omega_{uv}} \int_{\Delta\mu_u} \int_{\Delta\phi_v} F(\mu, \phi)\, d\mu\, d\phi\,. \end{aligned} \tag{1.34}$$

Note that $F(u, v)$ has the same dimensions as $F(\theta, \phi)$, e.g., units of radiance if F is a Stokes vector.

Scattering involves bi-directional functions of incident (μ', ϕ') and scattered (μ, ϕ) directions. The corresponding quad average of a bi-directional function $F(\mu', \phi' \rightarrow \mu, \phi)$ is (L&W 4.63)

$$F(r, s \rightarrow u, v) = \frac{1}{\Omega_{uv}} \int_{\Delta\mu_u} \int_{\Delta\phi_v} \left[\int_{\Delta\mu_r} \int_{\Delta\phi_s} F(\mu', \phi' \rightarrow \mu, \phi)\, d\mu'\, d\phi' \right] d\mu\, d\phi\,. \tag{1.35}$$

The bi-directional functions all describe scattering (either in the water column or at a boundary surface) and have units of sr^{-1}. Thus a phase matrix element

$P_{i,j}(\psi) = P_{i,j}(\mu', \phi' \to \mu, \phi)$ tells how much radiance is scattered from direction (μ', ϕ') to direction (μ, ϕ) and from polarization state j to *i per unit solid angle* in the scattered direction. The corresponding quad-averaged function $P_{i,j}(r, s \to u, v)$ is however non-dimensional. This is because $P_{i,j}(r, s \to u, v)$ tells how much radiance is scattered in total (not per unit solid angle) from quad $Q(r, s)$ into quad $Q(u, v)$.

As seen in Eq. (1.19), directional discretization of the VRTE involves quad average integrals of the form

$$\int_0^{2\pi} \int_0^{\pi} \underline{\tilde{P}}(\mu', \phi' \to \mu, \phi)\, \underline{S}(\mu', \phi') d\mu' d\phi' .$$

Quad averaging is applied to each matrix element or product of elements, e.g. $P_{i,j} S_j$, in such quantities. Dropping the matrix and element notation for brevity, the quad average of such an integral becomes

$$\begin{aligned}
&\frac{1}{\Omega_{uv}} \int_{\Delta\mu_u} \int_{\Delta\phi_v} \left\{ \int_0^{2\pi} \int_0^{\pi} P(\mu', \phi' \to \mu, \phi) S(\mu', \phi')\, d\mu'\, d\phi' \right\} d\mu\, d\phi \\
&= \frac{1}{\Omega_{uv}} \int_{\Delta\mu_u} \int_{\Delta\phi_v} \left\{ \sum_r \sum_s \int_{\Delta\mu_r} \int_{\Delta\phi_s} P(\mu', \phi' \to \mu, \phi) S(\mu', \phi')\, d\mu'\, d\phi' \right\} d\mu\, d\phi \\
&= \sum_r \sum_s \left\{ \frac{1}{\Omega_{uv}} \int_{\Delta\mu_u} \int_{\Delta\phi_v} \left[\int_{\Delta\mu_r} \int_{\Delta\phi_s} P(\mu', \phi' \to \mu, \phi) S(\mu', \phi')\, d\mu'\, d\phi' \right] d\mu\, d\phi \right\} \\
&= \sum_r \sum_s P(r, s \to u, v) S(r, s) .
\end{aligned}$$

In the second equation, the integral over all incident directions has been written as a sum of integrals over all quads $Q(r, s)$. The next equation interchanges the order of integration over the scattered-direction quad $Q(u, v)$ and the summation over incident direction quads. The final equation uses the definition of Eq. (1.35). Note that the quad average of the integrals reduces to sums over the quad-averaged integrand. The solid angle factors resulting from the differentials $d\mu\, d\phi$ are built into the quad-averaged terms; there are thus no explicit $\Delta\mu_u \Delta\phi_v$ factors in the final expression. For a given phase matrix, the integrations over the quads must be done numerically. However, those are one-time calculations for a given scattering matrix and quad partition. A library of discretized phase matrices therefore can be pre-computed and stored for use in solution of the VRTE.

It should be noted that the discretization of the phase matrix integral via quad averaging did not make any assumptions about the phase matrix other than isotropy of the scattering medium. In particular, it was not necessary to expand the phase matrix as an infinite series of Legendre polynomials, and then truncate that series to a finite number of terms, as is required by the discrete ordinates solution method (e.g., L&W Sect. 9.1). Such expansions can cause numerical difficulties for highly peaked phase functions (e.g., L&W Fig. 9.1), which are typical of ocean waters. One of the most important conceptual and numerical virtues of quad averaging is that it can handle arbitrary phase functions.

Discretization of the surface boundary conditions (1.30) and (1.31) follows in a similar fashion to that for the path integral in the VRTE. Libraries of discretized surface transfer functions likewise can be pre-computed for various wind speeds and surface indices of refraction.

1.3.2.1 Summary of the Discretized Mathematical Problem

The end result of the quad-averaging process gives the discretized VRTE:

$$\begin{aligned}\mu_u \frac{d\underline{S}(\zeta,u,v,j)}{d\zeta} &= -\underline{S}(\zeta,u,v,j) \\ &\quad + \underbrace{\omega_o(\zeta,j)}\sum_r\sum_s \underbrace{\tilde{P}(\zeta;r,s\to u,v;j)}\,\underline{S}(\zeta,r,s,j) \\ &\quad + \underbrace{\underline{\Sigma}(\zeta,u,v,j)}\ . \end{aligned} \tag{1.36}$$

Symbolically, $u = 1, ..., 2M$ and $v = 1, ..., 2N$ to cover all directions. The polar cap special cases will be considered below. The j index represents the jth wavelength band.

The mathematical problem can be stated as follows: Solve the time-independent, one-dimensional discretized VRTE (1.36) between the sea surface at depth $\zeta = 0$ and the bottom at depth $\zeta = m$, subject to the boundary conditions at the sea surface

$$\begin{aligned}\underline{S}(a,u,v,j) =& \sum_{\substack{r\\ r,s\in\Xi_d}}\sum_s \underbrace{\underline{r}_{aw}(r,s\to u,v;j)}\,\underbrace{\underline{S}(a,r,s,j)} \\ &+\sum_{\substack{r\\ r,s\in\Xi_u}}\sum_s \underbrace{\underline{t}_{wa}(r,s\to u,v;j)}\,\underline{S}(w,r,s,j) \quad \text{for (u, v)} \in \Xi_u \end{aligned} \tag{1.37}$$

and

$$\begin{aligned}\underline{S}(w,u,v,j) =& \sum_{\substack{r\\ r,s\in\Xi_u}}\sum_s \underbrace{\underline{r}_{wa}(r,s\to u,v;j)}\,\underline{S}(w,r,s,j) \\ &+\sum_{\substack{r\\ r,s\in\Xi_d}}\sum_s \underbrace{\underline{t}_{aw}(r,s\to u,v;j)}\,\underbrace{\underline{S}(a,r,s,j)} \quad \text{for (u, v)} \in \Xi_d\ , \end{aligned} \tag{1.38}$$

and at the bottom at depth $\zeta = m$

$$\underline{S}(m,u,v,j) = \sum_{\substack{r\\ r,s\in\Xi_d}}\sum_s \underbrace{\underline{r}_{mb}(r,s\to u,v;j)}\,\underline{S}(m,r,s,j) \quad \text{for (u, v)} \in \Xi_u \tag{1.39}$$

.

The quantities with a wavy underline are assumed known. These are the the water column IOPs, the internal source (if any), the sky radiance incident onto the air-side of the sea surface, the air-water surface reflectance and transmittance properties, and the bottom reflectance properties, all in discretized form. After solution of these equations, the Stokes vector $\underset{\sim}{S}(\zeta, u, v, j)$ is known within (depths $0 \leq \zeta \leq m$) and leaving (depth label $\zeta = a$) the water.

Finally, note that polar caps are always a special case because they have no ϕ dependence. Thus the sums represented symbolically by $\sum_r \sum_s$ must explicitly account for the polar caps $r = M$ and s undefined. In computer code, the polar cap values can be stored in array location $(r, s) = (M, 1)$, with the array elements $(r = M, s = 2, ..., 2N)$ being unused. Thus the path radiance term for scattering into $Q(u, v)$ can be evaluated in computer code as

$$\sum_{r} \sum_{s} \underset{\sim}{\tilde{P}}(\zeta; r, s \rightarrow u, v; j)\, \underset{\sim}{S}(\zeta, r, s, j) =$$
$$\sum_{r=1}^{M-1} \sum_{s=1}^{2N} \underset{\sim}{\tilde{P}}(\zeta; r, s \rightarrow u, v; j)\, \underset{\sim}{S}(\zeta, r, s, j) + \underset{\sim}{\tilde{P}}(\zeta; M, 1 \rightarrow u, v; j)\, \underset{\sim}{S}(\zeta, M, 1, j)\,. \tag{1.40}$$

1.4 The Air-Water Surface

The air-water surface boundary conditions seen in Eqs. (1.37) and (1.38) are stated in terms of four surface transfer functions, $\underset{\sim}{r}_{aw}$, $\underset{\sim}{t}_{aw}$, $\underset{\sim}{r}_{wa}$, and $\underset{\sim}{t}_{wa}$, which are assumed known. These functions are the inherent optical properties of the sea surface. They depend on the surface wave state and water index of refraction, but not on the light incident onto the surface from above or below. This section reviews the equations needed to compute these air-water surface transfer functions.

1.4.1 Reflection and Transmission by a Level Sea Surface

Consider a level or flat air-water surface. This simplest of environmental conditions is important for three reasons. First, a wind-blown sea surface can be modeled as a collection of randomly tilted, but locally flat, wave facets. Each tilted facet reflects and transmits light according to the laws for a flat surface. Thus a full understanding of how a flat surface reflects and transmits light is the foundation of modeling wind-blown surfaces. Second, the equations for a flat surface are analytically tractable and provide an important check on the numerical computations in the limit of zero wind speed. Third, level sea surfaces occasionally do occur in nature in the absence of wind-generated waves or swell.

Light can be incident onto this surface from the air, in which case part is reflected back to the air by the surface and part is transmitted through the surface into the water. Light can also be incident onto the underside of the sea surface, in which case part (or all) is reflected back to the water and part can be transmitted through the surface into the air. For either air- or water-incident light, $\underline{S}_i$ denotes the Stokes vector of the incident light, $\underline{S}_r$ is the reflected light, and $\underline{S}_t$ is the transmitted light. Angles θ_i, θ_r, and θ_t are the incident, reflected, and transmitted directions of the light propagation measured relative to the normal to the surface. $\underline{S}_i$, $\underline{S}_r$, and $\underline{S}_t$ all lie in the same plane.

Just as for scattering within the water, the surface scattering processes of reflection and refraction are described by 4×4 scattering matrices. There are four of these matrices: $\underline{R}_{aw}$ describes how air-incident light is reflected by the water surface back to the air, $\underline{T}_{aw}$ describes how air-incident light is transmitted through the surface into the water, $\underline{R}_{wa}$ reflects water-incident light back to the water, and $\underline{T}_{wa}$ transmits light from the water into the air. However, because $\underline{S}_i$, $\underline{S}_r$, and $\underline{S}_t$ are coplanar, scattering by the level surface does not involve rotation matrices as does scattering within the water body. (That is to say, the incident and final meridian planes and the scattering plane are all the same, the rotation angles α' and α are thus both 0, and the rotation matrices reduce to identity matrices.)

The reflection and (especially) transmission of polarized light by a dielectric surface such as a level water surface are rather complicated processes, and the literature contains a number of different (and, indeed, sometimes incorrect) mathematical formulations of the equations. The formulas given in Garcia (2012) are appropriate to the present needs. Note, however, that although the equations in Garcia (2012) are correct, some of his derivations and interpretations are incorrect, as explained by Zhai et al. (2012). Both papers must be used to understand the equations now presented. The equations in Garcia (2012) will be referenced by (G21) and so on; the corresponding equations in Zhai et al. (2012) will be referenced as (Z5), etc.

The reflectance and transmittance matrices have a general formulation for the interface between any two dielectric media a and b. Let n_a be the index of refraction of medium a and n_b be that of medium b. In general n_a and n_b are complex numbers, but for the air-water surface real indices of refraction $n_{\text{air}} = 1$ and $n_{\text{water}} \approx 1.34$ can be used. For reflection, the reflected angle θ_r equals the incident angle θ_i. For transmission from a to b, the transmitted angle is given by Snel's law, $n_a \sin\theta_a = n_b \sin\theta_b$, or

$$\theta_b = \sin^{-1}\left(\frac{n_a \sin\theta_i}{n_b}\right).$$

For water-incident light, $n_a = n_{\text{water}}$ and $n_b = n_{\text{air}}$, in which case the transmitted angle becomes undefined beyond the critical angle for total internal reflection, which for water is $\theta_{\text{tir}} = \sin^{-1}(1/n_{\text{water}}) \approx 48°$. For water-incident angles greater than θ_{tir} the incident light is totally reflected back to the water and no light is transmitted to the air.

Let $\underline{R}_{ab}$ denote the reflectance matrix for radiant *energy* (or power) incident from medium a and reflected back by medium b. $\underline{R}_{ab}$ thus represents either $\underline{R}_{aw}$ or $\underline{R}_{wa}$.

Likewise, let $\underline{T}_{ab}$ denote the transmittance matrix for light incident from medium a and transmitted through the surface into medium b. $\underline{T}_{ab}$ thus represents either $\underline{T}_{aw}$ or $\underline{T}_{wa}$.

With these preliminaries, the reflectance matrix $\underline{R}_{ab}$ is (G10)

$$\underline{R}_{ab} = \begin{bmatrix} \frac{1}{2}(R_\parallel R_\parallel^* + R_\perp R_\perp^*) & \frac{1}{2}(R_\parallel R_\parallel^* - R_\perp R_\perp^*) & 0 & 0 \\ \frac{1}{2}(R_\parallel R_\parallel^* - R_\perp R_\perp^*) & \frac{1}{2}(R_\parallel R_\parallel^* + R_\perp R_\perp^*) & 0 & 0 \\ 0 & 0 & \mathrm{Re}\{R_\parallel R_\perp^*\} & \mathrm{Im}\{R_\parallel R_\perp^*\} \\ 0 & 0 & -\mathrm{Im}\{R_\parallel R_\perp^*\} & \mathrm{Re}\{R_\parallel R_\perp^*\} \end{bmatrix}. \tag{1.41}$$

Here $\mathrm{Re}\{R_\parallel R_\perp^*\}$ denotes the real part of $R_\parallel R_\perp^*$ and $\mathrm{Im}\{R_\parallel R_\perp^*\}$ is the imaginary part.

The transmission matrix $\underline{T}_{ab}$ is (G11 or Z3)

$$\underline{T}_{ab} = f_T \begin{bmatrix} \frac{1}{2}(T_\parallel T_\parallel^* + T_\perp T_\perp^*) & \frac{1}{2}(T_\parallel T_\parallel^* - T_\perp T_\perp^*) & 0 & 0 \\ \frac{1}{2}(T_\parallel T_\parallel^* - T_\perp T_\perp^*) & \frac{1}{2}(T_\parallel T_\parallel^* + T_\perp T_\perp^*) & 0 & 0 \\ 0 & 0 & \mathrm{Re}\{T_\parallel T_\perp^*\} & \mathrm{Im}\{T_\parallel T_\perp^*\} \\ 0 & 0 & -\mathrm{Im}\{T_\parallel T_\perp^*\} & \mathrm{Re}\{T_\parallel T_\perp^*\} \end{bmatrix}. \tag{1.42}$$

The components of these equations are given by (G7):

$$R_\parallel = \frac{n_b \cos\theta_a - n_a \cos\theta_b}{n_b \cos\theta_a + n_a \cos\theta_b} \tag{1.43}$$

$$R_\perp = \frac{n_a \cos\theta_a - n_b \cos\theta_b}{n_a \cos\theta_a + n_b \cos\theta_b} \tag{1.44}$$

$$T_\parallel = \frac{2n_a \cos\theta_a}{n_b \cos\theta_a + n_a \cos\theta_b} \tag{1.45}$$

$$T_\perp = \frac{2n_a \cos\theta_a}{n_a \cos\theta_a + n_b \cos\theta_b}. \tag{1.46}$$

In general, the indices of refraction are complex numbers and these equations must be used. However, for real indices of refraction, as assumed here, the matrix elements can be simplified at the expense of having a special case for water-incident angles greater that the critical angle.

Define

$$n_{ab} = \frac{n_a}{n_b} \quad \text{and} \quad n_{ba} = \frac{n_b}{n_a}.$$

Then for the case of air-incident light, i.e., $n_a \le n_b$, or water-incident light with the incident angle less than the critical angle, i.e., $n_a > n_b$ and $\theta_a < \theta_{\mathrm{tir}}$, the equations yield the real forms (G14 and G15)

$$
\begin{aligned}
R_{\parallel} R_{\parallel}^{*} &= \left(\frac{\cos\theta_a - n_{ab}\cos\theta_b}{\cos\theta_a + n_{ab}\cos\theta_b} \right)^2 \\
R_{\perp} R_{\perp}^{*} &= \left(\frac{n_{ab}\cos\theta_a - \cos\theta_b}{n_{ab}\cos\theta_a + \cos\theta_b} \right)^2 \\
\mathrm{Re}\{R_{\parallel} R_{\perp}^{*}\} &= \left(\frac{\cos\theta_a - n_{ab}\cos\theta_b}{\cos\theta_a + n_{ab}\cos\theta_b} \right) \left(\frac{n_{ab}\cos\theta_a - \cos\theta_b}{n_{ab}\cos\theta_a + \cos\theta_b} \right) \\
\mathrm{Im}\{R_{\parallel} R_{\perp}^{*}\} &= 0 \\
T_{\parallel} T_{\parallel}^{*} &= \left(\frac{2 n_{ab}\cos\theta_a}{\cos\theta_a + n_{ab}\cos\theta_b} \right)^2 \\
T_{\perp} T_{\perp}^{*} &= \left(\frac{2 n_{ab}\cos\theta_a}{n_{ab}\cos\theta_a + \cos\theta_b} \right)^2 \\
\mathrm{Re}\{T_{\parallel} T_{\perp}^{*}\} &= \frac{4 n_{ab}^2 \cos^2\theta_a}{(\cos\theta_a + n_{ab}\cos\theta_b)(n_{ab}\cos\theta_a + \cos\theta_b)} \\
\mathrm{Im}\{T_{\parallel} T_{\perp}^{*}\} &= 0 \,.
\end{aligned}
$$

It should be noted that for the case of normal incidence, $\theta_i = 0$, both $R_{\parallel} R_{\parallel}^{*}$ and $R_{\perp} R_{\perp}^{*}$ reduce to

$$
R_{\parallel} R_{\parallel}^{*} = R_{\perp} R_{\perp}^{*} = \left(\frac{n_b - n_a}{n_b + n_a} \right)^2 \,.
$$

This gives a reflectance of $R_{ab}(\theta_i = 0) = 0.0209$ for $n_{\mathrm{water}} = 1.338$, for both air- and water-incident light.

For the case of total internal reflection, i.e., $n_a > n_b$ and $\theta_a \geq \theta_{\mathrm{tir}}$, the following equations are to be used (G22):

$$
\begin{aligned}
R_{\parallel} R_{\parallel}^{*} &= 1 \\
R_{\perp} R_{\perp}^{*} &= 1 \\
\mathrm{Re}\{R_{\parallel} R_{\perp}^{*}\} &= \frac{2\sin^4\theta_a}{1 - (1 + n_{ba}^2)\cos^2\theta_a} - 1 \\
\mathrm{Im}\{R_{\parallel} R_{\perp}^{*}\} &= -\frac{2\cos\theta_a \sin^2\theta_a \sqrt{\sin^2\theta_a - n_{ba}^2}}{1 - (1 + n_{ba}^2)\cos^2\theta_a}
\end{aligned}
$$

and all elements of the transmission matrix elements are 0:

$$
\underline{T}_{ab} = \underline{T}_{wa} = \underline{0}_{4\times 4} \,,
$$

where $\underline{0}_{4\times 4}$ is the 4×4 matrix of zeros.

Finally, the all-important transmission factor f_T in Eq. (1.42) requires discussion. The f_T factor is derived from conservation of energy across the surface. However, that conservation law can be expressed in terms of either radiance or irradiance. For

coherent Stokes vectors with units of irradiance, f_T is given by Zhai et al. (2012, Eq. 9), or Hecht (1989, Eq. 4.59)

$$f_T = \frac{n_b \cos\theta_b}{n_a \cos\theta_a} \,. \tag{1.47}$$

The cosine factors result from the different cross-sectional areas of the incident and transmitted beams due to the change in direction given by Snel's law. The index-of-refraction factors result from the different rates at which energy is transported toward or away from the surface in the two media, i.e., from the differences in the speeds of light in the two media. For diffuse Stokes vectors with units of radiance, as is the case in the present discussion, f_T is given by (Zhai et al. 2012, Eq. 5)

$$f_T = n_{ba}^2 \left(\frac{n_b \cos\theta_b}{n_a \cos\theta_a} \right) \,. \tag{1.48}$$

The extra factor of n_{ba}^2 in the radiance version results from the change in solid angle when crossing the surface, i.e., for the n^2 law for radiance. The f_T value given in Eq. (1.47) is the form to be used for Monte Carlo ray tracing of individual idealized photons treated as point particles that carry energy but do not have an associated solid angle. In that case, the n^2 law for radiance is built in to the radiance reflectance functions particle by particle as their directions change when crossing the interface according to Snel's law.

The preceding equations give everything needed to compute the reflection and transmission of light by a sea surface, either by analytical formulas in the case of a level surface, or by Monte Carlo simulation in the case of a wind-blown sea surface. The non-zero matrix elements of course depend on incident angle as seen above, but also depend weakly on the wavelength via the wavelength dependence of n_{water}.

As is usually the case, the literature contains a variety of notations, terminology, and formulations. The $R_\parallel$, $R_\perp$, $T_\parallel$, $T_\perp$ terms seen above specify how electric field *amplitudes* are reflected and transmitted for electric fields parallel and perpendicular to the dielectric interface. Snel's law can be used to obtain forms equivalent to Eqs. (1.43)–(1.46) in which the indices of refraction do not appear explicitly (e.g., Born and Wolf 1975, Eqs. 20a, 21a):

$$\begin{aligned}
R_\parallel &= \frac{\tan(\theta_a - \theta_b)}{\tan(\theta_a + \theta_b)} \\
R_\perp &= -\frac{\sin(\theta_a - \theta_b)}{\sin(\theta_a + \theta_b)} \\
T_\parallel &= \frac{2\sin\theta_b \cos\theta_a}{\sin(\theta_a + \theta_b)\cos(\theta_a - \theta_b)} \\
T_\perp &= \frac{2\sin\theta_b \cos\theta_a}{\sin(\theta_a + \theta_b)} \,.
\end{aligned}$$

The fractions of reflected and transmitted *energy* are obtained from the square of the electric field amplitudes. Hence these terms appear as products in Eqs. (1.41) and (1.42). The (1,1) element of $\underline{R}_{ab}$ gives the irradiance (energy) reflectance for air-incident unpolarized light. Using the alternate forms just above gives the Fresnel reflectance for unpolarized irradiance:

$$R_F = \frac{1}{2}(R_\parallel R_\parallel^* + R_\perp R_\perp^*) = \frac{1}{2}\left\{\left[\frac{\sin(\theta_a - \theta_b)}{\sin(\theta_a + \theta_b)}\right]^2 + \left[\frac{\tan(\theta_a - \theta_b)}{\tan(\theta_a + \theta_b)}\right]^2\right\} . \quad (1.49)$$

The corresponding transmitted irradiance for unpolarized light is

$$\begin{aligned} T_F &= f_T \frac{1}{2}(T_\parallel T_\parallel^* + T_\perp T_\perp^*) \\ &= \frac{n_b \cos\theta_b}{n_a \cos\theta_a}\frac{1}{2}\left\{\left[\frac{2\sin\theta_b \cos\theta_a}{\sin(\theta_a + \theta_b)\cos(\theta_a - \theta_b)}\right]^2 + \left[\frac{2\sin\theta_b \cos\theta_a}{\sin(\theta_a + \theta_b)}\right]^2\right\} = 1 - R_F , \end{aligned} \quad (1.50)$$

where the f_T of Eq. (1.47) is used for energy transfer.

The four $\underline{R}_{ab}$ and $\underline{T}_{ab}$ functions are plotted and discussed in further detail in Mobley (2014a). It suffices here to make only a few further comments. The reflectance curve for $R_{aw}(1, 1)$ is the same as seen in any elementary physics text (e.g. Eq. 1.49): it starts at 0.0209 for normal incidence (for $n_{\text{water}} = 1.338$) and rises to 1 at grazing incidence. When dealing with coherent Stokes vectors with units of irradiance, the f_T factor of Eq. (1.47) must be used. The transmittance for normal incidence then is then

$$T_{aw}(1, 1) = (4n_b)/(1 + n_b)^2 = 0.9791 ,$$

which with the reflectance sums to one (and also sums to one for all other incident angles).

When working with diffuse Stokes vectors, the the f_T factor of Eq. (1.48) must be used. The transmission value for $T_{aw}(1, 1)$ for normal incidence is then

$$T_{aw}(1, 1) = \frac{4n_b^3}{(1 + n_b)^2} = 1.7528 .$$

This value may look incorrect because it is greater than one. However, this value is indeed correct and because it describes transmission for a diffuse Stokes vector with units of radiance. When radiance travels from air to water, the solid angle in air is reduced by a factor of $1/n_{\text{water}}^2$ (a consequence of Snel's law), which increases the radiance by a factor of n_{water}^2 compared to the radiance for the same amount of power propagating in air. This is termed the n-squared law for radiance. When going from water to air, $\underline{R}_{wa}$ and $\underline{T}_{wa}$ are used. For angles less than the critical angle, the transmission is never more than about 0.54. This again shows the n-squared law for radiance. In going from water to air, the in-water radiance is decreased by a factor

of $1/n^2_{\text{water}}$ when crossing the surface because the solid angle in air is greater than that in water by a factor of n^2_{water}. The transmission function $\underline{T}_{wa}$ is zero beyond the critical angle because there is no transmission and total reflection.

Brewster's angle is given by $\theta_B = \tan^{-1}(n_b)$, which is 53.23° for $n_b = 1.338$. At this angle, $R_{aw}(1,2) = R_{aw}(2,1) = -R_{aw}(1,1)$, and $R_{aw}(3,3) = R_{aw}(4,4) = 0$. With the choice of "parallel" and "perpendicular" referring to the meridian plane, this R_{aw} has the form of a horizontal polarizer. For $n_b = 1.338$, $R_{aw}(1,1) \approx 0.04$ at θ_B, and the reflection process $\underline{S}_r = \underline{R}_{aw}(\theta_i = \theta_B)\underline{S}_i$ for unpolarized incident radiance ($\underline{S}_i = [1,0,0,0]^T$) gives $\underline{S}_r = [0.04I, -0.04I, 0, 0]^T$, which represents light that is totally linearly polarized in a direction parallel to the mean sea surface (i.e., horizontal polarization).

It should also be noted that the non-zero $T_{aw}(2,1)$ means that unpolarized radiance becomes partly linearly polarized upon transmission through the surface. Thus transmission by the air-water surface induces a partially polarized underwater light field, even if the sea surface is illuminated by unpolarized light.

The $\underline{R}_{aw}$ etc. matrices are non-dimensional reflectances and transmittances. The continuous-variable $\underline{r}_{aw}$ etc. seen in Eqs. 1.30 and 1.31 are radiance reflectances and transmittances with units of sr^{-1}. For a level surface, Dirac delta functions can be added to the $\underline{R}_{aw}$ etc. to indicate the the exact directions of the reflected and transmitted light, which also gives sr^{-1} dimensions to the result. Thus $\underline{r}_{aw} = \underline{R}_{aw}\,\delta(\boldsymbol{\xi} - \boldsymbol{\xi}_{r_{aw}})$, where $\boldsymbol{\xi}_{r_{aw}}$ is the direction of the air-incident light as reflected upward by the surface. Similar equations hold for the other three transfer functions. The quad averaged $\underline{r}_{aw}$ etc. are computed by numerical integration using these analytic function in integrals of the form (1.35).

1.4.2 *Reflection and Transmission by a Wind-Blown Sea Surface*

Wind-blown sea surfaces are vastly more complicated than level surfaces. The properties of the sea surface that determine its optical properties are the wave height and wave slope. The slope is of primary importance because the slope, along with the direction of an incident light ray, determines the angles of incidence and transmission seen in the Fresnel formulas. The wave height and slope in turn depend not just on the wind speed, but also on how long the wind has been blowing (the duration) and over what distance upwind from the point of interest (the fetch), on the direction of wave propagation relative to any current, on the possible presence of surface contaminants, on the depth in shallow water, and to a lesser extent even on air-sea temperature difference.

Techniques for generation of random sea surfaces and for polarized ray tracing with such surfaces can be found elsewhere, e.g. in Mobley (2014a, 2015). The four surface transfer functions $\underline{r}_{aw}$, ..., $\underline{t}_{aw}$ needed here can be estimated by Monte Carlo polarized ray tracing applied to numerous random realizations of the sea

surface corresponding to a fixed set of environmental conditions such as wind speed. Mobley (2015) shows examples of numerically computed transfer functions. However, these topics go beyond what is needed for the present development and need not be discussed further.

1.5 Solution of the VRTE Within the Water

The stage is now set for the primary goal of this chapter: the solution of the VRTE using invariant imbedding theory. The VRTE and its boundary conditions constitute a linear (in the radiances) two-point boundary value problem. That is, the linear VRTE must be solved subject to boundary conditions at both the sea surface and bottom. *The essence of invariant imbedding is that it converts a linear two-point boundary value problem into a pair of non-linear initial value problems.* In the present case, one of these initial value problems is solved starting with a boundary value at the sea surface and integrating downward, and the other is solved starting with a boundary value at the sea bottom and integrating upward.

Invariant imbedding is mathematically complicated and consequently tedious to program, but the payoff is that it is both much faster (at least for the SRTE) in run time than the mathematically simpler (and therefore widely used) Monte Carlo technique and more general in the allowed inputs (such as allowing arbitrary depth dependence of the IOPs) than some other techniques. In particular, the run time for invariant imbedding is linearly proportional to the optical depth, whereas run times increase exponentially with optical depth for Monte Carlo solutions. Invariant imbedding allows for arbitrary depth dependence of the IOPs, whereas some other techniques (e.g., discrete ordinates and adding-doubling methods) build up the water column as a stack of homogeneous layers, and run time is proportional to the number of layers used.

1.5.1 Physical Space Versus Fourier Space

Invariant imbedding theory can be used to solve directly for the discretized Stokes vectors $\underline{S}(k, u, v, j)$. There is an argument for doing this, namely that the mathematical development below is much easier and the quantities involved all have simple interpretations, e.g., as physical reflectances and transmittances that are bounded by 0 and 1. However, as will be seen in Sect. 1.5.5.4, a physical-space solution algorithm requires the simultaneous solution of a set of nonlinear ordinary differential equations (ODEs) of approximate size $128M^2N$ (ignoring internal source terms and special cases for the polar caps). For a nominal $10 \times 15^\circ$ quad partitioning with $M = 10$, $N = 12$, this gives a system of 153,600 equations. That is a formidable numerical problem.

A discrete function $f(v)$ of azimuthal angle ϕ_v, $v = 1, ..., 2N$, can be represented as sines and cosines via

$$f(v) = \sum_{\ell=0}^{N} \left[\hat{f}_1(\ell) \cos(\ell\phi_v) + \hat{f}_2(\ell) \sin(\ell\phi_v) \right] .$$

This results in $N + 1$ generally non-zero cosine amplitudes $\hat{f}_1(\ell)$ and $N - 1$ sine amplitudes $\hat{f}_2(\ell)$. When applied to the VRTE, Fourier analysis leads to the solution of $N + 1$ sets of ODEs of size $64M^2$ equations for cosine amplitudes, and $N - 1$ sets of $64M^2$ equations for sine amplitudes, for the same total of $128M^2N$ equations solved as for the physical-space formulation. However, with Fourier analysis the ODE solutions are made as a sequence of $N + 1$ independent "small" problems, rather than as one "large" problem.

Therefore this development uses a Fourier analysis of the VRTE in azimuthal angle to minimize computer run times. (The same is done in HydroLight for the SRTE.) The penalty paid for this increase in numerical efficiency is a corresponding increase in the complexity of the mathematical development and computer programming.

1.5.2 Recasting the VRTE as Upward and Downward Equations

Invariant imbedding requires that the VRTE be written as a pair of equations, one for upward directions and one for downward directions. In addition, the sea surface boundary conditions are formulated in terms of transmission and reflection operators that transfer radiance back and forth across the sea surface. As was shown in Fig. 1.2, the sea surface is described mathematically as a nonabsorbing layer of zero physical thickness that represents a discontinuity in the index of refraction and therefore causes scattering (reflection and refraction) of light incident onto the surface from above or below. A depth value of $\zeta = a$ denotes a location in the air just above the mean sea surface, and a depth value of $\zeta = w$ denotes a location at depth 0, but in the water just below the mean sea surface. The bottom at depth $\zeta = m$ is defined by its reflectance properties. Figure 1.6 shows a redrawn version of Fig. 1.2. As in Figs. 1.2 and 1.5, depth is measured positive downward from 0 at the mean sea surface. Because direction $+\zeta$ is downward, a superscript + will denote downwelling radiances (energy propagating downward); superscript − will denote upwelling radiances (energy heading upward). Thus in the figure, the downward arrow labeled $\underline{S}^+(a)$ represents downwelling radiance incident onto the sea surface from above, and $\underline{S}^-(a)$ represents upwelling radiance just above the surface, i.e., water-leaving radiance. $\underline{S}^+(w)$ represents radiance that has been transmitted through the surface and is incident onto the water below. The sphere of directions at depth ζ highlights a quad containing an upwelling radiance $\underline{S}^-(\zeta)$ (red quad and red arrow). The blue quad and blue arrow indicate downwelling direction and radiance.

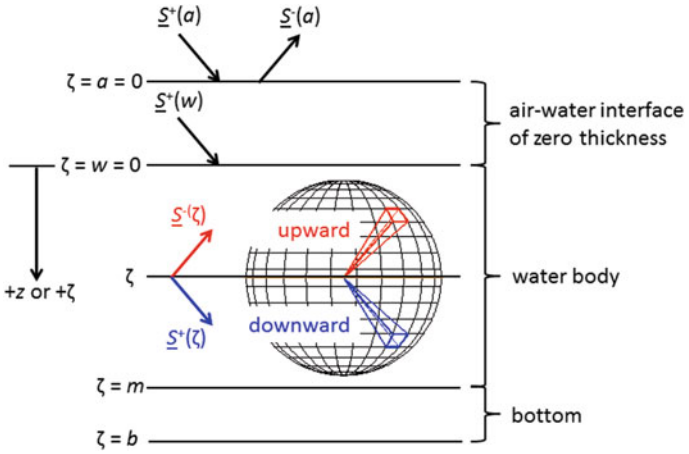

Fig. 1.6 The depth coordinate system used for invariant imbedding. Upwelling and downwelling directions are indicated, along with a few representative Stokes vectors

Recall also that $\mu_u = \cos\theta_u$ in Eq. (1.36) ranges from -1 when $\theta = 180$ (radiance heading straight up) to $+1$ when $\theta = 0$ (radiance heading straight down). It will be convenient to redefine μ_u as a positive number and incorporate the negative μ values for upwelling radiance via an explicit minus sign in the VRTE. With these conventions, and dropping the wavelength argument j for simplicity of notation, the VRTE (1.36) can be written as a pair of equations:

$$\begin{aligned} \mp\mu_u \frac{d\underline{S}^{\mp}(\zeta,u,v)}{d\zeta} &= -\underline{S}^{\mp}(\zeta,u,v) \\ &\quad + \omega_o(\zeta)\sum_r\sum_s \underline{\tilde{P}}^{\pm}(\zeta;r,s\to u,v)\,\underline{S}^{-}(\zeta,r,s) \\ &\quad + \omega_o(\zeta)\sum_r\sum_s \underline{\tilde{P}}^{\mp}(\zeta;r,s\to u,v)\,\underline{S}^{+}(\zeta,r,s) \\ &\quad + \underline{\Sigma}^{\mp}(\zeta,u,v)\ . \end{aligned} \tag{1.51}$$

Now $u = 1, ..., M$ in each hemisphere of directions (quad M is the polar cap), $v = 1, ..., 2N$, and $\mu_u > 0$. Note that now there are two radiances with the same (u, v) quad indices, but one is in the upper hemisphere ($-$ superscript) and one is in the lower ($+$ superscript). The $+$ superscript on the phase matrix $\underline{\tilde{P}}^{+}(\zeta; r, s \to u, v)$ denotes scattering between quads $Q(r, s)$ and $Q(u, v)$ that lie in the same hemisphere. The $-$ superscript on $\underline{\tilde{P}}^{-}(\zeta; r, s \to u, v)$ denotes scattering between quads in opposite hemispheres. This is illustrated in Fig. 1.7. Scattering from the green quad $Q(r, s)$ to the red quad $Q(u, v)$ is described by $\underline{\tilde{P}}^{+}(\zeta; r, s \to u, v)$; scattering from the green to the blue quad has a different scattering angle ψ and is described by

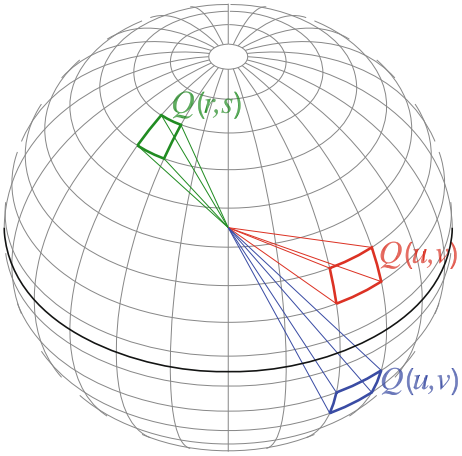

Fig. 1.7 Illustration of scattering within and between hemispheres of directions. Scattering within a hemisphere, e.g. from the green quad $Q(r, s)$ to the red quad $Q(u, v)$, is described by $\tilde{P}^{+}(r, s, u, v)$. Scattering between hemispheres, e.g. from the green quad $Q(r, s)$ to the blue quad $Q(u, v)$, is described by $\tilde{P}^{-}(r, s, u, v)$

$\underline{\tilde{P}}^{-}(\zeta; r, s \to u, v)$. Taking the top sign in Eq. (1.51) gives the VRTE for upwelling radiances $\underline{S}^{-}(\zeta, u, v)$, and the bottom sign gives the equation for downwelling radiances $\underline{S}^{+}(\zeta, u, v)$.

1.5.3 Fourier Decomposition of the Upward and Downward Equations

The next step is to Fourier decompose the upward and downward pair of VRTE of equations. Appendix A contains for reference the formulas for Fourier decomposition of discrete functions of one or two azimuthal angles, as required here.

The Stokes vectors are functions of a single azimuthal angle and are decomposed as (Eq. 1.191)

$$\underline{S}^{\pm}(\zeta, u, v) = \sum_{\ell=0}^{N} \left[\underline{\hat{S}}_1^{\pm}(\zeta, u|\ell) \cos(\ell\phi_v) + \underline{\hat{S}}_2^{\pm}(\zeta, u|\ell) \sin(\ell\phi_v) \right], \tag{1.52}$$

where $v = 1, 2, ..., 2N$. The vertical bar in the $(\zeta, u|\ell)$ argument list is used to separate the physical variables from the Fourier mode variable ℓ. The cosine and sine amplitudes are computed via Eqs. (1.192) and (1.194), respectively. A similar equation holds for the source term $\underline{\Sigma}$. In all amplitude variables here and below, subscript 1 denotes a cosine amplitude, and subscript 2 a sine amplitude.

As was noted in Sect. 1.2.4.1, the elements of the phase matrix depend on either $\cos(\phi_v - \phi_s)$ or $\sin(\phi_v - \phi_s)$. Rather than keep track of which element has which dependence, it is notationally convenient to expand each element of $\underline{P}^{\pm}(\zeta, r, s \to u, v)$ as

$$\underline{P}^{\pm}(\zeta, r, s \to u, v) =$$
$$\sum_{k=0}^{N} \left[\underline{\hat{P}}_1^{\pm}(\zeta, r, u|k) \cos[k(\phi_v - \phi_s)] + \underline{\hat{P}}_2^{\pm}(\zeta, r, u|k) \sin[k(\phi_v - \phi_s)] \right] . \quad (1.53)$$

However, when performing actual computations, either $\hat{P}_{1_{i,j}}^{\pm}$ or $\hat{P}_{2_{i,j}}^{\pm}$ will be zero for a particular $P_{i,j}^{\pm}$ element.

The non-zero amplitudes in Eq. (1.53) are computed by Eqs. (1.197) and (1.199). However, polar caps must always be treated as special cases (L&W 8.34(i)-8.34(iv)). Recall also that $s = 1$ or $\phi' = 0$ was chosen as the reference point to anchor the azimuthal dependence on $\phi - \phi'$. Then depending on which element of $\underline{P}^{\pm}(\zeta, r, s \to u, v)$ is being Fourier analyzed, either the cosine or sine formulas are used as follows (with the other amplitude set to zero):

(i) **Quad-to-quad scattering**: $u, r = 1, 2, ..., M - 1$. Equations (1.197) and (1.199) hold as written. Thus for given r, u values,

$$\underline{\hat{P}}_1^{\pm}(r, u|k) = \frac{1}{\epsilon_k} \sum_{v=1}^{2N} \underline{P}^{\pm}(r, s = 1 \to u, v) \cos(k\phi_v) \quad \text{for } k = 0, ..., N \quad (1.54a)$$

$$\underline{\hat{P}}_2^{\pm}(r, u|k) = \frac{1}{\gamma_k} \sum_{v=1}^{2N} \underline{P}^{\pm}(r, s = 1 \to u, v) \sin(k\phi_v) \quad \text{for } k = 1, ..., N - 1 \quad (1.54b)$$

$$\underline{\hat{P}}_2^{\pm}(r, u|k) = 0 \qquad \text{for } k = 0 \text{ or } N \quad (1.54c)$$

where ϵ_ℓ and γ_ℓ are defined by Eqs. (1.193) and (1.195).

(ii) **Polar cap-to-quad scattering**: $r = M$; $u = 1, 2, ..., M - 1$.

$$\underline{\hat{P}}_1^{\pm}(M, u|k) = \underline{P}^{\pm}(M, \cdot \to u, v) \quad \text{for } k = 0 \quad (1.54d)$$
$$\underline{\hat{P}}_1^{\pm}(M, u|k) = 0 \qquad \text{for } k = 1, ..., N \quad (1.54e)$$
$$\underline{\hat{P}}_2^{\pm}(M, u|k) = 0 \qquad \text{for } k = 0, ..., N \quad (1.54f)$$

(iii) **Quad-to-Polar cap scattering**: $r = 1, 2, ..., M - 1$; $u = M$.

$$\underline{\hat{P}}_1^{\pm}(r, M|k) = \underline{P}^{\pm}(r, s = 1 \to M, \cdot) \quad \text{for } k = 0 \quad (1.54g)$$
$$\underline{\hat{P}}_1^{\pm}(r, M|k) = 0 \qquad \text{for } k = 1, ..., N \quad (1.54h)$$
$$\underline{\hat{P}}_2^{\pm}(r, M|k) = 0 \qquad \text{for } k = 0, ..., N \quad (1.54i)$$

(iv) **Polar cap-to-Polar cap scattering**: $r = u = M$.

$$\underline{\hat{P}}_1^{\pm}(M, M|k) = \underline{P}^{\pm}(M, \cdot \to M, \cdot) \quad \text{for } k = 0 \tag{1.54j}$$

$$\underline{\hat{P}}_1^{\pm}(M, M|k) = 0 \quad \text{for } k = 1, ..., N \tag{1.54k}$$

$$\underline{\hat{P}}_2^{\pm}(M, M|k) = 0 \quad \text{for } k = 0, ..., N \tag{1.54l}$$

The dot argument for s or v for polar caps is a reminder that polar caps have no ϕ' or ϕ dependence. In the computer code, the polar cap values can be stored in the $r = M, s = 1$ or $u = M, v = 1$ array locations, with the remaining array elements for $r, u = M$ being unused and set to zero.

Inserting these Fourier decompositions into Eq. 1.51 gives (L&W 8.35)

$$\begin{aligned}
&\mp\mu_u \frac{d}{d\zeta} \sum_{\ell=0}^{N} \left[\underline{\hat{S}}_1^{\mp}(\zeta, u|\ell) \cos(\ell\phi_v) + \underline{\hat{S}}_2^{\mp}(\zeta, u|\ell) \sin(\ell\phi_v) \right] \\
&= -\sum_{\ell=0}^{N} \left[\underline{\hat{S}}_1^{\mp}(\zeta, u|\ell) \cos(\ell\phi_v) + \underline{\hat{S}}_2^{\mp}(\zeta, u|\ell) \sin(\ell\phi_v) \right] \\
&+ \omega_o(\zeta) \sum_{r=1}^{M-1} \sum_{s=1}^{2N} \left\{ \sum_{k=0}^{N} \left[\underline{\hat{P}}_1^{\pm}(\zeta, r, u|k) \cos[k(\phi_v - \phi_s)] + \underline{\hat{P}}_2^{\pm}(\zeta, r, u|k) \sin[k(\phi_v - \phi_s)] \right] \right\} \\
&\qquad \times \left\{ \sum_{\ell=0}^{N} \left[\underline{\hat{S}}_1^{-}(\zeta, r|\ell) \cos(\ell\phi_s) + \underline{\hat{S}}_2^{-}(\zeta, r|\ell) \sin(\ell\phi_s) \right] \right\} \\
&+ \omega_o(\zeta) \underline{\hat{P}}_1^{\pm}(\zeta, M, u|0) \underline{\hat{S}}_1^{-}(\zeta, M|0) \\
&+ \omega_o(\zeta) \sum_{r=1}^{M-1} \sum_{s=1}^{2n} \left\{ \sum_{k=0}^{N} \left[\underline{\hat{P}}_1^{\mp}(\zeta, r, u|k) \cos[k(\phi_v - \phi_s)] + \underline{\hat{P}}_2^{\mp}(\zeta, r, u|k) \sin[k(\phi_v - \phi_s)] \right] \right\} \\
&\qquad \times \left\{ \sum_{\ell=0}^{N} \left[\underline{\hat{S}}_1^{+}(\zeta, r|\ell) \cos(\ell\phi_s) + \underline{\hat{S}}_2^{+}(\zeta, r|\ell) \sin(\ell\phi_s) \right] \right\} \\
&+ \omega_o(\zeta) \underline{\hat{P}}_1^{\mp}(\zeta, M, u|0) \underline{\hat{S}}_1^{+}(\zeta, M|0) \\
&+ \sum_{\ell=0}^{N} \left[\underline{\hat{\Sigma}}_1^{\mp}(\zeta, u|\ell) \cos(\ell\phi_v) + \underline{\hat{\Sigma}}_2^{\mp}(\zeta, u|\ell) \sin(\ell\phi_v) \right] .
\end{aligned} \tag{1.55}$$

Here the generic sums over r and s have been written to explicitly separate out the polar cap quads, $r = M$, which have no azimuthal dependence. For polar caps, the Fourier cosine amplitude for $k = 0$ is just the value of the physical variable, the cosine amplitudes are zero for $k > 0$, and the sine amplitudes are zero for all k values.

Equation (1.55) is more complicated than its scalar equivalent (L&W 8.35) because it is now necessary to have both cosine ($\underline{\hat{P}}_1^{\pm}$) and sine ($\underline{\hat{P}}_2^{\pm}$) amplitudes to represent $\underline{P}^{\pm}$, whereas the phase function $\tilde{\beta}$ in the SRTE needed only cosine amplitudes $\hat{\beta}$, so no subscript was needed on $\hat{\beta}$ as seen in (L&W 8.35). The

necessity of separating cosine and sign amplitudes carries through to the local and global transmission and reflection operators, as will be seen below.

The path function terms in Eq. (1.55) have quantities of the form

$$\sum_{l=0}^{N}\sum_{k=0}^{N} f(k)g(\ell)\sum_{s=1}^{2N}\begin{Bmatrix}\cos[k(\phi_v-\phi_s)]\cos(\ell\phi_s)\\ \cos[k(\phi_v-\phi_s)]\sin(\ell\phi_s)\\ \sin[k(\phi_v-\phi_s)]\cos(\ell\phi_s)\\ \sin[k(\phi_v-\phi_s)]\sin(\ell\phi_s)\end{Bmatrix},$$

where $f(k)$ is a phase matrix element for a given ζ, r, u, and $g(\ell)$ is a Stokes vector element for a given ζ, r. These four combinations of cosine and sine terms are reduced as in Eqs. (1.203) to (1.206) to obtain

$$\sum_{l=0}^{N} f(\ell)g(\ell)\begin{Bmatrix}\epsilon_\ell\cos(\ell\phi_v)\\ \gamma_\ell\sin(\ell\phi_v)\\ \epsilon_\ell\sin(\ell\phi_v)\\ -\gamma_\ell\cos(\ell\phi_v)\end{Bmatrix},$$

where ϵ_ℓ and γ_ℓ are defined by Eqs. (1.193) and (1.195). (The minus sign in $-\gamma_\ell\cos(\ell\phi_v)$ results from the reduction of the $\sin[k(\phi_v-\phi_s)]\sin(\ell\phi_s)$ term and the minus sign in $\sin(a-b)=\sin a\cos b-\cos a\sin b$, so there's no escaping it. As will be seen, this sign carries through the rest of the derivation and causes a certain loss of symmetry in many equations, compared to the scalar equations in *Light and Water*.)

Equation (1.55) therefore reduces to (L&W 8.36)

$$\begin{aligned}
&\sum_{\ell=0}^{N}\left\{\mp\mu_u\frac{d}{d\zeta}\underline{\hat{S}}_1^{\mp}(\zeta,u|\ell)\cos(\ell\phi_v)\mp\mu_u\frac{d}{d\zeta}\underline{\hat{S}}_2^{\mp}(\zeta,u|\ell)\sin(\ell\phi_v)\right\}\\
&=-\sum_{\ell=0}^{N}\left\{\underline{\hat{S}}_1^{\mp}(\zeta,u|\ell)\cos(\ell\phi_v)+\underline{\hat{S}}_2^{\mp}(\zeta,u|\ell)\sin(\ell\phi_v)\right\}\\
&+\omega_o(\zeta)\sum_{\ell=0}^{N}\left\{\sum_{r=1}^{M-1}\epsilon_\ell\,\underline{\hat{P}}_1^{\pm}(\zeta,r,u|\ell)\underline{\hat{S}}_1^{-}(\zeta,r|\ell)+\delta_\ell\,\underline{\hat{P}}_1^{\pm}(\zeta,M,u|\ell)\underline{\hat{S}}_1^{-}(\zeta,M|\ell)\right\}\cos(\ell\phi_v)\\
&-\omega_o(\zeta)\sum_{\ell=0}^{N}\left\{\sum_{r=1}^{M-1}\gamma_\ell\,\underline{\hat{P}}_2^{\pm}(\zeta,r,u|\ell)\underline{\hat{S}}_2^{-}(\zeta,r|\ell)\right\}\cos(\ell\phi_v)\\
&+\omega_o(\zeta)\sum_{\ell=0}^{N}\left\{\sum_{r=1}^{M-1}\gamma_\ell\,\underline{\hat{P}}_1^{\pm}(\zeta,r,u|\ell)\underline{\hat{S}}_2^{-}(\zeta,r|\ell)+\sum_{r=1}^{M-1}\epsilon_\ell\,\underline{\hat{P}}_2^{\pm}(\zeta,r,u|\ell)\underline{\hat{S}}_1^{-}(\zeta,r|\ell)\right\}\sin(\ell\phi_v)\\
&+\omega_o(\zeta)\sum_{\ell=0}^{N}\left\{\sum_{r=1}^{M-1}\epsilon_\ell\,\underline{\hat{P}}_1^{\mp}(\zeta,r,u|\ell)\underline{\hat{S}}_1^{+}(\zeta,r|\ell)+\delta_\ell\,\underline{\hat{P}}_1^{\pm}(\zeta,M,u|\ell)\underline{\hat{S}}_1^{+}(\zeta,M|\ell)\right\}\cos(\ell\phi_v)\\
&-\omega_o(\zeta)\sum_{\ell=0}^{N}\left\{\sum_{r=1}^{M-1}\gamma_\ell\,\underline{\hat{P}}_2^{\mp}(\zeta,r,u|\ell)\underline{\hat{S}}_2^{+}(\zeta,r|\ell)\right\}\cos(\ell\phi_v)
\end{aligned}$$

$$+\omega_o(\zeta)\sum_{\ell=0}^{N}\left\{\sum_{r=1}^{M-1}\gamma_\ell\,\underline{\hat{P}}_1^{\mp}(\zeta,r,u|\ell)\underline{\hat{S}}_2^{+}(\zeta,r|\ell)+\sum_{r=1}^{M-1}\epsilon_\ell\,\underline{\hat{P}}_2^{\mp}(\zeta,r,u|\ell)\underline{\hat{S}}_1^{+}(\zeta,r|\ell)\right\}\sin(\ell\phi_v)$$
$$+\sum_{\ell=0}^{N}\left\{\underline{\hat{\Sigma}}_1^{\mp}(\zeta,u|\ell)\cos(\ell\phi_v)+\underline{\hat{\Sigma}}_2^{\mp}(\zeta,u|\ell)\sin(\ell\phi_v)\right\}\ .$$

The polar cap terms, which involve only $\ell=0$ cosine amplitudes, have been incorporated into the sums over ℓ by use of a Kronecker δ_ℓ function, which is given by Eq. (1.187). Note that the minus sign in resulting from the reduction of the sums over $\sin[k(\phi_v-\phi_s)]\sin(\ell\phi_s)$ gives a minus sign in all terms involving $\underline{\hat{P}}_2^{\pm}(\zeta,r,u|\ell)\underline{\hat{S}}_2^{\pm}(\zeta,r|\ell)$. This sign carries through the rest of the derivation and introduces a sign asymmetry in the cosine and sine equations, which does not occur in the corresponding scalar equations (because they do not require $\underline{\hat{P}}_2^{\pm}(\zeta,r,u|\ell)$ terms in the Fourier expansion of the phase function). The same sign appears, for example, in Eq. (16) of Hovenier (1971) in a decomposition of the continuous-variable (undiscretized) VRTE.

The linear independence of $\cos(\ell\phi_v)$ and $\sin(\ell\phi_v)$ for different ℓ values means that this last equation must separately hold true for each value of $\ell=0,\ldots,N$ for the $\underline{\hat{S}}_1$ amplitudes, and for $\ell=1,\ldots,N-1$ for the $\underline{\hat{S}}_2$ amplitudes. Collecting together the coefficients of $\cos(\ell\phi_v)$ then gives

$$\mp\mu_u\frac{d}{d\zeta}\underline{\hat{S}}_1^{\mp}(\zeta,u|\ell)=-\underline{\hat{S}}_1^{\mp}(\zeta,u|\ell)$$
$$+\sum_{r=1}^{M-1}\omega_o(\zeta)\,\epsilon_\ell\,\underline{\hat{P}}_1^{\pm}(\zeta,r,u|\ell)\underline{\hat{S}}_1^{-}(\zeta,r|\ell)+\omega_o(\zeta)\,\delta_\ell\,\underline{\hat{P}}_1^{\pm}(\zeta,M,u|\ell)\underline{\hat{S}}_1^{-}(\zeta,M|\ell)$$
$$-\sum_{r=1}^{M-1}\omega_o(\zeta)\,\gamma_\ell\,\underline{\hat{P}}_2^{\pm}(\zeta,r,u|\ell)\underline{\hat{S}}_2^{-}(\zeta,r|\ell)$$
$$+\sum_{r=1}^{M-1}\omega_o(\zeta)\,\epsilon_\ell\,\underline{\hat{P}}_1^{\mp}(\zeta,r,u|\ell)\underline{\hat{S}}_1^{+}(\zeta,r|\ell)+\omega_o(\zeta)\,\delta_\ell\,\underline{\hat{P}}_1^{\mp}(\zeta,M,u|\ell)\underline{\hat{S}}_1^{+}(\zeta,M|\ell)$$
$$-\sum_{r=1}^{M-1}\omega_o(\zeta)\,\gamma_\ell\,\underline{\hat{P}}_2^{\mp}(\zeta,r,u|\ell)\underline{\hat{S}}_2^{+}(\zeta,r|\ell)$$
$$+\underline{\hat{\Sigma}}_1^{\mp}(\zeta,u|\ell)\ . \tag{1.56}$$

The equation obtained from the coefficients of the $\sin(\ell\phi_v)$ terms is

$$\mp\mu_u\frac{d}{d\zeta}\underline{\hat{S}}_2^{\mp}(\zeta,u|\ell)=-\underline{\hat{S}}_2^{\mp}(\zeta,u|\ell)$$
$$+\sum_{r=1}^{M-1}\omega_o(\zeta)\,\gamma_\ell\,\underline{\hat{P}}_1^{\pm}(\zeta,r,u|\ell)\underline{\hat{S}}_2^{-}(\zeta,r|\ell)+\sum_{r=1}^{M-1}\omega_o(\zeta)\,\epsilon_\ell\,\underline{\hat{P}}_2^{\pm}(\zeta,r,u|\ell)\underline{\hat{S}}_1^{-}(\zeta,r|\ell)$$

$$+\sum_{r=1}^{M-1}\omega_o(\zeta)\,\gamma_\ell\,\underline{\hat{P}}_1^{\mp}(\zeta,r,u|\ell)\underline{\hat{S}}_2^{+}(\zeta,r|\ell)+\sum_{r=1}^{M-1}\omega_o(\zeta)\,\epsilon_\ell\,\underline{\hat{P}}_2^{\mp}(\zeta,r,u|\ell)\underline{\hat{S}}_1^{+}(\zeta,r|\ell)$$
$$+\underline{\hat{\Sigma}}_2^{\mp}(\zeta,u|\ell)\ . \tag{1.57}$$

Since $\underline{\hat{S}}_2$, $\underline{\hat{P}}_2$, and $\underline{\hat{\Sigma}}_2$ are all 0 when $\ell=0$ or $\ell=N$, for notational convenience this last equation can be regarded as holding for $\ell=0$ and N as well. Thus Eqs. (1.56) and (1.57) can both be regarded as holding for $u=1,...,M-1$ and $\ell=0,...,N$.

Note in Eqs. (1.56) and (1.57) that $\underline{\hat{P}}_1$ scatters $\underline{\hat{S}}_1$ into $\underline{\hat{S}}_1$ and $\underline{\hat{S}}_2$ into $\underline{\hat{S}}_2$, whereas $\underline{\hat{P}}_2$ scatters $\underline{\hat{S}}_1$ into $\underline{\hat{S}}_2$ and $\underline{\hat{S}}_2$ into $\underline{\hat{S}}_1$. That is, $\underline{\hat{P}}_2$ converts Stokes vector cosine amplitudes into sine amplitudes, and vice versa.

The equation for the polar cap $u=M$ involves only the cosine term for $\ell=0$ since there is no ϕ dependency. The vector equations are therefore have the same form as the scalar (L&W 8.39):

$$\mp\mu_u\frac{d}{d\zeta}\underline{\hat{S}}_1^{\mp}(\zeta,M|0)=-\underline{\hat{S}}_1^{\mp}(\zeta,M|0)$$
$$+\sum_{r=1}^{M-1}\omega_o(\zeta)\,\epsilon_0\,\underline{\hat{P}}_1^{\pm}(\zeta,r,M|0)\underline{\hat{S}}_1^{-}(\zeta,r|0)+\omega_o(\zeta)\,\underline{\hat{P}}_1^{\pm}(\zeta,M,M|0)\underline{\hat{S}}_1^{-}(\zeta,M|0)$$
$$+\sum_{r=1}^{M-1}\omega_o(\zeta)\,\epsilon_0\,\underline{\hat{P}}_1^{\mp}(\zeta,r,M|0)\underline{\hat{S}}_1^{-}(\zeta,r|0)+\omega_o(\zeta)\,\underline{\hat{P}}_1^{\mp}(\zeta,M,M|0)\underline{\hat{S}}_1^{+}(\zeta,M|0)$$
$$+\underline{\hat{\Sigma}}_1^{\mp}(\zeta,M|0)\ . \tag{1.58}$$

Taking the lower signs in Eq. (1.56) gives the equation for the downwelling radiance cosine amplitudes. The "attenuation" term $-\underline{\hat{S}}_1^{+}(\zeta,u|\ell)$ term can be incorporated into the sum over r by use of a Kronecker delta function. The result is (L&W 8.40)

$$\frac{d}{d\zeta}\underline{\hat{S}}_1^{+}(\zeta,u|\ell)=\sum_{r=1}^{M-1}\left[\frac{\omega_o(\zeta)\epsilon_\ell}{\mu_u}\underline{\hat{P}}_1^{-}(\zeta,r,u|\ell)\right]\underline{\hat{S}}_1^{-}(\zeta,r|\ell)$$
$$+\left[\frac{\omega_o(\zeta)\delta_\ell}{\mu_u}\underline{\hat{P}}_1^{-}(\zeta,M,u|\ell)\right]\underline{\hat{S}}_1^{-}(\zeta,M|\ell)$$
$$-\sum_{r=1}^{M-1}\left[\frac{\omega_o(\zeta)\gamma_\ell}{\mu_u}\underline{\hat{P}}_2^{-}(\zeta,r,u|\ell)\right]\underline{\hat{S}}_2^{-}(\zeta,r|\ell)$$
$$+\sum_{r=1}^{M-1}\left[\frac{\omega_o(\zeta)\epsilon_\ell}{\mu_u}\underline{\hat{P}}_1^{+}(\zeta,r,u|\ell)-\frac{\delta_{r-u}}{\mu_u}\underline{I}_{4\times4}\right]\underline{\hat{S}}_1^{+}(\zeta,r|\ell)$$
$$+\left[\frac{\omega_o(\zeta)\delta_\ell}{\mu_u}\underline{\hat{P}}_1^{+}(\zeta,M,u|\ell)\right]\underline{\hat{S}}_1^{+}(\zeta,M|\ell)$$

$$- \sum_{r=1}^{M-1} \left[\frac{\omega_o(\zeta)\gamma_\ell}{\mu_u} \underline{\hat{P}}_2^+(\zeta, r, u|\ell) \right] \underline{\hat{S}}_2^+(\zeta, r|\ell) + \frac{1}{\mu_u} \underline{\hat{\Sigma}}_1^+(\zeta, u|\ell) \,. \tag{1.59}$$

where $\underline{I}_{4\times4}$ is the 4×4 identity matrix. The corresponding equation for the upwelling radiance cosine amplitudes is (L&W 8.41)

$$\begin{aligned} -\frac{d}{d\zeta} \underline{\hat{S}}_1^-(\zeta, u|\ell) = & \sum_{r=1}^{M-1} \left[\frac{\omega_o(\zeta)\epsilon_\ell}{\mu_u} \underline{\hat{P}}_1^+(\zeta, r, u|\ell) - \frac{\delta_{r-u}}{\mu_u} \underline{I}_{4\times4} \right] \underline{\hat{S}}_1^-(\zeta, r|\ell) \\ & + \left[\frac{\omega_o(\zeta)\delta_\ell}{\mu_u} \underline{\hat{P}}_1^+(\zeta, M, u|\ell) \right] \underline{\hat{S}}_1^-(\zeta, M|\ell) \\ & - \sum_{r=1}^{M-1} \left[\frac{\omega_o(\zeta)\gamma_\ell}{\mu_u} \underline{\hat{P}}_2^+(\zeta, r, u|\ell) \right] \underline{\hat{S}}_2^-(\zeta, r|\ell) \\ & + \sum_{r=1}^{M-1} \left[\frac{\omega_o(\zeta)\epsilon_\ell}{\mu_u} \underline{\hat{P}}_1^-(\zeta, r, u|\ell) \right] \underline{\hat{S}}_1^+(\zeta, r|\ell) \\ & + \left[\frac{\omega_o(\zeta)\delta_\ell}{\mu_u} \underline{\hat{P}}_1^-(\zeta, M, u|\ell) \right] \underline{\hat{S}}_1^+(\zeta, M|\ell) \\ & - \sum_{r=1}^{M-1} \left[\frac{\omega_o(\zeta)\gamma_\ell}{\mu_u} \underline{\hat{P}}_2^-(\zeta, r, u|\ell) \right] \underline{\hat{S}}_2^+(\zeta, r|\ell) + \frac{1}{\mu_u} \underline{\hat{\Sigma}}_1^-(\zeta, u|\ell) \,. \end{aligned} \tag{1.60}$$

For these cosine amplitudes, $u = 1, ..., M-1$ and $\ell = 0, ..., N$.

The corresponding equation for the downwelling sine amplitudes is

$$\begin{aligned} \frac{d}{d\zeta} \underline{\hat{S}}_2^+(\zeta, u|\ell) = & \sum_{r=1}^{M-1} \left[\frac{\omega_o(\zeta)\gamma_\ell}{\mu_u} \underline{\hat{P}}_1^-(\zeta, r, u|\ell) \right] \underline{\hat{S}}_2^-(\zeta, r|\ell) \\ & + \sum_{r=1}^{M-1} \left[\frac{\omega_o(\zeta)\epsilon_\ell}{\mu_u} \underline{\hat{P}}_2^-(\zeta, r, u|\ell) \right] \underline{\hat{S}}_1^-(\zeta, r|\ell) \\ & + \sum_{r=1}^{M-1} \left[\frac{\omega_o(\zeta)\gamma_\ell}{\mu_u} \underline{\hat{P}}_1^+(\zeta, r, u|\ell) - \frac{\delta_{r-u}}{\mu_u} \underline{I}_{4\times4} \right] \underline{\hat{S}}_2^+(\zeta, r|\ell) \\ & + \sum_{r=1}^{M-1} \left[\frac{\omega_o(\zeta)\epsilon_\ell}{\mu_u} \underline{\hat{P}}_2^+(\zeta, r, u|\ell) \right] \underline{\hat{S}}_1^+(\zeta, r|\ell) + \frac{1}{\mu_u} \underline{\hat{\Sigma}}_2^+(\zeta, u|\ell) \,, \end{aligned} \tag{1.61}$$

and the equation for the upwelling sine amplitudes is

$$
\begin{aligned}
-\frac{d}{d\zeta}\underline{\hat{S}}_2^-(\zeta,u|\ell) = &\sum_{r=1}^{M-1}\left[\frac{\omega_o(\zeta)\gamma_\ell}{\mu_u}\underline{\hat{P}}_1^+(\zeta,r,u|\ell) - \frac{\delta_{r-u}}{\mu_u}\underline{I}_{4\times4}\right]\underline{\hat{S}}_2^-(\zeta,r|\ell) \\
&+ \sum_{r=1}^{M-1}\left[\frac{\omega_o(\zeta)\epsilon_\ell}{\mu_u}\underline{\hat{P}}_2^+(\zeta,r,u|\ell)\right]\underline{\hat{S}}_1^-(\zeta,r|\ell) \\
&+ \sum_{r=1}^{M-1}\left[\frac{\omega_o(\zeta)\gamma_\ell}{\mu_u}\underline{\hat{P}}_1^-(\zeta,r,u|\ell)\right]\underline{\hat{S}}_2^+(\zeta,r|\ell) \\
&+ \sum_{r=1}^{M-1}\left[\frac{\omega_o(\zeta)\epsilon_\ell}{\mu_u}\underline{\hat{P}}_2^-(\zeta,r,u|\ell)\right]\underline{\hat{S}}_1^+(\zeta,r|\ell) + \frac{1}{\mu_u}\underline{\hat{\Sigma}}_2^-(\zeta,u|\ell)\,.
\end{aligned}
\tag{1.62}
$$

For the sine amplitude equations, $u = 1, ..., M-1$ and $\ell = 1, ..., N-1$. Because $\epsilon_\ell = \gamma_\ell = N$ for $\ell = 1, ..., N-1$, the sine equations can be written using ϵ_ℓ.

1.5.3.1 Local Reflectance and Transmittance Matrices

The quantities in brackets in Eqs. (1.59)–(1.62) are the *local reflectance* and *local transmittance* functions for the radiance amplitudes. There are several things to note about these quantities:

- These quantities depend only the IOPs at depth ζ (and wavelength λ) and on the choice of quad partioning. They are thus "local" IOPs at a particular depth.
- The same quantities appear in both the downward and upward equations.
- The same quantities appear in both cosine and sine equations when $\ell = 1, ..., N-1$, in which case $\epsilon_\ell = \gamma_\ell$
- The terms involving $\underline{\hat{P}}_{1,2}^+$ "transmit" downwelling radiance into downwelling radiance, and upwelling into upwelling.
- The terms involving $\underline{\hat{P}}_{1,2}^-$ "reflect" downwelling radiance into upwelling radiance, and upwelling into downwelling.

It will prove convenient to cast the preceding equations into matrix form. Recall that the phase matrix amplitudes are 4×4 for a given (r, u) value. Next define $M \times M$ *composite* matrices $\underline{\underline{\hat{\rho}}}_{1,2}^{\pm}$ and $\underline{\underline{\hat{\tau}}}_{1,2}^{\pm}$ whose (row, column) elements (u, r) are obtained from the 4×4 $\underline{\hat{P}}_{1,2}^{\pm}(\zeta, r, u|\ell)$ matrices in the brackets. That is, $\underline{\underline{\hat{\rho}}}$ and $\underline{\underline{\hat{\tau}}}$ are $M \times M$ matrices, each of whose elements is a 4×4 matrix, so that the total size of $\underline{\underline{\hat{\rho}}}$ and $\underline{\underline{\hat{\tau}}}$ is $4M \times 4M$ when written out in full as real numbers. The double underline is a reminder that $\underline{\underline{\hat{\rho}}}$ and $\underline{\underline{\hat{\tau}}}$ are matrices of matrices. The 4×4 elements of these composite matrices are defined as follows:
Quad-to-quad case, $r, u = 1, ..., M-1$:

$$[\underline{\underline{\hat{\tau}}}_1(\zeta|\ell)]_{u,r} \equiv \frac{\omega_o(\zeta)\epsilon_\ell}{\mu_u}\underline{\hat{P}}_1^+(\zeta, r, u|\ell) - \frac{\delta_{r-u}}{\mu_u}\underline{I}_{4\times 4} \quad \text{for } \ell = 0, ..., N \tag{1.63a}$$

$$[\underline{\underline{\hat{\tau}}}_2(\zeta|\ell)]_{u,r} \equiv \frac{\omega_o(\zeta)\gamma_\ell}{\mu_u}\underline{\hat{P}}_2^+(\zeta, r, u|\ell) \quad \text{for } \ell = 1, ..., N-1 \tag{1.63b}$$

$$[\underline{\underline{\hat{\tau}}}_2(\zeta|\ell)]_{u,r} \equiv \underline{0}_{4\times 4} \quad \text{for } \ell = 0 \text{ or } N \tag{1.63c}$$

$$[\underline{\underline{\hat{\rho}}}_1(\zeta|\ell)]_{u,r} \equiv \frac{\omega_o(\zeta)\epsilon_\ell}{\mu_u}\underline{\hat{P}}_1^-(\zeta, r, u|\ell) \quad \text{for } \ell = 0, ..., N \tag{1.63d}$$

$$[\underline{\underline{\hat{\rho}}}_2(\zeta|\ell)]_{u,r} \equiv \frac{\omega_o(\zeta)\gamma_\ell}{\mu_u}\underline{\hat{P}}_2^-(\zeta, r, u|\ell) \quad \text{for } \ell = 1, ..., N-1 \tag{1.63e}$$

$$[\underline{\underline{\hat{\rho}}}_2(\zeta|\ell)]_{u,r} \equiv \underline{0}_{4\times 4} \quad \text{for } \ell = 0 \text{ or } N \tag{1.63f}$$

Polar cap-to-quad case, $r = M, u = 1, ..., M-1$:

$$[\underline{\underline{\hat{\tau}}}_1(\zeta|0)]_{u,M} \equiv \frac{\omega_o(\zeta)}{\mu_u}\underline{\hat{P}}_1^+(\zeta, M, u|0) \quad \text{for } \ell = 0 \tag{1.64a}$$

$$[\underline{\underline{\hat{\tau}}}_1(\zeta|\ell)]_{u,M} \equiv \underline{0}_{4\times 4} \quad \text{for } \ell = 1, ..., N \tag{1.64b}$$

$$[\underline{\underline{\hat{\tau}}}_2(\zeta|\ell)]_{u,M} \equiv \underline{0}_{4\times 4} \quad \text{for} \quad \text{for } \ell = 0, ..., N \tag{1.64c}$$

$$[\underline{\underline{\hat{\rho}}}_1(\zeta|0)]_{u,M} \equiv \frac{\omega_o(\zeta)}{\mu_u}\underline{\hat{P}}_1^-(\zeta, M, u|0) \quad \text{for } \ell = 0 \tag{1.64d}$$

$$[\underline{\underline{\hat{\rho}}}_1(\zeta|\ell)]_{u,M} \equiv \underline{0}_{4\times 4} \quad \text{for } \ell = 1, ..., N \tag{1.64e}$$

$$[\underline{\underline{\hat{\rho}}}_2(\zeta|\ell)]_{u,M} \equiv \underline{0}_{4\times 4} \quad \text{for} \quad \text{for } \ell = 0, ..., N \tag{1.64f}$$

Quad-to-polar cap case, $r = 1, ..., M-1, u = M$:

$$[\underline{\underline{\hat{\tau}}}_1(\zeta|0)]_{M,r} \equiv \frac{\omega_o(\zeta)\epsilon_0}{\mu_u}\underline{\hat{P}}_1^+(\zeta, r, M|0) \tag{1.65a}$$

$$[\underline{\underline{\hat{\tau}}}_1(\zeta|\ell)]_{M,r} \equiv \underline{0}_{4\times 4} \quad \text{for } \ell = 1, ..., N \tag{1.65b}$$

$$[\underline{\underline{\hat{\tau}}}_2(\zeta|\ell)]_{M,r} \equiv \underline{0}_{4\times 4} \quad \text{for } \ell = 0, ..., N \tag{1.65c}$$

$$[\underline{\underline{\hat{\rho}}}_1(\zeta|0)]_{M,r} \equiv \frac{\omega_o(\zeta)\epsilon_0}{\mu_u}\underline{\hat{P}}_1^-(\zeta, r, M|0) \tag{1.65d}$$

$$[\underline{\underline{\hat{\rho}}}_1(\zeta|\ell)]_{M,r} \equiv \underline{0}_{4\times4} \quad \text{for } \ell = 1, ..., N \tag{1.65e}$$

$$[\underline{\underline{\hat{\rho}}}_2(\zeta|\ell)]_{M,r} \equiv \underline{0}_{4\times4} \quad \text{for } \ell = 0, ..., N \tag{1.65f}$$

Polar cap-to-polar cap case, $r = M, u = M$:

$$[\underline{\underline{\hat{\tau}}}_1(\zeta|0)]_{M,M} \equiv \frac{\omega_o(\zeta)\epsilon_0}{\mu_u}\underline{\hat{P}}_1^+(\zeta, M, M|0) - \frac{1}{\mu_M}\underline{I}_{4\times4} \tag{1.66a}$$

$$[\underline{\underline{\hat{\tau}}}_1(\zeta|\ell)]_{M,M} \equiv \underline{0}_{4\times4} \quad \text{for } \ell = 1, ..., N \tag{1.66b}$$

$$[\underline{\underline{\hat{\tau}}}_2(\zeta|\ell)]_{M,M} \equiv \underline{0}_{4\times4} \quad \text{for } \ell = 0, ..., N \tag{1.66c}$$

$$[\underline{\underline{\hat{\rho}}}_1(\zeta|0)]_{M,M} \equiv \frac{\omega_o(\zeta)\epsilon_0}{\mu_u}\underline{\hat{P}}_1^-(\zeta, M, M|0) \tag{1.66d}$$

$$[\underline{\underline{\hat{\rho}}}_1(\zeta|\ell)]_{M,M} \equiv \underline{0}_{4\times4} \quad \text{for } \ell = 1, ..., N \tag{1.66e}$$

$$[\underline{\underline{\hat{\rho}}}_2(\zeta|\ell)]_{M,M} \equiv \underline{0}_{4\times4} \quad \text{for } \ell = 0, ..., N \tag{1.66f}$$

Note that the elements of the Mth row and Mth column of $\underline{\underline{\hat{\rho}}}_1$ and $\underline{\underline{\hat{\tau}}}_1$ are $\underline{0}_{4\times4}$ except when $\ell = 0$; $\underline{0}_{4\times4}$ is the 4×4 matrix of zeros.

With these definitions, Eqs. (1.59) and (1.60) can be written

$$\begin{aligned}\frac{d}{d\zeta}\underline{\hat{S}}_1^+(\zeta, u|\ell) = &\sum_{r=1}^{M}[\underline{\underline{\hat{\rho}}}_1(\zeta|\ell)]_{u,r}\,\underline{\hat{S}}_1^-(\zeta, r|\ell) + \sum_{r=1}^{M}[\underline{\underline{\hat{\tau}}}_1(\zeta|\ell)]_{u,r}\,\underline{\hat{S}}_1^+(\zeta, r|\ell)\\ &- \sum_{r=1}^{M}[\underline{\underline{\hat{\rho}}}_2(\zeta|\ell)]_{u,r}\,\underline{\hat{S}}_2^-(\zeta, r|\ell) - \sum_{r=1}^{M}[\underline{\underline{\hat{\tau}}}_2(\zeta|\ell)]_{u,r}\,\underline{\hat{S}}_2^+(\zeta, r|\ell)\\ &+ \frac{1}{\mu_u}\underline{\hat{\Sigma}}_1^+(\zeta, u|\ell)\,,\end{aligned} \tag{1.67}$$

and

$$\begin{aligned}-\frac{d}{d\zeta}\underline{\hat{S}}_1^-(\zeta, u|\ell) = &\sum_{r=1}^{M}[\underline{\underline{\hat{\rho}}}_1(\zeta|\ell)]_{u,r}\,\underline{\hat{S}}_1^+(\zeta, r|\ell) + \sum_{r=1}^{M}[\underline{\underline{\hat{\tau}}}_1(\zeta|\ell)]_{u,r}\,\underline{\hat{S}}_1^-(\zeta, r|\ell)\\ &- \sum_{r=1}^{M}[\underline{\underline{\hat{\rho}}}_2(\zeta|\ell)]_{u,r}\,\underline{\hat{S}}_2^+(\zeta, r|\ell) - \sum_{r=1}^{M}[\underline{\underline{\hat{\tau}}}_2(\zeta|\ell)]_{u,r}\,\underline{\hat{S}}_2^-(\zeta, r|\ell)\\ &+ \frac{1}{\mu_u}\underline{\hat{\Sigma}}_1^-(\zeta, u|\ell)\,.\end{aligned} \tag{1.68}$$

These equations are valid for $\ell = 0, ..., N$ and $u = 1, ..., M-1$. The polar caps in Eq. (1.58) result in a similar pair of equations involving $\underline{\underline{\hat{\rho}}}_1(\zeta, 0)_{M,r}$ and $\underline{\underline{\hat{\tau}}}_1(\zeta, 0)_{M,r}$.

The corresponding equations for the sine amplitudes are

$$\begin{aligned}\frac{d}{d\zeta}\underline{\hat{S}}_2^+(\zeta, u|\ell) = &\sum_{r=1}^{M}[\underline{\underline{\hat{\rho}}}_1(\zeta|\ell)]_{u,r}\,\underline{\hat{S}}_2^-(\zeta, r|\ell) + \sum_{r=1}^{M}[\underline{\underline{\hat{\tau}}}_1(\zeta|\ell)]_{u,r}\,\underline{\hat{S}}_2^+(\zeta, r|\ell)\\ &+ \sum_{r=1}^{M}[\underline{\underline{\hat{\rho}}}_2(\zeta|\ell)]_{u,r}\,\underline{\hat{S}}_1^-(\zeta, r|\ell) + \sum_{r=1}^{M}[\underline{\underline{\hat{\tau}}}_2(\zeta|\ell)]_{u,r}\,\underline{\hat{S}}_1^+(\zeta, r|\ell)\\ &+ \frac{1}{\mu_u}\underline{\hat{\Sigma}}_2^+(\zeta, u|\ell)\,, \end{aligned} \tag{1.69}$$

and

$$\begin{aligned}-\frac{d}{d\zeta}\underline{\hat{S}}_2^-(\zeta, u|\ell) = &\sum_{r=1}^{M}[\underline{\underline{\hat{\rho}}}_1(\zeta|\ell)]_{u,r}\,\underline{\hat{S}}_2^+(\zeta, r|\ell) + \sum_{r=1}^{M}[\underline{\underline{\hat{\tau}}}_1(\zeta|\ell)]_{u,r}\,\underline{\hat{S}}_2^-(\zeta, r|\ell)\\ &+ \sum_{r=1}^{M}[\underline{\underline{\hat{\rho}}}_2(\zeta|\ell)]_{u,r}\,\underline{\hat{S}}_1^+(\zeta, r|\ell)\\ &+ \sum_{r=1}^{M}[\underline{\underline{\hat{\tau}}}_2(\zeta|\ell)]_{u,r}\,\underline{\hat{S}}_1^-(\zeta, r|\ell) + \frac{1}{\mu_u}\underline{\hat{\Sigma}}_2^-(\zeta, u|\ell)\,. \end{aligned} \tag{1.70}$$

These equations are valid for $\ell = 1, ..., N-1$ and $u = 1, ..., M-1$. The sine amplitudes are zero for the polar caps $u = M$.

It can now be seen that the (u, r) = (row, column) order in the definitions of $[\underline{\underline{\hat{\rho}}}_{1,2}(\zeta|\ell)]_{u,r}$ and $[\underline{\underline{\hat{\tau}}}_{1,2}(\zeta|\ell)]_{u,r}$ was chosen to facilitate converting the summations over r in Eqs. (1.59) to (1.62) to matrix multiplications. Continuing in this direction, now "stack up" the Stokes vectors to define $4M \times 1$ composite column vectors:

$$\underline{\underline{\hat{S}}}_{1,2}^{\pm}(\zeta|\ell) = \begin{bmatrix} \underline{\hat{S}}_{1,2}^{\pm}(\zeta, 1|\ell) \\ \underline{\hat{S}}_{1,2}^{\pm}(\zeta, 2|\ell) \\ \vdots \\ \underline{\hat{S}}_{1,2}^{\pm}(\zeta, M-1|\ell) \\ \underline{\hat{S}}_{1,2}^{\pm}(\zeta, M|\ell) \end{bmatrix}. \tag{1.71}$$

Similarly, arrange the 4×4 elements $[\underline{\underline{\hat{\rho}}}_{1,2}(\zeta|\ell)]_{u,r}$ of the composite $\underline{\underline{\hat{\rho}}}_{1,2}$ matrices into $4M \times 4M$ matrices

$$\underline{\underline{\hat{\rho}}}_{1,2}(\zeta|\ell) = \begin{bmatrix} [\underline{\underline{\hat{\rho}}}_{1,2}(\zeta|\ell)]_{1,1} & [\underline{\underline{\hat{\rho}}}_{1,2}(\zeta|\ell)]_{1,2} & \cdots & [\underline{\underline{\hat{\rho}}}_{1,2}(\zeta|\ell)]_{1,M} \\ [\underline{\underline{\hat{\rho}}}_{1,2}(\zeta|\ell)]_{2,1} & [\underline{\underline{\hat{\rho}}}_{1,2}(\zeta|\ell)]_{2,2} & \cdots & [\underline{\underline{\hat{\rho}}}_{1,2}(\zeta|\ell)]_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ [\underline{\underline{\hat{\rho}}}_{1,2}(\zeta|\ell)]_{M,1} & [\underline{\underline{\hat{\rho}}}_{1,2}(\zeta|\ell)]_{M,2} & \cdots & [\underline{\underline{\hat{\rho}}}_{1,2}(\zeta|\ell)]_{M,M} \end{bmatrix}. \tag{1.72}$$

with a similar equation for $\underline{\underline{\hat{\tau}}}_{1,2}(\zeta|\ell)$.

1.5.3.2 Matrix Form of the Local Interaction Equations

Definitions (1.71) and (1.72) allow Eqs. (1.67)–(1.70) to be written as matrix equations (L&W 8.43 and 8.44):

$$\begin{aligned} \frac{d}{d\zeta}\underline{\underline{\hat{S}}}_1^+(\zeta|\ell) &= \underline{\underline{\hat{\rho}}}_1(\zeta|\ell)\underline{\underline{\hat{S}}}_1^-(\zeta|\ell) + \underline{\underline{\hat{\tau}}}_1(\zeta|\ell)\underline{\underline{\hat{S}}}_1^+(\zeta|\ell) \\ &\quad - \underline{\underline{\hat{\rho}}}_2(\zeta|\ell)\underline{\underline{\hat{S}}}_2^-(\zeta|\ell) - \underline{\underline{\hat{\tau}}}_2(\zeta|\ell)\underline{\underline{\hat{S}}}_2^+(\zeta|\ell) + \underline{\underline{\hat{\Sigma}}}_1^+(\zeta|\ell) \qquad (1.73) \\ -\frac{d}{d\zeta}\underline{\underline{\hat{S}}}_1^-(\zeta|\ell) &= \underline{\underline{\hat{\rho}}}_1(\zeta|\ell)\underline{\underline{\hat{S}}}_1^+(\zeta|\ell) + \underline{\underline{\hat{\tau}}}_1(\zeta|\ell)\underline{\underline{\hat{S}}}_1^-(\zeta|\ell) \\ &\quad - \underline{\underline{\hat{\rho}}}_2(\zeta|\ell)\underline{\underline{\hat{S}}}_2^+(\zeta|\ell) - \underline{\underline{\hat{\tau}}}_2(\zeta|\ell)\underline{\underline{\hat{S}}}_2^-(\zeta|\ell) + \underline{\underline{\hat{\Sigma}}}_1^-(\zeta|\ell) \qquad (1.74) \end{aligned}$$

$$\begin{aligned} \frac{d}{d\zeta}\underline{\underline{\hat{S}}}_2^+(\zeta|\ell) &= \underline{\underline{\hat{\rho}}}_1(\zeta|\ell)\underline{\underline{\hat{S}}}_2^-(\zeta|\ell) + \underline{\underline{\hat{\tau}}}_1(\zeta|\ell)\underline{\underline{\hat{S}}}_2^+(\zeta|\ell) \\ &\quad + \underline{\underline{\hat{\rho}}}_2(\zeta|\ell)\underline{\underline{\hat{S}}}_1^-(\zeta|\ell) + \underline{\underline{\hat{\tau}}}_2(\zeta|\ell)\underline{\underline{\hat{S}}}_1^+(\zeta|\ell) + \underline{\underline{\hat{\Sigma}}}_2^+(\zeta|\ell) \qquad (1.75) \\ -\frac{d}{d\zeta}\underline{\underline{\hat{S}}}_2^-(\zeta|\ell) &= \underline{\underline{\hat{\rho}}}_1(\zeta|\ell)\underline{\underline{\hat{S}}}_2^+(\zeta|\ell) + \underline{\underline{\hat{\tau}}}_1(\zeta|\ell)\underline{\underline{\hat{S}}}_2^-(\zeta|\ell) \\ &\quad + \underline{\underline{\hat{\rho}}}_2(\zeta|\ell)\underline{\underline{\hat{S}}}_1^+(\zeta|\ell) + \underline{\underline{\hat{\tau}}}_2(\zeta|\ell)\underline{\underline{\hat{S}}}_1^-(\zeta|\ell) + \underline{\underline{\hat{\Sigma}}}_2^-(\zeta|\ell) \qquad (1.76) \end{aligned}$$

These equations are the composite matrix form of the *local interaction equations* for the radiance amplitudes. The minus signs on the $\underline{\underline{\hat{\rho}}}_2\underline{\underline{\hat{S}}}_2^{\pm}$ and $\underline{\underline{\hat{\tau}}}_2\underline{\underline{\hat{S}}}_2^{\pm}$ terms in Eqs. (1.73) and (1.74) trace back to the minus sign in Eq. (1.206) These equations hold for each value of $\ell = 0, 1, ..., N$, although some terms may be zero for special cases, e.g., sine terms are zero when $\ell = 0$ or N. Although the minus signs in the $\underline{\underline{\hat{\rho}}}_2\underline{\underline{\hat{S}}}_2^{\pm}$ and $\underline{\underline{\hat{\tau}}}_2\underline{\underline{\hat{S}}}_2^{\pm}$ terms in Eqs. (1.73) and (1.73) may at first seem some way philosophically wrong because they destroy the symmetry between the $d\underline{\underline{\hat{S}}}_1^{\pm}/d\zeta$ and the $d\underline{\underline{\hat{S}}}_2^{\pm}/d\zeta$ equations, and because it looks like there are "negative" contributions by reflectance and transmission of the sine amplitudes. However, up to this point the derivation is just straightforward algebra, and the signs are what they are. In any case, minus signs

are physically allowed because these quantities are all Fourier amplitudes and there is no requirement that a Fourier amplitude "reflectance" for a Fourier amplitude Stokes vector be non-negative. Had the derivation stayed in physical space, there would be no Fourier decomposition, hence no minus signs, and reflectances and transmittances would be more easily interpreted non-negative physical quantities (although the matrices would be much larger).

To summarize, the development so far has

1. Written the discretized VRTE as upward and downward sets of equations,
2. Fourier decomposed the upward and downward equations,
3. Written the resulting equations in matrix form.

The local interaction equations are therefore just a rewritten version of the VRTE.

1.5.4 Global Interaction Equations

The local interaction equations show how radiance is "reflected" and "transmitted" (i.e., scattered) by an infinitesimally thin later of water at a particular depth ζ. Note that "reflected" is not synonymous with "backscattered" because downwelling radiance scattered back upward can result from either forward or backward scattering, even for single scattering. Moreover, the "reflectance" described here includes all orders of multiple scattering. Likewise, "transmitted" is not the same as "forward scattered."

The next step of developing the invariant imbedding solution of the VRTE is to formulate *global interaction equations*, which show how light is reflected and transmitted by finitely thick layers or slabs of water. In particular, the global interaction equations show how the unknown response radiances (radiances leaving the slab) are related to the known incident radiances falling onto the slab from above and below.

Recall the coordinate system of Fig. 1.6. The water column from just beneath the sea surface at depth $\zeta = w = 0$ to the bottom at depth $\zeta = m$ is conceptually divided into an upper layer from the surface at $\zeta = w$ to an arbitrary depth ζ, and a lower layer from ζ to the bottom at $\zeta = m$. These layers generally have depth dependent IOPs. The layer from the surface to depth ζ will be denoted as the slab $[w, \zeta]$, and the layer from ζ to m is slab $[\zeta, m]$. The entire water column is the union of these two slabs: $[w, m] = [w, \zeta] \cup [\zeta, m]$.

The linear interaction principle of electromagnetic theory (Preisendorfer 1976) states that the response radiances leaving a layer of water are linear functions of the incident radiances impinging onto the layer from above and below. As shown in *Light and Water* Sect. 8.6, it is possible to develop the global interaction equations beginning with the local interaction equations and developing *fundamental* and *transport* solutions for the radiance amplitudes. However, it is also possible to write down the interaction equations by inspection, given the guidance of Eqs. (1.73)–(1.76) and the corresponding form of the scalar theory seen in *Light and Water*.

Consider first the upper layer or slab of water $[w, \zeta]$. Let $\underline{R}(w, \zeta|r, s, u, v)$ denote the matrix that "reflects" or converts downwelling Stokes vectors incident onto the upper surface of the slab at depth w, $\underline{S}^+(w, r, s)$, into upwelling vectors $\underline{S}^-(w, u, v)$ that are leaving the upper surface of the slab. In physical space, the radiance (first element of the Stokes vector) is non-negative, and the corresponding reflectance is a non-negative quantity.

Now, however, the formulation is in terms of Fourier amplitudes, which can be positive or negative even when the physical quantity is positive. Moreover, there are different reflectance operators for cosine and sine amplitudes, as was seen in the local interaction equations. Let $\underline{\underline{\hat{R}}}_1(w, \zeta|\ell)$ denote the matrix that converts downwelling radiance amplitudes incident onto the upper side of the slab, $\underline{\hat{S}}^+_{1,2}(w|\ell)$, into the upwelling radiance amplitudes $\underline{\hat{S}}^-_{1,2}(w|\ell)$ that are leaving the upper surface of the slab, and which leaves radiance cosine (sine) amplitudes as cosine (sine) amplitudes. Let $\underline{\underline{\hat{T}}}_1(w, \zeta|\ell)$ denote the matrix that transmits or converts downwelling radiance amplitudes incident onto the upper surface of the slab at depth w into downwelling radiance amplitudes $\underline{\hat{S}}^+_{1,2}(\zeta|\ell)$ that are leaving the lower surface of the slab at depth ζ. Likewise, let $\underline{\underline{\hat{R}}}_2(w, \zeta|\ell)$ and $\underline{\underline{\hat{T}}}_2(w, \zeta|\ell)$ denote matrices that reflect and transmit radiance amplitudes, but which convert cosine (sine) radiance amplitudes to sine (cosine) amplitudes. Thus $\underline{\underline{\hat{R}}}_1(w, \zeta|\ell)$ functions like the local reflection matrix $\underline{\hat{\rho}}_1(\zeta|\ell)$, $\underline{\underline{\hat{T}}}_2(w, \zeta|\ell)$ functions like $\underline{\hat{\tau}}_2(\zeta|\ell)$, and so on, except that $\underline{\underline{\hat{R}}}_1(w, \zeta|\ell)$ etc. describe the reflectance and transmittance properties of the entire slab of water $[w, \zeta]$. Following this notation, $\underline{\underline{\hat{R}}}_1(\zeta, w|\ell)$ is the matrix the reflects upwelling radiance amplitudes incident onto the slab from below at depth ζ, $\underline{\hat{S}}^-_{1,2}(\zeta|\ell)$, into downwelling radiance amplitudes exiting the slab at ζ; $\underline{\underline{\hat{T}}}_2(\zeta, w|\ell)$ transmits upwelling radiance amplitudes from depth ζ to w and interchanges cosine and sine amplitudes, etc. These $\underline{\underline{\hat{R}}}_1$, $\underline{\underline{\hat{R}}}_2$, $\underline{\underline{\hat{T}}}_1$ and $\underline{\underline{\hat{T}}}_2$ matrices are called the (Fourier amplitude versions of the) *standard reflectance* and *standard transmittance* matrices (or operators) for the slab.

Finally, let $\underline{\hat{\Sigma}}^{-t}_{1,2}(\zeta, w|\ell)$ be the contribution of internal sources to the upwelling radiance amplitudes $\underline{\hat{S}}^-_{1,2}(w|\ell)$ exiting the top of the slab. This radiance is built up within the water column starting at depth ζ and increasing upward. The t in the superscript indicates the *transport* solution, following the notation of *Light and Water*. Similarly, $\underline{\hat{\Sigma}}^{+t}_{1,2}(w, \zeta|\ell)$ is the contribution of internal sources to the downwelling radiance amplitudes $\underline{\hat{S}}^+_{1,2}(\zeta|\ell)$ exiting the bottom of the slab. That radiance is built up within the water column starting at the surface w. Figure 1.8 visually summarizes all of the reflectance and transmittance operations that convert incident radiances into response radiances for the slab $[w, \zeta]$. The notation has been simplified for clarity of the figure: $R_1(w, \zeta)$ represents the matrix $\underline{\underline{\hat{R}}}_1(w, \zeta|\ell)$, etc.

Figure 1.8 shows the various reflectance and transmittance operations. The mathematical statement of those operations must be written to conform to the structure

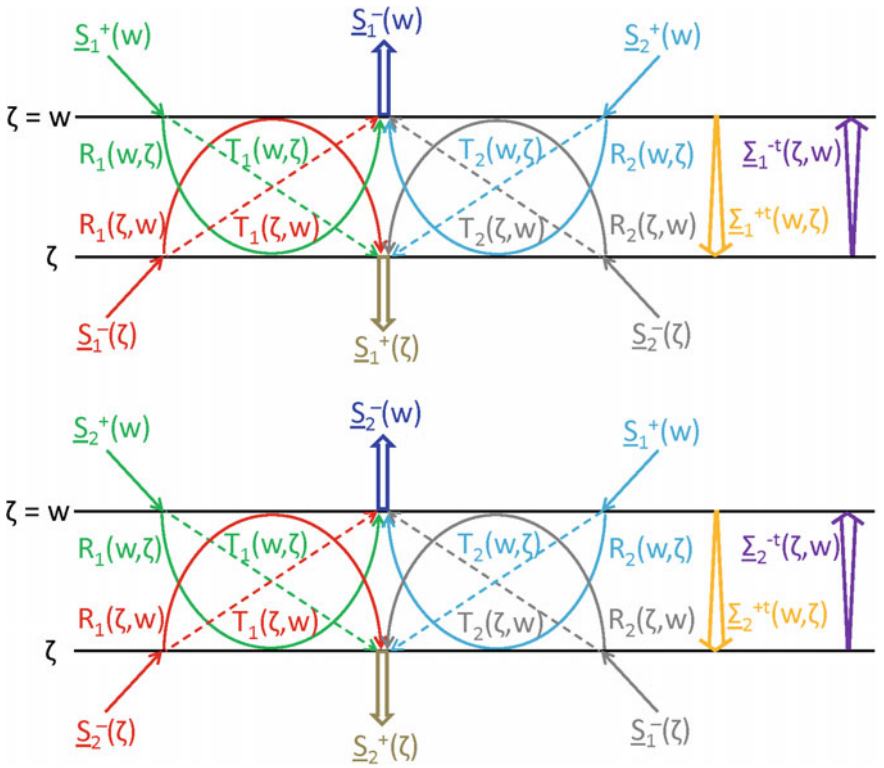

Fig. 1.8 Global interaction principles for the slab $[w, \zeta]$. The upper figure is for the cosine amplitudes of the Stokes vector; the lower figure is for the sine amplitudes. Single-width arrows are incident radiances; double arrows are response radiances. Dotted lines represent transmittances, solid lines reflectances, and variable-width arrows internal sources. Colors code the different radiances and the associated reflectance and transmittance operators

of the local interaction Eqs. (1.73)–(1.76). In particular, terms involving $\underline{\underline{\hat{R}}}_2\underline{\hat{S}}_2^{\pm}$ and $\underline{\underline{\hat{T}}}_2\underline{\hat{S}}_2^{\pm}$ must be given minus signs, as seen in Eqs. (1.73) and (1.74) With this guidance, the global interaction equations for the cosine radiance amplitudes for slab $[w, \zeta]$ (L&W 8.91, 8.92) are

$$\begin{aligned}\underline{\hat{S}}_1^-(w|\ell) &= \underline{\underline{\hat{R}}}_1(w,\zeta|\ell)\underline{\hat{S}}_1^+(w|\ell) + \underline{\underline{\hat{T}}}_1(\zeta,w|\ell)\underline{\hat{S}}_1^-(\zeta|\ell) \\ &\quad - \underline{\underline{\hat{R}}}_2(w,\zeta|\ell)\underline{\hat{S}}_2^+(w|\ell) - \underline{\underline{\hat{T}}}_2(\zeta,w|\ell)\underline{\hat{S}}_2^-(\zeta|\ell) + \underline{\hat{\Sigma}}_1^{-t}(\zeta,w|\ell) \qquad (1.77)\\ \underline{\hat{S}}_1^+(\zeta|\ell) &= \underline{\underline{\hat{R}}}_1(\zeta,w|\ell)\underline{\hat{S}}_1^-(\zeta|\ell) + \underline{\underline{\hat{T}}}_1(w,\zeta|\ell)\underline{\hat{S}}_1^+(w|\ell) \\ &\quad - \underline{\underline{\hat{R}}}_2(\zeta,w|\ell)\underline{\hat{S}}_2^-(\zeta|\ell) - \underline{\underline{\hat{T}}}_2(w,\zeta|\ell)\underline{\hat{S}}_2^+(w|\ell) + \underline{\hat{\Sigma}}_1^{+t}(w,\zeta|\ell) \qquad (1.78)\end{aligned}$$

A similar pair of equations can be written for the sine amplitudes:

$$\begin{aligned}\underline{\underline{\hat{S}}}_2^-(w|\ell) &= \underline{\underline{\hat{R}}}_1(w,\zeta|\ell)\underline{\underline{\hat{S}}}_2^+(w|\ell) + \underline{\underline{\hat{T}}}_1(\zeta,w|\ell)\underline{\underline{\hat{S}}}_2^-(\zeta|\ell) \\ &\quad + \underline{\underline{\hat{R}}}_2(w,\zeta|\ell)\underline{\underline{\hat{S}}}_1^+(w|\ell) + \underline{\underline{\hat{T}}}_2(\zeta,w|\ell)\underline{\underline{\hat{S}}}_1^-(\zeta|\ell) + \underline{\underline{\hat{\Sigma}}}_2^{-t}(\zeta,w|\ell) \qquad (1.79)\\ \underline{\underline{\hat{S}}}_2^+(\zeta|\ell) &= \underline{\underline{\hat{R}}}_1(\zeta,w|\ell)\underline{\underline{\hat{S}}}_2^-(\zeta|\ell) + \underline{\underline{\hat{T}}}_1(w,\zeta|\ell)\underline{\underline{\hat{S}}}_2^+(w|\ell) \\ &\quad + \underline{\underline{\hat{R}}}_2(\zeta,w|\ell)\underline{\underline{\hat{S}}}_1^-(\zeta|\ell) + \underline{\underline{\hat{T}}}_2(w,\zeta|\ell)\underline{\underline{\hat{S}}}_1^+(w|\ell) + \underline{\underline{\hat{\Sigma}}}_2^{+t}(w,\zeta|\ell) \qquad (1.80)\end{aligned}$$

Note that these equations give the response radiances leaving the top and bottom of the slab as functions of the incident radiances. Just as with $\underline{\underline{\hat{\rho}}}_{1,2}$ and $\underline{\underline{\hat{\tau}}}_{1,2}$ in the local interaction equations, the global interaction cosine and sine equations share a common set of standard reflectance and transmittance matrices, but the same matrix operates on difference radiance amplitudes in the cosine and sine pairs of equations.

A tally of the "information content" of the local and global interaction equations is worthwhile. The local interaction Eqs. (1.73)–(1.76) involve 4 matrices of size $4M \times 4M$: $\underline{\underline{\hat{\rho}}}_{1,2}(\zeta|\ell)$ and $\underline{\underline{\hat{\tau}}}_{1,2}(\zeta|\ell)$. However, the corresponding global interaction Eqs. (1.77)–(1.80) involve 8 matrices of size $4M \times 4M$: $\underline{\underline{\hat{R}}}_1(w,\zeta|\ell)$,..., $\underline{\underline{\hat{T}}}_2(\zeta,w|\ell)$. Thus it seems that the global equations some way contain twice as much information as the local equations. This is indeed correct. The local equations contain information about only one depth, ζ. The global equations on the other hand contain information about two depths, w and ζ. Another way to view this is that an infinitesimally thin layer of water reflects and transmits radiance the same for radiance incident from either above or below. However, the IOPs generally depend on depth, in which case finitely thick layers of water do not reflect downwelling radiance from above the layer the same as upwelling radiance from below the layer (ditto for transmission of downwelling vs upwelling radiance). Thus $\underline{\underline{\hat{R}}}_1(w,\zeta|\ell) \neq \underline{\underline{\hat{R}}}_1(\zeta,w|\ell)$, and so on. For finitely thick layers, reflectance and transmittance depend on the direction of the incident light, and there are thus twice as many reflectance and transmittance functions as for an infinitesimal layer.

In one viewpoint, Eqs. (1.77)–(1.80) can be regarded as *definitions* of the standard operators, although they can be derived from the local interaction equations as is done in *Light and Water* Sect. 8.6. In any case, the linear interaction principle guarantees the existence of these quantities.

The lower slab $[\zeta, m]$ has a corresponding set of layer standard reflectances and transmittances. These matrices convert radiances incident onto the slab from above at depth ζ and below at depth m into radiances exiting the slab at ζ and m. These operations are illustrated in Fig. 1.9. Note that $\underline{\underline{\hat{S}}}_1^+(\zeta|\ell)$ was a response radiance for slab $[w, \zeta]$, but it is an incident radiance for slab $[\zeta, m]$, etc. The corresponding global interaction equations are

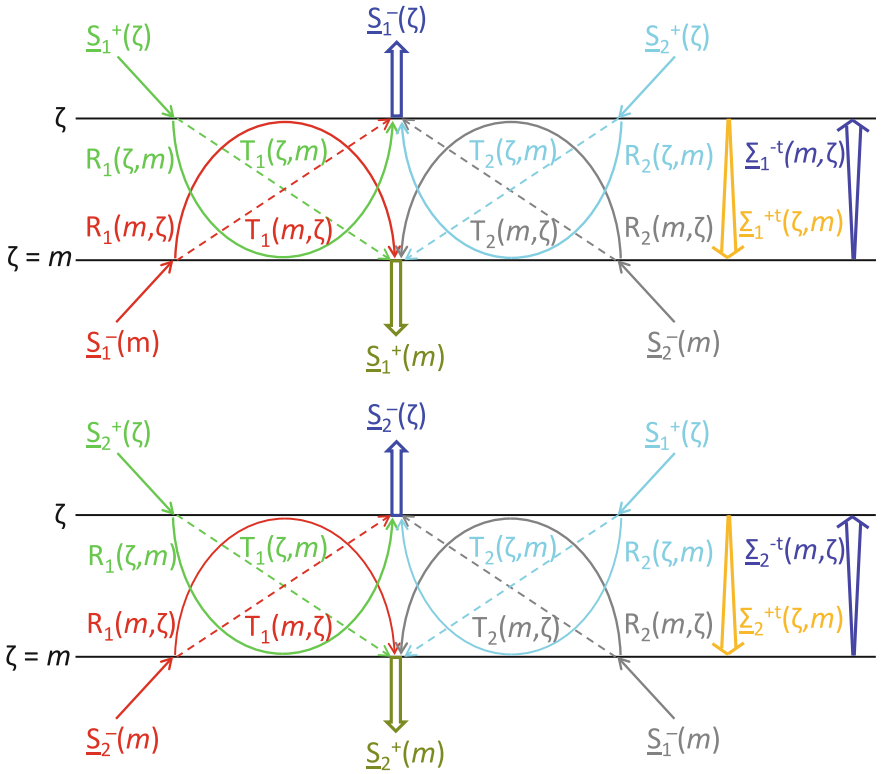

Fig. 1.9 Global interaction principles for the slab $[\zeta, m]$

$$\hat{\underline{\underline{S}}}_1^-(\zeta|\ell) = \hat{\underline{\underline{R}}}_1(\zeta, m|\ell)\hat{\underline{\underline{S}}}_1^+(\zeta|\ell) + \hat{\underline{\underline{T}}}_1(m, \zeta|\ell)\hat{\underline{\underline{S}}}_1^-(m|\ell)$$
$$- \hat{\underline{\underline{R}}}_2(\zeta, m|\ell)\hat{\underline{\underline{S}}}_2^+(\zeta|\ell) - \hat{\underline{\underline{T}}}_2(m, \zeta|\ell)\hat{\underline{\underline{S}}}_2^-(m|\ell) + \hat{\underline{\underline{\Sigma}}}_1^{-t}(m, \zeta|\ell) \quad (1.81)$$

$$\hat{\underline{\underline{S}}}_1^+(m|\ell) = \hat{\underline{\underline{R}}}_1(m, \zeta|\ell)\hat{\underline{\underline{S}}}_1^-(m|\ell) + \hat{\underline{\underline{T}}}_1(\zeta, m|\ell)\hat{\underline{\underline{S}}}_1^+(\zeta|\ell)$$
$$- \hat{\underline{\underline{R}}}_2(m, \zeta|\ell)\hat{\underline{\underline{S}}}_2^-(m|\ell) - \hat{\underline{\underline{T}}}_2(\zeta, m|\ell)\hat{\underline{\underline{S}}}_2^+(\zeta|\ell) + \hat{\underline{\underline{\Sigma}}}_1^{+t}(\zeta, m|\ell) \quad (1.82)$$

$$\hat{\underline{\underline{S}}}_2^-(\zeta|\ell) = \hat{\underline{\underline{R}}}_1(\zeta, m|\ell)\hat{\underline{\underline{S}}}_2^+(\zeta|\ell) + \hat{\underline{\underline{T}}}_1(m, \zeta|\ell)\hat{\underline{\underline{S}}}_2^-(m|\ell)$$
$$+ \hat{\underline{\underline{R}}}_2(\zeta, m|\ell)\hat{\underline{\underline{S}}}_1^+(\zeta|\ell) + \hat{\underline{\underline{T}}}_2(m, \zeta|\ell)\hat{\underline{\underline{S}}}_1^-(m|\ell) + \hat{\underline{\underline{\Sigma}}}_2^{-t}(m, \zeta|\ell) \quad (1.83)$$

$$\hat{\underline{\underline{S}}}_2^+(m|\ell) = \hat{\underline{\underline{R}}}_1(m, \zeta|\ell)\hat{\underline{\underline{S}}}_2^-(m|\ell) + \hat{\underline{\underline{T}}}_1(\zeta, m|\ell)\hat{\underline{\underline{S}}}_2^+(\zeta|\ell)$$
$$+ \hat{\underline{\underline{R}}}_2(m, \zeta|\ell)\hat{\underline{\underline{S}}}_1^-(m|\ell) + \hat{\underline{\underline{T}}}_2(\zeta, m|\ell)\hat{\underline{\underline{S}}}_1^+(\zeta|\ell) + \hat{\underline{\underline{\Sigma}}}_2^{+t}(\zeta, m|\ell) \quad (1.84)$$

1.5.5 *Differential Equations for the Standard Matrices*

The standard matrices describe how radiance is reflected and transmitted by a layer of water, which in turn is determined by the water absorption and scattering properties. *The standard matrices are therefore spatially integrated inherent optical properties.* The interaction principle guarantees the existence of the standard matrices. Unfortunately, it does not specify how to compute these quantities. The next step in the solution algorithm development is therefore to develop equations that can be solved to obtain the standard matrices, given the water-column IOPs. This is done as follows:

1. The global interaction equations are differentiated with respect to depth.
2. The resulting depth derivatives of the radiance are replaced using the local interaction equations.
3. The global interaction equations are used to replace response radiances by incident radiances.
4. The terms of the resulting equations are grouped into factors multiplying the incident radiances, and a group of terms involving the internal sources.
5. The arbitrariness of the incident radiances and internal sources is used to obtain sets of differential equations for the standard matrices.

1.5.5.1 Downward Sweep Equations

To illustrate this process, consider, for example, $\underline{\underline{\hat{R}}}_1(\zeta, w|\ell)$. In this development, the Fourier mode index ℓ is omitted for brevity. This matrix occurs in Eq. (1.78). Differentiating that equation with respect to depth gives

$$\begin{aligned}\frac{d}{d\zeta}\underline{\underline{\hat{S}}}_1^+(\zeta) = &\left[\frac{d}{d\zeta}\underline{\underline{\hat{R}}}_1(\zeta, w)\right]\underline{\underline{\hat{S}}}_1^-(\zeta) + \underline{\underline{\hat{R}}}_1(\zeta, w)\frac{d}{d\zeta}\underline{\underline{\hat{S}}}_1^-(\zeta) + \left[\frac{d}{d\zeta}\underline{\underline{\hat{T}}}_1(w, \zeta)\right]\underline{\underline{\hat{S}}}_1^+(w)\\ &- \left[\frac{d}{d\zeta}\underline{\underline{\hat{R}}}_2(\zeta, w)\right]\underline{\underline{\hat{S}}}_2^-(\zeta) - \underline{\underline{\hat{R}}}_2(\zeta, w)\frac{d}{d\zeta}\underline{\underline{\hat{S}}}_2^-(\zeta) - \left[\frac{d}{d\zeta}\underline{\underline{\hat{T}}}_2(w, \zeta)\right]\underline{\underline{\hat{S}}}_2^+(w)\\ &+ \frac{d}{d\zeta}\underline{\underline{\hat{\Sigma}}}_1^{+t}(w, \zeta)\,. \end{aligned} \tag{1.85}$$

Note that $\underline{\underline{\hat{S}}}_{1,2}^+(w)$ is the incident radiance at a fixed location w; it therefore does not have a depth derivative. Quantities such as $\underline{\underline{\hat{S}}}_{1,2}^{\pm}(\zeta)$, on the other hand, have depth derivatives because the lower boundary depth of slab $[w, \zeta]$ is allowed to vary.

The local interaction Eqs. (1.73), (1.74), and (1.76) are next used to replace the three depth derivatives of the radiances in the preceding equation. The result is

$$
\begin{aligned}
&\underline{\underline{\hat{\rho}}}_1\underline{\underline{\hat{S}}}_1^-(\zeta)+\underline{\underline{\hat{\tau}}}_1\underline{\underline{\hat{S}}}_1^+(\zeta)-\underline{\underline{\hat{\rho}}}_2\underline{\underline{\hat{S}}}_2^-(\zeta)-\underline{\underline{\hat{\tau}}}_2\underline{\underline{\hat{S}}}_2^+(\zeta)+\underline{\underline{\hat{\Sigma}}}_1^+(\zeta)\\
&\quad=\left[\frac{d}{d\zeta}\underline{\underline{\hat{R}}}_1(\zeta,w)\right]\underline{\underline{\hat{S}}}_1^-(\zeta)\\
&\quad+\underline{\underline{\hat{R}}}_1(\zeta,w)\left\{-\underline{\underline{\hat{\rho}}}_1\underline{\underline{\hat{S}}}_1^+(\zeta)-\underline{\underline{\hat{\tau}}}_1\underline{\underline{\hat{S}}}_1^-(\zeta)+\underline{\underline{\hat{\rho}}}_2\underline{\underline{\hat{S}}}_2^+(\zeta)+\underline{\underline{\hat{\tau}}}_2\underline{\underline{\hat{S}}}_2^-(\zeta)-\underline{\underline{\hat{\Sigma}}}_1^-(\zeta)\right\}\\
&\quad+\left[\frac{d}{d\zeta}\underline{\underline{\hat{T}}}_1(w,\zeta)\right]\underline{\underline{\hat{S}}}_1^+(w)-\left[\frac{d}{d\zeta}\underline{\underline{\hat{R}}}_2(\zeta,w)\right]\underline{\underline{\hat{S}}}_2^-(\zeta)\\
&\quad-\underline{\underline{\hat{R}}}_2(\zeta,w)\left\{-\underline{\underline{\hat{\rho}}}_1\underline{\underline{\hat{S}}}_2^+(\zeta)-\underline{\underline{\hat{\tau}}}_1\underline{\underline{\hat{S}}}_2^-(\zeta)-\underline{\underline{\hat{\rho}}}_2\underline{\underline{\hat{S}}}_1^+(\zeta)-\underline{\underline{\hat{\tau}}}_2\underline{\underline{\hat{S}}}_1^-(\zeta)-\underline{\underline{\hat{\Sigma}}}_2^-(\zeta)\right\}\\
&\quad-\left[\frac{d}{d\zeta}\underline{\underline{\hat{T}}}_2(w,\zeta)\right]\underline{\underline{\hat{S}}}_2^+(w)+\frac{d}{d\zeta}\underline{\underline{\hat{\Sigma}}}_1^{+t}(w,\zeta)
\end{aligned}
\tag{1.86}
$$

The common $(\zeta|\ell)$ arguments of the $\underline{\underline{\hat{\rho}}}_{1,2}$ and $\underline{\underline{\hat{\tau}}}_{1,2}$ matrices seen in Eqs. (1.73)–(1.76) have been omitted for brevity. However, the depth argument must be retained on the radiances to identify which radiances are incident and which are responses.

The response radiances $\underline{\underline{\hat{S}}}_{1,2}^+(\zeta)$ in this last equation are now replaced with incident radiances using the global interaction Eqs. (1.78) and (1.80). The terms are then grouped into factors multiplying the incident radiances $\underline{\underline{\hat{S}}}_{1,2}^+(w)$ and $\underline{\underline{\hat{S}}}_{1,2}^-(\zeta)$, plus the remaining terms involving the internal sources. The resulting equation is

$$
\begin{aligned}
0=&\left\{\frac{d}{d\zeta}\underline{\underline{\hat{R}}}_1(\zeta,w)-\underline{\underline{\hat{\rho}}}_1-\underline{\underline{\hat{R}}}_1(\zeta,w)\underline{\underline{\hat{\tau}}}_1+\underline{\underline{\hat{R}}}_2(\zeta,w)\underline{\underline{\hat{\tau}}}_2\right.\\
&\qquad-\left[\underline{\underline{\hat{\tau}}}_1+\underline{\underline{\hat{R}}}_1(\zeta,w)\underline{\underline{\hat{\rho}}}_1-\underline{\underline{\hat{R}}}_2(\zeta,w)\underline{\underline{\hat{\rho}}}_2\right]\underline{\underline{\hat{R}}}_1(\zeta,w)\\
&\qquad\left.+\left[\underline{\underline{\hat{\tau}}}_2+\underline{\underline{\hat{R}}}_1(\zeta,w)\underline{\underline{\hat{\rho}}}_2+\underline{\underline{\hat{R}}}_2(\zeta,w)\underline{\underline{\hat{\rho}}}_1\right]\underline{\underline{\hat{R}}}_2(\zeta,w)\right\}\underline{\underline{\hat{S}}}_1^-(\zeta)\\
+&\left\{-\frac{d}{d\zeta}\underline{\underline{\hat{R}}}_2(\zeta,w)+\underline{\underline{\hat{\rho}}}_2+\underline{\underline{\hat{R}}}_1(\zeta,w)\underline{\underline{\hat{\tau}}}_2+\underline{\underline{\hat{R}}}_2(\zeta,w)\underline{\underline{\hat{\tau}}}_1\right.\\
&\qquad+\left[\underline{\underline{\hat{\tau}}}_1+\underline{\underline{\hat{R}}}_1(\zeta,w)\underline{\underline{\hat{\rho}}}_1-\underline{\underline{\hat{R}}}_2(\zeta,w)\underline{\underline{\hat{\rho}}}_2\right]\underline{\underline{\hat{R}}}_2(\zeta,w)\\
&\qquad\left.+\left[\underline{\underline{\hat{\tau}}}_2+\underline{\underline{\hat{R}}}_1(\zeta,w)\underline{\underline{\hat{\rho}}}_2+\underline{\underline{\hat{R}}}_2(\zeta,w)\underline{\underline{\hat{\rho}}}_1\right]\underline{\underline{\hat{R}}}_1(\zeta,w)\right\}\underline{\underline{\hat{S}}}_2^-(\zeta)\\
&\left\{\frac{d}{d\zeta}\underline{\underline{\hat{T}}}_1(w,\zeta)-\left[\underline{\underline{\hat{\tau}}}_1+\underline{\underline{\hat{R}}}_1(\zeta,w)\underline{\underline{\hat{\rho}}}_1-\underline{\underline{\hat{R}}}_2(\zeta,w)\underline{\underline{\hat{\rho}}}_2\right]\underline{\underline{\hat{T}}}_1(w,\zeta)\right.\\
&\qquad\left.+\left[\underline{\underline{\hat{\tau}}}_2+\underline{\underline{\hat{R}}}_1(\zeta,w)\underline{\underline{\hat{\rho}}}_2+\underline{\underline{\hat{R}}}_2(\zeta,w)\underline{\underline{\hat{\rho}}}_1\right]\underline{\underline{\hat{T}}}_2(w,\zeta)\right\}\underline{\underline{\hat{S}}}_1^+(w)\\
+&\left\{-\frac{d}{d\zeta}\underline{\underline{\hat{T}}}_2(w,\zeta)+\left[\underline{\underline{\hat{\tau}}}_1+\underline{\underline{\hat{R}}}_1(\zeta,w)\underline{\underline{\hat{\rho}}}_1-\underline{\underline{\hat{R}}}_2(\zeta,w)\underline{\underline{\hat{\rho}}}_2\right]\underline{\underline{\hat{T}}}_2(w,\zeta)\right.\\
&\qquad\left.+\left[\underline{\underline{\hat{\tau}}}_2+\underline{\underline{\hat{R}}}_1(\zeta,w)\underline{\underline{\hat{\rho}}}_2+\underline{\underline{\hat{R}}}_2(\zeta,w)\underline{\underline{\hat{\rho}}}_1\right]\underline{\underline{\hat{T}}}_1(w,\zeta)\right\}\underline{\underline{\hat{S}}}_2^+(w)
\end{aligned}
$$

$$+\left\{\frac{d}{d\zeta}\hat{\underline{\underline{\Sigma}}}_1^{+t}(w,\zeta)-\hat{\underline{\underline{\Sigma}}}_1^{+}(\zeta)-\hat{\underline{\underline{R}}}_1(\zeta,w)\hat{\underline{\underline{\Sigma}}}_1^{-}(\zeta)+\hat{\underline{\underline{R}}}_2(\zeta,w)\hat{\underline{\underline{\Sigma}}}_2^{-}(\zeta)\right.$$
$$-\left[\hat{\underline{\tau}}_1+\hat{\underline{\underline{R}}}_1(\zeta,w)\hat{\underline{\rho}}_1-\hat{\underline{\underline{R}}}_2(\zeta,w)\hat{\underline{\rho}}_2\right]\hat{\underline{\underline{\Sigma}}}_1^{+t}(w,\zeta)$$
$$\left.+\left[\hat{\underline{\tau}}_2+\hat{\underline{\underline{R}}}_1(\zeta,w)\hat{\underline{\rho}}_2+\hat{\underline{\underline{R}}}_2(\zeta,w)\hat{\underline{\rho}}_1\right]\hat{\underline{\underline{\Sigma}}}_2^{+t}(w,\zeta)\right\} \tag{1.87}$$

The first four groups of terms in brackets multiply incident radiances. The last group of terms contains the internal sources.

The four incident radiances and the internal sources have arbitrary values that must be specified as known inputs to the radiative transfer problem. For example, all of the input radiances except for $\hat{\underline{\underline{S}}}_1^{+}(w)$ can be zero, and all of the internal sources can be zero. In that case, the terms in brackets multiplying $\hat{\underline{\underline{S}}}_1^{+}(w)$ must sum to zero. Similarly, all of the incident radiances can be zero, but there can be internal sources generating the light field within the water. In that case, the last group of bracketed terms must sum to zero. Arguments of this type show that each group of terms in brackets must be zero. Setting the individual groups of bracketed terms to zero then gives a set of five coupled nonlinear ordinary differential equations (ODEs) (L&W 8.74, 8.75, and 8.78):

$$\frac{d}{d\zeta}\hat{\underline{\underline{R}}}_1(\zeta,w)=\hat{\underline{\rho}}_1+\hat{\underline{\underline{R}}}_1(\zeta,w)\hat{\underline{\tau}}_1-\hat{\underline{\underline{R}}}_2(\zeta,w)\hat{\underline{\tau}}_2$$
$$+\left[\hat{\underline{\tau}}_1+\hat{\underline{\underline{R}}}_1(\zeta,w)\hat{\underline{\rho}}_1-\hat{\underline{\underline{R}}}_2(\zeta,w)\hat{\underline{\rho}}_2\right]\hat{\underline{\underline{R}}}_1(\zeta,w)$$
$$-\left[\hat{\underline{\tau}}_2+\hat{\underline{\underline{R}}}_1(\zeta,w)\hat{\underline{\rho}}_2+\hat{\underline{\underline{R}}}_2(\zeta,w)\hat{\underline{\rho}}_1\right]\hat{\underline{\underline{R}}}_2(\zeta,w) \tag{1.88}$$

$$\frac{d}{d\zeta}\hat{\underline{\underline{R}}}_2(\zeta,w)=\hat{\underline{\rho}}_2+\hat{\underline{\underline{R}}}_1(\zeta,w)\hat{\underline{\tau}}_2+\hat{\underline{\underline{R}}}_2(\zeta,w)\hat{\underline{\tau}}_1$$
$$+\left[\hat{\underline{\tau}}_1+\hat{\underline{\underline{R}}}_1(\zeta,w)\hat{\underline{\rho}}_1-\hat{\underline{\underline{R}}}_2(\zeta,w)\hat{\underline{\rho}}_2\right]\hat{\underline{\underline{R}}}_2(\zeta,w)$$
$$+\left[\hat{\underline{\tau}}_2+\hat{\underline{\underline{R}}}_1(\zeta,w)\hat{\underline{\rho}}_2+\hat{\underline{\underline{R}}}_2(\zeta,w)\hat{\underline{\rho}}_1\right]\hat{\underline{\underline{R}}}_1(\zeta,w) \tag{1.89}$$

$$\frac{d}{d\zeta}\hat{\underline{T}}_1(w,\zeta)=\left[\hat{\underline{\tau}}_1+\hat{\underline{\underline{R}}}_1(\zeta,w)\hat{\underline{\rho}}_1-\hat{\underline{\underline{R}}}_2(\zeta,w)\hat{\underline{\rho}}_2\right]\hat{\underline{T}}_1(w,\zeta)$$
$$-\left[\hat{\underline{\tau}}_2+\hat{\underline{\underline{R}}}_1(\zeta,w)\hat{\underline{\rho}}_2+\hat{\underline{\underline{R}}}_2(\zeta,w)\hat{\underline{\rho}}_1\right]\hat{\underline{T}}_2(w,\zeta) \tag{1.90}$$

$$\frac{d}{d\zeta}\hat{\underline{T}}_2(w,\zeta)=\left[\hat{\underline{\tau}}_1+\hat{\underline{\underline{R}}}_1(\zeta,w)\hat{\underline{\rho}}_1-\hat{\underline{\underline{R}}}_2(\zeta,w)\hat{\underline{\rho}}_2\right]\hat{\underline{T}}_2(w,\zeta)$$
$$+\left[\hat{\underline{\tau}}_2+\hat{\underline{\underline{R}}}_1(\zeta,w)\hat{\underline{\rho}}_2+\hat{\underline{\underline{R}}}_2(\zeta,w)\hat{\underline{\rho}}_1\right]\hat{\underline{T}}_1(w,\zeta) \tag{1.91}$$

$$\begin{aligned}\frac{d}{d\zeta}\hat{\underline{\underline{\Sigma}}}_1^{+t}(w,\zeta) &= \hat{\underline{\underline{\Sigma}}}_1^{+}(\zeta) + \hat{\underline{\underline{R}}}_1(\zeta,w)\hat{\underline{\underline{\Sigma}}}_1^{-}(\zeta) - \hat{\underline{\underline{R}}}_2(\zeta,w)\hat{\underline{\underline{\Sigma}}}_2^{-}(\zeta) \\ &+ \left[\hat{\underline{\underline{\tau}}}_1 + \hat{\underline{\underline{R}}}_1(\zeta,w)\hat{\underline{\underline{\rho}}}_1 - \hat{\underline{\underline{R}}}_2(\zeta,w)\hat{\underline{\underline{\rho}}}_2\right]\hat{\underline{\underline{\Sigma}}}_1^{+t}(w,\zeta) \\ &- \left[\hat{\underline{\underline{\tau}}}_2 + \hat{\underline{\underline{R}}}_1(\zeta,w)\hat{\underline{\underline{\rho}}}_2 + \hat{\underline{\underline{R}}}_2(\zeta,w)\hat{\underline{\underline{\rho}}}_1\right]\hat{\underline{\underline{\Sigma}}}_2^{+t}(w,\zeta) \qquad (1.92)\end{aligned}$$

This self-contained set of ODEs can be solved given the IOPs $\hat{\underline{\underline{\rho}}}_{1,2}(\zeta)$, $\hat{\underline{\underline{\tau}}}_{1,2}(\zeta)$, and initial conditions. The initial conditions are deduced as follows. If $\zeta \to w = 0$, there will be no reflectance by a slab $[w, w]$ of zero thickness, and all of the radiance will be transmitted. Likewise, there will be no contribution by internal sources in a slab of zero thickness. Thus the initial conditions at $\zeta = w$ are simply

$$\hat{\underline{\underline{R}}}_1(w,w) = \hat{\underline{\underline{R}}}_2(w,w) = \underline{\underline{0}}_{4M\times 4M} \qquad (1.93)$$

$$\hat{\underline{\underline{T}}}_1(w,w) = \hat{\underline{\underline{T}}}_2(w,w) = \underline{\underline{I}}_{4M\times 4M} \qquad (1.94)$$

$$\hat{\underline{\underline{\Sigma}}}_1^{+t}(w,w) = \underline{0}_{4M\times 1} \qquad (1.95)$$

where $\underline{\underline{0}}_{4M\times 4M}$ is a $4M \times 4M$ matrix of zeros, $\underline{\underline{I}}_{4M\times 4M}$ is a $4M \times 4M$ identity matrix, and $\underline{0}_{4M\times 1}$ is a $4M \times 1$ column vector of zeros.

Equations (1.88)–(1.92) can therefore be integrated in a "downward sweep" from depth $w = 0$ to any depth ζ, starting with the initial conditions of Eqs. (1.93)–(1.95).

It should be noted that the reflectance and transmittance properties of the slab $[w, \zeta]$ *can be computed without explicit consideration of the air-water surface* $[a, w]$. Slab $[w, \zeta]$ is therefore referred to as a "bare" slab, i.e., a water body without an air-water surface. This is physically equivalent to a water body with the same index of refraction as the air, which would not reflect or refract light passing though such a surface. The effects of the actual air-sea surface will be incorporated into the solution in the next section.

Differentiating global interaction Eq. (1.77) and repeating this process gives the following set of equations (L&W 8.76, 8.77, 8.78):

$$\begin{aligned}\frac{d}{d\zeta}\hat{\underline{\underline{R}}}_1(w,\zeta) &= \left[\hat{\underline{\underline{T}}}_1(\zeta,w)\hat{\underline{\underline{\rho}}}_1 - \hat{\underline{\underline{T}}}_2(\zeta,w)\hat{\underline{\underline{\rho}}}_2\right]\hat{\underline{\underline{T}}}_1(w,\zeta) \\ &- \left[\hat{\underline{\underline{T}}}_1(\zeta,w)\hat{\underline{\underline{\rho}}}_2 + \hat{\underline{\underline{T}}}_2(\zeta,w)\hat{\underline{\underline{\rho}}}_1\right]\hat{\underline{\underline{T}}}_2(w,\zeta) \qquad (1.96) \\ \frac{d}{d\zeta}\hat{\underline{\underline{R}}}_2(w,\zeta) &= \left[\hat{\underline{\underline{T}}}_1(\zeta,w)\hat{\underline{\underline{\rho}}}_1 - \hat{\underline{\underline{T}}}_2(\zeta,w)\hat{\underline{\underline{\rho}}}_2\right]\hat{\underline{\underline{T}}}_2(w,\zeta) \\ &+ \left[\hat{\underline{\underline{T}}}_1(\zeta,w)\hat{\underline{\underline{\rho}}}_2 + \hat{\underline{\underline{T}}}_2(\zeta,w)\hat{\underline{\underline{\rho}}}_1\right]\hat{\underline{\underline{T}}}_1(w,\zeta) \qquad (1.97)\end{aligned}$$

$$\begin{aligned}\frac{d}{d\zeta}\underline{\underline{\hat{T}}}_1(\zeta,w) &= \underline{\underline{\hat{T}}}_1(\zeta,w)\underline{\underline{\hat{\tau}}}_1 - \underline{\underline{\hat{T}}}_2(\zeta,w)\underline{\underline{\hat{\tau}}}_2 \\ &+ \left[\underline{\underline{\hat{T}}}_1(\zeta,w)\underline{\underline{\hat{\rho}}}_1 - \underline{\underline{\hat{T}}}_2(\zeta,w)\underline{\underline{\hat{\rho}}}_2\right]\underline{\underline{\hat{R}}}_1(\zeta,w) \\ &- \left[\underline{\underline{\hat{T}}}_1(\zeta,w)\underline{\underline{\hat{\rho}}}_2 + \underline{\underline{\hat{T}}}_2(\zeta,w)\underline{\underline{\hat{\rho}}}_1\right]\underline{\underline{\hat{R}}}_2(\zeta,w) \qquad (1.98)\end{aligned}$$

$$\begin{aligned}\frac{d}{d\zeta}\underline{\underline{\hat{T}}}_2(\zeta,w) &= \underline{\underline{\hat{T}}}_1(\zeta,w)\underline{\underline{\hat{\tau}}}_2 + \underline{\underline{\hat{T}}}_2(\zeta,w)\underline{\underline{\hat{\tau}}}_1 \\ &+ \left[\underline{\underline{\hat{T}}}_1(\zeta,w)\underline{\underline{\hat{\rho}}}_1 - \underline{\underline{\hat{T}}}_2(\zeta,w)\underline{\underline{\hat{\rho}}}_2\right]\underline{\underline{\hat{R}}}_2(\zeta,w) \\ &+ \left[\underline{\underline{\hat{T}}}_1(\zeta,w)\underline{\underline{\hat{\rho}}}_2 + \underline{\underline{\hat{T}}}_2(\zeta,w)\underline{\underline{\hat{\rho}}}_1\right]\underline{\underline{\hat{R}}}_1(\zeta,w) \qquad (1.99)\end{aligned}$$

$$\begin{aligned}\frac{d}{d\zeta}\underline{\underline{\hat{\Sigma}}}_1^{-t}(\zeta,w) &= \underline{\underline{\hat{T}}}_1(\zeta,w)\underline{\underline{\hat{\Sigma}}}_1^{-}(\zeta) - \underline{\underline{\hat{T}}}_2(\zeta,w)\underline{\underline{\hat{\Sigma}}}_2^{-}(\zeta)) \\ &+ \left[\underline{\underline{\hat{T}}}_1(\zeta,w)\underline{\underline{\hat{\rho}}}_1 - \underline{\underline{\hat{T}}}_2(\zeta,w)\underline{\underline{\hat{\rho}}}_2\right]\underline{\underline{\hat{\Sigma}}}_1^{+t}(w,\zeta) \\ &- \left[\underline{\underline{\hat{T}}}_1(\zeta,w)\underline{\underline{\hat{\rho}}}_2 + \underline{\underline{\hat{T}}}_2(\zeta,w)\underline{\underline{\hat{\rho}}}_1\right]\underline{\underline{\hat{\Sigma}}}_2^{+t}(w,\zeta) \qquad (1.100)\end{aligned}$$

The preceding development gives ODEs for all of the reflectance and transmittance matrices seen in Fig. 1.8, and for the cosine amplitudes of the internal source terms. Equations for the remaining internal source sine amplitudes can be obtained from Eqs. (1.79) and (1.80). Differentiating Eq. (1.79) and repeating the above process gives the equation for $\underline{\underline{\hat{\Sigma}}}_2^{-t}(\zeta,w)$; starting with Eq. (1.80) gives the ODE for $\underline{\underline{\hat{\Sigma}}}_2^{+t}(\zeta,w)$. Those derivations also re-derive the equations for $\underline{\underline{R}}_{1,2}(\zeta,w)$, $\underline{\underline{T}}_{1,2}(w,\zeta)$, $\underline{\underline{R}}_{1,2}(w,\zeta)$ and $\underline{\underline{T}}_{1,2}(\zeta,w)$, which serves as a check on those equations. The results for the remaining source terms are

$$\begin{aligned}\frac{d}{d\zeta}\underline{\underline{\hat{\Sigma}}}_2^{-t}(\zeta,w) &= \underline{\underline{\hat{T}}}_1(\zeta,w)\underline{\underline{\hat{\Sigma}}}_2^{-}(\zeta) + \underline{\underline{\hat{T}}}_2(\zeta,w)\underline{\underline{\hat{\Sigma}}}_1^{-}(\zeta) \\ &+ \left[\underline{\underline{\hat{T}}}_1(\zeta,w)\underline{\underline{\hat{\rho}}}_2 + \underline{\underline{\hat{T}}}_2(\zeta,w)\underline{\underline{\hat{\rho}}}_1\right]\underline{\underline{\hat{\Sigma}}}_1^{+t}(w,\zeta) \\ &+ \left[\underline{\underline{\hat{T}}}_1(\zeta,w)\underline{\underline{\hat{\rho}}}_1 - \underline{\underline{\hat{T}}}_2(\zeta,w)\underline{\underline{\hat{\rho}}}_2\right]\underline{\underline{\hat{\Sigma}}}_2^{+t}(w,\zeta) \qquad (1.101)\end{aligned}$$

$$\begin{aligned}\frac{d}{d\zeta}\underline{\underline{\hat{\Sigma}}}_2^{+t}(\zeta,w) &= \underline{\underline{\hat{\Sigma}}}_2^{+}(\zeta) + \underline{\underline{\hat{R}}}_1(\zeta,w)\underline{\underline{\hat{\Sigma}}}_2^{-}(\zeta) + \underline{\underline{\hat{R}}}_2(\zeta,w)\underline{\underline{\hat{\Sigma}}}_1^{-}(\zeta)) \\ &+ \left[\underline{\underline{\hat{\tau}}}_2 + \underline{\underline{\hat{R}}}_1(\zeta,w)\underline{\underline{\hat{\rho}}}_2 + \underline{\underline{\hat{R}}}_2(\zeta,w)\underline{\underline{\hat{\rho}}}_1\right]\underline{\underline{\hat{\Sigma}}}_1^{+t}(w,\zeta) \\ &+ \left[\underline{\underline{\hat{\tau}}}_1 + \underline{\underline{\hat{R}}}_1(\zeta,w)\underline{\underline{\hat{\rho}}}_1 - \underline{\underline{\hat{R}}}_2(\zeta,w)\underline{\underline{\hat{\rho}}}_2\right]\underline{\underline{\hat{\Sigma}}}_1^{+t}(w,\zeta) \qquad (1.102)\end{aligned}$$

These equations can be integrated from $\zeta = w$ downward, starting with the initial conditions

$$\underline{\underline{\hat{R}}}_1(w,w) = \underline{\underline{\hat{R}}}_2(w,w) = \underline{\underline{0}}_{4M\times 4M} \tag{1.103}$$

$$\underline{\underline{\hat{T}}}_1(w,w) = \underline{\underline{\hat{T}}}_2(w,w) = \underline{\underline{I}}_{4M\times 4M} \tag{1.104}$$

$$\underline{\underline{\hat{\Sigma}}}_1^{-t}(w,w) = \underline{\underline{0}}_{4M\times 1} \tag{1.105}$$

$$\underline{\underline{\hat{\Sigma}}}_2^{\pm t}(w,w) = \underline{\underline{0}}_{4M\times 1} \tag{1.106}$$

Note that the set of Eqs. (1.88)–(1.92) is self contained, whereas Eqs. (1.96)–(1.100) contain terms from the first set. The set (1.88)–(1.92) is therefore called the "major" set, and (1.96)–(1.100) the "minor" set. Equation (1.102) belongs to the major set, and (1.101) to the minor set. In any case, these equations must be solved simultaneously.

1.5.5.2 Upward Sweep Equations

The slab $[\zeta, m]$ illustrated in Fig. 1.9 has a corresponding set of ODEs for the standard matrices shown in the figure. The derivations follow the same procedure as for slab $[w, \zeta]$, so the details do not need to be presented. The resulting equations are as follows:

The major set, obtained starting with Eq. (1.81) and (1.83):

$$\begin{aligned}
-\frac{d}{d\zeta}\underline{\underline{\hat{R}}}_1(\zeta,m) &= \underline{\underline{\hat{\rho}}}_1 + \underline{\underline{\hat{R}}}_1(\zeta,m)\underline{\underline{\hat{\tau}}}_1 - \underline{\underline{\hat{R}}}_2(\zeta,m)\underline{\underline{\hat{\tau}}}_2 \\
&\quad + \left[\underline{\underline{\hat{\tau}}}_1 + \underline{\underline{\hat{R}}}_1(\zeta,m)\underline{\underline{\hat{\rho}}}_1 - \underline{\underline{\hat{R}}}_2(\zeta,m)\underline{\underline{\hat{\rho}}}_2\right]\underline{\underline{\hat{R}}}_1(\zeta,m) \\
&\quad - \left[\underline{\underline{\hat{\tau}}}_2 + \underline{\underline{\hat{R}}}_1(\zeta,m)\underline{\underline{\hat{\rho}}}_2 + \underline{\underline{\hat{R}}}_2(\zeta,m)\underline{\underline{\hat{\rho}}}_1\right]\underline{\underline{\hat{R}}}_2(\zeta,m)
\end{aligned} \tag{1.107}$$

$$\begin{aligned}
-\frac{d}{d\zeta}\underline{\underline{\hat{R}}}_2(\zeta,m) &= \underline{\underline{\hat{\rho}}}_2 + \underline{\underline{\hat{R}}}_1(\zeta,m)\underline{\underline{\hat{\tau}}}_2 + \underline{\underline{\hat{R}}}_2(\zeta,m)\underline{\underline{\hat{\tau}}}_1 \\
&\quad + \left[\underline{\underline{\hat{\tau}}}_1 + \underline{\underline{\hat{R}}}_1(\zeta,m)\underline{\underline{\hat{\rho}}}_1 - \underline{\underline{\hat{R}}}_2(\zeta,m)\underline{\underline{\hat{\rho}}}_2\right]\underline{\underline{\hat{R}}}_2(\zeta,m) \\
&\quad + \left[\underline{\underline{\hat{\tau}}}_2 + \underline{\underline{\hat{R}}}_1(\zeta,m)\underline{\underline{\hat{\rho}}}_2 + \underline{\underline{\hat{R}}}_2(\zeta,m)\underline{\underline{\hat{\rho}}}_1\right]\underline{\underline{\hat{R}}}_1(\zeta,m)
\end{aligned} \tag{1.108}$$

$$\begin{aligned}
-\frac{d}{d\zeta}\underline{\underline{\hat{T}}}_1(m,\zeta) &= \left[\underline{\underline{\hat{\tau}}}_1 + \underline{\underline{\hat{R}}}_1(\zeta,m)\underline{\underline{\hat{\rho}}}_1 - \underline{\underline{\hat{R}}}_2(\zeta,m)\underline{\underline{\hat{\rho}}}_2\right]\underline{\underline{\hat{T}}}_1(m,\zeta) \\
&\quad - \left[\underline{\underline{\hat{\tau}}}_2 + \underline{\underline{\hat{R}}}_1(\zeta,m)\underline{\underline{\hat{\rho}}}_2 + \underline{\underline{\hat{R}}}_2(\zeta,m)\underline{\underline{\hat{\rho}}}_1\right]\underline{\underline{\hat{T}}}_2(m,\zeta)
\end{aligned} \tag{1.109}$$

$$\begin{aligned}
-\frac{d}{d\zeta}\underline{\underline{\hat{T}}}_2(m,\zeta) &= \left[\underline{\underline{\hat{\tau}}}_1 + \underline{\underline{\hat{R}}}_1(\zeta,m)\underline{\underline{\hat{\rho}}}_1 - \underline{\underline{\hat{R}}}_2(\zeta,m)\underline{\underline{\hat{\rho}}}_2\right]\underline{\underline{\hat{T}}}_2(m,\zeta) \\
&\quad + \left[\underline{\underline{\hat{\tau}}}_2 + \underline{\underline{\hat{R}}}_1(\zeta,m)\underline{\underline{\hat{\rho}}}_2 + \underline{\underline{\hat{R}}}_2(\zeta,m)\underline{\underline{\hat{\rho}}}_1\right]\underline{\underline{\hat{T}}}_1(m,\zeta)
\end{aligned} \tag{1.110}$$

$$
\begin{aligned}
-\frac{d}{d\zeta}\underline{\underline{\hat{\Sigma}}}_1^{-t}(m,\zeta) &= \underline{\underline{\hat{\Sigma}}}_1^{-}(\zeta) + \underline{\underline{\hat{R}}}_1(\zeta,m)\underline{\underline{\hat{\Sigma}}}_1^{+}(\zeta) - \underline{\underline{\hat{R}}}_2(\zeta,m)\underline{\underline{\hat{\Sigma}}}_2^{+}(\zeta) \\
&+ \left[\underline{\underline{\hat{\tau}}}_1 + \underline{\underline{\hat{R}}}_1(\zeta,m)\underline{\underline{\hat{\rho}}}_1 - \underline{\underline{\hat{R}}}_2(\zeta,m)\underline{\underline{\hat{\rho}}}_2\right]\underline{\underline{\hat{\Sigma}}}_1^{-t}(m,\zeta) \\
&- \left[\underline{\underline{\hat{\tau}}}_2 + \underline{\underline{\hat{R}}}_1(\zeta,m)\underline{\underline{\hat{\rho}}}_2 + \underline{\underline{\hat{R}}}_2(\zeta,m)\underline{\underline{\hat{\rho}}}_1\right]\underline{\underline{\hat{\Sigma}}}_2^{-t}(m,\zeta) \qquad (1.111)\\
-\frac{d}{d\zeta}\underline{\underline{\hat{\Sigma}}}_2^{-t}(m,\zeta) &= \underline{\underline{\hat{\Sigma}}}_2^{-}(\zeta) + \underline{\underline{\hat{R}}}_1(\zeta,m)\underline{\underline{\hat{\Sigma}}}_2^{+}(\zeta) + \underline{\underline{\hat{R}}}_2(\zeta,m)\underline{\underline{\hat{\Sigma}}}_1^{+}(\zeta)) \\
&+ \left[\underline{\underline{\hat{\tau}}}_1 + \underline{\underline{\hat{R}}}_1(\zeta,m)\underline{\underline{\hat{\rho}}}_1 - \underline{\underline{\hat{R}}}_2(\zeta,m)\underline{\underline{\hat{\rho}}}_2\right]\underline{\underline{\hat{\Sigma}}}_2^{-t}(m,\zeta) \\
&+ \left[\underline{\underline{\hat{\tau}}}_2 + \underline{\underline{\hat{R}}}_1(\zeta,m)\underline{\underline{\hat{\rho}}}_2 + \underline{\underline{\hat{R}}}_2(\zeta,m)\underline{\underline{\hat{\rho}}}_1\right]\underline{\underline{\hat{\Sigma}}}_1^{-t}(m,\zeta) \qquad (1.112)
\end{aligned}
$$

The minor set, obtained starting with Eqs. (1.82) and (1.84):

$$
\begin{aligned}
-\frac{d}{d\zeta}\underline{\underline{\hat{R}}}_1(m,\zeta) &= \left[\underline{\underline{\hat{T}}}_1(\zeta,m)\underline{\underline{\hat{\rho}}}_1 - \underline{\underline{\hat{T}}}_2(\zeta,m)\underline{\underline{\hat{\rho}}}_2\right]\underline{\underline{\hat{T}}}_1(m,\zeta) \\
&- \left[\underline{\underline{\hat{T}}}_1(\zeta,m)\underline{\underline{\hat{\rho}}}_2 + \underline{\underline{\hat{T}}}_2(\zeta,m)\underline{\underline{\hat{\rho}}}_1\right]\underline{\underline{\hat{T}}}_2(m,\zeta) \qquad (1.113)\\
-\frac{d}{d\zeta}\underline{\underline{\hat{R}}}_2(m,\zeta) &= \left[\underline{\underline{\hat{T}}}_1(\zeta,m)\underline{\underline{\hat{\rho}}}_1 - \underline{\underline{\hat{T}}}_2(\zeta,m)\underline{\underline{\hat{\rho}}}_2\right]\underline{\underline{\hat{T}}}_2(m,\zeta) \\
&+ \left[\underline{\underline{\hat{T}}}_1(\zeta,m)\underline{\underline{\hat{\rho}}}_2 + \underline{\underline{\hat{T}}}_2(\zeta,m)\underline{\underline{\hat{\rho}}}_1\right]\underline{\underline{\hat{T}}}_1(m,\zeta) \qquad (1.114)\\
-\frac{d}{d\zeta}\underline{\underline{\hat{T}}}_1(\zeta,m) &= \underline{\underline{\hat{T}}}_1(\zeta,m)\underline{\underline{\hat{\tau}}}_1 - \underline{\underline{\hat{T}}}_2(\zeta,m)\underline{\underline{\hat{\tau}}}_2 \\
&+ \left[\underline{\underline{\hat{T}}}_1(\zeta,m)\underline{\underline{\hat{\rho}}}_1 - \underline{\underline{\hat{T}}}_2(\zeta,m)\underline{\underline{\hat{\rho}}}_2\right]\underline{\underline{\hat{R}}}_1(\zeta,m) \\
&- \left[\underline{\underline{\hat{T}}}_1(\zeta,m)\underline{\underline{\hat{\rho}}}_2 + \underline{\underline{\hat{T}}}_2(\zeta,m)\underline{\underline{\hat{\rho}}}_1\right]\underline{\underline{\hat{R}}}_2(\zeta,m) \qquad (1.115)\\
-\frac{d}{d\zeta}\underline{\underline{\hat{T}}}_2(\zeta,m) &= \underline{\underline{\hat{T}}}_1(\zeta,m)\underline{\underline{\hat{\tau}}}_2 + \underline{\underline{\hat{T}}}_2(\zeta,m)\underline{\underline{\hat{\tau}}}_1 \\
&+ \left[\underline{\underline{\hat{T}}}_1(\zeta,m)\underline{\underline{\hat{\rho}}}_1 - \underline{\underline{\hat{T}}}_2(\zeta,m)\underline{\underline{\hat{\rho}}}_2\right]\underline{\underline{\hat{R}}}_2(\zeta,m) \\
&+ \left[\underline{\underline{\hat{T}}}_1(\zeta,m)\underline{\underline{\hat{\rho}}}_2 + \underline{\underline{\hat{T}}}_2(\zeta,m)\underline{\underline{\hat{\rho}}}_1\right]\underline{\underline{\hat{R}}}_1(\zeta,m) \qquad (1.116)\\
-\frac{d}{d\zeta}\underline{\underline{\hat{\Sigma}}}_1^{+t}(\zeta,m) &= \underline{\underline{\hat{T}}}_1(\zeta,m)\underline{\underline{\hat{\Sigma}}}_1^{+}(\zeta) - \underline{\underline{\hat{T}}}_2(\zeta,m)\underline{\underline{\hat{\Sigma}}}_2^{+}(\zeta)) \\
&+ \left[\underline{\underline{\hat{T}}}_1(\zeta,m)\underline{\underline{\hat{\rho}}}_1 - \underline{\underline{\hat{T}}}_2(\zeta,m)\underline{\underline{\hat{\rho}}}_2\right]\underline{\underline{\hat{\Sigma}}}_1^{-t}(m,\zeta) \\
&- \left[\underline{\underline{\hat{T}}}_1(\zeta,m)\underline{\underline{\hat{\rho}}}_2 + \underline{\underline{\hat{T}}}_2(\zeta,m)\underline{\underline{\hat{\rho}}}_1\right]\underline{\underline{\hat{\Sigma}}}_2^{-t}(m,\zeta) \qquad (1.117)
\end{aligned}
$$

$$
\begin{aligned}
-\frac{d}{d\zeta}\underline{\underline{\hat{\Sigma}}}_2^{+t}(\zeta,m) &= \underline{\underline{\hat{T}}}_1(\zeta,m)\underline{\underline{\hat{\Sigma}}}_2^{+}(\zeta) + \underline{\underline{\hat{T}}}_2(\zeta,m)\underline{\underline{\hat{\Sigma}}}_1^{+}(\zeta)) \\
&+ \left[\underline{\underline{\hat{T}}}_1(\zeta,w)\underline{\underline{\hat{\rho}}}_2 + \underline{\underline{\hat{T}}}_2(\zeta,m)\underline{\underline{\hat{\rho}}}_1\right]\underline{\underline{\hat{\Sigma}}}_2^{-t}(m,\zeta) \\
&+ \left[\underline{\underline{\hat{T}}}_1(\zeta,w)\underline{\underline{\hat{\rho}}}_1 - \underline{\underline{\hat{T}}}_2(\zeta,m)\underline{\underline{\hat{\rho}}}_2\right]\underline{\underline{\hat{\Sigma}}}_1^{-t}(m,\zeta) \qquad (1.118)
\end{aligned}
$$

The initial conditions for the upward sweep are

$$\underline{\underline{\hat{R}}}_1(m,m) = \underline{\underline{\hat{R}}}_2(m,m) = \underline{\underline{0}}_{4M\times 4M} \qquad (1.119)$$

$$\underline{\underline{\hat{T}}}_1(m,m) = \underline{\underline{\hat{T}}}_2(m,m) = \underline{\underline{I}}_{4M\times 4M} \qquad (1.120)$$

$$\underline{\underline{\hat{\Sigma}}}_{1,2}^{\pm t}(m,m) = \underline{0}_{4M\times 1} \qquad (1.121)$$

Just as for the bare slab surface boundary conditions of Eqs. (1.103)–(1.105), these initial conditions are general for a bare or fully transparent bottom. However, unlike for the surface boundary, Eqs. (1.119) and (1.120) will be revised in Sect. 1.6.2 to incorporate the actual bottom reflectance and transmittance as the initial condition for the upward integration sweep.

1.5.5.3 Checks on the ODEs

If the phase matrix is zero except for the (1,1) element, the VRTE reduces to the SRTE. The Fourier expansion of the phase function then contains only cosine amplitudes, which means that all of the $\underline{\underline{\hat{\rho}}}_2$, $\underline{\underline{\hat{\tau}}}_2$, $\underline{\underline{\hat{R}}}_2$ and $\underline{\underline{\hat{T}}}_2$ terms are zero. In that case Eqs. (1.88)–(1.117) all reduce to the corresponding Eqs. (8.74)–(8.85) seen in *Light and Water* for the case of no polarization. Those equations have proven to be correct and numerically efficient via their use in HydroLight. Unfortunately, this comparison does not help to verify the correctness of the newly derived equations for the various sine amplitudes. (Strictly speaking, the VRTE matrix equations reduce to the transposes of the matrix equations seen in *Light and Water*. This is because L&W started by writing the path function as $\int L\,\tilde{\beta}\,d\Omega$, rather than as $\int \tilde{\beta}\,L\,d\Omega$ corresponding to the vector version $\int \underline{\tilde{P}}\,\underline{S}\,d\Omega$ as seen in Eq. (1.19). This led to the scalar equivalent of Eq. (1.71) being written as a row vector, and the scalar versions of $\underline{\underline{\hat{\rho}}}_1$ being the transpose of Eq. (1.72), etc.)

1.5.5.4 Computational Issues

Recalling that the 8 standard matrices are each $4M \times 4M$ in size, and the 4 internal source vectors are $4M \times 1$, this gives a set of $N_{\text{eq}} = 8 \times 4M \times 4M + 4 \times 4M = 128M^2 + 16M = 12,960$ scalar ODEs for a quad discretization with $M = 10$. The unpolarized case has no sine terms, which reduces the number of standard operators

equations from 8 to 4, with 2 source-term equations, and each standard operator matrix is $M \times M$ since there is only a phase function (or (1,1) element of the phase matrix). The total number of scalar ODEs to be solved is then just $N_{\text{eq}} = 4 \times M \times M + 2 \times M = 420$. Thus there are roughly 30 times as many scalar ODEs to be solved for the VRTE as for the SRTE, namely twice as many Fourier coefficients and 4×4 phase matrices matrices rather than scalar phase functions. In both cases these sets of ODEs must be solved for each Fourier mode, $\ell = 0, ..., N$.

In practice, many quantities can be pre-computed. In particular, it should be remembered from Fig. 1.4 that the phase matrix elements have zero sine amplitudes for the upper left and lower right 2×2 blocks, and zero cosine elements for the upper right and lower left 2×2 blocks. These zeros carry through the evaluation of the $\underline{\underline{\hat{\rho}}}_{1,2}$ and $\underline{\underline{\hat{\tau}}}_{1,2}$ matrices, which reduces some of the computations needed to evaluate the right hand sides of the ODEs for the standard operators. However, the ODEs for $\underline{\underline{\hat{R}}}_1$ etc. involve both cosine and sine components of the $\underline{\underline{\hat{\rho}}}_{1,2}$ and $\underline{\underline{\hat{\tau}}}_{1,2}$ matrices, so that all elements of $\underline{\underline{\hat{R}}}_1$ etc. will in general be non-zero. If source terms are not included in the run, then the equations for $\underline{\underline{\hat{\Sigma}}}^{\pm t}_{1,2}$ can be omitted, resulting in a slightly smaller set of ODEs to be solved.

Now that the laborious development of the ODEs for the Fourier amplitudes of the standard matrices has been seen, it is worthwhile to consider the corresponding development in physical space. The upward and downward pair of of Eqs. (1.51) would be recast into local interaction equations of the form (ignoring internal sources)

$$\frac{d}{d\zeta}\underline{\underline{S}}^{+} = \underline{\underline{\tau}}\,\underline{\underline{S}}^{+} + \underline{\underline{\rho}}\,\underline{\underline{S}}^{-}$$
$$-\frac{d}{d\zeta}\underline{\underline{S}}^{-} = \underline{\underline{\rho}}\,\underline{\underline{S}}^{+} + \underline{\underline{\tau}}\,\underline{\underline{S}}^{-}$$

Here $\underline{\underline{S}}^{\pm}$ are each $8MN \times 1$ composite vectors of 4×1 Stokes vectors $\underline{S}^{\pm}(\zeta, u, v)$ for $u = 1, ..., M$ and $v = 1, ..., 2N$. $\underline{\underline{\rho}}$ and $\underline{\underline{\tau}}$ are consequently of size $8MN \times 8MN$. However, recall that the phase matrix depends on $\phi - \phi'$, i.e. on $\phi_v - \phi_s$, so that $\phi' = 0$ or $s = 1$ can be chosen. This reduces the number of unique elements of the $\underline{\underline{\rho}}$ and $\underline{\underline{\tau}}$ matrices by a factor of N. Thus the two $\underline{\underline{\rho}}$ and $\underline{\underline{\tau}}$ matrices each have an "information content" of $64M^2N$ independent numbers. The local interaction equations would lead to global interaction equations with four standard matrices: $\underline{\underline{R}}(w, \zeta)$, $\underline{\underline{R}}(\zeta, w)$, $\underline{\underline{T}}(w, \zeta)$, and $\underline{\underline{T}}(\zeta, w)$, with a total information content of $128M^2N$. It would be necessary to solve for the elements of these standard matrices via a coupled system of $128M^2N$ ODEs. For a quad partioning with $M = 10$, $N = 12$ (again ignoring the internal source terms), this is a system of 153,600 equations.

Taking the Fourier-space route gave global interaction equations with 4 matrices for Fourier amplitudes, $\underline{\underline{\hat{R}}}_1(w, \zeta|\ell)$, $\underline{\underline{\hat{R}}}_1(\zeta, w|\ell)$, $\underline{\underline{\hat{T}}}_1(w, \zeta|\ell)$, and $\underline{\underline{\hat{T}}}_1(\zeta, w|\ell)$, with each matrix being $4M \times 4M$ and $\ell = 0, ..., N$. The information content of these cosine-amplitude equations is thus $64M^2(N + 1)$. There are 4 matrices for sine amplitudes, $\underline{\underline{\hat{R}}}_2(w, \zeta|\ell)$, $\underline{\underline{\hat{R}}}_2(\zeta, w|\ell)$, $\underline{\underline{\hat{T}}}_2(w, \zeta|\ell)$, and $\underline{\underline{\hat{T}}}_2(\zeta, w|\ell)$, with each matrix being $4M \times 4M$ and $\ell = 1, ..., N - 1$. The information content of these sine-amplitude equations is thus $64M^2(N - 1)$. The total number of unique matrix elements being computed is then $128M^2N$, the same as for the physical-space formulation. However, the Fourier approach allows this total to be obtained by a sequence of solutions of much smaller sets of ODEs. Even with the Fourier decomposition, the ODE sets for the VRTE are still about 13,000 equations vs 420 for the SRTE. ODE solvers generally have run times proportional to the number of equations to be solved, so a thirty-fold increase in run can be anticipated. Additional computational issues such as array storage are given in Mobley (2014a).

1.6 Incorporation of the Boundary Conditions

The matrices $\underline{\underline{\hat{R}}}_1(w, z_k|\ell)$ etc. computed from the ODE solutions give the reflectances and transmittances for subslabs $[w, z_k]$ and $[z_k, m]$ of the "bare-slab" water column $[w, m]$. These will be used to compute the Stokes vectors at the user-defined output depths z_k. However, to obtain the Stokes vectors, it is necessary to incorporate the surface and bottom boundary conditions for the particular incident lighting, surface wave conditions, and bottom reflectance of interest. Thus the air-water surface $[a, w]$ must now be merged with the water body $[w, m]$.

1.6.1 Fourier Decomposition of the Surface Boundary Conditions

Recall the discretized surface boundary conditions of Eqs. (1.37) and(1.38). Using the $\pm$ superscripts to denote downwelling and upwelling directions as in Eq. (1.51), the surface boundary conditions can be written

$$\begin{aligned} \underline{S}^-(a, u, v) = &\sum_r \sum_s \underline{r}_{aw}(r, s \to u, v)\, \underline{S}^+(a, r, s) \\ &+ \sum_r \sum_s \underline{t}_{wa}(r, s \to u, v)\, \underline{S}^-(w, r, s) \end{aligned} \tag{1.122}$$

$$\underline{S}^{+}(w,u,v) = \sum_r \sum_s \underline{r}_{wa}(r,s \rightarrow u,v)\,\underline{S}^{-}(w,r,s) + \sum_r \sum_s \underline{t}_{aw}(r,s \rightarrow u,v)\,\underline{S}^{+}(a,r,s) \tag{1.123}$$

where the wavelength index has been omitted for brevity. The superscripts on $\underline{S}^{\pm}$ indicate whether u, v and r, s are in the upward or downward hemispheres. The 4×4 matrices $\underline{r}_{aw}(r, s \rightarrow u, v)$ etc. play the same role for surface scattering as does the phase matrix for scattering in the interior of the water body. In particular, $\underline{r}_{aw}(r, s \rightarrow u, v)$ etc. include the effects of the rotation matrices.

As with the VRTE, the boundary conditions are Fourier decomposed. However, the surface transfer functions $\underline{r}_{aw}(r, s \rightarrow u, v)$ etc. do not in general have the $\phi - \phi' = \phi_v - \phi_s$ dependence on azimuthal angle that the phase matrix has. This is because the wind-blown sea surface can have different wave-slope statistics in the along-wind and cross-wind directions. The only assumption made here is that the surface statistics have elliptical symmetry in ϕ; that is, there is no difference in the upwind and downwind directions. If azimuthally averaged slope statistics are used, the problem simplifies considerably. In particular, the Fourier modes then decouple. However, to retain the desired generality in modeling the sea surface, it is necessary to do a full Fourier decomposition in ϕ_s and ϕ_v as shown in Eq. (1.200). Thus, for given r and u values, $\underline{r}_{aw}(r, s \rightarrow u, v)$ becomes

$$\begin{aligned}\underline{r}_{aw}(r,s \rightarrow u,v) = &\sum_{k=0}^{N}\sum_{\ell=0}^{N} \hat{\underline{r}}_{11}(a,w|r,u|k,\ell)\cos(k\phi_s)\cos(\ell\phi_v) \\ &+ \sum_{k=0}^{N}\sum_{\ell=0}^{N} \hat{\underline{r}}_{12}(a,w|r,u|k,\ell)\cos(k\phi_s)\sin(\ell\phi_v) \\ &+ \sum_{k=0}^{N}\sum_{\ell=0}^{N} \hat{\underline{r}}_{21}(a,w|r,u|k,\ell)\sin(k\phi_s)\cos(\ell\phi_v) \\ &+ \sum_{k=0}^{N}\sum_{\ell=0}^{N} \hat{\underline{r}}_{22}(a,w|r,u|k,\ell)\sin(k\phi_s)\sin(\ell\phi_v)\,. \end{aligned} \tag{1.124}$$

To find $\hat{\underline{r}}_{11}$, multiply Eq. (1.124) by $\cos(k'\phi_s)\cos(\ell'\phi_v)$, sum over s and v, and apply the orthogonality relations (1.188). The other three amplitudes are found in an analogous manner. The results are as shown in Eq. (1.201):

$$\hat{\underline{r}}_{11}(a,w|r,u|k,\ell) = \frac{1}{\epsilon_k \epsilon_\ell} \sum_{s=1}^{2N}\sum_{v=1}^{2N} \underline{r}_{aw}(r,s \rightarrow u,v)\cos(k\phi_s)\cos(\ell\phi_v)\,, \tag{1.125a}$$

$$\hat{\underline{r}}_{12}(a,w|r,u|k,\ell) = \frac{1}{\epsilon_k \gamma_\ell} \sum_{s=1}^{2N}\sum_{v=1}^{2N} \underline{r}_{aw}(r,s \rightarrow u,v)\cos(k\phi_s)\sin(\ell\phi_v)\,, \tag{1.125b}$$

$$\hat{\underline{r}}_{21}(a, w|r, u|k, \ell) = \frac{1}{\gamma_k \epsilon_\ell} \sum_{s=1}^{2N} \sum_{v=1}^{2N} \underline{r}_{aw}(r, s \to u, v) \sin(k\phi_s) \cos(\ell\phi_v) \,, \quad (1.125c)$$

$$\hat{\underline{r}}_{22}(a, w|r, u|k, \ell) = \frac{1}{\gamma_k \gamma_\ell} \sum_{s=1}^{2N} \sum_{v=1}^{2N} \underline{r}_{aw}(r, s \to u, v) \sin(k\phi_s) \sin(\ell\phi_v) \,. \quad (1.125d)$$

The sine amplitudes are zero for various special cases:

$$\begin{aligned}
&\hat{\underline{r}}_{12}(a, w|r, u|k, 0) = \hat{\underline{r}}_{12}(a, w|k, N) = 0 \quad \text{for } k = 0, ..., N \,, \\
&\hat{\underline{r}}_{21}(a, w|r, u|0, \ell) = \hat{\underline{r}}_{21}(a, w|N, \ell) = 0 \quad \text{for } \ell = 0, ..., N \,, \\
&\hat{\underline{r}}_{22}(a, w|r, u|0, 0) = \hat{\underline{r}}_{22}(a, w|r, u|0, N) \\
&\qquad\qquad = \hat{\underline{r}}_{22}(a, w|r, u|N, 0) = \hat{\underline{r}}_{22}(a, w|r, u|N, N) = 0 \,. \quad (1.126)
\end{aligned}$$

These special cases allow the exclusion of any k or ℓ values in Eqs. (1.125b)–(1.125d) that would result in division by zero resulting from the γ_k and γ_ℓ factors.

This expansion allows for the modeling of any sea surface. However, wave spectra generally have symmetry about the upwind-downwind direction. That is, waves may be propagating in all directions, but with equal probability to the left and right of the wind direction. Such wave spectra generally have elliptical symmetry, with the major axis of the ellipse in the along-wind direction, and the minor axis in the cross-wind direction. Thus the surface transfer functions from, say, $\phi' = \phi_s = 30°$ to $\phi = \phi_v = 110$ would be the same as for $\phi' = 330$ to $\phi = 250$, but not the same as from $\phi' = 70$ to $\phi = 150$.

If the coordinate system is chosen (as in HydroLight) so that an azimuthal direction of zero is in the downwind direction, then the statistics of the sea surface waves propagating in some direction ϕ are the same as for direction $2\pi - \phi$. In the quad indexing scheme of Fig. 1.5, direction ϕ_w is symmetric about the downwind direction with ϕ_{2N+2-w}. Now consider $\hat{\underline{r}}_{12}(a, w|r, u|k, \ell)$, which with these symmetries can be written

$$\hat{\underline{r}}_{12}(a, w|r, u|k, \ell) = \frac{1}{\epsilon_k \gamma_\ell} \sum_{s=1}^{2N} \sum_{v=1}^{2N} \underline{r}_{aw}(r, 2N + 2 - s \to u, 2N + 2 - v) \cos(k\phi_s) \sin(\ell\phi_v) \,.$$

Changing the summation indices to $s' = 2N + 2 - s$ and $v' = 2N + 2 - v$ gives

$$\hat{\underline{r}}_{12}(a, w|r, u|k, \ell) = \frac{1}{\epsilon_k \gamma_\ell} \sum_{s'=2N+1}^{2} \sum_{v'=2N+1}^{2} \underline{r}_{aw}(r, s' \to u, v') \cos(k\phi_{2N+2-s'}) \sin(\ell\phi_{2N+2-v'}) \,.$$

Noting once again that $\phi_{2M+2-w} = 2\pi - \phi_w$, using the evenness of the cosine and oddness of the sine, remembering that quad $w = 2N + 1$ is the same as quad $w = 1$, and reordering the sums gives

$$\begin{aligned}\underline{\hat{r}}_{12}(a, w|r, u|k, \ell) &= -\frac{1}{\epsilon_k \gamma_\ell} \sum_{s'=1}^{2N} \sum_{v'=1}^{2N} \underline{r}_{aw}(r, s' \to u, v') \cos(k\phi_{s'}) \sin(\ell\phi_{v'}) \\ &= -\underline{\hat{r}}_{12}(a, w|r, u|k, \ell)\,.\end{aligned}$$

It therefore follows that

$$\underline{\hat{r}}_{12}(a, w|r, u|k, \ell) = \underline{0} \quad \text{for } r, u = 1, ..., M \text{ and } k, \ell = 0, ..., N\,.$$

The same analysis gives $\underline{\hat{r}}_{21}(a, w|r, u|k, \ell) = \underline{0}$, with the same result for the other three surface transfer functions. Thus *elliptical symmetry of the surface about the wind direction eliminates two of the four terms in Eq. (1.124). This is the reason for choosing a wind-centered azimuthal coordinate system.*

A similar analysis based on the elliptical symmetry between ϕ and $\pi - \phi$ gives

$$\begin{aligned}\underline{\hat{r}}_{11}(a, w|r, u|k, \ell) &= (-1)^{k+\ell} \frac{1}{\epsilon_k \epsilon_\ell} \sum_{s'=1}^{2N} \sum_{v'=1}^{2N} \underline{r}_{aw}(r, s' \to u, v') \cos(k\phi_{s'}) \sin(\ell\phi_{v'}) \\ &= (-1)^{k+\ell}\, \underline{\hat{r}}_{11}(a, w|k, \ell)\,.\end{aligned}$$

Therefore it follows that

$$\underline{\hat{r}}_{11}(a, w|r, u|k, \ell) = \underline{0} \quad \text{if } k + \ell \text{ is odd},$$

for $r, u = 1, ..., M$ and $k, \ell = 0, ..., N$. Corresponding results are obtained for $\underline{\hat{r}}_{22}(a, w|r, u|k, \ell)$ and the other three surface transfer functions.

We can therefor simplify the notation to one subscript for the $\underline{\hat{r}}$ and $\underline{\hat{t}}$ terms:

$$\begin{aligned}\underline{\hat{r}}_1(a, w|r, u|k, \ell) &= \underline{\hat{r}}_{11}(a, w|r, u|k, \ell) \\ \underline{\hat{r}}_2(a, w|r, u|k, \ell) &= \underline{\hat{r}}_{22}(a, w|r, u|k, \ell)\end{aligned}$$

since the $\underline{\hat{r}}_{12}$ and $\underline{\hat{r}}_{21}$ terms are always zero. The surface transfer functions for elliptically symmetric sea surfaces thus can be decomposed as

$$\underline{r}_{aw}(r, s \to u, v) = \sum_{\substack{k=0 \\ (k+\ell \text{ even})}}^{N} \sum_{\ell=0}^{N} \underline{\hat{r}}_1(a, w|r, u|k, \ell) \cos(k\phi_s) \cos(\ell\phi_v)$$

$$+\sum_{\substack{k=0\\(k+\ell\ \text{even})}}^{N}\sum_{\ell=0}^{N}\underline{\hat{r}}_2(a,w|r,u|k,\ell)\sin(k\phi_s)\sin(\ell\phi_v)\,. \tag{1.127}$$

where

$$\underline{\hat{r}}_1(a,w|r,u|k,\ell)=\begin{cases}\frac{1}{\epsilon_k\epsilon_\ell}\sum_{s=1}^{2N}\sum_{v=1}^{2N}\underline{r}_{aw}(r,s\to u,v)\cos(k\phi_s)\cos(\ell\phi_v)\\ \quad\text{for } k,\ell=0,...,N \text{ and } k+\ell \text{ even}\\ \underline{0}_{4\times4} \text{ if } k+\ell \text{ is odd}\end{cases}$$

and

$$\underline{\hat{r}}_2(a,w|r,u|k,\ell)=\begin{cases}\frac{1}{\gamma_k\gamma_\ell}\sum_{s=1}^{2N}\sum_{v=1}^{2N}\underline{r}_{aw}(r,s\to u,v)\cos(k\phi_s)\cos(\ell\phi_v)\\ \quad\text{for } k,\ell=1,...,N-1 \text{ and } k+\ell \text{ even}\\ \underline{0}_{4\times4} \text{ if } k=0 \text{ and } \ell=0,...,N\\ \underline{0}_{4\times4} \text{ if } \ell=0 \text{ and } k=0,...,N\\ \underline{0}_{4\times4} \text{ if } k+\ell \text{ is odd}\end{cases}$$

Equations (1.52) and (1.127) can now be used to Fourier decompose the surface boundary conditions. The process is the same as before, using the orthogonality relations (1.188)–(1.190) to reduce double sums to single, and noting the linear independence of sines and cosines. The result for the cosine amplitudes of Eq. (1.122) is

$$\begin{aligned}&\underline{S}_1^-(a,u|\ell)\\ &=\sum_{\substack{k=0\\(k+\ell\ \text{even})}}^{N}\left[\sum_{r=1}^{M-1}\epsilon_k\,\underline{\hat{r}}_1(a,w|r,u|k,\ell)\underline{S}_1^+(a,r|k)+\delta_k\,\underline{\hat{r}}_1(a,w|M,u|k,\ell)\,\underline{S}_1^+(a,M|k)\right]\\ &+\sum_{\substack{k=0\\(k+\ell\ \text{even})}}^{N}\left[\sum_{r=1}^{M-1}\epsilon_k\,\underline{\hat{t}}_1(w,a|r,u|k,\ell)\,\underline{S}_1^-(w,r|k)+\delta_k\,\underline{\hat{t}}_1(w,a|M,u|k,\ell)\,\underline{S}_1^-(w,M|k)\right]\end{aligned} \tag{1.128}$$

which holds for $u=1,...,M$ and $\ell=0,...,N$. As always, the polar cap $r=M$ is a special case with no ϕ dependence, so that only the $k=0$ cosine amplitude is nonzero. However, the polar cap terms can be incorporated into the sum over k by use of a Kronecker δ_k factor, as shown here. The sine amplitudes are given by

$$\underline{S}_2^-(a,u|\ell) = \sum_{\substack{k=0\\(k+\ell \text{ even})}}^{N} \left[\sum_{r=1}^{M-1} \gamma_k \, \underline{\hat{r}}_2(a,w|r,u|k,\ell) \, \underline{S}_2^+(a,r|k) \right] + \sum_{\substack{k=0\\(k+\ell \text{ even})}}^{N} \left[\sum_{r=1}^{M-1} \gamma_k \, \underline{\hat{t}}_2(w,a|r,u|k,\ell) \, \underline{S}_2^-(w,r|k) \right] \tag{1.129}$$

where now $\ell = 1, ..., N$. As always, sine amplitudes are zero for polar caps. A similar result holds for Eq. (1.123).

It is important to note that *the Fourier modes do not decouple for the surface transfer functions*. That is, the equations for $\underline{S}_{1,2}^-(a,u|\ell)$ involve sums over k. This is the reason that the surface boundary conditions were not used as the initial conditions for mode-by-mode integration of the ODEs to obtain the standard reflection and transmission matrices for the interior of the water body.

These equations can be placed in a more compact form by defining $4M \times 4M$ composite matrices $\underline{\underline{\hat{r}}}_{1,2}$ and $\underline{\underline{\hat{t}}}_{1,2}$ with (row, column) = u, r elements as follows:

$$[\underline{\underline{\hat{r}}}_1(a,w|k,\ell)]_{u,r} = \begin{cases} \epsilon_k \, \underline{\hat{r}}_1(a,w|r,u|k,\ell) & \text{for } r = 1, ..., M-1 \\ \delta_k \, \underline{\hat{r}}_1(a,w|M,u|k,\ell) & \text{for } r = M \end{cases}$$

which holds for $u = 1, ..., M$ and $k, \ell = 0, ..., N$. The corresponding definition for the sine term is

$$[\underline{\underline{\hat{r}}}_2(a,w|k,\ell)]_{u,r} = \begin{cases} \gamma_k \, \underline{\hat{r}}_2(a,w|r,u|k,\ell) & \text{for } r = 1, ..., M-1 \text{ and } \ell = 1, ..., N-1 \\ \underline{0}_{4\times 4} & \text{for } r = M \text{ or } \ell = 0 \text{ or } N \end{cases}$$

which holds for which holds for $u = 1, ..., M$ and $k = 0, ..., N$. The $\underline{\hat{t}}_1$ and $\underline{\hat{t}}_2$ functions have corresponding definitions.

Recalling the composite matrix form (1.71) for Stokes vectors, Eqs. (1.128) and (1.129) can now be combined and written as

$$\underline{\underline{\hat{S}}}_p^-(a|\ell) = \sum_{\substack{k=0\\(k+\ell \text{ even})}}^{N} \underline{\underline{\hat{r}}}_p(a,w|k,\ell) \, \underline{\underline{\hat{S}}}_p^+(a|\ell) + \sum_{\substack{k=0\\(k+\ell \text{ even})}}^{N} \underline{\underline{\hat{t}}}_p(w,a|k,\ell) \, \underline{\underline{\hat{S}}}_p^-(w|\ell) \, . \tag{1.130}$$

Equation (1.123) similarly becomes

$$\underline{\underline{\hat{S}}}_p^+(w|\ell) = \sum_{\substack{k=0\\(k+\ell \text{ even})}}^{N} \underline{\underline{\hat{r}}}_p(w,a|k,\ell) \, \underline{\underline{\hat{S}}}_p^-(w|\ell) + \sum_{\substack{k=0\\(k+\ell \text{ even})}}^{N} \underline{\underline{\hat{t}}}_p(a,w|k,\ell) \, \underline{\underline{\hat{S}}}_p^+(a|\ell) \, , \tag{1.131}$$

where $p = 1$ or 2 and $\ell = 0, ..., N$.

Equations (1.130) and (1.131) are the desired Fourier-amplitude forms of the surface boundary conditions. These equations are at a notational level equivalent to the local interaction equations (1.73)–(1.76).

Equations (1.130) and (1.131) can be notationally simplified still further by defining $4M(N+1) \times 4M(N+1)$ composite matrices that contain all Fourier modes. The $4M \times 4M$ cosine amplitude matrices $\underline{\underline{\hat{r}}}_1(a, w|k, \ell)$ for $k, \ell = 0, 1, ..., N$ can be combined into one matrix as follows (L&W 8.86):

$$\underline{\underline{\hat{r}}}_1(a, w) =$$

$$\begin{bmatrix}
\underline{\underline{\hat{r}}}_1(a, w|0, 0) & \underline{\underline{0}} & \underline{\underline{\hat{r}}}_1(a, w|2, 0) & \underline{\underline{0}} & \dots & \underline{\underline{\hat{r}}}_1(a, w|N, 0) \\
\underline{\underline{0}} & \underline{\underline{\hat{r}}}_1(a, w|1, 1) & \underline{\underline{0}} & \underline{\underline{\hat{r}}}_1(a, w|3, 1) & \dots & \underline{\underline{0}} \\
\underline{\underline{\hat{r}}}_1(a, w|0, 2) & \underline{\underline{0}} & \underline{\underline{\hat{r}}}_1(a, w|2, 2) & \underline{\underline{0}} & \dots & \underline{\underline{\hat{r}}}_1(a, w|N, 2) \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\underline{\underline{0}} & \underline{\underline{\hat{r}}}_1(a, w|1, N-1) & \underline{\underline{0}} & \underline{\underline{\hat{r}}}_1(a, w|3, N-1) & \dots & \underline{\underline{0}} \\
\underline{\underline{\hat{r}}}_1(a, w|0, N) & \underline{\underline{0}} & \underline{\underline{\hat{r}}}_1(a, w|2, N) & \underline{\underline{0}} & \dots & \underline{\underline{\hat{r}}}_1(a, w|N, N)
\end{bmatrix} \tag{1.132}$$

The checkerboard structure of this matrix incorporates the previous result that $\underline{\underline{\hat{r}}}_1(a, w|k, \ell) \neq \underline{\underline{0}}$ only when $k + \ell$ is even. Recall than N is always even. The $\underline{\underline{0}}$ submatrices are all $4M \times 4M$. Similarly, the sine amplitude matrices $\underline{\underline{\hat{r}}}_2(a, w|k, \ell)$ can be combined as

$$\underline{\underline{\hat{r}}}_2(a, w) = \begin{bmatrix}
\underline{\underline{0}} & \underline{\underline{0}} & \underline{\underline{0}} & \underline{\underline{0}} & \dots & \underline{\underline{0}} \\
\underline{\underline{0}} & \underline{\underline{\hat{r}}}_2(a, w|1, 1) & \underline{\underline{0}} & \underline{\underline{\hat{r}}}_2(a, w|3, 1) & \dots & \underline{\underline{0}} \\
\underline{\underline{0}} & \underline{\underline{0}} & \underline{\underline{\hat{r}}}_2(a, w|2, 2) & \underline{\underline{0}} & \dots & \underline{\underline{0}} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\underline{\underline{0}} & \underline{\underline{\hat{r}}}_2(a, w|1, N-1) & \underline{\underline{0}} & \underline{\underline{\hat{r}}}_2(a, w|3, N-1) & \dots & \underline{\underline{0}} \\
\underline{\underline{0}} & \underline{\underline{0}} & \underline{\underline{0}} & \underline{\underline{0}} & \dots & \underline{\underline{0}}
\end{bmatrix}. \tag{1.133}$$

The $\underline{\underline{0}}$ first and last row and column show that sine amplitudes are zero for $k, \ell = 0$ or N. The $\underline{\underline{\hat{t}}}_{1,2}$ matrices have corresponding definitions. The Stokes vector amplitudes $\underline{\underline{\hat{S}}}^{\pm}_{1,2}(\zeta|\ell)$ are likewise combined into one $4M(N+1) \times 1$ column vector containing all Fourier modes (L&W 8.87):

$$\underline{\underline{\hat{S}}}^{\pm}_{1,2}(\zeta) = \begin{bmatrix}
\underline{\underline{\hat{S}}}^{\pm}_{1,2}(\zeta|0) \\
\underline{\underline{\hat{S}}}^{\pm}_{1,2}(\zeta|1) \\
\vdots \\
\underline{\underline{\hat{S}}}^{\pm}_{1,2}(\zeta|N)
\end{bmatrix}. \tag{1.134}$$

It is worth pausing to note that the $\underline{\underline{\hat{S}}}^{\pm}_{1,2}(\zeta)$ are matrices whose elements $\underline{\underline{\hat{S}}}^{\pm}_{1,2}(\zeta|\ell)$ are by Eq. (1.71) $4M \times 1$ matrices, whose elements in turn are 4×1 Stokes vectors. In other words, these equations involve matrices whose elements are matrices whose elements are matrices.

Equations (1.132)–(1.134) allow the surface boundary conditions (1.130) and (1.131) to be written as matrix equations:

$$\underline{\underline{\hat{S}}}^{-}_{p}(a) = \underline{\underline{\hat{r}}}_{p}(a, w)\underline{\underline{\hat{S}}}^{+}_{p}(a) + \underline{\underline{\hat{t}}}_{p}(w, a)\underline{\underline{\hat{S}}}^{-}_{p}(w) \tag{1.135}$$

and

$$\underline{\underline{\hat{S}}}^{+}_{p}(w) = \underline{\underline{\hat{r}}}_{p}(w, a)\underline{\underline{\hat{S}}}^{-}_{p}(w) + \underline{\underline{\hat{t}}}_{p}(a, w)\underline{\underline{\hat{S}}}^{+}_{p}(a)\,. \tag{1.136}$$

It should be noted that $\underline{\underline{\hat{r}}}_{1,2}$ and $\underline{\underline{\hat{t}}}_{1,2}$ are in effect the standard reflection and transmission matrices for the air-water surface. Unlike the standard matrices for the interior of the water body $[w, \zeta]$, which must be found by depth integrations of the ODEs derived in the preceding chapter, the surface $[a, w]$ has no thickness and standard matrices are determined from the reflection and transmission matrices for the sea surface, as seen in Sect. 1.4.

Finally, the standard reflectance and transmittance matrices for the water body $[w, \zeta]$, which were computed ℓ mode by ℓ mode when solving the ODE systems, can be combined into matrices containing all Fourier modes, e.g.:

$$\underline{\underline{\hat{R}}}_{1}(\zeta, w) = \begin{bmatrix} \underline{\underline{\hat{R}}}_{1}(\zeta, w|0) & \underline{\underline{0}} & \underline{\underline{0}} \ldots & \underline{\underline{0}} \\ \underline{\underline{0}} & \underline{\underline{\hat{R}}}_{1}(\zeta, w|1)) & \underline{\underline{0}} \ldots & \underline{\underline{0}} \\ \vdots & \vdots & \vdots \ddots & \vdots \\ \underline{\underline{0}} & \underline{\underline{0}} & \underline{\underline{0}} \ldots & \underline{\underline{\hat{R}}}_{1}(\zeta, w|N) \end{bmatrix}$$

with corresponding definitions for the other standard matrices.

These definitions allow the global interaction Eqs. (1.77)–(1.80) for slab $[w, \zeta]$ to be written in a form that contains all Fourier modes:

$$\begin{aligned} \underline{\underline{\hat{S}}}^{-}_{1}(w) &= \underline{\underline{\hat{R}}}_{1}(w, \zeta)\underline{\underline{\hat{S}}}^{+}_{1}(w) + \underline{\underline{\hat{T}}}_{1}(\zeta, w)\underline{\underline{\hat{S}}}^{-}_{1}(\zeta) \\ &\quad - \underline{\underline{\hat{R}}}_{2}(w, \zeta)\underline{\underline{\hat{S}}}^{+}_{2}(w) - \underline{\underline{\hat{T}}}_{2}(\zeta, w)\underline{\underline{\hat{S}}}^{-}_{2}(\zeta) + \underline{\underline{\hat{\Sigma}}}^{-t}_{1}(\zeta, w) \end{aligned} \tag{1.137}$$

$$\begin{aligned} \underline{\underline{\hat{S}}}^{+}_{1}(\zeta) &= \underline{\underline{\hat{R}}}_{1}(\zeta, w)\underline{\underline{\hat{S}}}^{-}_{1}(\zeta) + \underline{\underline{\hat{T}}}_{1}(w, \zeta)\underline{\underline{\hat{S}}}^{+}_{1}(w) \\ &\quad - \underline{\underline{\hat{R}}}_{2}(\zeta, w)\underline{\underline{\hat{S}}}^{-}_{2}(\zeta) - \underline{\underline{\hat{T}}}_{2}(w, \zeta)\underline{\underline{\hat{S}}}^{+}_{2}(w) + \underline{\underline{\hat{\Sigma}}}^{+t}_{1}(w, \zeta) \end{aligned} \tag{1.138}$$

$$\begin{aligned} \underline{\underline{\hat{S}}}^{-}_{2}(w) &= \underline{\underline{\hat{R}}}_{1}(w, \zeta)\underline{\underline{\hat{S}}}^{+}_{2}(w) + \underline{\underline{\hat{T}}}_{1}(\zeta, w)\underline{\underline{\hat{S}}}^{-}_{2}(\zeta) \\ &\quad + \underline{\underline{\hat{R}}}_{2}(w, \zeta)\underline{\underline{\hat{S}}}^{+}_{1}(w) + \underline{\underline{\hat{T}}}_{2}(\zeta, w)\underline{\underline{\hat{S}}}^{-}_{1}(\zeta) + \underline{\underline{\hat{\Sigma}}}^{-t}_{2}(\zeta, w) \end{aligned} \tag{1.139}$$

$$\underline{\underline{\hat{S}}}^{+}_{2}(\zeta) = \underline{\underline{\hat{R}}}_{1}(\zeta, w)\underline{\underline{\hat{S}}}^{-}_{2}(\zeta) + \underline{\underline{\hat{T}}}_{1}(w, \zeta)\underline{\underline{\hat{S}}}^{+}_{2}(w)$$

$$+ \underline{\underline{\hat{R}}}_2(\zeta, w)\underline{\underline{\hat{S}}}_1^-(\zeta) + \underline{\underline{\hat{T}}}_2(w, \zeta)\underline{\underline{\hat{S}}}_1^+(w) + \underline{\underline{\hat{\Sigma}}}_2^{+t}(w, \zeta) \tag{1.140}$$

Here the internal source terms for each Fourier mode have been combined as in Eq. (1.134).

1.6.2 Fourier Decomposition of the Bottom Boundary Conditions

As noted in Sect. 1.2.6, the bottom boundary condition has the form (dropping the wavelength dependence for brevity)

$$\begin{aligned}\underline{S}(m, \theta, \phi) &= \iint_{2\pi_d} \underline{VBRDF}(\theta', \theta, \phi - \phi') \cos\theta' \, \underline{S}(m, \theta', \phi') \, d\Omega(\theta', \phi') \\ &= \iint_{2\pi_d} \underline{r}_{mb}(\theta', \theta, \phi - \phi') \, \underline{S}(m, \theta', \phi') \, d\Omega(\theta', \phi') \quad \text{for } (\theta, \phi) \in 2\pi_u \ .\end{aligned}$$

A $\underline{VBRDF}$ or $\underline{r}_{mb}$ of this form can be expanded as Fourier cosine series. As was seen in Eq. (1.39), the corresponding quad-averaged bottom boundary condition is

$$\underline{S}(m, u, v) = \sum_{\substack{r \\ r,s \in \Xi_d}} \sum_{s} \underline{r}_{mb}(r, s \to u, v) \, \underline{S}(m, r, s) \quad \text{for (u, v)} \in \Xi_{\text{u}} \ . \tag{1.141}$$

The Fourier decomposition of the bottom boundary condition conceptually follows in parallel to the decomposition of the surface boundary condition, but the equations are much simpler because of the $\phi - \phi'$ symmetry and the absence of transmission terms as discussed in Sect. 1.2.6. The Stokes vectors are expanded as in Eq. (1.52). $\underline{r}_{mb}(r, s \to u, v) = \underline{r}_{mb}(r, u, |v - s|)$ is expanded as a cosine series using (1.196) (L&W 8.54):

$$\underline{r}_{mb}(r, u, |v - s|) = \sum_{k=0}^{N} \underline{\hat{r}}_1(m, b|r, u|k) \cos[k(\phi_v - \phi_s)] \, , \tag{1.142}$$

where from (1.197) (L&W 8.55)

$$\underline{\hat{r}}_1(m, b|r, u|k) = \frac{1}{\epsilon_k \cos(k\phi_s)} \sum_{v=1}^{2N} \underline{r}_{mb}(r, u, |v - s|) \cos(k\phi_v) \, .$$

We are again free to choose $s = 1$, or $\phi_s = 0$, which gives

$$\hat{\underline{r}}_1(m,b|r,u|k) = \frac{1}{\epsilon_k}\sum_{v=1}^{2N} \underline{r}_{mb}(r,u,|v-1|)\cos(k\phi_v)\,. \tag{1.143}$$

Substituting these expansions into Eq. (1.141) gives

$$\begin{aligned}
&\sum_{\ell=0}^{N}\left[\hat{\underline{S}}_1^-(m,u|\ell)\cos(\ell\phi_v) + \hat{\underline{S}}_2^-(m,u|\ell)\sin(\ell\phi_v)\right] \\
&= \sum_r \sum_s \sum_{k=0}^{N} \hat{\underline{r}}_1(m,b|r,u|k)\cos[k(\phi_v-\phi_s)] \\
&\times \left\{\sum_{\ell=0}^{N}\left[\hat{\underline{S}}_1^+(m,r|\ell)\cos(\ell\phi_s) + \hat{\underline{S}}_2^+(m,r|\ell)\sin(\ell\phi_s)\right]\right\}
\end{aligned} \tag{1.144}$$

Expanding $\cos[k(\phi_v-\phi_s)]$ and grouping terms gives

$$\begin{aligned}
&\sum_{\ell=0}^{N}\left[\hat{\underline{S}}_1^-(m,u|\ell)\cos(\ell\phi_v) + \hat{\underline{S}}_2^-(m,u|\ell)\sin(\ell\phi_v)\right] \\
&= \sum_{\ell=0}^{N}\sum_r\sum_{k=1}^{N} \hat{\underline{r}}_1(m,b|r,u|k)\cos(k\phi_v)\,\hat{\underline{S}}_1^+(m,r|\ell)\sum_s \cos(k\phi_s)\cos(\ell\phi_s) \\
&= \sum_{\ell=0}^{N}\sum_r\sum_{k=1}^{N} \hat{\underline{r}}_1(m,b|r,u|k)\cos(k\phi_v)\,\hat{\underline{S}}_2^+(m,r|\ell)\sum_s \cos(k\phi_s)\sin(\ell\phi_s) \\
&= \sum_{\ell=0}^{N}\sum_r\sum_{k=1}^{N} \hat{\underline{r}}_1(m,b|r,u|k)\sin(k\phi_v)\,\hat{\underline{S}}_1^+(m,r|\ell)\sum_s \sin(k\phi_s)\cos(\ell\phi_s) \\
&= \sum_{\ell=0}^{N}\sum_r\sum_{k=1}^{N} \hat{\underline{r}}_1(m,b|r,u|k)\sin(k\phi_v)\,\hat{\underline{S}}_2^+(m,r|\ell)\sum_s \sin(k\phi_s)\sin(\ell\phi_s)
\end{aligned} \tag{1.145}$$

Applying the orthogonality relations (1.188)–(1.190) to the sums over s and then observing as before that $k=\ell$ follows, this reduces to

$$\begin{aligned}
&\sum_{\ell=0}^{N}\left[\hat{\underline{S}}_1^-(m,u|\ell)\cos(\ell\phi_v) + \hat{\underline{S}}_2^-(m,u|\ell)\sin(\ell\phi_v)\right] \\
&= \sum_{\ell=0}^{N}\sum_r \hat{\underline{r}}_1(m,b|r,u|k)\,\hat{\underline{S}}_1^+(m,r|\ell)\,\epsilon_\ell\cos(\ell\phi_v) \\
&+ \sum_{\ell=0}^{N}\sum_r \hat{\underline{r}}_1(m,b|r,u|k)\,\hat{\underline{S}}_2^+(m,r|\ell)\,\gamma_\ell\sin(\ell\phi_v)
\end{aligned} \tag{1.146}$$

where ϵ_ℓ and γ_ℓ are given by Eqs. (1.193) and (1.195), respectively. As always, the polar caps $u, r = M$ are special cases with no ϕ dependence.

Invoking the linear independence of $\sin(\ell\phi_v)$ and $\cos(\ell\phi_v)$ gives two separate equations relating $\underline{\hat{S}}_1^{\pm}(m, u|\ell)$ and $\underline{\hat{S}}_2^{\pm}(m, u|\ell)$. These can be put into matrix form by defining $4M \times 4M$ composite matrices with 4×4 elements (L&W 8.57):

$$[\underline{\underline{\hat{r}}}_1(m, b|\ell)]_{u,r} = \begin{cases} \epsilon_\ell \, \underline{\hat{r}}_1(m, b|r, u|\ell) & \text{if } r = 1, ..., M-1 \\ \delta_\ell \, \underline{\hat{r}}_1(m, b|r, u|\ell) & \text{if } r = M \end{cases}$$

for $u = 1, ..., M$ and $\ell = 0, ..., N$; and

$$[\underline{\underline{\hat{r}}}_2(m, b|\ell)]_{u,r} = \begin{cases} \gamma_\ell \, \underline{\hat{r}}_2(m, b|r, u|\ell) & \text{if } r = 1, ..., M-1 \text{ and } \ell = 1, ...N-1 \\ 0 & \text{if } r = M \text{ or } \ell = 0 \text{ or } \ell = N \end{cases}$$

Equation (1.146) thus becomes

$$\underline{\hat{S}}_1^-(m, u|\ell) = \sum_r [\underline{\underline{\hat{r}}}_1(m, b|\ell)]_{u,r} \, \underline{\hat{S}}_1^+(m, r|\ell)$$
$$\underline{\hat{S}}_2^-(m, u|\ell) = \sum_r [\underline{\underline{\hat{r}}}_2(m, b|\ell)]_{u,r} \, \underline{\hat{S}}_2^+(m, r|\ell)$$

In composite matrix form these equations are (L&W 8.56)

$$\underline{\underline{\hat{S}}}_1^-(m|\ell) = \underline{\underline{\hat{r}}}_1(m, b|\ell) \underline{\underline{\hat{S}}}_1^+(m|\ell)$$
$$\underline{\underline{\hat{S}}}_2^-(m|\ell) = \underline{\underline{\hat{r}}}_2(m, b|\ell) \underline{\underline{\hat{S}}}_2^+(m|\ell)$$

where $\underline{\underline{\hat{S}}}_{1,2}^{\pm}$ are $4M \times 1$ composite column vectors defined in Eq. (1.71).

Note that, unlike for the surface boundary conditions, *the ℓ modes decouple for the bottom boundary condition because of the restriction to a bottom VBRDF dependence on* $\cos(\phi - \phi')$. *This allows the initial condition (1.119) for the upward sweep of ODEs to be revised to incorporate the bottom boundary reflectance as the initial condition for integration of the upward set of equations.* Thus Eq. (1.119),

$$\underline{\underline{\hat{R}}}_1(m, m) = \underline{\underline{\hat{R}}}_2(m, m) = \underline{\underline{0}}_{4M \times 4M} \, ,$$

is replaced by (L&W 8.94)

$$\underline{\underline{\hat{R}}}_1(m, m) = \underline{\underline{\hat{r}}}_1(m, b|\ell)$$
$$\underline{\underline{\hat{R}}}_2(m, m) = \underline{\underline{\hat{r}}}_2(m, b|\ell)$$

for each ℓ value of the upward sweep integrations. Similarly, the bottom boundary is assumed to be opaque, so that no light is transmitted upward through the bottom boundary. The initial conditions for the upward sweep integration for a reflecting, opaque, source-free lower boundary are then

$$\underline{\underline{\hat{R}}}_1(m,b) = \underline{\underline{\hat{r}}}_1(m,b|\ell) \tag{1.147}$$

$$\underline{\underline{\hat{R}}}_2(m,b) = \underline{\underline{\hat{r}}}_2(m,b|\ell) \tag{1.148}$$

$$\underline{\underline{\hat{T}}}_1(m,b) = \underline{\underline{\hat{T}}}_2(m,b) = \underline{\underline{0}}_{4M\times 4M} \tag{1.149}$$

$$\underline{\underline{\hat{\Sigma}}}^{\pm t}_{1,2}(m,b) = \underline{\underline{0}}_{4M\times 1} \tag{1.150}$$

Finally, recall from Eq. (1.32) that for a Lambertian depolarizing bottom only the (1,1) element of $\underline{\underline{r}}_{mb}$ is nonzero, and that $\underline{\underline{r}}_{mb}(1,1) = [R(\lambda)\cos\theta']/\pi$. This implies that only the (1,1), $\ell = 0$ element of $\underline{\underline{\hat{r}}}_1(m,b|\ell)$ is nonzero, and that $\underline{\underline{\hat{r}}}_2(m,b|\ell) = 0$ for all ℓ values. In particular,

$$[\underline{\underline{\hat{r}}}_1(m,b|0)]_{1,1} = \frac{R(\lambda)}{\pi}\mu_r\Omega_{r1} = \underline{\underline{r}}_{mb}(r,1\rightarrow u,v)\,.$$

1.6.3 Combining the Sea Surface with the Water Body

We now have the Fourier amplitude forms of the standard reflectance and transmittance matrices for the air-water surface $[a,w]$, $\underline{\underline{\hat{r}}}_1(a,w)$, $\underline{\underline{\hat{t}}}_1(a,w)$, etc., from Sect. 1.6.1. The standard matrices $\underline{\underline{\hat{R}}}_1(w,\zeta)$, $\underline{\underline{\hat{T}}}_1(w,\zeta)$, etc. for the bare-slab water body $[w,\zeta]$ are known from the solutions of the ODEs in Sect. 1.5.5. The next step is to combine these two slabs to obtain matrices for the air-water surface plus the water body. That is, the "real" sea surface must be added to the top of the bare-slab water body.

The goal is to obtain the response radiances at the sea surface, $\underline{\underline{\hat{S}}}^{-}_{1,2}(a)$, and at depth ζ, $\underline{\underline{\hat{S}}}^{+}_{1,2}(\zeta)$, in terms of the known incident radiance at the surface, $\underline{\underline{\hat{S}}}^{+}_{1,2}(a)$, and at the bottom, $\underline{\underline{\hat{S}}}^{-}_{1,2}(\zeta)$. This is a two-step process:

1. The first step is to obtain the internal radiances $\underline{\underline{\hat{S}}}^{\pm}_{1,2}(w)$ at depth w within the combined slab $[a,\zeta] = [a,w]\cup[w,\zeta]$ as functions of the incident radiances at a and ζ.
2. Those results will then be used to eliminate $\underline{\underline{\hat{S}}}^{\pm}_{1,2}(w)$ in the surface boundary condition (1.135) and the global interaction Eqs. (1.138) and (1.140).

The possibility of internal sources within $[a,w]$ is included to obtain a general result for the union of any two slabs. When $[a,w]$ is just a discontinuity in optical properties,

as for an air-water surface, those sources are zero. However, the surface $[a, w]$ also could represent a very thin layer of oil, which can fluoresce. In that case, the surface slab $[a, w]$ would include an internal source.

Surface boundary condition (1.136) for $[a, w]$ becomes, for $p = 1, 2$ and with internal source terms added for complete generality,

$$\underline{\underline{\hat{S}}}_1^+(w) = \underline{\underline{\hat{r}}}_1(w, a)\underline{\underline{\hat{S}}}_1^-(w) + \underline{\underline{\hat{t}}}_1(a, w)\underline{\underline{\hat{S}}}_1^+(a) + \underline{\underline{\hat{\sigma}}}_1^{+t}(a, w)$$
$$\underline{\underline{\hat{S}}}_2^+(w) = \underline{\underline{\hat{r}}}_2(w, a)\underline{\underline{\hat{S}}}_2^-(w) + \underline{\underline{\hat{t}}}_2(a, w)\underline{\underline{\hat{S}}}_2^+(a) + \underline{\underline{\hat{\sigma}}}_2^{+t}(a, w)\,.$$

The global interaction Eqs. (1.137) and (1.139) for $[w, \zeta]$ are

$$\begin{aligned}\underline{\underline{\hat{S}}}_1^-(w) &= \underline{\underline{\hat{R}}}_1(w, \zeta)\underline{\underline{\hat{S}}}_1^+(w) + \underline{\underline{\hat{T}}}_1(\zeta, w)\underline{\underline{\hat{S}}}_1^-(\zeta)\\ &\quad - \underline{\underline{\hat{R}}}_2(w, \zeta)\underline{\underline{\hat{S}}}_2^+(w) - \underline{\underline{\hat{T}}}_2(\zeta, w)\underline{\underline{\hat{S}}}_2^-(\zeta) + \underline{\underline{\hat{\Sigma}}}_1^{-t}(\zeta, w)\\ \underline{\underline{\hat{S}}}_2^-(w) &= \underline{\underline{\hat{R}}}_1(w, \zeta)\underline{\underline{\hat{S}}}_2^+(w) + \underline{\underline{\hat{T}}}_1(\zeta, w)\underline{\underline{\hat{S}}}_2^-(\zeta)\\ &\quad + \underline{\underline{\hat{R}}}_2(w, \zeta)\underline{\underline{\hat{S}}}_1^+(w) + \underline{\underline{\hat{T}}}_2(\zeta, w)\underline{\underline{\hat{S}}}_1^-(\zeta) + \underline{\underline{\hat{\Sigma}}}_2^{-t}(\zeta, w)\,.\end{aligned}$$

These four equations can be solved for the four internal radiances $\underline{\underline{\hat{S}}}_{1,2}^{\pm}(w)$ in terms of the four incident radiances $\underline{\underline{\hat{S}}}_{1,2}^{+}(a)$ and $\underline{\underline{\hat{S}}}_{1,2}^{-}(\zeta)$ and the given internal sources. For simplicity of notation in obtaining this solution, let $r_p = \underline{\underline{\hat{r}}}_p(w, a)$, $t_p = \underline{\underline{\hat{t}}}_p(a, w)$, $R_1 = \underline{\underline{\hat{R}}}_1(w, \zeta)$,..., $T_2 = \underline{\underline{\hat{T}}}_2(\zeta, w)$, $\sigma_p = \underline{\underline{\hat{\sigma}}}_p^{+t}(a, w)$ and $\Sigma_p = \underline{\underline{\hat{\Sigma}}}_p^{-t}(\zeta, w)$. Keep in mind that r_1 etc. are all $4M(N+1) \times 4M(N+1)$ composite matrices, and that σ_p and Σ_p are $4M(N+1) \times 1$. These equations then can be placed in matrix form as

$$\begin{bmatrix}\underline{\underline{\hat{S}}}_1^+(w)\\ \underline{\underline{\hat{S}}}_2^+(w)\\ \underline{\underline{\hat{S}}}_1^-(w)\\ \underline{\underline{\hat{S}}}_2^-(w)\end{bmatrix} = \begin{bmatrix}0 & 0 & r_1 & 0\\ 0 & 0 & 0 & r_2\\ R_1 & -R_2 & 0 & 0\\ R_2 & R_1 & 0 & 0\end{bmatrix}\begin{bmatrix}\underline{\underline{\hat{S}}}_1^+(w)\\ \underline{\underline{\hat{S}}}_2^+(w)\\ \underline{\underline{\hat{S}}}_1^-(w)\\ \underline{\underline{\hat{S}}}_2^-(w)\end{bmatrix} + \begin{bmatrix}t_1 & 0 & 0 & 0\\ 0 & t_2 & 0 & 0\\ 0 & 0 & T_1 & -T_2\\ 0 & 0 & T_2 & T_1\end{bmatrix}\begin{bmatrix}\underline{\underline{\hat{S}}}_1^+(a)\\ \underline{\underline{\hat{S}}}_2^+(a)\\ \underline{\underline{\hat{S}}}_1^-(\zeta)\\ \underline{\underline{\hat{S}}}_2^-(\zeta)\end{bmatrix} + \begin{bmatrix}\sigma_1\\ \sigma_2\\ \Sigma_1\\ \Sigma_2\end{bmatrix}.$$

The solution is then

$$\begin{bmatrix}\underline{\underline{\hat{S}}}_1^+(w)\\ \underline{\underline{\hat{S}}}_2^+(w)\\ \underline{\underline{\hat{S}}}_1^-(w)\\ \underline{\underline{\hat{S}}}_2^-(w)\end{bmatrix} = \begin{bmatrix}I & 0 & -r_1 & 0\\ 0 & I & 0 & -r_2\\ -R_1 & R_2 & I & 0\\ -R_2 & -R_1 & 0 & I\end{bmatrix}^{-1}\left\{\begin{bmatrix}t_1 & 0 & 0 & 0\\ 0 & t_2 & 0 & 0\\ 0 & 0 & T_1 & -T_2\\ 0 & 0 & T_2 & T_1\end{bmatrix}\begin{bmatrix}\underline{\underline{\hat{S}}}_1^+(a)\\ \underline{\underline{\hat{S}}}_2^+(a)\\ \underline{\underline{\hat{S}}}_1^-(\zeta)\\ \underline{\underline{\hat{S}}}_2^-(\zeta)\end{bmatrix} + \begin{bmatrix}\sigma_1\\ \sigma_2\\ \Sigma_1\\ \Sigma_2\end{bmatrix}\right\}. \tag{1.151}$$

The needed inverse matrix can be obtained using the block matrix inversion formula

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix} \tag{1.152}$$

where

$$A = D = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}, \; B = \begin{bmatrix} -r_1 & 0 \\ 0 & -r_2 \end{bmatrix}, \text{ and } C = \begin{bmatrix} -R_1 & R_2 \\ -R_2 & -R_1 \end{bmatrix}.$$

The result is

$$\begin{bmatrix} I & B \\ C & I \end{bmatrix}^{-1} = \begin{bmatrix} (I - BC)^{-1} & -B(I - CB)^{-1} \\ -C(I - BC)^{-1} & (I - CB)^{-1} \end{bmatrix} \tag{1.153}$$

To return to the 4×4 form of Eq. (1.151), apply Eq. (1.152) to each of the blocks of the inverse:

$$\begin{aligned}
(I - BC)^{-1} &= \begin{bmatrix} I - r_1R_1 & r_1R_2 \\ -r_2R_2 & I - r_2R_1 \end{bmatrix}^{-1} \\
&= \begin{bmatrix} [I - r_1R_1 + r_1R_2(I - r_2R_1)^{-1}r_2R_2]^{-1} & \\ (I - r_2R_1)^{-1}r_2R_2[I - r_1R_1 + r_1R_2(I - r_2R_1)^{-1}r_2R_2]^{-1} & \\ -(I - r_1R_1)^{-1}r_1R_2[I - r_2R_1 + r_2R_2(I - r_1R_1)^{-1}r_1R_2]^{-1} & \\ [I - r_2R_1 + r_2R_2(I - r_1R_1)^{-1}r_1R_2]^{-1} & \end{bmatrix} \\
&\equiv \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}
\end{aligned} \tag{1.154}$$

$$-C(I - BC)^{-1} = \begin{bmatrix} R_1 & -R_2 \\ R_2 & R_1 \end{bmatrix} \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} = \begin{bmatrix} R_1m_{11} - R_2m_{21} & R_1m_{12} - R_2m_{22} \\ R_2m_{11} + R_1m_{21} & R_2m_{12} + R_1m_{22} \end{bmatrix}$$

$$\begin{aligned}
(I - CB)^{-1} &= \begin{bmatrix} I - R_1r_1 & R_2r_2 \\ -R_2r_1 & I - R_1r_2 \end{bmatrix}^{-1} \\
&= \begin{bmatrix} [I - R_1r_1 + R_2r_2(I - R_1r_2)^{-1}R_2r_1]^{-1} & \\ (I - R_1r_2)^{-1}R_2r_1[I - R_1r_1 + R_2r_2(I - R_1r_2)^{-1}R_2r_1]^{-1} & \\ -(I - R_1r_1)^{-1}R_2r_2[I - R_1r_2 + R_2r_1(I - R_1r_1)^{-1}R_2r_2]^{-1} & \\ [I - R_1r_2 + R_2r_1(I - R_1r_1)^{-1}R_2r_2]^{-1} & \end{bmatrix} \\
&\equiv \begin{bmatrix} m_{33} & m_{34} \\ m_{43} & m_{44} \end{bmatrix}
\end{aligned} \tag{1.155}$$

and

$$-B(I - CB)^{-1} = \begin{bmatrix} r_1 & 0 \\ 0 & r_2 \end{bmatrix} \begin{bmatrix} m_{33} & m_{34} \\ m_{43} & m_{44} \end{bmatrix} = \begin{bmatrix} r_1m_{33} & r_1m_{34} \\ r_2m_{43} & r_2m_{44} \end{bmatrix}$$

Equation (1.151) now becomes

$$\begin{bmatrix} \hat{\underline{\underline{S}}}_1^+(w) \\ \hat{\underline{\underline{S}}}_2^+(w) \\ \hat{\underline{\underline{S}}}_1^-(w) \\ \hat{\underline{\underline{S}}}_2^-(w) \end{bmatrix} = \begin{bmatrix} m_{11}t_1 & m_{12}t_2 \\ m_{21}t_1 & m_{22}t_2 \\ (R_1m_{11} - R_2m_{21})t_1 & (R_1m_{12} - R_2m_{22})t_2 \\ (R_2m_{11} + R_1m_{21})t_1 & (R_2m_{12} + R_1m_{22})t_2 \end{bmatrix}$$

$$\begin{bmatrix} r_1m_{33}T_1 + r_1m_{34}T_2 & -r_1m_{33}T_2 + r_1m_{34}T_1 \\ r_2m_{43}T_1 + r_2m_{44}T_2 & -r_2m_{43}T_2 + r_2m_{44}T_1 \\ m_{33}T_1 + m_{34}T_2 & -m_{33}T_2 + m_{34}T_1 \\ m_{43}T_1 + m_{44}T_2 & -m_{43}T_2 + m_{44}T_1 \end{bmatrix} \begin{bmatrix} \hat{\underline{\underline{S}}}_1^+(a) \\ \hat{\underline{\underline{S}}}_2^+(a) \\ \hat{\underline{\underline{S}}}_1^-(\zeta) \\ \hat{\underline{\underline{S}}}_2^-(\zeta) \end{bmatrix}$$

$$+ \begin{bmatrix} m_{11}\sigma_1 + m_{12}\sigma_2 + r_1m_{33}\Sigma_1 + r_1m_{34}\Sigma_2 \\ m_{21}\sigma_1 + m_{22}\sigma_2 + r_2m_{43}\Sigma_1 + r_2m_{44}\Sigma_2 \\ (R_1m_{11} - R_2m_{21})\sigma_1 + (R_1m_{12} - R_2m_{22})\sigma_2 + m_{33}\Sigma_1 + m_{34}\Sigma_2 \\ (R_2m_{11} + R_1m_{21})\sigma_1 + (R_2m_{12} + R_1m_{22})\sigma_2 + m_{43}\Sigma_1 + m_{44}\Sigma_2 \end{bmatrix} \tag{1.156}$$

1.6.3.1 Invariant Imbedding Relations and Imbed Rules for the Surface Plus the Water Body

Following the notation of *Light and Water* (L&W 7.65–7.72 and 8.95–8.97), the matrix elements of the Eq. (1.156) are rewritten as (L&W 7.65–7.66)

$$\begin{bmatrix} \hat{\underline{\underline{S}}}_1^+(w) \\ \hat{\underline{\underline{S}}}_2^+(w) \\ \hat{\underline{\underline{S}}}_1^-(w) \\ \hat{\underline{\underline{S}}}_2^-(w) \end{bmatrix} \equiv \begin{bmatrix} \mathfrak{T}_{11}(a,w,\zeta) & \mathfrak{T}_{12}(a,w,\zeta) & \mathfrak{R}_{11}(\zeta,w,a) & \mathfrak{R}_{12}(\zeta,w,a) \\ \mathfrak{T}_{21}(a,w,\zeta) & \mathfrak{T}_{22}(a,w,\zeta) & \mathfrak{R}_{21}(\zeta,w,a) & \mathfrak{R}_{22}(\zeta,w,a) \\ \mathfrak{R}_{11}(a,w,\zeta) & \mathfrak{R}_{12}(a,w,\zeta) & \mathfrak{T}_{11}(\zeta,w,a) & \mathfrak{T}_{12}(\zeta,w,a) \\ \mathfrak{R}_{21}(a,w,\zeta) & \mathfrak{R}_{22}(a,w,\zeta) & \mathfrak{T}_{21}(\zeta,w,a) & \mathfrak{T}_{22}(\zeta,w,a) \end{bmatrix} \begin{bmatrix} \hat{\underline{\underline{S}}}_1^+(a) \\ \hat{\underline{\underline{S}}}_2^+(a) \\ \hat{\underline{\underline{S}}}_1^-(\zeta) \\ \hat{\underline{\underline{S}}}_2^-(\zeta) \end{bmatrix}$$

$$+ \begin{bmatrix} \mathfrak{s}_1(a,w,\zeta) \\ \mathfrak{s}_2(a,w,\zeta) \\ \mathfrak{S}_1(\zeta,w,a) \\ \mathfrak{S}_2(\zeta,w,a) \end{bmatrix} \tag{1.157}$$

The equations contained in (1.157) are the *invariant imbedding relations* for $[a, \zeta]$. These relations hold for any internal depth w, $a \le w \le \zeta$. As a mnemonic aid, note that depth w, which appears on the left side of the equation, in imbedded within (lies between) the boundary depths a and ζ. *Invariant imbedding relations show how to compute the radiances at any internal depth, given the radiances incident onto a slab from above and below.*

The $\mathfrak{T}_{ij}(a, w, \zeta)$ are known as the *complete transmittances for downwelling radiance* for the combined slab $[a, \zeta]$. Note that these matrices transmit downwelling radiances from the surface at depth a to depth w, accounting for the infinite series of internal reflections within the combined slabs. $\mathfrak{R}_{ij}(a, w, \zeta)$ are the *complete reflectances for downwelling radiance* incident onto the sea surface. Likewise, $\mathfrak{T}_{ij}(\zeta, w, a)$ are the complete transmittances for upwelling radiance incident onto the bottom of the combined slab, and $\mathfrak{R}_{ij}(\zeta, w, a)$ are the corresponding complete reflectances for upwelling radiance. The $\mathfrak{s}_{1,2}(a, w, \zeta)$ are the *complete source-induced downwelling radiances*, and $\mathfrak{S}_{1,2}(\zeta, w, a)$ are the *complete source-induced upwelling radiances*. The order of (a, w, ζ) or (ζ, w, a) shows the direction of the incident or internal-source radiance. The subscripts tell how cosine and sine amplitudes are transformed as they are transmitted or reflected. Thus $\mathfrak{T}_{11}(a, w, \zeta)$ transmits cosine amplitudes from a to w, $\mathfrak{T}_{12}(a, w, \zeta)$ transmits from a to w and turns sine amplitudes into cosine amplitudes in the process, $\mathfrak{T}_{21}(\zeta, w, a)$ transmits from ζ to w and converts cosine amplitudes to sine amplitudes, and so on.

Writing out the complete reflectances, transmittances, and source matrices in terms of the standard matrices gives (L&W 7.67–7.69)

$$\begin{aligned}\mathfrak{T}_{11}(\zeta, w, a) &= [I - R_1 r_1 + R_2 r_2 (I - R_1 r_2)^{-1} R_2 r_1]^{-1} T_1 \\ &\quad -(I - R_1 r_1)^{-1} R_2 r_2 [I - R_1 r_2 + R_2 r_1 (I - R_1 r_1)^{-1} R_2 r_2]^{-1} T_2 \end{aligned} \tag{1.158a}$$

$$\begin{aligned}\mathfrak{T}_{12}(\zeta, w, a) &= -[I - R_1 r_1 + R_2 r_2 (I - R_1 r_2)^{-1} R_2 r_1]^{-1} T_2 \\ &\quad -(I - R_1 r_1)^{-1} R_2 r_2 [I - R_1 r_2 + R_2 r_1 (I - R_1 r_1)^{-1} R_2 r_2]^{-1} T_1 \end{aligned} \tag{1.158b}$$

$$\begin{aligned}\mathfrak{T}_{21}(\zeta, w, a) &= (I - R_1 r_2)^{-1} R_2 r_1 [I - R_1 r_1 + R_2 r_2 (I - R_1 r_2)^{-1} R_2 r_1]^{-1} T_1 \\ &\quad +[I - R_1 r_2 + R_2 r_1 (I - R_1 r_1)^{-1} R_2 r_2]^{-1} T_2 \end{aligned} \tag{1.158c}$$

$$\begin{aligned}\mathfrak{T}_{22}(\zeta, w, a) &= -(I - R_1 r_2)^{-1} R_2 r_1 [I - R_1 r_1 + R_2 r_2 (I - R_1 r_2)^{-1} R_2 r_1]^{-1} T_2 \\ &\quad +[I - R_1 r_2 + R_2 r_1 (I - R_1 r_1)^{-1} R_2 r_2]^{-1} T_1 \end{aligned} \tag{1.158d}$$

$$\mathfrak{R}_{11}(\zeta, w, a) = r_1 \mathfrak{T}_{11}(\zeta, w, a) \tag{1.158e}$$

$$\mathfrak{R}_{12}(\zeta, w, a) = r_1 \mathfrak{T}_{12}(\zeta, w, a) \tag{1.158f}$$

$$\mathfrak{R}_{21}(\zeta, w, a) = r_2 \mathfrak{T}_{21}(\zeta, w, a) \tag{1.158g}$$

$$\mathfrak{R}_{22}(\zeta, w, a) = r_2 \mathfrak{T}_{22}(\zeta, w, a) \tag{1.158h}$$

$$\begin{aligned}\mathfrak{S}_1(\zeta, w, a) &= (R_1 m_{11} - R_2 m_{21}) \sigma_1 \\ &\quad +(R_1 m_{12} - R_2 m_{22}) \sigma_2 + m_{33} \Sigma_1 + m_{34} \Sigma_2 \end{aligned} \tag{1.158i}$$

$$\begin{aligned}\mathfrak{S}_2(\zeta, w, a) &= (R_2 m_{11} + R_1 m_{21}) \sigma_1 \\ &\quad +(R_2 m_{12} + R_1 m_{22}) \sigma_2 + m_{43} \Sigma_1 + m_{44} \Sigma_2 \end{aligned} \tag{1.158j}$$

and (L&W 7.70–7.72):

$$\mathfrak{T}_{11}(a,w,\zeta) = [I - r_1R_1 + r_1R_2(I - r_2R_1)^{-1}r_2R_2]^{-1}t_1 \tag{1.159a}$$

$$\mathfrak{T}_{12}(a,w,\zeta) = -(I - r_1R_1)^{-1}r_1R_2[I - r_2R_2 + r_2R_2(I - r_1R_1)^{-1}r_1R_2]^{-1}t_2 \tag{1.159b}$$

$$\mathfrak{T}_{21}(a,w,\zeta) = [(I - r_2R_1)^{-1}r_2R_2[I - r_1R_1 + r_1R_2(I - r_2R_1)^{-1}r_2R_2]t_1 \tag{1.159c}$$

$$\mathfrak{T}_{22}(a,w,\zeta) = [I - r_2R_1 + r_2R_2(I - r_1R_1)^{-1}r_1R_2]^{-1}t_2 \tag{1.159d}$$

$$\mathfrak{R}_{11}(a,w,\zeta) = R_1\mathfrak{T}_{11}(a,w,\zeta) - R_2\mathfrak{T}_{21}(a,w,\zeta) \tag{1.159e}$$

$$\mathfrak{R}_{12}(a,w,\zeta) = R_1\mathfrak{T}_{12}(a,w,\zeta) - R_2\mathfrak{T}_{22}(a,w,\zeta) \tag{1.159f}$$

$$\mathfrak{R}_{21}(a,w,\zeta) = R_2\mathfrak{T}_{11}(a,w,\zeta) + R_1\mathfrak{T}_{21}(a,w,\zeta) \tag{1.159g}$$

$$\mathfrak{R}_{22}(a,w,\zeta) = R_2\mathfrak{T}_{12}(a,w,\zeta) + R_1\mathfrak{T}_{22}(a,w,\zeta) \tag{1.159h}$$

$$\mathfrak{s}_1(a,w,\zeta) = m_{11}\sigma_1 + m_{12}\sigma_2 + r_1m_{33}\Sigma_1 + r_1m_{34}\Sigma_2 \tag{1.159i}$$

$$\mathfrak{s}_2(a,w,\zeta) = m_{21}\sigma_1 + m_{22}\sigma_2 + r_2m_{43}\Sigma_1 + r_2m_{44}\Sigma_2 \tag{1.159j}$$

The source terms are left in terms of the m_{ij} matrices simply because these equations are too lengthy to write out in full.

These equations are the *imbed rules* for the combined slabs $[a, w]$ and $[w, \zeta]$. *Imbed rules show how standard operators for two slabs are combined to generate complete operators for the composite slab.*

1.6.3.2 Interpretation of the Complete Operators

It is worthwhile to consider the physical interpretation of the complete operators defined in Eqs. (1.158) and (1.159). Consider, for example, $\mathfrak{T}_{11}(a,w,\zeta)$, which according to Eq. (1.157) converts $\hat{\underline{\underline{S}}}_1^+(a)$ into $\hat{\underline{\underline{S}}}_1^+(w)$. Grouping the terms of Eq. (1.159a) as

$$\mathfrak{T}_{11}(a,w,\zeta) = \Big(\underbrace{I - r_1R_1}_{A} + \underbrace{r_1R_2(I - r_2R_1)^{-1}r_2R_2}_{B}\Big)^{-1}t_1\,, \tag{1.160}$$

this inverse can be expanded using the formula

$$(A + B)^{-1} = A^{-1} - A^{-1}B(I + A^{-1}B)^{-1}A^{-1}\,.$$

for the inverse of the sum of two matrices. Applying this expansion to $\mathfrak{T}_{11}(a,w,\zeta)$ with A and B as identified in Eq. (1.160) gives

$$\begin{aligned}\mathfrak{T}_{11}(a,w,\zeta) = {} & (I - r_1R_1)^{-1}t_1 \\ & - (I - r_1R_1)^{-1}r_1R_2(I - r_2R_1)^{-1}r_2R_2 \\ & \times \Big(I + (I - r_1R_1)^{-1}r_1R_2(I - r_2R_1)^{-1}r_2R_2\Big)(I - r_1R_1)^{-1}t_1\,.\end{aligned} \tag{1.161}$$

Now consider the first term on the right hand side of this equation. Expanding the inverse of $(I - r_1 R_1)$ using the formula

$$(I - A)^{-1} = I + A + AA + AAA + \cdots$$

for the inverse of the identity minus a matrix, and reinserting the depth arguments, gives

$$\begin{aligned}\Big(I - r_1(a,w)R_1(w,\zeta)\Big)^{-1} t_1(a,w) \\ = t_1(a,w) + r_1(a,w)R_1(w,\zeta)t_1(a,w) \\ + r_1(a,w)R_1(w,\zeta)r_1(a,w)R_1(w,\zeta)t_1(a,w) + \cdots \end{aligned} \tag{1.162}$$

The upper panel of Fig. 1.10 shows the meaning of these terms when $\mathfrak{T}_{11}(a,w,\zeta)$ operates on the incident radiance $\underline{\underline{\hat{S}}}_1^+(a)$. (This figure is similar to L&W Fig. 7.1, which includes the bottom boundary.) The first term of the expansion, $t_1(a,w)$ simply transmits $\underline{\underline{\hat{S}}}_1^+(a)$ from above the sea surface at depth a to just below the surface at depth w. The second term describes radiance that is transmitted through the surface by $t_1(a,w)$, then reflected upward by the water between w and depth ζ by the operation of $R_1(w,\zeta)$. This upward reflection creates upwelling radiance $\underline{\underline{\hat{S}}}_1^-(w)$ at depth w. That radiance is then reflected back downward by the air-water surface via the operation of $r_1(w,a)$. The end result is another contribution to $\underline{\underline{\hat{S}}}_1^+(w)$. The next term of the expansion represents another cycle of upward reflection by the water body followed by downward reflection by the sea surface.

Now consider the expansion of the second term on the right hand side of Eq. (1.161). Expanding the $(I - r_i R_j)^{-1}$ terms leads to the sequence of terms

$$\begin{aligned}\text{rhs second term} = -r_1(w,a)R_2(w,\zeta)r_2(w,a)R_2(w,\zeta)t_1(a,w) \\ - r_1(w,a)R_2(w,\zeta)r_2(w,a)R_1(w,\zeta)r_1(w,a)R_2(w,\zeta)r_2(w,a)R_1(w,\zeta)t_1(a,w) + \cdots\end{aligned}$$

The bottom panel of Fig. 1.10 shows the meaning of the first group of these terms. First $t_1(a,w)$ transmits $\underline{\underline{\hat{S}}}_1^+(a)$ through the sea surface as before. Then reflection by $R_2(w,\zeta)$ converts the downwelling cosine amplitude $\underline{\underline{\hat{S}}}_1^+(w)$ into an upwelling sine amplitude $\underline{\underline{\hat{S}}}_2^-(w)$; recall that the standard operators for the water body R_2 and T_2 convert cosine amplitudes to sine amplitudes and vice versa. $\underline{\underline{\hat{S}}}_2^-(w)$ is then reflected back downward by the water surface by $r_2(w,a)$ to create $\underline{\underline{\hat{S}}}_2^+(w)$. That downwelling sine amplitude is then reflected back upward by the water body via $R_2(w,\zeta)$, which converts the sine back to a cosine amplitude $\underline{\underline{\hat{S}}}_1^-(w)$, which is reflected back downward by $r_1(w,a)$. The end result is another contribution to $\underline{\underline{\hat{S}}}_1^+(w)$. These terms do not occur is the SRTE because $R_2 = 0$, but *in the VRTE these terms describe conversions between various components of the Stokes vectors by the elements of the phase matrix*.

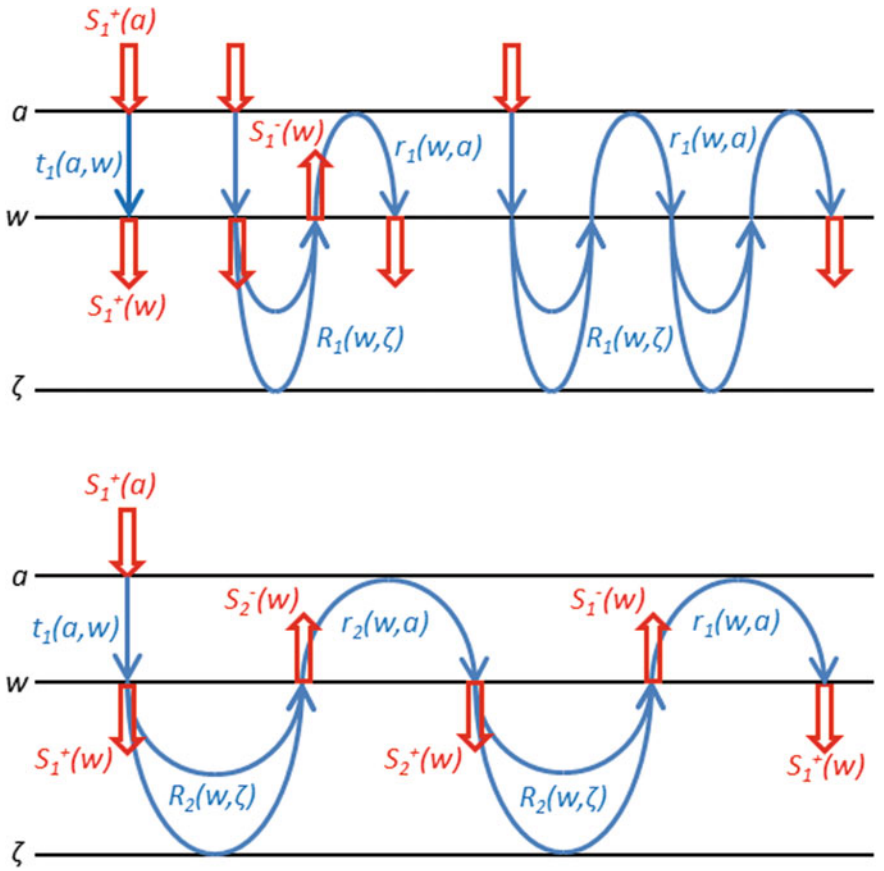

Fig. 1.10 Graphical interpretation of the leading terms in the expansion of $\mathfrak{T}_{11}(a, w, \zeta)$

As another example, consider $\mathfrak{R}_{11}(\zeta, w, a)$, which according to Eq. (1.157) converts $\underline{\underline{\hat{S}}}_1^-(\zeta)$ into $\underline{\underline{\hat{S}}}_1^+(w)$. Performing a similar expansion of terms gives

$$\begin{aligned}\mathfrak{R}_{11}(\zeta, w, a) &= r_1[I - R_1 r_1 + R_2 r_2 (I - R_1 r_2)^{-1} R_2 r_1]^{-1} T_1 \\ &= r_1(w,a)T_1(\zeta,w) + r_1(w,a)R_1(w,\zeta)r_1(w,a)T_1(\zeta,w) + \cdots \\ &\quad - r_1(w,a)R_2(\zeta,w)r_2(w,a)T_2(\zeta,w) \\ &\quad - r_1(w,a)R_2(w,\zeta)r_2(w,a)R_1(w,\zeta)r_2(w,a)T_2(\zeta,w) + \cdots\end{aligned}$$

The first series of terms is equivalent to the corresponding terms in the SRTE: $T_1(\zeta, w)$ transmits upwelling radiance $\underline{\underline{\hat{S}}}_1^-(\zeta)$ from ζ to the bottom of the sea surface, where it is reflected back down by $r_1(w, a)$ to create $\underline{\underline{\hat{S}}}_1^+(w)$, and so on. Figure 1.11 shows the effect of the second group of terms involving $T_2(\zeta, w)$. Now $T_2(\zeta, w)$ transmits

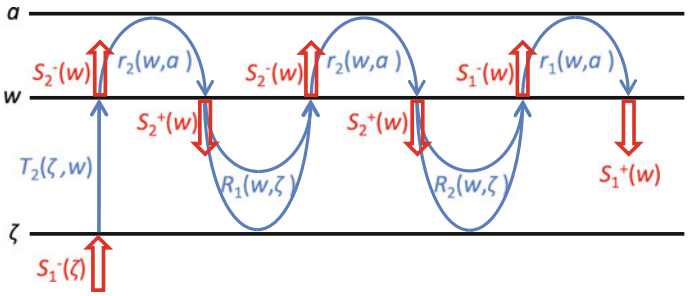

Fig. 1.11 Graphical interpretation of one group of terms in the expansion of $\mathfrak{R}_{11}(\zeta, w, a)$

radiance from ζ to w, but converts cosine amplitudes to sine amplitudes. These are then reflected back down by $r_2(w, a)$. $R_1(\zeta, w)$ then reflects the downwelling $\underline{\underline{\hat{S}}}_2^+(w)$ back upward, leaving it as a sine amplitude. Another reflection by $r_2(w, a)$ gives a downwelling sign amplitude, which is the reflected back upward by the water body and converted back to a cosine amplitude in the process. A final reflection by the sea surface gives another contribution to $\underline{\underline{\hat{S}}}_1^+(w)$. This process does not occur in the SRTE because $T_2 = 0$.

The other complete operators have similar interpretations. *These infinite series of inter-reflections by the water body and the sea surface shows how invariant imbedding theory automatically accounts for all orders of multiple scattering within a water body and between the water and the surface (or bottom) boundary.* This result can be contrasted with the successive-order-scattering solution method. That method solves first for terms that represent unscattered radiance, then radiance scattered once, then radiance scattered twice, and so on. The solution process is then terminated after some number n of scatterings (often n is just 2 or 3). This gives an approximate solution of the RTE that represents the radiance due to a finite number of scatterings. That is, the successive-order-of-scattering solution approximates an infinite series of scatterings by adding up the first n terms. Invariant imbedding, on the other hand, solves directly for the analytical sum of the infinite series of scatterings. Thus during computations, the matrix seen on the right hand side of Eq. (1.160) is inverted. The expansion of $(I - A)^{-1}$ as an infinite series is used here only as an aid to the physical interpretation of these matrices; the inverse is not computed from a series expansion. The same holds true for the other matrices on the right had sides of Eqs. (1.158) and (1.159).

1.6.3.3 Global Interaction Equations and Union Rules for the Surface Plus Water Body

We can now take the second step of the present derivation by using Eq. (1.157) to replace the internal radiances in Eq. (1.135) with $p = 1, 2$, (1.138), and (1.140),

which are (including for complete generality the possibility of internal sources in $[a, w]$)

$$\begin{aligned}
\underline{\underline{\hat{S}}}_1^-(a) &= \underline{\underline{\hat{r}}}_1(a,w)\underline{\underline{\hat{S}}}_1^+(a) + \underline{\underline{\hat{t}}}_1(w,a)\underline{\underline{\hat{S}}}_1^-(w) + \underline{\underline{\hat{\sigma}}}_1^{-t}(w,a)\\
\underline{\underline{\hat{S}}}_2^-(a) &= \underline{\underline{\hat{r}}}_2(a,w)\underline{\underline{\hat{S}}}_2^+(a) + \underline{\underline{\hat{t}}}_2(w,a)\underline{\underline{\hat{S}}}_2^-(w) + \underline{\underline{\hat{\sigma}}}_2^{-t}(w,a)\\
\underline{\underline{\hat{S}}}_1^+(\zeta) &= \underline{\underline{\hat{R}}}_1(\zeta,w)\underline{\underline{\hat{S}}}_1^-(\zeta) + \underline{\underline{\hat{T}}}_1(w,\zeta)\underline{\underline{\hat{S}}}_1^+(w)\\
&\quad - \underline{\underline{\hat{R}}}_2(\zeta,w)\underline{\underline{\hat{S}}}_2^-(\zeta) - \underline{\underline{\hat{T}}}_2(w,\zeta)\underline{\underline{\hat{S}}}_2^+(w) + \underline{\underline{\hat{\Sigma}}}_1^{+t}(w,\zeta)\\
\underline{\underline{\hat{S}}}_2^+(\zeta) &= \underline{\underline{\hat{R}}}_1(\zeta,w)\underline{\underline{\hat{S}}}_2^-(\zeta) + \underline{\underline{\hat{T}}}_1(w,\zeta)\underline{\underline{\hat{S}}}_2^+(w)\\
&\quad + \underline{\underline{\hat{R}}}_2(\zeta,w)\underline{\underline{\hat{S}}}_1^-(\zeta) + \underline{\underline{\hat{T}}}_2(w,\zeta)\underline{\underline{\hat{S}}}_1^+(w) + \underline{\underline{\hat{\Sigma}}}_2^{+t}(w,\zeta)
\end{aligned}$$

Placing these equations in matrix form (again omitting the full matrix amplitude notation for brevity, so that $t_1(w,a)$ here represents $\underline{\underline{\hat{t}}}_1(w,a)$, etc.) gives

$$\begin{bmatrix} \underline{\underline{\hat{S}}}_1^-(a)\\ \underline{\underline{\hat{S}}}_2^-(a)\\ \underline{\underline{\hat{S}}}_1^+(\zeta)\\ \underline{\underline{\hat{S}}}_2^+(\zeta)\end{bmatrix} = \begin{bmatrix} 0 & 0 & t_1(w,a) & 0\\ 0 & 0 & 0 & t_2(w,a)\\ T_1(w,\zeta) & -T_2(w,\zeta) & 0 & 0\\ T_2(w,\zeta) & T_1(w,\zeta) & 0 & 0\end{bmatrix}\begin{bmatrix} \underline{\underline{\hat{S}}}_1^+(w)\\ \underline{\underline{\hat{S}}}_2^+(w)\\ \underline{\underline{\hat{S}}}_1^-(w)\\ \underline{\underline{\hat{S}}}_2^-(w)\end{bmatrix}$$
$$+\begin{bmatrix} r_1(a,w) & 0 & 0 & 0\\ 0 & r_2(a,w) & 0 & 0\\ 0 & 0 & R_1(\zeta,w) & -R_2(\zeta,w)\\ 0 & 0 & R_2(\zeta,w) & R_1(\zeta,w)\end{bmatrix}\begin{bmatrix} \underline{\underline{\hat{S}}}_1^+(a)\\ \underline{\underline{\hat{S}}}_2^+(a)\\ \underline{\underline{\hat{S}}}_1^-(\zeta)\\ \underline{\underline{\hat{S}}}_2^-(\zeta)\end{bmatrix} + \begin{bmatrix} \underline{\underline{\hat{\sigma}}}_1^{-t}(w,a)\\ \underline{\underline{\hat{\sigma}}}_2^{-t}(w,a)\\ \underline{\underline{\hat{\Sigma}}}_1^{+t}(w,\zeta)\\ \underline{\underline{\hat{\Sigma}}}_2^{+t}(w,\zeta)\end{bmatrix}$$

The invariant imbedding Eq. (1.157) is now used to replace the column vector of internal radiances at depth w in this equation. Carrying out the matrix multiplications gives (L&W 7.73, 7.74)

$$\begin{bmatrix} \underline{\underline{\hat{S}}}_1^-(a)\\ \underline{\underline{\hat{S}}}_2^-(a)\\ \underline{\underline{\hat{S}}}_1^+(\zeta)\\ \underline{\underline{\hat{S}}}_2^+(\zeta)\end{bmatrix} = \begin{bmatrix} t_1(w,a)\mathfrak{R}_{11}(a,w,\zeta) + r_1(a,w)\\ t_2(w,a)\mathfrak{R}_{21}(a,w,\zeta)\\ T_1(w,\zeta)\mathfrak{T}_{11}(a,w,\zeta) - T_2(w,\zeta)\mathfrak{T}_{21}(a,w,\zeta)\\ T_2(w,\zeta)\mathfrak{T}_{11}(a,w,\zeta) + T_1(w,\zeta)\mathfrak{T}_{21}(a,w,\zeta)\end{bmatrix}$$

$$\begin{bmatrix} t_1(w,a)\mathfrak{R}_{12}(a,w,\zeta)\\ t_2(w,a)\mathfrak{R}_{22}(a,w,\zeta) + r_2(a,w)\\ T_1(w,\zeta)\mathfrak{T}_{12}(a,w,\zeta) - T_2(w,\zeta)\mathfrak{T}_{22}(a,w,\zeta)\\ T_2(w,\zeta)\mathfrak{T}_{12}(a,w,\zeta) + T_1(w,\zeta)\mathfrak{T}_{22}(a,w,\zeta)\end{bmatrix}$$

$$
\begin{bmatrix}
t_1(w,a)\mathfrak{T}_{11}(\zeta,w,a) \\
t_2(w,a)\mathfrak{T}_{21}(\zeta,w,a) \\
T_1(w,\zeta)\mathfrak{R}_{11}(\zeta,w,a) - T_2(w,\zeta)\mathfrak{R}_{21}(\zeta,w,a) + R_1(\zeta,w) \\
T_2(w,\zeta)\mathfrak{R}_{11}(\zeta,w,a) + T_1(w,\zeta)\mathfrak{R}_{21}(\zeta,w,a) + R_2(\zeta,w)
\end{bmatrix}
$$

$$
\begin{bmatrix}
t_1(w,a)\mathfrak{T}_{12}(\zeta,w,a) \\
t_2(w,a)\mathfrak{T}_{22}(\zeta,w,a) \\
T_1(w,\zeta)\mathfrak{R}_{12}(\zeta,w,a) - T_2(w,\zeta)\mathfrak{R}_{22}(\zeta,w,a) - R_2(\zeta,w) \\
T_2(w,\zeta)\mathfrak{R}_{12}(a,w,\zeta) + T_1(w,\zeta)\mathfrak{R}_{22}(a,w,\zeta) + R_1(\zeta,w)
\end{bmatrix}
\begin{bmatrix}
\hat{\underline{\underline{S}}}_1^+(a) \\
\hat{\underline{\underline{S}}}_2^+(a) \\
\hat{\underline{\underline{S}}}_1^-(\zeta) \\
\hat{\underline{\underline{S}}}_2^-(\zeta)
\end{bmatrix}
$$

$$
+
\begin{bmatrix}
t_1(w,a)\mathfrak{S}_1(\zeta,w,a) + \hat{\underline{\underline{\sigma}}}_1^{-t}(w,a) \\
t_2(w,a)\mathfrak{S}_2(\zeta,w,a) + \hat{\underline{\underline{\sigma}}}_2^{-t}(w,a) \\
T_1(w,\zeta)\mathfrak{s}_1(a,w,\zeta) - T_2(w,\zeta)\mathfrak{s}_2(a,w,\zeta) + \hat{\underline{\underline{\Sigma}}}_1^{+t}(w,\zeta) \\
T_2(w,\zeta)\mathfrak{s}_1(a,w,\zeta) + T_1(w,\zeta)\mathfrak{s}_2(a,w,\zeta) + \hat{\underline{\underline{\Sigma}}}_2^{+t}(w,\zeta)
\end{bmatrix}
\tag{1.163}
$$

These equations show how the composite slab $[a,\zeta]$ responds to incident radiances; they are therefore the *global interaction equations for the composite slab*. Global interaction equations are written in terms of standard operators, e.g. as in Eqs. (1.137)–(1.140). The last equation can therefore be rewritten as

$$
\begin{bmatrix}
\hat{\underline{\underline{S}}}_1^-(a) \\
\hat{\underline{\underline{S}}}_2^-(a) \\
\hat{\underline{\underline{S}}}_1^+(\zeta) \\
\hat{\underline{\underline{S}}}_2^+(\zeta)
\end{bmatrix}
\equiv
\begin{bmatrix}
R_{11}(a,\zeta) & R_{12}(a,\zeta) & T_{11}(\zeta,a) & T_{12}(\zeta,a) \\
R_{21}(a,\zeta) & R_{22}(a,\zeta) & T_{21}(\zeta,a) & T_{22}(\zeta,a) \\
T_{11}(a,\zeta) & T_{12}(a,\zeta) & R_{11}(\zeta,a) & R_{12}(\zeta,a) \\
T_{21}(a,\zeta) & T_{22}(a,\zeta) & R_{21}(\zeta,a) & R_{22}(\zeta,a)
\end{bmatrix}
\begin{bmatrix}
\hat{\underline{\underline{S}}}_1^+(a) \\
\hat{\underline{\underline{S}}}_2^+(a) \\
\hat{\underline{\underline{S}}}_1^-(\zeta) \\
\hat{\underline{\underline{S}}}_2^-(\zeta)
\end{bmatrix}
+
\begin{bmatrix}
\hat{\underline{\underline{\sigma}}}_1^{-t}(\zeta,a) \\
\hat{\underline{\underline{\sigma}}}_2^{-t}(\zeta,a) \\
\hat{\underline{\underline{\Sigma}}}_1^{+t}(a,\zeta) \\
\hat{\underline{\underline{\Sigma}}}_2^{+t}(a,\zeta)
\end{bmatrix}
\tag{1.164}
$$

The equations defining the standard operators for the combined slab are (L&W 7.75–7.77),

$$R_{11}(a,\zeta) = t_1(w,a)\mathfrak{R}_{11}(a,w,\zeta) + r_1(a,w) \tag{1.165a}$$
$$R_{12}(a,\zeta) = t_1(w,a)\mathfrak{R}_{12}(a,w,\zeta) \tag{1.165b}$$
$$R_{21}(a,\zeta) = t_2(w,a)\mathfrak{R}_{21}(a,w,\zeta) \tag{1.165c}$$
$$R_{22}(a,\zeta) = t_2(w,a)\mathfrak{R}_{22}(a,w,\zeta) + r_2(a,w) \tag{1.165d}$$
$$T_{11}(\zeta,a) = t_1(w,a)\mathfrak{T}_{11}(\zeta,w,a) \tag{1.165e}$$
$$T_{12}(\zeta,a) = t_1(w,a)\mathfrak{T}_{12}(\zeta,w,a) \tag{1.165f}$$
$$T_{21}(\zeta,a) = t_2(w,a)\mathfrak{T}_{21}(\zeta,w,a) \tag{1.165g}$$
$$T_{22}(\zeta,a) = t_2(w,a)\mathfrak{T}_{22}(\zeta,w,a) \tag{1.165h}$$
$$\hat{\underline{\underline{\sigma}}}_1^{-t}(\zeta,a) = t_1(w,a)\mathfrak{S}_1(\zeta,w,a) + \hat{\underline{\underline{\sigma}}}_1^{-t}(w,a) \tag{1.165i}$$
$$\hat{\underline{\underline{\sigma}}}_2^{-t}(\zeta,a) = t_2(w,a)\mathfrak{S}_2(\zeta,w,a) + \hat{\underline{\underline{\sigma}}}_2^{-t}(w,a) \tag{1.165j}$$

and (L&W 7.78–7.80)

$$R_{11}(\zeta,a) = T_1(w,\zeta)\mathfrak{R}_{11}(\zeta,w,a) - T_2(w,\zeta)\mathfrak{R}_{21}(\zeta,w,a) + R_1(\zeta,w) \quad (1.166a)$$
$$R_{12}(\zeta,a) = T_1(w,\zeta)\mathfrak{R}_{12}(\zeta,w,a) - T_2(w,\zeta)\mathfrak{R}_{22}(\zeta,w,a) - R_2(\zeta,w) \quad (1.166b)$$
$$R_{21}(\zeta,a) = T_2(w,\zeta)\mathfrak{R}_{11}(\zeta,w,a) + T_1(w,\zeta)\mathfrak{R}_{21}(\zeta,w,a) + R_2(\zeta,w) \quad (1.166c)$$
$$R_{22}(\zeta,a) = T_2(w,\zeta)\mathfrak{R}_{12}(\zeta,w,a) + T_1(w,\zeta)\mathfrak{R}_{22}(\zeta,w,a) + R_1(\zeta,w) \quad (1.166d)$$
$$T_{11}(a,\zeta) = T_1(w,\zeta)\mathfrak{T}_{11}(a,w,\zeta) - T_2(w,\zeta)\mathfrak{T}_{21}(a,w,\zeta) \quad (1.166e)$$
$$T_{12}(a,\zeta) = T_1(w,\zeta)\mathfrak{T}_{12}(a,w,\zeta) - T_2(w,\zeta)\mathfrak{T}_{22}(a,w,\zeta) \quad (1.166f)$$
$$T_{21}(a,\zeta) = T_2(w,\zeta)\mathfrak{T}_{11}(a,w,\zeta) + T_1(w,\zeta)\mathfrak{T}_{21}(a,w,\zeta) \quad (1.166g)$$
$$T_{22}(a,\zeta) = T_2(w,\zeta)\mathfrak{T}_{12}(a,w,\zeta) + T_1(w,\zeta)\mathfrak{T}_{22}(a,w,\zeta) \quad (1.166h)$$

$$\hat{\underline{\underline{\Sigma}}}_1^{+t}(a,\zeta) = T_1(w,\zeta)\mathfrak{s}_1(a,w,\zeta) - T_2(w,\zeta)\mathfrak{s}_2(a,w,\zeta) + \hat{\underline{\underline{\Sigma}}}_1^{+t}(w,\zeta) \quad (1.166i)$$
$$\hat{\underline{\underline{\Sigma}}}_2^{+t}(a,\zeta) = T_2(w,\zeta)\mathfrak{s}_1(a,w,\zeta) + T_1(w,\zeta)\mathfrak{s}_2(a,w,\zeta) + \hat{\underline{\underline{\Sigma}}}_2^{+t}(w,\zeta) \quad (1.166j)$$

Recalling that the imbed rules (1.158) and (1.159) define the complete operators in terms of standard operators, these equations show how the standard operators for two slabs are combined to obtain the standard operators for a composite slab. One can imagine holding a and w fixed and "constructing" the water body by letting ζ increase downward starting at $\zeta = w$. These relations are therefore known as the *downward union rules* for the composite slab.

As always, the VRTE equations should reduce to their SRTE equivalents. Thus $T_{11}(\zeta,a)$ reduces to (the transpose of) (L&W 7.75), $R_{11}(a,\zeta)$ is the equivalent of (L&W 7.76), $R_{12}(a,\zeta) = R_{21}(a,\zeta) = 0$ and have no scalar equivalents, and $\hat{\underline{\underline{\sigma}}}_1^{-t}(\zeta,a)$ reduces to (L&W 7.77). Likewise, $R_{11}(\zeta,a)$ reduces to (L&W 7.79), $T_{11}(a,\zeta)$ is equivalent to (L&W 7.78), and so on.

A review of the preceding developments is warranted.

- The transfer functions (standard operators) for the air-water surface, $r_{1,2}(a,w)$, $r_{1,2}(w,a)$, $t_{1,2}(a,w)$, and $t_{1,2}(w,a)$, are known from the surface boundary conditions.
- The standard operators for the bare slab $[w,\zeta]$, $R_1(w,\zeta)$, etc. are known from the downward sweep integrations of the ODEs (1.96–1.102) beginning with initial conditions (1.103)–(1.106)
- These two sets of standard operators were combined via the imbed rules (1.157) and (1.158)–(1.159). Those equations give the radiances at depth w, $S_{1,2}^{\pm}(w)$, in terms of the incident radiances at depths a and ζ. However, the radiances at ζ are not yet known.
- The global interaction Eqs. (1.163), (1.164) and union rules (1.165j), (1.166j) give the standard operators $R_{11}(a,\zeta)$ etc. for the combined slabs $[a,w] \cup [w,\zeta]$.
- In general the above process would be repeated to combine the bare slab $[\zeta,m]$ with the lower boundary $[m,b]$. However, the lower boundary was assumed to be an

azimuthally isotropic reflecting layer, in which case the Fourier modes decoupled ℓ value by ℓ value. This allowed the standard operators $R_{1,2}(\zeta, b)$ etc. for the slab $[\zeta, m] \cup [m, b]$ to be obtained by the upward sweep integrations of the ODEs (1.107)–(1.118) with initial conditions (1.147)–(1.150) for each ℓ value. Thus the standard operators $R_1(\zeta, b)$ etc. are known from the upward sweep ODE solutions and the known bottom reflectance.

1.6.4 Computing the Radiances at Depth w

The next step is to combine slabs $[a, \zeta]$, and $[\zeta, b]$, both of whose properties are now known and incorporate the surface and bottom boundary conditions. It will be convenient first to let $\zeta = w$ so as to obtain the upward and downward radiances at depth w. Those values and the standard operators for slabs $[w, \zeta]$ and $[\zeta, b]$ will then be used to obtain the radiances at any depth ζ within the water body. It will then be possible to obtain the water-leaving radiances. The final step of the solution is then to compute the physical-space Stokes vectors from the Fourier amplitudes.

The invariant imbedding equation for the depth triplet (a, w, b) can be written with the same form as Eq. (1.157):

$$\begin{bmatrix} \underline{\underline{\hat{S}}}_1^+(w) \\ \underline{\underline{\hat{S}}}_2^+(w) \\ \underline{\underline{\hat{S}}}_1^-(w) \\ \underline{\underline{\hat{S}}}_2^-(w) \end{bmatrix} \equiv \begin{bmatrix} \mathfrak{T}_{11}(a,w,b) & \mathfrak{T}_{12}(a,w,b) & \mathfrak{R}_{11}(b,w,a) & \mathfrak{R}_{12}(b,w,a) \\ \mathfrak{T}_{21}(a,w,b) & \mathfrak{T}_{22}(a,w,b) & \mathfrak{R}_{21}(b,w,a) & \mathfrak{R}_{22}(b,w,a) \\ \mathfrak{R}_{11}(a,w,b) & \mathfrak{R}_{12}(a,w,b) & \mathfrak{T}_{11}(b,w,a) & \mathfrak{T}_{12}(b,w,a) \\ \mathfrak{R}_{21}(a,w,b) & \mathfrak{R}_{22}(a,w,b) & \mathfrak{T}_{21}(b,w,a) & \mathfrak{T}_{22}(b,w,a) \end{bmatrix} \begin{bmatrix} \underline{\underline{\hat{S}}}_1^+(a) \\ \underline{\underline{\hat{S}}}_2^+(a) \\ \underline{\underline{\hat{S}}}_1^-(b) \\ \underline{\underline{\hat{S}}}_2^-(b) \end{bmatrix} + \begin{bmatrix} \mathfrak{s}_1(a,w,b) \\ \mathfrak{s}_2(a,w,b) \\ \mathfrak{S}_1(b,w,a) \\ \mathfrak{S}_2(b,w,a) \end{bmatrix} \tag{1.167}$$

Now, however, the upwelling radiance at depth b is zero because no light enters the water from below the bottom boundary. Thus $\underline{\underline{\hat{S}}}_{1,2}^-(b) = 0$, and these equations reduce to (L&W 8.95 and 8.99 with $p = 1$ and 2)

$$\underline{\underline{\hat{S}}}_1^+(w) = \mathfrak{T}_{11}(a,w,b)\underline{\underline{\hat{S}}}_1^+(a) + \mathfrak{T}_{12}(a,w,b)\underline{\underline{\hat{S}}}_2^+(a) + \mathfrak{s}_1(a,w,b) \tag{1.168}$$

$$\underline{\underline{\hat{S}}}_2^+(w) = \mathfrak{T}_{21}(a,w,b)\underline{\underline{\hat{S}}}_1^+(a) + \mathfrak{T}_{22}(a,w,b)\underline{\underline{\hat{S}}}_2^+(a) + \mathfrak{s}_2(a,w,b) \tag{1.169}$$

$$\underline{\underline{\hat{S}}}_1^-(w) = \mathfrak{R}_{11}(a,w,b)\underline{\underline{\hat{S}}}_1^+(a) + \mathfrak{R}_{12}(a,w,b)\underline{\underline{\hat{S}}}_2^+(a) + \mathfrak{S}_1(b,w,a) \tag{1.170}$$

$$\underline{\underline{\hat{S}}}_2^-(w) = \mathfrak{R}_{21}(a,w,b)\underline{\underline{\hat{S}}}_1^+(a) + \mathfrak{R}_{22}(a,w,b)\underline{\underline{\hat{S}}}_2^+(a) + \mathfrak{S}_2(b,w,a) \tag{1.171}$$

The complete operators $\mathfrak{T}_{11}(a, w, b)$ etc. are given by equations of the same form as Eqs. (1.154)–(1.156). Thus, in analogy to Eq. (1.159a),

$$\mathfrak{T}_{11}(a, w, b) = \Big[I - r_1(w, a)R_1(w, b) + r_1(w, a)R_2(w, b)(I - r_2(w, a)R_1(w, b))^{-1} r_2(w, a)R_2(w, b) \Big]^{-1} t_1(a, w) \tag{1.172}$$

and so on.

For the case of no polarization, $R_2(w, b) = 0$ and Eq. (1.168) reduces to (L&W 8.95–8.97 for $p = 1$)

$$\underline{\underline{\hat{S}}}_1^+(w) = [I - r_1(w, a)R_1(w, b)]^{-1} t_1(a, w)\underline{\underline{\hat{S}}}_1^+(a) + [I - r_1(w, a)R_1(w, b)]^{-1}\sigma(a, w) + r_1(w, a)[I - R_1(w, b)r_1(w, a)]^{-1}\Sigma(b, w)$$

Moreover, in this equation $\sigma(a, w) = 0$ because it is assumed that the air-water surface itself is source free.

Equations (1.168)–(1.171) give the radiances at depth w, $\underline{\underline{\hat{S}}}_{1,2}^{\pm}(w)$, in terms of the known incident radiances onto the sea surface, $\underline{\underline{\hat{S}}}_{1,2}^{+}(a)$, and the internal sources in the water body, $\mathfrak{S}_{1,2}(b, w, a)$.

1.6.5 Computing the Radiances at Depth ζ

Now that the radiances at depth w are known, they can be used along with the known standard operators for slab $[w, \zeta]$ and for slab $[\zeta, b]$ to write down global interaction equations for the combined slab $[w, \zeta] \cup [\zeta, b]$. These equations have the same form as the global interaction equations seen before in Eqs. (1.78), (1.80), (1.81), and (1.83), with the exception that depth m is replaced by b. The result is

$$\underline{\underline{\hat{S}}}_1^+(\zeta|\ell) = \underline{\underline{\hat{R}}}_1(\zeta, w|\ell)\underline{\underline{\hat{S}}}_1^-(\zeta|\ell) + \underline{\underline{\hat{T}}}_1(w, \zeta|\ell)\underline{\underline{\hat{S}}}_1^+(w|\ell) - \underline{\underline{\hat{R}}}_2(\zeta, w|\ell)\underline{\underline{\hat{S}}}_2^-(\zeta|\ell) - \underline{\underline{\hat{T}}}_2(w, \zeta|\ell)\underline{\underline{\hat{S}}}_2^+(w|\ell) + \underline{\underline{\hat{\Sigma}}}_1^{+t}(w, \zeta|\ell) \tag{1.173}$$

$$\underline{\underline{\hat{S}}}_2^+(\zeta|\ell) = \underline{\underline{\hat{R}}}_1(\zeta, w|\ell)\underline{\underline{\hat{S}}}_2^-(\zeta|\ell) + \underline{\underline{\hat{T}}}_1(w, \zeta|\ell)\underline{\underline{\hat{S}}}_2^+(w|\ell) + \underline{\underline{\hat{R}}}_2(\zeta, w|\ell)\underline{\underline{\hat{S}}}_1^-(\zeta|\ell) + \underline{\underline{\hat{T}}}_2(w, \zeta|\ell)\underline{\underline{\hat{S}}}_1^+(w|\ell) + \underline{\underline{\hat{\Sigma}}}_2^{+t}(w, \zeta|\ell) \tag{1.174}$$

$$\underline{\underline{\hat{S}}}_1^-(\zeta|\ell) = \underline{\underline{\hat{R}}}_1(\zeta, b|\ell)\underline{\underline{\hat{S}}}_1^+(\zeta|\ell) + \underline{\underline{\hat{T}}}_1(b, \zeta|\ell)\underline{\underline{\hat{S}}}_1^-(b|\ell) - \underline{\underline{\hat{R}}}_2(\zeta, b|\ell)\underline{\underline{\hat{S}}}_2^+(\zeta|\ell) - \underline{\underline{\hat{T}}}_2(b, \zeta|\ell)\underline{\underline{\hat{S}}}_2^-(b|\ell) + \underline{\underline{\hat{\Sigma}}}_1^{-t}(b, \zeta|\ell) \tag{1.175}$$

$$\underline{\underline{\hat{S}}}_2^-(\zeta|\ell) = \underline{\underline{\hat{R}}}_1(\zeta, m|\ell)\underline{\underline{\hat{S}}}_2^+(\zeta|\ell) + \underline{\underline{\hat{T}}}_1(m, \zeta|\ell)\underline{\underline{\hat{S}}}_2^-(m|\ell)$$

$$+ \underline{\underline{\hat{R}}}_2(\zeta, b|\ell)\underline{\underline{\hat{S}}}_1^+(\zeta|\ell) + \underline{\underline{\hat{T}}}_2(b, \zeta|\ell)\underline{\underline{\hat{S}}}_1^-(b|\ell) + \underline{\underline{\hat{\Sigma}}}_2^{-t}(b, \zeta|\ell) \tag{1.176}$$

These four equations can be solved for each ℓ value for the internal radiances at any depth ζ in terms of the known incident radiances at depths w and b. For this development, drop the ℓ argument and simplify the notation as before, so that $R_1(\zeta, w) = \underline{\underline{\hat{R}}}_1(\zeta, w|\ell)$, etc. Placing these equations in matrix form then gives the solution, which is similar in form to Eq. (1.151):

$$\begin{bmatrix} \underline{\underline{\hat{S}}}_1^+(\zeta) \\ \underline{\underline{\hat{S}}}_2^+(\zeta) \\ \underline{\underline{\hat{S}}}_1^-(\zeta) \\ \underline{\underline{\hat{S}}}_2^-(\zeta) \end{bmatrix} = \begin{bmatrix} I & 0 & -R_1(\zeta, w) & R_2(\zeta, w) \\ 0 & I & -R_2(\zeta, w) & -R_1(\zeta, w) \\ -R_1(\zeta, b) & R_2(\zeta, b) & I & 0 \\ -R_2(\zeta, b) & -R_1(\zeta, b) & 0 & I \end{bmatrix}^{-1} \times$$

$$\left\{ \begin{bmatrix} T_1(w, \zeta) & -T_2(w, \zeta) & 0 & 0 \\ T_2(w, \zeta) & T_1(w, \zeta) & 0 & 0 \\ 0 & 0 & T_1(b, \zeta) & -T_2(b, \zeta) \\ 0 & 0 & T_2(b, \zeta) & T_1(b, \zeta) \end{bmatrix} \begin{bmatrix} \underline{\underline{\hat{S}}}_1^+(w) \\ \underline{\underline{\hat{S}}}_2^+(w) \\ \underline{\underline{\hat{S}}}_1^-(b) \\ \underline{\underline{\hat{S}}}_2^-(b) \end{bmatrix} + \begin{bmatrix} \Sigma_1^{+t}(w, \zeta) \\ \Sigma_2^{+t}(w, \zeta) \\ \Sigma_1^{-t}(b, \zeta) \\ \Sigma_2^{-t}(b, \zeta) \end{bmatrix} \right\} \tag{1.177}$$

Repeated application of the block matrix inversion formulas (1.152) and (1.153) can again be used to obtain the inverse of the 4×4 block matrix. Letting B and C be the upper right and lower left 2×2 blocks, respectively, gives

$$\begin{aligned} (I - BC)^{-1} = &\begin{bmatrix} I - R_1(\zeta, w)R_1(\zeta, b) + R_2(\zeta, w)R_2(\zeta, b) \\ -R_2(\zeta, w)R_1(\zeta, b) - R_1(\zeta, w)R_2(\zeta, b) \end{bmatrix} \\ &\begin{bmatrix} R_1(\zeta, w)R_2(\zeta, b) + R_2(\zeta, w)R_1(\zeta, b) \\ I + R_2(\zeta, w)R_2(\zeta, b) - R_1(\zeta, w)R_1(\zeta, b) \end{bmatrix}^{-1} \equiv \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}^{-1} \\ = &\begin{bmatrix} (\alpha - \beta\delta^{-1}\gamma)^{-1} & -\alpha^{-1}\beta(\delta - \gamma\alpha^{-1}\beta)^{-1} \\ \delta^{-1}\gamma(\alpha - \beta\delta^{-1}\gamma)^{-1} & (\delta - \gamma\alpha^{-1}\beta)^{-1} \end{bmatrix} \equiv \begin{bmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{bmatrix} \end{aligned} \tag{1.178}$$

$$-C(I - BC)^{-1} = \begin{bmatrix} -R_1(\zeta, b)n_{11} + R_2(\zeta, b)n_{21} & -R_1(\zeta, b)n_{12} + R_2(\zeta, b)n_{22} \\ -R_2(\zeta, b)n_{11} - R_2(\zeta, b)n_{21} & -R_2(\zeta, b)n_{12} + R_1(\zeta, b)n_{22} \end{bmatrix} \tag{1.179}$$

$$\begin{aligned} (I - CB)^{-1} = &\begin{bmatrix} I - R_1(\zeta, b)R_1(\zeta, w) + R_2(\zeta, b)R_2(\zeta, w) \\ -R_2(\zeta, b)R_1(\zeta, w) - R_1(\zeta, b)R_2(\zeta, w) \end{bmatrix} \\ &\begin{bmatrix} R_1(\zeta, b)R_2(\zeta, w) + R_2(\zeta, b)R_1(\zeta, 2) \\ I + R_2(\zeta, b)R_2(\zeta, w) - R_1(\zeta, b)R_1(\zeta, w) \end{bmatrix}^{-1} \equiv \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} \end{aligned}$$

$$= \begin{bmatrix} (a - bd^{-1}c)^{-1} & -a^{-1}b(d - ca^{-1}b)^{-1} \\ d^{-1}c(a - bd^{-1}c)^{-1} & (d - ca^{-1}b)^{-1} \end{bmatrix} \equiv \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \tag{1.180}$$

$$-B(I - CB)^{-1} = \begin{bmatrix} -R_1(\zeta, w)p_{11} + R_2(\zeta, w)p_{21} & -R_1(\zeta, w)p_{12} + R_2(\zeta, w)p_{22} \\ -R_2(\zeta, w)p_{11} - R_2(\zeta, w)p_{21} & -R_2(\zeta, w)p_{12} + R_1(\zeta, w)p_{22} \end{bmatrix} \tag{1.181}$$

Equation (1.153) now gives the inverse in Eq. (1.177) as

$$\begin{bmatrix} I & 0 & -R_1(\zeta, w) & R_2(\zeta, w) \\ 0 & I & -R_2(\zeta, w) & -R_1(\zeta, w) \\ -R_1(\zeta, b) & R_2(\zeta, b) & I & 0 \\ -R_2(\zeta, b) & -R_1(\zeta, b) & 0 & I \end{bmatrix}^{-1} =$$

$$\begin{bmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \\ -R_1(\zeta, b)n_{11} + R_2(\zeta, b)n_{21} & -R_1(\zeta, b)n_{12} + R_2(\zeta, b)n_{22} \\ -R_2(\zeta, b)n_{11} - R_2(\zeta, b)n_{21} & -R_2(\zeta, b)n_{12} - R_1(\zeta, b)n_{22} \end{bmatrix}$$

$$\begin{bmatrix} -R_1(\zeta, w)p_{11} + R_2(\zeta, w)p_{21} & -R_1(\zeta, w)p_{12} + R_2(\zeta, w)p_{22} \\ -R_2(\zeta, w)p_{11} - R_2(\zeta, w)p_{21} & -R_2(\zeta, w)p_{12} + R_1(\zeta, w)p_{22} \\ p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \tag{1.182}$$

Equation (1.177) now becomes

$$\begin{bmatrix} \hat{\underline{\underline{S}}}_1^+(\zeta) \\ \hat{\underline{\underline{S}}}_2^+(\zeta) \\ \hat{\underline{\underline{S}}}_1^-(\zeta) \\ \hat{\underline{\underline{S}}}_2^-(\zeta) \end{bmatrix} =$$

$$\begin{bmatrix} n_{11}T_1(w, \zeta) + n_{12}T_2(w, \zeta) \\ n_{12}T_1(w, \zeta) + n_{22}T_2(w, \zeta) \\ [-R_1(\zeta, b)n_{11} + R_2(\zeta, b)n_{21}]T_1(w, \zeta) + [-R_1(\zeta, b)n_{12} + R_2(\zeta, b)n_{22}]T_2(w, \zeta) \\ [-R_2(\zeta, b)n_{11} - R_1(\zeta, b)n_{21}]T_1(w, \zeta) + [-R_2(\zeta, b)n_{12} - R_2(\zeta, b)n_{22}]T_2(w, \zeta) \end{bmatrix}$$

$$\begin{bmatrix} n_{11}T_2(w, \zeta) + n_{12}T_1(w, \zeta) \\ -n_{12}T_2(w, \zeta) + n_{22}T_1(w, \zeta) \\ [-R_1(\zeta, b)n_{11} + R_2(\zeta, b)n_{21}][-T_2(w, \zeta)] + [-R_1(\zeta, b)p_{12} + R_2(\zeta, b)p_{22}]T_2(w, \zeta) \\ [-R_2(\zeta, b)n_{11} - R_1(\zeta, b)n_{21}][-T_2(w, \zeta)] + [-R_1(\zeta, b)p_{12} + R_2(\zeta, b)p_{22}]T_2(w, \zeta) \end{bmatrix}$$

$$\begin{bmatrix} [-R_1(\zeta, w)p_{11} + R_2(\zeta, w)p_{21}]T_1(b, \zeta) + [-R_1(\zeta, w)p_{12} + R_2(\zeta, w)p_{22}]T_2(b, \zeta) \\ [-R_2(\zeta, w)p_{11} - R_1(\zeta, w)p_{21}]T_1(b, \zeta) + ([-R_2(\zeta, w)p_{12} - R_2(\zeta, w)p_{22}]T_2(b, \zeta) \\ p_{11}T_1(b, \zeta) + p_{12}T_2(b, \zeta) \\ p_{21}T_1(b, \zeta) + p_{22}T_2(b, \zeta) \end{bmatrix}$$

$$
\begin{bmatrix}
[-R_1(\zeta,w)p_{11}+R_2(\zeta,w)p_{21}][-T_2(b,\zeta)]+[-R_1(\zeta,w)p_{12}+R_2(\zeta,w)p_{22}]T_1(b,\zeta)\\
[-R_2(\zeta,w)p_{11}-R_1(\zeta,w)p_{21}][-T_2(b,\zeta)]+[-R_2(\zeta,w)p_{12}-R_2(\zeta,w)p_{22}]T_1(b,\zeta)\\
-p_{11}T_2(b,\zeta)+p_{12}T_1(b,\zeta)\\
-p_{21}T_2(b,\zeta)+p_{22}T_1(b,\zeta)
\end{bmatrix}
\times
\begin{bmatrix}
\hat{\underline{\underline{S}}}_1^{+}(w)\\
\hat{\underline{\underline{S}}}_2^{+}(w)\\
\hat{\underline{\underline{S}}}_1^{-}(b)\\
\hat{\underline{\underline{S}}}_2^{-}(b)
\end{bmatrix}
$$

$$
+
\begin{bmatrix}
n_{11}\Sigma_1^{+t}(w,\zeta)+n_{12}\Sigma_2^{+t}(w,\zeta)\\
n_{12}\Sigma_1^{+t}(w,\zeta)+n_{22}\Sigma_2^{+t}(w,\zeta)\\
[-R_1(\zeta,b)n_{11}+R_2(\zeta,b)n_{21}]\Sigma_1^{+t}(w,\zeta)+[-R_2(\zeta,b)n_{11}-R_1(\zeta,b)n_{21}]\Sigma_1^{+t}(w,\zeta)\\
[-R_2(\zeta,b)n_{11}-R_1(\zeta,b)n_{21}]\Sigma_1^{+t}(w,\zeta)+[-R_2(\zeta,b)n_{12}-R_1(\zeta,b)n_{22}]\Sigma_1^{+t}(w,\zeta)
\end{bmatrix}
$$

$$
\begin{bmatrix}
[-R_1(\zeta,w)p_{11}+R_2(\zeta,w)p_{21}]\Sigma_1^{-t}(b,\zeta)+[-R_1(\zeta,w)p_{12}+R_2(\zeta,w)p_{22}]\Sigma_2^{-t}(b,\zeta)\\
[-R_2(\zeta,w)p_{11}-R_1(\zeta,w)p_{21}]\Sigma_1^{-t}(b,\zeta)+[-R_2(\zeta,w)p_{12}-R_1(\zeta,w)p_{22}]\Sigma_2^{-t}(b,\zeta)\\
p_{11}\Sigma_1^{-t}(b,\zeta)+p_{12}\Sigma_2^{-t}(b,\zeta)\\
p_{21}\Sigma_1^{-t}(b,\zeta)+p_{22}\Sigma_2^{-t}(b,\zeta)
\end{bmatrix}
\tag{1.183}
$$

Writing this in terms of complete operators gives

$$
\begin{bmatrix}
\hat{\underline{\underline{S}}}_1^{+}(\zeta)\\
\hat{\underline{\underline{S}}}_2^{+}(\zeta)\\
\hat{\underline{\underline{S}}}_1^{-}(\zeta)\\
\hat{\underline{\underline{S}}}_2^{-}(\zeta)
\end{bmatrix}
\equiv
\begin{bmatrix}
\mathfrak{T}_{11}(w,\zeta,b) & \mathfrak{T}_{12}(w,\zeta,b) & \mathfrak{R}_{11}(b,\zeta,w) & \mathfrak{R}_{12}(b,\zeta,w)\\
\mathfrak{T}_{21}(w,\zeta,b) & \mathfrak{T}_{22}(w,\zeta,b) & \mathfrak{R}_{21}(b,\zeta,w) & \mathfrak{R}_{22}(b,\zeta,w)\\
\mathfrak{R}_{11}(w,\zeta,b) & \mathfrak{R}_{12}(w,\zeta,b) & \mathfrak{T}_{11}(b,\zeta,w) & \mathfrak{T}_{12}(b,\zeta,w)\\
\mathfrak{R}_{21}(w,\zeta,b) & \mathfrak{R}_{22}(w,\zeta,b) & \mathfrak{T}_{21}(b,\zeta,w) & \mathfrak{T}_{22}(b,\zeta,w)
\end{bmatrix}
\begin{bmatrix}
\hat{\underline{\underline{S}}}_1^{+}(w)\\
\hat{\underline{\underline{S}}}_2^{+}(w)\\
\hat{\underline{\underline{S}}}_1^{-}(b)\\
\hat{\underline{\underline{S}}}_2^{-}(b)
\end{bmatrix}
$$

$$
+
\begin{bmatrix}
\mathfrak{S}_1(w,\zeta,b)\\
\mathfrak{S}_2(w,\zeta,b)\\
\mathfrak{S}_1(b,\zeta,w)\\
\mathfrak{S}_2(b,\zeta,w)
\end{bmatrix}
\tag{1.184}
$$

These equations are the invariant imbedding relationships for depths (w,ζ,b). The incident radiances at depth w and b are now known, so these equations allow the radiances at any depth ζ in the water body $[w,m]$ to be computed, including all effects of the air-water surface, the bottom reflectance, and the internal sources. These equations are evaluated for each Fourier ℓ mode. It should be noted that the incident radiance at b is zero for an opaque reflecting bottom, so in practice only the $\mathfrak{T}_{ij}(w,\zeta,b)$ and $\mathfrak{R}_{ij}(w,\zeta,b)$ terms need to be computed. In this case, for example,

$$
\hat{\underline{\underline{S}}}_1^{+}(\zeta)=\mathfrak{T}_{11}(w,\zeta,b)\hat{\underline{\underline{S}}}_1^{+}(w)+\mathfrak{T}_{12}(w,\zeta,b)\hat{\underline{\underline{S}}}_1^{+}(w)+\mathfrak{S}_1(w,\zeta,b)
\tag{1.185}
$$

For the case of no polarization, $\mathfrak{T}_{12}(w,\zeta,b)=0$, and this equation reduces to (the transpose of) L&W (8.103) with $p=1$.

1.6.6 Computing the Water-Leaving Radiances

The only radiances still unknown are the water-leaving radiances $\hat{\underline{\underline{S}}}^{-}_{1,2}(a)$. Now that the upwelling radiances at depth w are known, the water-leaving radiances can be obtained from the surface boundary condition (1.135) (L&W 8.107):

$$\hat{\underline{\underline{S}}}^{-}_{p}(a) = \hat{\underline{\underline{r}}}_{p}(a, w)\hat{\underline{\underline{S}}}^{+}_{p}(a) + \hat{\underline{\underline{t}}}_{p}(w, a)\hat{\underline{\underline{S}}}^{-}_{p}(w) \tag{1.186}$$

with $p = 1, 2$. Because the surface-boundary Fourier modes do not decouple, this equation must be evaluated simultaneously for all ℓ modes via the $4M(N+1) \times 4M(N+1)$ composite matrices as seen in Eqs. (1.132) and (1.133), and the $4M(N+1) \times 1$ vectors of Eq. (1.134).

1.6.7 Synthesis of the Physical-Space Stokes Vectors

We have now obtained the Fourier amplitudes of the Stokes vectors at all depths. The final step of the solution of the VRTE is to reconstitute the quad-averaged Stokes vectors in physical space. The $4M \times 1$ column vectors $\hat{\underline{\underline{S}}}^{-\pm}_{1,2}$ are "unstacked" as in Eq. (1.71) to obtain the 4×1 Stokes vector amplitudes $\hat{\underline{S}}^{\pm}_{1,2}(\zeta, u|\ell)$ for each u and ℓ value. The physical Stokes vectors are then obtained via Eq. (1.52) (L&W 8.108). Evaluation of that equation at each depth a, $w = \zeta_1, \zeta_2, ..., \zeta_K = m$, for each value of $u = 1, ..., M$ and $v = 1, 2, ..., 2N$, gives the desired quad-averaged Stokes vectors.

This completes the solution of the vector radiative transfer equation.

Acknowledgements This work was supported by NASA Contract NNH12CD06C titled *Radiative Transfer Modeling for Improved Ocean Color Remote Sensing*. I thank three anonymous reviewers for their helpful comments on the manuscript.

A Fourier Analysis of Discrete Functions

The invariant imbedding solution algorithm for the directionally discretized VRTE employs a Fourier decomposition of the equations in azimuthal direction. This appendix collects for reference various results for the Fourier decomposition of discrete functions of azimuthal angle, as needed in Sect. 1.5.

A.1 Functions of One Azimuthal Angle

Let $f_v \equiv f(\phi_v)$ be any discrete function of the azimuthal angle ϕ, i.e., f_v is defined only at the discrete values $\phi_v = (v-1)\Delta\phi$, for $v = 1, 2, ..., 2N$. Here $\Delta\phi = 2\pi/2N$

is the angular width of a quad as shown in Fig. 1.5; $\Delta\phi$ is the same for each quad. Let δ_k be the Kronecker delta function defined by

$$\delta_k \equiv \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{if } k \neq 0 \end{cases} \quad \text{or equivalently} \quad \delta_{k-\ell} \equiv \begin{cases} 1 & \text{if } k = \ell \\ 0 & \text{if } k \neq \ell \end{cases} \tag{1.187}$$

where k and ℓ are integers.

The discrete orthogonality relations for sines and cosines can then be written

$$\sum_{s=1}^{2N} \cos(k\phi_s)\cos(\ell\phi_s) = \begin{cases} 0 & \text{if } k \neq \ell \\ N & \text{if } k = \ell \text{ and } \ell = 1, 2, ..., N-1 \\ 2N & \text{if } k = \ell \text{ and } \ell = 0 \text{or} \ell = N \end{cases}$$
$$= N(\delta_{k+\ell} + \delta_{k-\ell} + \delta_{k+\ell-2N}) \tag{1.188}$$

$$\sum_{s=1}^{2N} \sin(k\phi_s)\sin(\ell\phi_s) = \begin{cases} 0 & \text{if } k \neq \ell \\ N & \text{if } k = \ell \text{ and } \ell = 1, 2, ..., N-1 \\ 0 & \text{if } k = \ell \text{ and } \ell = 0 \text{ or} \ell = N \end{cases}$$
$$= N(\delta_{k-\ell} - \delta_{k+\ell} - \delta_{k+\ell-2N}) \tag{1.189}$$

and

$$\sum_{s=1}^{2N} \cos(k\phi_s)\sin(\ell\phi_s) = 0 \quad \text{for all } k \text{ and } \ell. \tag{1.190}$$

A discrete function of one azimuthal angle (e.g., a discretized Stokes vector $\underline{S}(\zeta, u, v, j)$ with the depth ζ, polar angle u, and wavelength j being held constant) has the Fourier spectral decomposition

$$f(v) = \sum_{\ell=0}^{N} \left[\hat{f}_1(\ell)\cos(\ell\phi_v) + \hat{f}_2(\ell)\sin(\ell\phi_v) \right] \tag{1.191}$$

where $v = 1, 2, ..., 2N$. The notation $\hat{f}$ denotes a Fourier amplitude of the corresponding physical variable f; subscript 1 denotes cosine amplitudes and subscript 2 denotes sine amplitudes. The cosine amplitudes $\hat{f}_1(\ell)$ are obtained by multiplying Eq. (1.191) by $\cos(k\phi_v)$, summing over v, and applying the orthogonality relations to get

$$\hat{f}_1(\ell) = \frac{1}{\epsilon_l} \sum_{v=1}^{2N} f(v)\cos(\ell\phi_v) \quad \text{for } \ell = 0, 1, ..., N \tag{1.192}$$

where

$$\epsilon_\ell = N(1 + \delta_{2\ell} + \delta_{2\ell-2N}) = \begin{cases} 2N & \text{if } \ell = 0 \text{ or } \ell = N \\ N & \text{if } \ell = 1, 2, ..., N-1 \end{cases} \tag{1.193}$$

Note that the $\hat{f}_1(0)$ amplitude is just the average value of $f(v)$. The sine amplitudes $\hat{f}_2(\ell)$ are obtained in a similar way by multiplying (1.191) by $\sin(k\phi_v)$:

$$\hat{f}_2(\ell) = \frac{1}{\gamma_l} \sum_{v=1}^{2N} f(v) \cos(\ell\phi_v) \quad \text{for } \ell = 1, ..., N-1 \tag{1.194}$$

where

$$\gamma_\ell = N(1 - \delta_{2\ell} - \delta_{2\ell-2N}) = \begin{cases} 0 & \text{if } \ell = 0 \text{ or } \ell = N \\ N & \text{if } \ell = 1, 2, ..., N-1 \end{cases} \tag{1.195}$$

Note that values of $\gamma_0 = \gamma_N = 0$ do not occur in Eq. (1.194) for the sine amplitudes.

If $f(v)$ is a constant f then $\hat{f}_1(0) = f$ and all other cosine and sine components are zero. This is the case for a polar cap radiance, which has no azimuthal dependence. In general, the $2N$ values of $f(v)$ are determined exactly by the $N+1$ cosine amplitudes and the $N-1$ sine amplitudes; the information content of the physical and Fourier representations is the same. In the Fourier decomposition of discrete functions all terms must be included; there can be no truncation of these summations.

A.2 Functions of the Difference of Two Azimuthal Angles

Let $g_{\cos}(v, s) = g[\cos(\phi_v - \phi_s)]$ be a discrete cosine function of $\phi_v - \phi_s$, where ϕ_v and ϕ_s are two azimuthal angles and $v, s = 1, ..., 2N$. Then $g_{\cos}(v, s)$ has the Fourier expansion (L&W 8.24)

$$g_{\cos}(v, s) = \sum_{k=0}^{N} \hat{g}_1(k) \cos[k(\phi_v - \phi_s)]\,. \tag{1.196}$$

Multiplying this equation by $\cos(\ell\phi_v)$, summing over v, expanding the cosine, and using the orthogonality relations gives (L&W 8.25)

$$\hat{g}_1(\ell) = \frac{1}{\epsilon_\ell \cos(\ell\phi_s)} \sum_{v=1}^{2N} g_{\cos}(v, s) \cos \ell\phi_v) \quad \text{for } \ell = 0, ..., N\,. \tag{1.197}$$

Similarly, let $g_{\sin}(v,s) = g[\sin(\phi_v - \phi_s)]$ be a discrete sine function of $\phi_v - \phi_s$. Then $g_{\sin}(v,s)$ has the Fourier expansion

$$g_{\sin}(v,s) = \sum_{k=1}^{N-1} \hat{g}_2(k) \sin[k(\phi_v - \phi_s)] \,. \tag{1.198}$$

Multiplying this equation by $\sin(\ell\phi_v)$, summing over v, expanding the sine, and using the orthogonality relations gives

$$\hat{g}_2(\ell) = \frac{1}{\gamma_\ell \cos(\ell\phi_s)} \sum_{v=1}^{2N} g_{\sin}(v,s) \sin \ell\phi_v) \quad \text{for } \ell = 1, ..., N-1 \,. \tag{1.199}$$

These expansions will be used for the elements of the phase matrix, which depend on either the cosine or the sine of $\phi_v - \phi_s$. Note that in Eqs. (1.197) and (1.199), the choice $\phi_s = \phi_1 = 0$ can be made without loss of generality. This merely anchors the difference $\phi_v - \phi_s$ to $\phi_1 = 0$. The computer code then needs to compute and store phase function elements only for the range of $v = 1, ..., N+1$, which generates all discretized scattering angles ψ as $\phi_v = 0$ to 180°. The phase matrix element $\underline{P}(r, s \to u, v)$ is then obtained from the stored value of $\underline{P}(r, 1 \to u, v - s + 1)$, which can be stored as a three-index array $\underline{P}(r, u, v)$.

A.3 Functions of Two Azimuthal Angles

Finally, let $h(s,v) = h(\phi_s, \phi_v)$ be an arbitrary function of two azimuthal angles. Then $h(s,v)$ can be represented as

$$\begin{aligned} h(s,v) = &\sum_{k=0}^{N}\sum_{\ell=0}^{N} \hat{h}_{11}(k,\ell) \cos(k\phi_s)\cos(\ell\phi_v) \\ &+ \sum_{k=0}^{N}\sum_{\ell=0}^{N} \hat{h}_{12}(k,\ell) \cos(k\phi_s)\sin(\ell\phi_v) \\ &+ \sum_{k=0}^{N}\sum_{\ell=0}^{N} \hat{h}_{21}(k,\ell) \sin(k\phi_s)\cos(\ell\phi_v) \\ &+ \sum_{k=0}^{N}\sum_{\ell=0}^{N} \hat{h}_{22}(k,\ell) \sin(k\phi_s)\sin(\ell\phi_v) \,. \end{aligned} \tag{1.200}$$

To find $\hat{h}_{11}$, multiply Eq. (1.200) by $\cos(k'\phi_s)\cos(\ell'\phi_v)$, sum over s and v, and apply the orthogonality relations. The other three amplitudes are found in an analogous manner. The results are

$$\hat{h}_{11}(k,\ell) = \frac{1}{\epsilon_k \epsilon_\ell} \sum_{s=1}^{2N} \sum_{v=1}^{2N} h(s,v) \cos(k\phi_s) \cos(\ell\phi_v)\,,$$
$$\hat{h}_{12}(k,\ell) = \frac{1}{\epsilon_k \gamma_\ell} \sum_{s=1}^{2N} \sum_{v=1}^{2N} h(s,v) \cos(k\phi_s) \sin(\ell\phi_v)\,,$$
$$\hat{h}_{21}(k,\ell) = \frac{1}{\gamma_k \epsilon_\ell} \sum_{s=1}^{2N} \sum_{v=1}^{2N} h(s,v) \sin(k\phi_s) \cos(\ell\phi_v)\,,$$
$$\hat{h}_{22}(k,\ell) = \frac{1}{\gamma_k \gamma_\ell} \sum_{s=1}^{2N} \sum_{v=1}^{2N} h(s,v) \sin(k\phi_s) \sin(\ell\phi_v)\,. \tag{1.201}$$

The arbitrary zero sine amplitudes $\hat{f}_2(0)$ and $\hat{f}_2(N)$ have their counterparts here:

$$\hat{h}_{12}(k,0) = \hat{h}_{12}(k,N) = 0 \quad \text{for } k = 0, ..., N\,,$$
$$\hat{h}_{21}(0,\ell) = \hat{h}_{21}(N,\ell) = 0 \quad \text{for } \ell = 0, ..., N\,,$$
$$\hat{h}_{22}(0,0) = \hat{h}_{22}(0,N) = \hat{h}_{22}(N,0) = \hat{h}_{22}(N,N) = 0\,. \tag{1.202}$$

These special cases allow the exclusion of any k or ℓ values in Eq. (1.201) that would result in division by zero resulting from the γ_k and γ_ℓ factors.

Equations (1.200)–(1.202) are applicable to the air-water surface transfer functions. However, the symmetries of those functions result in considerable simplification. In particular, all $\hat{h}_{12}(k,\ell)$ and $\hat{h}_{21}(k,\ell)$ are zero for the surface transfer functions.

The equations of Sect. 1.5.3 contain sums over k and ℓ having the form

$$S1 = \sum_{l=0}^{N} \sum_{k=0}^{N} f(k) g(\ell) \left[\sum_{s=1}^{2N} \cos[k(\phi_v - \phi_s)] \cos(\ell\phi_s) \right]$$

where $f(k)$ and $g(\ell)$ are discrete functions of k and ℓ for $k, \ell = 0, ..., N$. Application of the the trigonometric formula for the cosine of the difference of two angles and the previous equations reduces this to

$$\begin{aligned} S1 = &\sum_{l=0}^{N} \sum_{k=0}^{N} f(k) g(\ell) \left\{ \left[\sum_{s=1}^{2N} \cos(k\phi_s) \cos(\ell\phi_s) \right] \cos(k\phi_v) \right. \\ &\left. + \left[\sum_{s=1}^{2N} \sin(k\phi_s) \cos(\ell\phi_s) \right] \sin(k\phi_v) \right\} \\ = &\sum_{l=0}^{N} \sum_{k=0}^{N} f(k) g(\ell)\ N(\delta_{k+\ell} + \delta_{k-\ell} + \delta_{k+\ell-2N}) \cos(k\phi_v) \end{aligned}$$

$$= \sum_{l=0}^{N} f(\ell) g(\ell)\ N(1 + \delta_{2\ell} + \delta_{2\ell-2N}) \cos(\ell\phi_v)$$
$$= \sum_{l=0}^{N} \epsilon_\ell f(\ell) g(\ell) \cos(\ell\phi_v)\,. \tag{1.203}$$

The same process gives

$$S2 = \sum_{l=0}^{N} \sum_{k=0}^{N} f(k) g(\ell) \left[\sum_{s=1}^{2N} \cos[k(\phi_v - \phi_s)] \sin(\ell\phi_s) \right]$$
$$= \sum_{l=0}^{N} \sum_{k=0}^{N} f(k) g(\ell)\ N(\delta_{k-\ell} - \delta_{k+\ell} - \delta_{k+\ell-2N}) \sin(k\phi_v)$$
$$= \sum_{l=0}^{N} f(\ell) g(\ell)\ N(1 - \delta_{2\ell} - \delta_{2\ell-2N}) \sin(\ell\phi_v)$$
$$= \sum_{l=0}^{N} \gamma_\ell f(\ell) g(\ell) \sin(\ell\phi_v)\,. \tag{1.204}$$

Likewise

$$S3 = \sum_{l=0}^{N} \sum_{k=0}^{N} f(k) g(\ell) \left[\sum_{s=1}^{2N} \sin[k(\phi_v - \phi_s)] \cos(\ell\phi_s) \right]$$
$$= \sum_{l=0}^{N} \epsilon_\ell f(\ell) g(\ell) \sin(\ell\phi_v)\,, \tag{1.205}$$

and

$$S4 = \sum_{l=0}^{N} \sum_{k=0}^{N} f(k) g(\ell) \left[\sum_{s=1}^{2N} \sin[k(\phi_v - \phi_s)] \sin(\ell\phi_s) \right]$$
$$= -\sum_{l=0}^{N} \gamma_\ell f(\ell) g(\ell) \cos(\ell\phi_v)\,. \tag{1.206}$$

Note the minus sign in the last equation.

These results for the Fourier decomposition of discrete functions of one or two azimuthal angles provide all of the tools necessary for converting the discretized physical-space VRTE into discretized equations in Fourier space.

References

Ambartsumian VA (1943) On the issue of diffuse reflection of light by a turbid medium. Dokl Akad Nauk SSSR (in Russian) 38:229–232

Bellman R, Kalaba R (1956) On the principle of invariant imbedding and propagation through inhomogeneous media. Proc Natl Acad Sci 42:629–632

Bellman R, Kalaba R, Wing GM (1960) Invariant imbedding and the reduction of two-point boundary value problems to initial value problems. Proc Natl Acad Sci 46(12):1646–1649

Bohren C, Huffman D (1983) Absorption and Scattering of Light by Small Particles. Wiley, New York

Born M, Wolf E (1975) Principles of optics, 5th edn. Pergamon, Oxford

Chandrasekhar S (1950) Radiative transfer. Oxford University Press, Oxford

Chowdhary J, Cairns B, Travis LD (2006) Contribution of water-leaving radiances to multiangle, multispectral polarimetric observations over the open ocean: bio-optical model results for case 1 waters. Appl Opt 45(22):5542–5567

Garcia RDM (2012) Fresnel boundary and interface conditions for polarized radiative transfer in a multilayer medium. J Quant Spectros Radiat Trans 113:306–307

Hecht E (1989) Optics, 2nd edn. Addison-Wesley, Boston

Hovenier JW (1969) Symmetry relationships for scattering of polarized light in a slab of randomly oriented particles. J Atmos Sci 26:488–499

Hovenier JW (1971) Multiple scattering of polarized light in planetary atmospheres. Astron Astrophys 13:7–29

Hu Y-X, Winker D, Yang P, Baum B, Poole L, Vann L (2001) Identification of cloud phase from PICASSO-CENA lidar depolarization: a multiple scattering sensitivity study. J Quant Spectros Radiat Trans 70:569–579

Kattawar GW (1994) Polarization of light in the ocean. In: Spinrad RW, Carder KL, Perry MJ (eds) Ocean Optics. Oxford University Press, pp. 200–225

Kattawar GW, Adams CN (1989) Stokes vector calculations of the submarine light field in an atmosphere-ocean with scattering according to a Rayleigh phase matrix: effect of interface refractive index on radiance and polarization. Limnol Oceanogr 34:1453–1472

Mishchenko MI (2007) Electromagnetic scattering by a fixed finite object embedded in an absorbing medium. Opt Express 15:13188–13202

Mishchenko MI (2008a) Multiple scattering by particles embedded in an absorbing medium. 2. Radiative transfer equation. J Quant Spectrosc Radiat Trans 109:2386–2390

Mishchenko MI (2008b) Multiple scattering, radiative transfer, and weak localization in discrete random media: unified microphysical approach. Rev Geophys 46:1–33

Mishchenko MI (2013a) 125 years of radiative transfer: Enduring triumphs and persistent misconceptions. In: Cahalan RF, Fischer J (eds) Radiation processes in the atmosphere and ocean, vol 1531. AIP Publishing, pp. 11–18. https://doi.org/10.1063/1.4804696

Mishchenko MI (2013b) Radiative transfer theory verified by controlled laboratory experiments. Opt Lett 38:3522–3525

Mishchenko MI (2014) Directional radiometry and radiative transfer: the convoluted path from centuries-old phenomenology to physical optics. J Quant Spect Radiat Trans 146:4–33

Mishchenko MI, Travis LD, Lacis AA (2002) Scattering, absorption, and emission of light by small particles. Cambridge University Press, Cambridge

Mishchenko MI, Dlugach JM, Yurkin MA, Bi L, Cairns B, Liu L, Panetta RL, Travis LD, Yang P, Zakharova NT (2016) First-principles modeling of electromagnetic scattering by discrete and discretely heterogeneous random media. Phys Rep 632:1–75

Mobley CD (1994) Light and Water: Radiative Transfer in Natural Waters. Academic Press, Cambridge

Mobley CD (2014a) Hydropol Mathematical documentation: invariant imbedding theory for the vector radiative transfer equation. Technical report, Sequoia Scientific, Inc. http://www.oceanopticsbook.info/view/publications/references

Mobley CD (2014b) Modeling sea surfaces: a tutorial on fourier transform techniques. Technical report, Sequoia Scientific, Inc. http://www.oceanopticsbook.info/view/publications/references
Mobley CD (2015) Polarized reflectance and transmittance properties of windblown sea surfaces. Appl Opt 54:4828–4849
Preisendorfer RW (1958) Functional relations for the r and t operators on plane-parallel media. Proc Natl Acad Sci 44:323–237
Preisendorfer RW (1965) Radiative transfer on discrete spaces. Pergamon Press, Oxford
Preisendorfer RW (1976) Hydrologic optics. U.S. Department of Commerce, Pacific Marine Environ. Lab. In 6 volumes. www.oceanopticsbook.info/view/introduction/level_2/text_books_relevant_to_ocean_optics
Schott JR (1999) Fundamentals of Polarimetric Remote Sensing. SPIE Tutorial Text TT81. SPIE Press, Bellingham
Sobolev VV (1956) The transmission of radiation through an inhomogeneous medium. Dokl Akad Nauk SSSR (in Russian) 111:1000–1003
Stokes GG (1862) On the intensity of the light reflected from or transmitted through a pile of plates. Proc R Soc Lond 11:545–556
van de Hulst HC (1980) Multiple light scattering: tables, formulas, and applications. Academic Press, Cambridge
Zhai P-W, Kattawar GW, Yang P (2008) Impulse response solution of the three-dimensional vector radiative transfer equation in atmosphere-ocean systems I. Monte Carlo Method Appl Opt 47:1037–1047
Zhai P-W, Kattawar GW, Hu Y (2012) Comment on the transmission matrix for a dielectric interface. J Quant Spectrosc Radiat Trans 113:1981–1984

Chapter 2
Multiple Scattering of Light in Ordered Particulate Media

Valery A. Loiko and Alexander A. Miskevich

2.1 Introduction

The particulate structures with short- and long-range order in spatial arrangement of scatterers are distinguished among the huge amount of optically inhomogeneous (scattering) media. In these structures strong coherent light scattering and interference effects are implemented.

Snow, ice, foam, biological tissues, photographic layers, milk glasses, dispersion filters, island films, luminescent screens, paper, etc. are examples of short-range ordered media.

The crystal-like structures with periodic and near to periodic spatial particle location relate to class of the long-range ordered systems. There is great number of photonic and optoelectronic objects and devices belonging to this class. For example, some types of metamaterials, photonic crystals, antireflection coatings, light filters, reflectors, waveguides, resonators, light emitting diodes, solar cells, detectors, lasers, communication systems, multiplexers, microantenna substrates, biological objects, etc. are based on ordered particulate structures. Pronounced crystal-like structure of spatially ordered particulate systems determines the choice of approaches for their studying. Most complete theoretical description of their optical properties yield quantum-mechanical methods and methods of the theory of multiple scattering of waves since they takes into account effects of multiple scattering, diffraction and interference of waves. Quantum-mechanical methods are based on the extension of the theory developed in quantum physics to analyze scattering (diffraction) of

V. A. Loiko (✉) · A. A. Miskevich
Institute of Physics, National Academy of Sciences of Belarus,
68-2 Nezalezhnastsi avenue, 220072 Minsk, Belarus
e-mail: loiko@ifanbel.bas-net.by

A. A. Miskevich
e-mail: miskevic@ifanbel.bas-net.by

A. Kokhanovsky (ed.), *Springer Series in Light Scattering*, Springer Series
in Light Scattering, https://doi.org/10.1007/978-3-319-70796-9_2

elementary particles in crystals to the case of photons. They are convenient to describe scattering in the structures with perfect lattices.

Statistical methods of the theory of multiple scattering of waves (TMSW) are generally used to determine the optical properties of ensembles with random and partially ordered arrangement of particles. Under these methods the ensemble structure is described by many-particle distribution functions characterizing the probability of relative spatial position of particles. Knowledge of N-particle distribution function allows one to obtain rigorous solution of scattering problem for ensemble of N particles. However, its determination is a very difficult task, especially for large N. Therefore, as a rule, within the TMSW one finds the approximate solution that requires knowledge of only the s-particle ($s << N$) function. To simulate densely packed particulate media the two-particle distribution function is usually used. For statistically homogeneous ensembles is used the radial distribution function. It characterizes the probability of the location of any two particles in space at a certain distance from each other.

Well-developed mathematical apparatus of the TMSW and methods for finding the radial distribution functions to describe the particulate media with a short-range order (partially ordered media) on the one hand and quantum-mechanical approach for media with long-range order on the other hand led to the accumulation of vast theoretical data on scattering by ensembles with random, partially ordered, and perfect (strictly regular) arrangement of particles.

Much less attention is paid to structures with imperfect lattices. The quantitative account of imperfection is an important scientific and practical problem, because its solution allows one to optimize optical characteristics of the ordered structures. Deviations from the perfection are usually (except for specially created lattice defects) have a random (statistical) nature. Therefore, to simulate the interaction of light with such structures it is reasonable to use a statistical approach of the TMSW.

This chapter is mainly devoted to review of the works on light scattering by short- and long-range ordered ensembles of particles. The described results can be used for solution of a number of practical problems of optics, photonics and optoelectronics.

2.2 Basic Approaches to Describe Wave Propagation and Scattering in Particulate Media

2.2.1 Interaction of Waves with Ensembles of Discrete Scatterers

In this section we consider the most common methods to describe the interaction of waves with ensembles (three-dimensional in the general case) of discrete scatterers. There are two main approaches to solve the problem of multiple scattering of waves in such ensembles: the radiative transfer theory and the rigorous analytic theory (Ishimaru 1978a).

The *theory of the radiative transfer,* describing the energy transfer in a particulate medium, was developed in the late 19th - early 20th centuries by Lommel (1887), Chwolson (1889), and Schuster (1905). This theory is based on the transport equation operating with the wave intensities and phenomenological characteristics of the medium (coefficients of scattering, absorption, and extinction and the phase function (also known as the scattering indicatrix)). This theory assumes no correlation in the particle positions. It is successfully used to solve the problems of the atmosphere and ocean optics, analysis of the propagation of radiation in the atmospheres of planets, stars and galaxies, and other random media (Ishimaru 1978a, b; Sobolev 1956b; Chandrasekhar 1960; Ivanov et al. 1988; Apresian and Kravtsov 1983; Zege et al. 1991a; Mishchenko et al. 2006). The theory is valid when the distance between particles are much larger than their size and the wavelength of the incident light.

A rigorous theory (also known as the *wave theory)* or the *theory of multiple scattering of waves* (TMSW) is based on the solution of the fundamental equations for fields of waves scattered by ensembles of particles. It takes into account effects of multiple scattering, diffraction and interference. This theory is used to solve the problem of wave scattering by particulate media, when the contribution of multiple scattering in the *resulting field* has to be taken into account.

The problem is solved analytically only for simplest systems. In 1935, Trinks solved it for two small spheres symmetrically arranged with respect to the direction of illumination (Trinks 1935). Germogenova (1964) generalized this approach for the case of spheres of different sizes and arbitrary direction of illumination. Twersky (1962c) applied the "successive-scattering" procedure to describe multiple scattering on two identical scatterers (in fact - point scatterers) at normal (to the line connecting the centers of the particles) illumination by a plane wave. Liang and Lo (1967) and Bruning and Lo (1971a, b) solved the problem of multiple scattering of electromagnetic (EM) waves on two spheres with arbitrary sizes and refractive indices. Fuller and Kattawar (1988a, b) proposed technique to calculate scattering characteristics of two and more spheres. It is based on the expansion of the field of the scattered wave in terms of orders of scattering. Cooray and Ciric (1989) solved the problem of EM wave scattering by system of two perfectly conducting arbitrary oriented prolate spheroids.

When the more complex systems are considered, e.g. ensembles consisting of large number of arbitrary arranged particles with arbitrary size, shape, orientation, etc., commonly used approach is based on the *self-consistency* of equations describing the field in the observation point. Self-consistency procedure is based on the assumption that the wave emitted by each scatterer is determined by the *effective field* acting on the scatterer. This field is the sum of field of incident wave and fields of waves emitted by the other scatterers (the effective field includes all scattering orders). Self-consistency procedure, also called as the *self-consistent approach* or the *self-consistent field method* (see., e.g., Lax 1951; Vereshchagin et al. 1991; Ponyavina and Silvanovich 1994; Bogomolov et al. 1997; Zamkovets et al. 2003b; Ponyavina et al. 2004 where the method was used to find the field in the systems composed of particulate monolayers), was suggested by Ewald to describe X-ray scattering by crystals (Ewald 1916). Foldy applied it to solve the problem of multiple scattering

of scalar waves by media with randomly distributed point scatterers (Foldy 1945). Later the procedure became the basis for most studies within the TMSW.

Consider the equations of the TMSW and the procedure of their self-consistency (Foldy 1945; Lax 1951) on the example of system of point scatterers in the scalar field. Let there be a medium consisting of $N >> 1$ particles placed in the scalar field ψ^i of the incident wave satisfying the wave equation (Helmholtz equation): $\nabla^2\psi + k^2\psi = 0$, where k is the wave number in the medium without of particles. The particles are distributed in the volume V, their positions are specified by the radius-vectors $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N$. Field $\psi_\mathbf{r}$ at some point $\mathbf{r}$ outside the particles is determined by the sum of the filed $\psi^i_\mathbf{r}$ of incident wave and the fields of waves scattered by all the particles:

$$\psi_\mathbf{r} = \psi^i_\mathbf{r} + \sum_{j=1}^{N} t_{\mathbf{r}j}\psi^e_j, \tag{2.2.1}$$

where the $t_{\mathbf{r}j}\psi^e_j$ quantity describes field scattered into the point $\mathbf{r}$ by j-th particle acted by the effective field ψ^e_j, $t_{\mathbf{r}j}$ is the scattering operator. The effective field acting on a particle (e.g., k-th) is the sum of the filed ψ^i_k of incident wave and fields of waves scattered by other particles:

$$\psi^e_k = \psi^i_k + \sum_{j=1, j\neq k}^{N} t_{kj}\psi^e_j. \tag{2.2.2}$$

Formulae (2.2.1) and (2.2.2) are the basic equations of the TMSW.

Field $\psi_\mathbf{r}$, i.e. the solution of Eq. (2.2.1), can be found by the order-of-scattering expansion of fields included in (2.2.1) and (2.2.2) (see Appendix). However, for arrays consisting of huge number ($N \to \infty$) of arbitrary arranged particles this method is impractical due to computational complexity.

It is convenient to express the unknown field $\psi_\mathbf{r}$ as the sum of coherent $\langle\psi_\mathbf{r}\rangle$ (average field) and incoherent $v_\mathbf{r}$ (fluctuating field) components: $\psi_\mathbf{r} = \langle\psi_\mathbf{r}\rangle + v_\mathbf{r}$ and use statistical approach for these fields determination. The procedure of finding the $\langle\psi_\mathbf{r}\rangle$ and $v_\mathbf{r}$ fields is described, for example, in the works of Lax (1951, 1952).

Let us consider derivation of the equations for the average field $\langle\psi_\mathbf{r}\rangle$. Assume that an ensemble consists of $N \to \infty$ particles. Their positions and states (size, orientation, optical constants, etc.) in general are arbitrary. The average field at the point $\mathbf{r}$ can be obtained by averaging the field $\psi_\mathbf{r}$, found for all possible configurations of the positions and states of particles. That is practically impossible-doing task when $N \to \infty$. Therefore, the other approach, developed by Foldy (1945), is used. It is based on averaging the total ($\psi_\mathbf{r}$) and effective (ψ^e) fields in Eqs. (2.2.1) and (2.2.2) and finding the solutions for average quantities.

Averaging is carried out by the integration over the volume V where the scatterers are located using the multiparticle probability distribution functions describing the distribution of particles over the spatial position and state. The quantity averaged in such a manner is known as the *configurational average* (Foldy 1945). N-particle

probability distribution function has the form: $p(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N; s_1, s_2, \ldots, s_N)$, where $\mathbf{r}_i$ and s_i describe spatial position and state of i-th particle. The probability of finding the first particle in the elementary volume $\mathrm{d}\mathbf{r}_1$ in the neighborhood of $\mathbf{r}_1$ in the state $\mathrm{d}s_1$, the second particle in the elementary volume $\mathrm{d}\mathbf{r}_2$ in the neighborhood of $\mathbf{r}_2$ in the state $\mathrm{d}s_2$ etc. is defined as follows: $\mathrm{d}w = p(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N; s_1, s_2, \ldots, s_N)\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2 \ldots \mathrm{d}\mathbf{r}_N\mathrm{d}s_1\mathrm{d}s_2 \ldots \mathrm{d}s_N$.

The probability distribution function is normalized:

$$\int_V p(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N; s_1, s_2, \ldots, s_N)\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2 \ldots \mathrm{d}\mathbf{r}_N\mathrm{d}s_1\mathrm{d}s_2 \ldots \mathrm{d}s_N = 1. \quad (2.2.3)$$

Here and below is used a short notation for the integrals:

$$\int_V f(\mathbf{r}_1, \ldots, \mathbf{r}_N; s_1, \ldots, s_N)\mathrm{d}\mathbf{r}_1 \ldots \mathrm{d}\mathbf{r}_N\mathrm{d}s_1 \ldots \mathrm{d}s_N$$
$$\equiv \int_V \mathrm{d}\mathbf{r}_1 \ldots \int_V \mathrm{d}\mathbf{r}_N \int_V \mathrm{d}s_1 \ldots \int_V \mathrm{d}s_N f(\mathbf{r}_1, \ldots, \mathbf{r}_N; s_1, \ldots, s_N).$$

The expression for the field, averaged over the ensemble, can be written as:

$$\langle \psi_{\mathbf{r}} \rangle = \int_V \psi_{\mathbf{r}}(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N; s_1, s_2, \ldots, s_N) p(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N; s_1, s_2, \ldots, s_N)$$
$$\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2 \ldots \mathrm{d}\mathbf{r}_N\mathrm{d}s_1\mathrm{d}s_2 \ldots \mathrm{d}s_N. \quad (2.2.4)$$

If the position and state of, for example, the first particle is known (in other words, if the particle is fixed in the point $\mathbf{r}_1$ and state s_1), the field averaged over all other particles can be written as (Lax 1951):

$$\langle \psi_{\mathbf{r}} \rangle_1 = \int_V \psi_{\mathbf{r}}(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N; s_1, s_2, \ldots, s_N) p(\mathbf{r}_2, \ldots, \mathbf{r}_N; s_2, \ldots, s_N | \mathbf{r}_1; s_1)$$
$$\mathrm{d}\mathbf{r}_2 \ldots \mathrm{d}\mathbf{r}_N\mathrm{d}s_2 \ldots \mathrm{d}s_N, \quad (2.2.5)$$

where the subscript "1" in $\langle \psi_{\mathbf{r}} \rangle_1$ denotes averaging over the positions and states of all the particles but for the first one, the $p(\mathbf{r}_2, \ldots, \mathbf{r}_N; s_2, \ldots, s_N | \mathbf{r}_1; s_1)\mathrm{d}\mathbf{r}_2 \ldots \mathrm{d}\mathbf{r}_N\mathrm{d}s_2 \ldots \mathrm{d}s_N$ describes the conditional probability of finding all particles, except for the first one, at the points $\mathbf{r}_2, \ldots, \mathbf{r}_N$ in the states $s_2, \ldots, s_N$, with condition that the first one is located at $\mathbf{r}_1$ in the state s_1.

If two particles are fixed at $\mathbf{r}_1$ and $\mathbf{r}_2$ points in s_1 and s_2 states, then

$$\langle\psi_{\mathbf{r}}\rangle_{1,2} = \int\limits_V \psi_{\mathbf{r}}(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N; s_1, s_2, \ldots, s_N) p(\mathbf{r}_3, \ldots, \mathbf{r}_N; s_3, \ldots, s_N | \mathbf{r}_1, \mathbf{r}_2; s_1, s_2) \mathrm{d}\mathbf{r}_3 \ldots \mathrm{d}\mathbf{r}_N \mathrm{d}s_3 \ldots \mathrm{d}s_N \tag{2.2.6}$$

The probability distribution is associated with the distribution of particle concentration (density) (Lax 1951):

$$\rho(\mathbf{r}_1; s_1) = Np(\mathbf{r}_1; s_1), \tag{2.2.7}$$

$$\rho(\mathbf{r}_1, \mathbf{r}_2; s_1, s_2) = N^2 p(\mathbf{r}_1, \mathbf{r}_2; s_1, s_2). \tag{2.2.8}$$

Here $\rho(\mathbf{r}_1; s_1)$ is the concentration (density) of particles in the state s_1 in a neighborhood of $\mathbf{r}_1$, $\rho(\mathbf{r}_1, \mathbf{r}_2; s_1, s_2)$ is the concentration of particles in the state s_1 in a neighborhood of $\mathbf{r}_1$ and particles in the state s_2 in a neighborhood of $\mathbf{r}_2$.

Using Eqs. (2.2.3)–(2.2.6) we can write the equation for the average field $\langle\psi_{\mathbf{r}}\rangle$ at some point $\mathbf{r}$ outside the particles. The expression (2.2.1) takes the form:

$$\langle\psi_{\mathbf{r}}\rangle = \psi^i_{\mathbf{r}} + \sum_{j=1}^{N} \int\limits_V t_{\mathbf{r}j} \left\langle \psi^e_j \right\rangle_j p(\mathbf{r}_j; s_j) \mathrm{d}\mathbf{r}_j \mathrm{d}s_j. \tag{2.2.9}$$

The right side of Eq. (2.2.9) contains two averaging operations. One of them describes the averaging the effective field acting on the j-th particle located at point $\mathbf{r}_j$ in state s_j, over all possible positions and states of other particles: $\langle\psi^e_j\rangle_j$. This field is differed from the average field $\langle\psi_{\mathbf{r}}\rangle$ by contribution from the j-th particle. Another one describes the field scattered by the j-th particle into the point $\mathbf{r}$, averaged over the position and the state of the particle: $\int\limits_V t_{\mathbf{r}j} \left\langle \psi^e_j \right\rangle_j p(\mathbf{r}_j; s_j) \mathrm{d}\mathbf{r}_j \mathrm{d}s_j$. This integral gives the contribution to the average field $\langle\psi_{\mathbf{r}}\rangle$ from the particle averaged over all possible locations and states in the volume V. From this it follows that for any $j = [1 \ldots N]$ (i.e. any "averaged" particle of the ensemble) the integral has the same value (at that index j is the summation index over all particles rather than specific particle number, and $\mathbf{r}_j$ and s_j are the dummy variables of integration). Accordingly, in Eq. (2.2.9) are summed up N identical integrals: $\langle\psi_{\mathbf{r}}\rangle = \psi^i_{\mathbf{r}} + N \int\limits_V t_{\mathbf{r}j} \left\langle \psi^e_j \right\rangle_j p(\mathbf{r}_j; s_j) \mathrm{d}\mathbf{r}_j \mathrm{d}s_j$. Using Eq. (2.2.7), we obtain:

$$\langle\psi_{\mathbf{r}}\rangle = \psi^i_{\mathbf{r}} + \int\limits_V t_{\mathbf{r}j} \left\langle \psi^e_j \right\rangle_j \rho(\mathbf{r}_j; s_j) \mathrm{d}\mathbf{r}_j \mathrm{d}s_j. \tag{2.2.10}$$

If the states of particles do not depend on particle locations, then $\rho(\mathbf{r}_j; s_j) = \rho(\mathbf{r}_j)\rho(s_j)$ and Eq. (2.2.10) takes the form: $\langle\psi_{\mathbf{r}}\rangle = \psi^i_{\mathbf{r}} + \int\limits_V t_{\mathbf{r}j}\left\langle\psi^e_j\right\rangle_j \rho(\mathbf{r}_j)\rho(s_j)\mathrm{d}\mathbf{r}_j \mathrm{d}s_j$. If the states of all particles are the same, and particle concentration is constant throughout the volume under consideration ($\rho(\mathbf{r}_j) = \rho_0 = \text{const}$), then $\langle\psi_{\mathbf{r}}\rangle = \psi^i_{\mathbf{r}} + \rho_0\int\limits_V t_{\mathbf{r}j}\left\langle\psi^e_j\right\rangle_j \mathrm{d}\mathbf{r}_j$.

A rigorous solution of Eq. (2.2.10) can be obtained as follows. Averaging Eq. (2.2.2) by analogy with Eq. (2.2.1) yields (Lax 1951, 1952):

$$\left\langle\psi^e_k\right\rangle_k = \psi^i_k + \int\limits_V t_{kj}\left\langle\psi^e_j\right\rangle_{jk}\rho(\mathbf{r}_j; s_j|\mathbf{r}_k; s_k)\mathrm{d}\mathbf{r}_j\mathrm{d}s_j. \tag{2.2.11}$$

Thus, the effective field $\left\langle\psi^e_j\right\rangle_j$ with one fixed particle (Eq. (2.2.10)) is expressed in terms of effective field $\left\langle\psi^e_j\right\rangle_{jk}$ with two fixed particles, which, in turn, is expressed in terms of effective field $\left\langle\psi^e_j\right\rangle_{jkl}$ with three fixed particles:

$$\left\langle\psi^e_k\right\rangle_{kl} = \psi^i_k + \int\limits_V t_{kj}\left\langle\psi^e_j\right\rangle_{jkl}\rho(\mathbf{r}_j; s_j|\mathbf{r}_k, \mathbf{r}_l; s_k, s_l)\mathrm{d}\mathbf{r}_j\mathrm{d}s_j. \tag{2.2.12}$$

Expressing in this manner the effective field with n fixed particles through the effective field with $n+1$ fixed particles one can obtain hierarchy of N equations for finding the avarage field; Eqs. (2.2.10)–(2.2.12) are the first three of them. The solution of N equations of the hierarchy allows one to find the average field at a point $\mathbf{r}$ outside the particles for an N-particle ensemble.

If all particles are located in the sites of perfect crystal lattice, then to obtain a rigorous solution only two Eqs. (2.2.10) and (2.2.11) are needed. In this case, any fixed particle and two-particle distribution function completely describe structure of the ensemble. Fixation of one or two particles yields the same result for the effective field: $\left\langle\psi^e_j\right\rangle_j = \left\langle\psi^e_j\right\rangle_{jk}$. That allows one to write the integral equation

$$\left\langle\psi^e_k\right\rangle_k = \psi^i_k + \int\limits_V t_{kj}\left\langle\psi^e_j\right\rangle_j \rho(\mathbf{r}_j; s_j|\mathbf{r}_k; s_k)\mathrm{d}\mathbf{r}_j\mathrm{d}s_j. \tag{2.2.13}$$

Its solution gives the effective field.

In general, the accuracy of the solution of Eq. (2.2.10) depends on the number of Eqs. (2.2.10)–(2.2.12), etc. taken into account in calculation. In some cases, the approximations yield a good accuracy when the number of equations much smaller than N.

The simplest in the framework of this "hierarchical" approach is the single scattering approximation (SSA) in which the effective field (in Eq. (2.2.10)) is replaced by the incident wave field $\left\langle \psi_j^e \right\rangle_j \approx \psi_j^i$:

$$\langle \psi_{\mathbf{r}} \rangle = \psi_{\mathbf{r}}^i + \int_V t_{\mathbf{r}_j} \psi_j^i \rho(\mathbf{r}_j; s_j) \mathrm{d}\mathbf{r}_j \mathrm{d}s_j. \tag{2.2.14}$$

This replacement means that each particle is illuminated only by the incident wave: contributions from the other particles are not taken into account. In this case only interference of singly scattered waves at the observation point $\mathbf{r}$ is taken into account, and the reillumination between particles is neglected. For thin layers and monolayers of particles, this approximation can be successfully used at low concentrations, for large (in comparison with the wavelength) sizes of particles and for particles with a strongly elongated phase functions (Ivanov et al. 1988).

For ensemble of point scatterers Foldy suggested approximation, in which the effective field is replaced by the average field (Foldy 1945): $\left\langle \psi_j^e \right\rangle_j \approx \langle \psi_j \rangle$. In this case Eq. (2.2.10) takes the form:

$$\langle \psi_{\mathbf{r}} \rangle = \psi_{\mathbf{r}}^i + \int_V t_{\mathbf{r}_j} \langle \psi_j \rangle \rho(\mathbf{r}_j; s_j) \mathrm{d}\mathbf{r}_j \mathrm{d}s_j, \tag{2.2.15}$$

where $\langle \psi_j \rangle$ is the field at the point $\mathbf{r}_j$, averaged over the positions and states of all the particles, except for the j-th particle (i.e., the average field, which would exist at the point $\mathbf{r}_j$ in the absence of the j-th particle, created by all other particles). In this approximation it is neglected the contribution from the j-th particle in the effective field at point $\mathbf{r}_j$. Since the effective field at this point is differed from the average field by contribution from a single particle, the *Foldy approximation*, also known as the *effective field approximation* (EFA) (Tsang and Kong 1980; Wang et al. 1994; Siqueira and Sarabandi 1996; Guérin et al. 2006; Barrera et al. 2007) gives the more accurate results, the larger number N of particles is considered. A number of studies have shown that the Foldy approximation is valid for sparse media when correlation in spatial arrangement of particles is negligible (Wang et al. 1994; Waterman and Truell 1961; Brown 1980; Javanaud and Thomas 1988). Thus, according to the estimation made in Javanaud and Thomas (1988), the range of its applicability is limited to the volume filling factor (volume fraction of particles in the considered space region) $\eta_V \approx 0.01$.

The discrepancy of the results obtained in the Foldy approximation with the experimental data (see, for example, Wang et al. 1994, West et al. 1994) for the densely packed media can also be interpreted from general physical considerations. Indeed, with particle concentration increasing, main contribution to the effective field acting on some particle, e.g. j-th one, is produced by its neighbors. This occurs because the particle in question is "shielded" by a "shell" from neighboring particles, which

diminish its interaction with the other (located outside the "shell") ones. Thus, the total field is mainly formed by a relatively small number of closely spaced particles. It means that each of them contributes significantly to the formation of the average field, and the effective field acting on j-th particle is significantly differed from the average field produced by all these particles including j-th one. Correspondingly, the replacement of the effective field by the average one leads to incorrect results.

At considering the densely packed particulate media (when finiteness of particle sizes is the essential factor) the approximations of higher orders should be used. To describe scattering in such media Lax suggested *quasicrystalline approximation* (QCA) (Lax 1952). In this approach it is assumed that the effective field at one fixed particle approximately equals to the effective field at two fixed particles: $\left\langle \psi_j^e \right\rangle_j \approx \left\langle \psi_j^e \right\rangle_{jk}$. At this assumption hierarchy of Eqs. (2.2.10)–(2.2.12), etc. reduces to two equations: (2.2.10) and (2.2.13). In the case of perfect crystals the QCA gives the rigorous solution (see Eq. (2.2.13)), since fixation of one or two particles leads to the same result for the effective field. In other words, the structure of ensemble is completely described by a single fixed particle and two-particle distribution function (fixation of the second particle does not give addition information about the ensemble). Accordingly, for any two particles $\left\langle \psi_j^e \right\rangle_j = \left\langle \psi_j^e \right\rangle_{jk}$, and Eq. (2.2.11) includes the effective fields of the same type (Lax 1952). In the case of partially ordered ensembles the QCA gives an approximate solution which is the more accurate the closer the structure of medium to the structure of perfect crystal. Therefore, the approximation had been called as "quasicrystalline" (Lax 1952). The QCA is widely used in study of the densely packed particulate media (Tsang and Kong 1980, 1982a, 1983, 1992b; West et al. 1994; Varadan et al. 1979, 1983, 1984, 1985a, b, 1987, 1989; Ma et al. 1988, 1990; Tsang and Ishimaru 1985b; Tsang et al. 1982, 1992, 2000, 2007; Ding and Tsang 1988; Varadan and Varadan 1980; Bringi et al. 1982; Davis and Schwartz 1985; Ao and Kong 2002; Tishkovets and Jockers 2006; Tishkovets 2007, 2010; Tishkovets et al. 2011; Dick and Ivanov 1999; Maurel et al. 2010; Parnell and Abrahams 2010; Parnell and Martin 2011), because it provides a good agreement with the experimental data (Ponyavina and Silvanovich 1994; West et al. 1994; Varadan et al. 1983, 1985a, b, 1989; Tsang and Kong 1983, 1992b; Dick and Ivanov 1999; Hong 1980; Ponyavina and Sil'vanovlch 1990). With regard to the partially ordered monolayers of particles (including the study of multilayer systems consisting of such monolayers), the QCA was used in Vereshchagin et al. (1991), Ponyavina and Silvanovich (1994), Bogomolov et al. (1997), Zamkovets et al. (2003b), Ponyavina et al. (2004), Hong (1980), Ponyavina and Sil'vanovlch (1990), Vereshchagin et al. (1990), Loiko and Molochko (1995, 1996), Loiko et al. (1998, 2000, 2005b), Ponyavina (1998), Kachan and Ponyavina (2002a, b), Kachan et al. (2006), Kinnan et al. (2009), Loiko and Miskevich (2004, 2005a, b, 2013), Miskevich and Loiko (2011a, b, 2013a, b, 2014b, c, 2015b).

Further efforts in the TMSW development were directed to solve the problems of scattering by ensembles of particles with an arbitrary size, shape, orientation, etc.; to solve the problem of electromagnetic (vector) wave scttering; to obtain various approximate solutions and determine their applicability domains; to create and develope calculation methods and obtain theoretical results for particular scattering media.

So, Waterman, Truell and Fikioris developed formalism of the TMSW to describe scattering of scalar waves by ensembles of finite-sized spheres (Waterman and Truell 1961; Fikioris and Waterman 1964). Mathur and Yeh developed the theory to describe scattering of the EM waves by ensembles of particles of arbitrary size. Their solution is based on the multipole expansion of the vector fields (Mathur and Yeh 1964). It should be noted that the work of Fikioris and Waterman (2013) devoted to the generalization of TMSW to the case of EM waves scattering was prepared in 1964, but it was published only in 2013 in a special issue of the "Journal of Quantitative Spectroscopy & Radiative Transfer", dedicated to scientific heritage of P. Waterman.

Twersky studied the various aspects of multiple scattering of waves by a rough surface, ensembles of randomly and regularly arranged cylinders and other particles (Twersky 1952a, b, c, 1962a, b, c, d, 1964, 1967, 1970a, b, 1975a, b, c, 1978). He considered the media in which the particles are located far enough from each other, that allowed him to simplify the solutions. In particular, expanding fields in terms of the scattering orders (see Appendix) he used approximation (today known as the *Twersky approximation* (Mishchenko et al. 2006; Mishchenko 2008a) which takes into account only the "trajectories" of scattering in which the wave passes each particle only once, i.e., the cyclic trajectories (including "forward-backward" scattering) are neglected (see, for example, Twersky 1964):

$$\psi_{\mathbf{r}} = \psi^i_{\mathbf{r}} + \sum_{j=1}^{N} t_{\mathbf{r}j}\psi^i_j + \sum_{j=1}^{N}\sum_{\substack{k=1\\k\neq j}}^{N} t_{\mathbf{r}j}t_{jk}\psi^i_k + \sum_{j=1}^{N}\sum_{\substack{k=1\\k\neq j}}^{N}\sum_{\substack{l=1\\l\neq j\,l\neq k}}^{N} t_{\mathbf{r}j}t_{jk}t_{kl}\psi^i_l + \sum_{j=1}^{N}\sum_{\substack{k=1\\k\neq j\,l\neq k,m\neq j}}^{N}\sum_{\substack{l=1\\l\neq j\,l\neq k}}^{N}\sum_{\substack{m=1\\m\neq k,m\neq j}}^{N} t_{\mathbf{r}j}t_{jk}t_{kl}t_{lm}\psi^i_m + \cdots . \quad (2.2.16)$$

The expansion (2.2.16) for the average field at $N \to \infty$ can be written as:

$$\langle\psi_{\mathbf{r}}\rangle = \psi^i_{\mathbf{r}} + \int_V t_{\mathbf{r}j}\psi^i_j\rho(\mathbf{r}_j,s_j)\mathrm{d}\mathbf{r}_j\mathrm{d}s_j + \int_V t_{\mathbf{r}j}t_{jk}\psi^i_k\rho(\mathbf{r}_j,s_j)\rho(\mathbf{r}_k,s_k)\mathrm{d}\mathbf{r}_j\mathrm{d}\mathbf{r}_k\mathrm{d}s_j\mathrm{d}s_k + \int_V t_{\mathbf{r}j}t_{jk}t_{kl}\psi^i_l\rho(\mathbf{r}_j,s_j)\rho(\mathbf{r}_k,s_k)\rho(\mathbf{r}_l,s_l)\mathrm{d}\mathbf{r}_j\mathrm{d}\mathbf{r}_k\mathrm{d}\mathbf{r}_l\mathrm{d}s_j\mathrm{d}s_k\mathrm{d}s_l + \int_V t_{\mathbf{r}j}t_{jk}t_{kl}t_{lm}\psi^i_m\rho(\mathbf{r}_j,s_j)\rho(\mathbf{r}_k,s_k)\rho(\mathbf{r}_l,s_l)\rho(\mathbf{r}_m,s_m)\mathrm{d}\mathbf{r}_j\mathrm{d}\mathbf{r}_k\mathrm{d}\mathbf{r}_l\mathrm{d}\mathbf{r}_m\mathrm{d}s_j\mathrm{d}s_k\mathrm{d}s_l\mathrm{d}s_m + \cdots . \quad (2.2.17)$$

Twersky showed that the order-of-scattering expansion of the Foldy integral Eq. (2.2.15) is equivalent to the expansion (2.2.17) (Twersky 1964). Indeed, from the analysis of Eqs. (2.2.16) and (2.2.17) it follows that when the effective field is formed in point $\mathbf{r}_j$ the contribution of j-th particle is excluded. Thus Twersky established physical interpretation of the Foldy integral equation: for ensembles of finite-size particles the effective field can be approximated by the average field at small concentration of particles when the cyclic trajectories can be neglected. Moreover, Brown showed (Brown 1980) that the domain of applicability of the Foldy equation (2.2.15), also known as the Foldy–Twersky equation (Ishimaru 1978b), is limited by media in which the contribution of multiple scattering to the coherent field formation is negligible. In (Ishimaru 1978b; Mishchenko et al. 2006) the relationship between Twersky approximation and the radiative transfer theory is considered.

The most common approach to solve the problems of multiple scattering of waves by ensembles of finite-sized particles is the expansion of the incident and scattered wave fields in terms of wave functions (modes, harmonics or multipoles), followed by finding a relationships between them.

Waterman developed a method (Waterman 1969) which allows one to express the expansion coefficients of the scattered wave via the expansion coefficients of the incident wave with the help of the so-called T-matrix (transition matrix). He used the T-matrix method for solving the problems of scattering of scalar (Waterman 1969) and vector (Waterman 1971) waves by an individual homogeneous particle. Peterson and Ström generalized the method to the case of multilayer particles and arbitrary clusters of non-spherical particles (Peterson and Ström 1973, 1974). The main advantage of the method is that if the matrix $\mathbf{T}$ and expansion coefficients of the incident wave are determined, then the scattered wave field can be found at any point in space outside the sphere circumscribing the scattering object (which generally may consist of a group of individual scatterers). Moreover, the matrix $\mathbf{T}$ depends only on the properties of the object, and does not depend on propagation directions and polarizations of the incident and scattered waves. Generally, this matrix has infinite size, which must be truncated in the calculations. The T-matrix method is typically used to calculate the scattering characteristics of small groups (clusters) of the particles, including the non-spherical ones. Active use of the method began since the 1990s with the development of computers and approaches to improve its stability and convergence (Mishchenko et al. 2004, 2006, 2014).

The multipole expansions of the fields contain an infinite number of terms (modes). In the case of small ("Rayleigh") particles (with size parameter $x << 1$; for spherical particle with diameter D the size parameter is defined as: $x = \pi D/\lambda$, where λ is the length of incident wave), it is enough to take into account only the first terms of the expansions to provide closed form solution with good accuracy. In many studies the formalism of solutions is developed in general, and the closed form solutions are obtained for such special cases.

So, Varadan et al. developed a method to solve the problem of multiple scattering of EM waves based on the use of T-matrix for coupling the expansion coefficients of the scattered by particle and acting on it (effective) fields (Varadan et al. 1979). To find the effective field the configurational averaging and the QCA were used. The

analytical solutions and numerical results are obtained for the phase velocity and attenuation coefficient in systems of spherical and oriented spheroidal particles at $x << 1$. In (Varadan and Varadan 1980) the method was generalized to the case of randomly oriented spheroids.

An important step in solving the problems of multiple scattering is to find the correlation functions describing the probability of the mutual arrangement of particles in space. For statistically uniform ensembles of particles is usually used the radial distribution function (RDF) $g(R) = Vp(\mathbf{r}_i|\mathbf{r}_j)$ which characterizes the probability of mutual location of i-th and j-th particles in the volume V at a distance $R = |\mathbf{r}_i - \mathbf{r}_j|$ between their centers (Ziman 1979; Percus and Yevick 1958; Loiko et al. 1985). In most of the above considered works it was assumed that $g(R) = 0$ if $R < 2a$, and $g(R) = 1$ otherwise (a is the radius of the sphere circumscribing the particle). This RDF describes the so-called "hole correction": equiprobable arrangement of particles in space regardless of the distance (if it is greater or equal $2a$) between their centers. This function can be used only at low particle concentrations (in particular, in Varadan and Varadan 1980, at volume filling factor $\eta_V \leq 0.05$ calculated for the spheres circumscribing the particles). For moderate and high concentrations it is necessary to use more sophisticated distribution functions.

With regard to the development of the method described in Varadan et al. (1979), Varadan and Varadan (1980) it was done for the ensembles of spheroids with random and specified orientations in Bringi et al. (1982), Varadan et al. (1984, 1987). In Bringi et al. (1982), Varadan et al. (1984) was used the pair correlation function (PCF) calculated in the Percus–Yevick approximation (Percus and Yevick 1958) for hard (mutually nonpenetrating) spheres. The method to calculate this function allowed the authours to consider systems with volume filling factor $\eta_V \leq 0.26$ (Bringi et al. 1982). The influence of "non-spherical statistics" on the coherent field in ensemble of oriented spheroids was studied in Varadan et al. (1987). It was shown that the use of spheres to simulate spheroids may lead to an underestimation or overestimation of the attenuation of the coherent field as compare with the use of actual geometry of particles. Approximation by spheres of the equivalent volume gives better results than approximation by the circumscribing spheres. Experimental verification of the method developed in Varadan et al. (1979), Varadan and Varadan (1980), Bringi et al. (1982) showed good agreement of the results for $x \leq 10$ and $\eta_V \leq 0.4$ (Varadan et al. 1983). Furthermore, in Varadan et al. (1983) it was found that multiple scattering effects must be taken into account if $\eta_V \geq\sim 0.01$. It was noted that the shape of the PCF is very important when $x \leq 10$. In Varadan et al. (1985b) the method based on the formalism of Green's functions was developed. A good agreement of calculation and experimental results was obtained for the intensity and the effective propagation constant for the considered volume filling factors: $\eta_V <\sim 0.48$. The method was also developed to find the incoherent component of light (Ma et al. 1988), the absorption of EM waves (Ma et al. 1990), and to describe the propagation of elastic waves (Varadan et al. 1985a, 1989) in ensembles of randomly distributed scatterers.

A significant contribution to the development of the TMSW is made by Tsang and co-authors (Tsang and Kong 1980, 1983, 1992a, b, 2001; Tsang and Ishimaru 1984, 1985a, b; Ding and Tsang 1988; Tsang et al. 1992, 2000; Ao and Kong 2002;

Ishimaru and Tsang 1988; Wen et al. 1990; Tsang et al. 1992, 2000, 2001a, 2007). They obtained the solution of the multiple scattering problem for ensembles of particles with sizes smaller than the wavelength (Tsang and Kong 1980), ensembles of correlated polydisperse particles (Ding and Tsang 1988), ensembles of randomly distributed oriented spheroids (Ao and Kong 2002). The formalism was developed to find the field of wave reflected from the semi-infinite particulate medium at oblique illumination (Tsang and Kong 1983, 1992b). It was described (Tsang and Ishimaru 1984, 1985a; Ishimaru and Tsang 1988) the experimentally observed effect of the *backscattering enhancement* (Kuga and Ishimaru 1984), also known as the effect of *weak localization* (Albada and Lagendijk 1985; Wolf and Maret 1985). On the basis of the strict TMSW was developed the *dense media radiative transfer theory* (Tsang and Ishimaru 1985b; Tsang et al. 1992, 2000, 2007; Tsang and Kong 1992a; Wen et al. 1990). It allows one to describe attenuation of the coherent field and the angular distribution of the intensity of wave scattered in the random particulate medium. The theory was applied, in particular, for the solution of the inverse problems to characterize the state of snow (Tsang et al. 1992, 2000, 2007) using the data of microwave remote sensing. In 2000–2001 various aspects of TMSW and its practical application were published in the three-volume monograph of authors (Tsang et al. 2000, 2001a; Tsang and Kong 2001).

It is important to define the applicability limits of the used approximations (Siqueira and Sarabandi 1996; Guérin et al. 2006; Barrera et al. 2007; Davis and Schwartz 1985; Dick and Ivanov 1999; Kelly and Wu 1993; Vries et al. 1998; Zaccanti et al. 2003; Cassier and Hazard 2013). Thus, in Davis and Schwartz (1985) it was found that when $x \to 0$ the QCA is equivalent to the Maxwell–Garnett approximation (Maxwell Garnett 1904; Markel 2016) which does not take into account correlations in the mutual arrangement of particles. It was concluded that in calculation of the effective permittivity of the disordered systems the only available approach taking into account the short-range order when $x \to 0$ is the *effective medium approximation* (EMA) (Roth 1974; Huisman et al. 1981; Schwartz and Plona 1984). The domains of applicability of the QCA and the Foldy approximation were investigated in Siqueira and Sarabandi (1996). In Cassier and Hazard (2013) it was given the mathematical justification of some known asymptotic models, including the of Foldy–Lax one.

Since 1990s, the intensive development and wide spread occurrence of a computer technology, and sufficiently well developed TMSW to that time made especially urgent a creation of effective algorithms for the numerical solution of the multiple scattering problem.

So, Mackowski analyzed multiple scattering of waves by an arbitrary configuration of spheres (Mackowski 1991). He rederived the addition theorems for vector spherical wave functions (VSWF) and developed simple recurrence relations for the addition coefficients, which allowed him to obtain an effective "order of scattering" solution to the resultant field. The approach is especially effective for ensembles (clusters), consisting of small the number of particles and ensembles of small particles. In Mackowski (1994) was developed a method to calculate and the numerical values were obtained for the extinction, scattering and absorption cross sections of clusters of spheres for both fixed and random orientations. It was noted that in the

case of the conductive spheres of small size ($x \leq 0.1$), the number N_s of expansion coefficients given by the widely used formula $N_s = x + 4x^{1/3} + 2$ may be not sufficient to obtain the proper solution for the extincion and absorption cross sections of pair of the touching spheres.

Xu in a series of papers studied in details the multiple scattering of waves by arbitrary clusters of spheres of an arbitrary size (Xu 1995, 1997, 1998a, b; Xu and Wang 1998). In Xu (1995), he obtained the solution based on the generalization of the Mie theory to the case of the illumination of particle by the EM wave of arbitrary profile. The computational problems of the method of the resulting field re-expansion in terms of the VSWF used in Xu (1995) were discussed in Xu (1997). Instead of such re-expansion, to find the far field was proposed to use the asymptotic expressions. They allow one to avoid convergence problems and significantly accelerate the calculations due to reducing the number of the expansion coefficients. He obtained analytical expressions for the amplitude scattering matrix. Efficient algorithms for calculating the VSWF-expansion coefficients were proposed in Xu (1998a). A good agreement of the calculation results based on the theory of Xu (1997) with the measured data were obtained in Xu and Wang (1998). It was noted that the strongest multiple scattering effects occur when the particles in the cluster are aligned along the direction of the incident wave propagation. The analytical expressions for the asymmetry parameter of the phase function (Bohren and Huffman 1983) and the scattering cross section of an arbitrary cluster of spheres, sizes of which are comparable with the length of the incident wave, were found in Xu (1998b). The good agreement between the results of calculation and experiment was demonstrated.

In Dufva et al. (2008) the addition theorems for scalar and vector spherical wave functions were derived. The derivation is simpler than the one obtained previously by other authors. As a result, the simpler expressions for the translation coefficients were obtained.

In Mishchenko et al. (2007), Okada and Kokhanovsky (2009) scattering by clusters of particles with $x = 4$ for $\eta_V \leq 0.3$ was investigated using the T-matrix method. It was shown that multiple scattering can make a noticeable contribution to the resulting field even at significant distances between particles (Mishchenko et al. 2007). Accounting the multiple scattering may increase extinction for systems of optically soft particles, and always leads to decreasing absorption as compared with the case of the independent scattering regime (Okada and Kokhanovsky 2009).

In recent decades, the TMSW was mainly extended to generalize the previously developed approaches for the ever greater range of heterogeneous media. In addition, the methods for solving the inverse problems were developed.

So, in Felbacq et al. (1994) the method of self-consistency was generalized to the case of scattering by ensemble consisting of a finite number of randomly positioned cylinders with arbitrary cross-sections. The method is especially suitable to study the phenomenon of enhanced backscattering by a set of arbitrary-shaped random rods. Generalization of the Foldy–Lax equations to describe multiple scattering of waves by a cluster of particles arranged in an absorbing isotropic medium was implemented in Mishchenko (2008a). Using these results, the transport equation for the coherent field of the sparse ensemble of particles was obtained (Mishchenko 2008a).

In Mishchenko (2008b) the full radiative transfer equation was derived. The Foldy–Lax equations are generalized (Huang et al. 2010a) to the case of multiple scattering of scalar waves by a two-dimensional system consisting of a large object (with the wavelength and overwavelength size), surrounded by point scatterers, when all particles are in far-field zones of each other. In Huang and Li (2010b) this approach was extended to the case of a three-dimensional system, and in Huang et al. (2013) to a system containing several large objects. In Hu et al. (2014) were considered the direct and inverse problems of scattering of scalar waves by a system consisting of point scatterers surrounding the object the size of which is comparable to or larger than the wavelength. In contrast to Huang et al. (2010a, 2013), Huang and Li (2010b), here the solution of the algebraic rather than integral equations was used. In Liao and Ji (2014) for ensembles of spheres with $x << 1$ the "extended" Foldy–Lax approximation taking into account the dipole effects and the effects of self-interaction was proposed.

Direct and inverse problems of multiple scattering of EM waves by the three-dimensional ensemble consisting of a finite number of point scatterers are discussed in Challa et al. (2014). To solve the direct problem the Foldy model and the method of regularization of the tensor Green's function were used. The Born approximation (single-scattering approximation) and the "intermediate" between the Born and Foldy approximations taking into account different number of scattering orders were considered. They were obtained by the iteration metod with the initial Born approximation. In the limit of the infinite scattering order they describe the Foldy model. The solution of the inverse problem was applied to determine the locations of sctterers and their scattering coefficients using data of the far-field measurements. The effect of multiple scattering on finding these quantities is discussed.

The equations for the coherent (direct) and incoherent (diffuse) components of light reflected from the plane-parrallel close-packed layer of finite thickness, consisting of randomly arranged non-spherical particles were obtained in Tishkovets (2007).

In Parnell et al. (2010) was found the dispersion integral equation for the effective wave number k_{eff} in sparse media. In the long-wave limit the results coincide with those obtained in the Foldy approximation. Multiple scattering of waves in a thin perforated plate was examined in Parnell and Martin (2011). An expression for k_{eff} was obtained. It was noted that the developed theory can be used for solving problems of the non-destructive testing of composite plates and thickness of the sea ice.

The reviews dedicated to the theory of multiple scattering can be found in Tishkovets et al. (2011), Lagendijk and van Tiggelen (1996). In Lagendijk and van Tiggelen (1996) was made a comparison of the theories dealed with the resonant multiple scattering of waves and electrons. In Tishkovets et al. (2011) was reviewed the current state of the TMSW. Some approaches are described and the expressions are given to solve the problems of scattering by particulate media. The cases of wave interaction with a semi-infinite random particulate medium, a system of two spheres, and closely packed clusters of particles are considered. The near-field and shadowing (shielding) effects were investigated. The consideration was focused on remote sensing problem, scattering by cosmic objects, etc. A comparison with the exact solution

of Maxwell's equations showed that the radiative transfer theory can be applied to ensembles of non-absorbing and weakly absorbing particles at $\eta_V \leq \sim 0.3$. It was noted that a quantitative theoretical interpretation of the measurement results for the intensity of radiation reflected from the densely packed media remains problematic. Therefore, for such media are commonly used approach is the numerical solution of the Maxwell's equations. However, its applicability is still limited by systems although with large but finite number of scatterers.

2.2.2 *Partially Ordered Monolayers and Systems of Monolayers*

Among ensembles of discrete scatterers one can distinguish the monolayers: the structures in which the particle centers are located in the same plane. One of the first solutions of multiple scattering problem for a densely packed monolayer of partially ordered polydisperse spherical particles normally illuminated by the plane EM wave was obtained by Hong (1980). His approach is based on the multipole expansion of the fields and using the QCA to determine the coherent field. A good agreement of calculation and experimental results for optical density of monolayers of small selenium particles was demonstrated.

The developed in Hong (1980) formalism was used in a number of works to study partially ordered particulate monolayers, i.e., monolayers with short-range order in the spatial arrangement of particles (Ponyavina and Sil'vanovlch 1990; Vereshchagin et al. 1990; Loiko and Molochko 1996; Loiko et al. 1998, 2000; Ponyavina 1998; Kachan and Ponyavina 2002a; Kinnan et al. 2009; Loiko and Miskevich 2005a, b), and layererd systems (multilayers, stacks) consisting of such monolayers (Vereshchagin et al. 1991; Ponyavina and Silvanovich 1994; Bogomolov et al. 1997; Zamkovets et al. 2003a, b; Ponyavina et al. 2004; Ponyavina 1998; Kachan and Ponyavina 2002b; Kachan et al. 2006; Loiko and Miskevich 2004, 2013; Loiko et al. 2005b; Miskevich and Loiko 2013a, b, 2014b).

So, in Vereshchagin et al. (1990) was studied the cooperative effects of the first kind (interference of fields scattered by ensemble of particles in the far field zone) and second kind (re-illumination of particles) on the formation of a coherent field of a monolayer. The dependences of the direct transmission coefficient of the normally illuminated monolayer (with a surface filling factor (the ratio of particle projections on the layer plane to the area where they are distributed) $\eta = 0.624$) of monodisperse spheres on their size parameter x was investigated. Calculations were carried out with (in the framework of the formalism of Hong 1980) and without (in the SSA) taking into account multiple scattering. It was shown that multiple scattering should be taken into account if $x < 2 \div 3$ and values of the relative refractive index of particles $n \geq 1.4$. Comparison of experimental and calculated direct transmittance T_c and specular reflectance R_c spectra of monolayer of polydisperse Al_2O_3 particles in polyethylene showed that the QCA gives better agreement with the measured data (Ponyavina and

Sil'vanovlch 1990) than the SSA. In Kachan and Ponyavina (2002a) was studied influence of concentration ($\eta = 0.2 \div 0.6$) of silver nanospheres (diameter $D = 2 \div 10$ nm) on T_c and R_c spectra of monolayer in the plasmon resonance region. The size dependence of refractive index n of the nanosphere particles was taken into account. It was found that size increasing results in R_c increasing and T_c decreasing in the plasmon resonance region. Increasing the monolayer filling factor η and the refractive index of the matrix where the particles are located leads to enhancement of resonance and to the long-wave shift of the plasmon resonance frequency. In Kinnan et al. (2009) the approach was used to find the optimal value of the refractive index of the dielectric matrix in order to obtain a sharp increase in extinction due to plasmon resonance.

Simple analytical expressions for T_c and R_c of the normally illuminated partially ordered monolayer of "Rayleigh" particles were obtained in Loiko and Molochko (1996) in the framework of dipole approximation. Comparison of the results with those obtained in the QCA revealed that the values of T_c and R_c for monolayer with $\eta = 0.5$ calculated with the simple formulas and with formulas of the QCA are the same at $x \leq 0.4$ for relative complex particle refractive index $m = 1.6 + 0.001i$ and at $x \leq 0.25$ for $m = 1.6 + 3i$. It was found that the results of calculations in the QCA and SSA are close for monolayers of large particles even at high particle concentration. A more detailed comparison of the QCA and the SSA is made in Loiko et al. (1998), where monolayers with $x \leq 10$, $\eta \leq 0.6$, and particle refractive indices $m \leq 1.4$ were considered. It was found that with x increasing and m decreasing the differences of calculation results obtained in the QCA and the SSA are reduced. Thus for $m = 1.1$ the difference is less than a few tenths of a percent, for $m = 1.2$ they are less than 1%, and for $m = 1.4$ are less than 5% at $\eta \leq 0.4$. In Loiko et al. (2000) the influence of multiple scattering on the interference *quenching effect* (Ivanov et al. 1988) for the coherent transmission coefficient of monolayer was investigated. It was shown that multiple scattering leads to the occurrence of the effect at larger sizes and concentrations of particles. The results are in good agreement with the experimental data for monolayers of latex particles in water. In Loiko and Molochko (1995) formalism developed in Hong (1980) was generalized to the case of oblique illumination of the monolayer. A method to estimate the applicability of the the QCA was proposed.

Among other approaches developed in the framework of the TMSW we single out here the methods described in Linton and Martin (2005), García-Valenzuela et al. (2012). In Linton and Martin (2005) the expression for the effective wave number in sparse random array of identical cylinders was derived using the quasicrystalline approximation. The expression is a "two-dimensional version" of Lloyd–Berry formula for the three-dimensional systems (Lloyd and Berry 1967). In García-Valenzuela et al. (2012), using the QCA, simple analytical expressions to calculate coherent transmission and reflection coefficients of the monolayer of randomly located spheres were derived. They can be used at any illumination and size of the particles. The simplification was achieved by replacing the actual effective field by the sum of two plane waves: incident and specularly reflected. The approach is applicable at low concentrations of particles.

In some cases it is convenient to consider a three-dimensional particulate medium as set (multilayer, stack) of plane-parallel monolayers. For example, in Vereshchagin et al. (1991) was developed an approach for finding the T_c and R_c coefficients of dispersion filters. Such filters consist of a set of periodically arranged plane-parallel monolayers of particles in a dielectric matrix. The calculation was conducted in two stages. First, in the QCA (Hong 1980) the amplitude direct transmittance and specular reflectance of an individual monolayer were calculated. Then, using the self-consistency procedure, they were used to find the transmittance and reflectance of a multilayer. In Ponyavina and Silvanovich (1994) the approach was applied to study the layered systems consisting of monolayers of polydisperse Al_2O_3 particles. Comparison of the calculation and the experimental data showed good agreement on the transmission coefficient of both single monolayer and the multilayer. In Bogomolov et al. (1997), Ponyavina (1998) the approach (Vereshchagin et al. 1991) was applied to calculate transmission and reflection spectra of three-dimensional photonic crystal (3D PhC): opal-like structure from spherical SiO_2 particles. In particular, it was determined the dependence of the spectral position λ_{PBG} of the photonic band gap (PBG) minimum on the refractive index n_{env} of environment in which the PhC was located. A satisfactory agreement of the theoretical and experimental results was observed (experimental and theoretical values of $\lambda_{\mathrm{PBG}}(n_{env})$ are close for small n_{env} and diverge with n_{env} growing, see Fig. 2.34c of Sect. 2.4.3.2). The discrepancy may be caused by several factors, in particular: (i) difference between the simulated and the actual structures of a single monolayer: actual monolayer had although imperfect, but rather highly ordered triangular lattice, while in theory the model of the partially ordered monolayer was used; (ii) mismatching the real and model filling factors of an individual monolayer: calculations were made at $\eta = 0.6$, while the filling factor of actual monolayer was close to the maximum (for monolayer with triangular lattice from monodisperse spherical particles $\eta_{\max} = \pi/(2\sqrt{3}) \approx 0.9069$); (iii) usage of the model of "statistically independent" monolayers, in which it is assumed that the adjacent monolayers are sufficiently far from each other, whereas in a real structure they are at a short (comparable with the wavelength) distance; (iv) usage of the model of homogeneous monodisperse spherical particles while in experiment the heterogeneity (porosity) of the structure of particles and their polydispersity take place.

In Zamkovets et al. (2003b), Kachan and Ponyavina (2002b) the approach described in Vereshchagin et al. (1991), Ponyavina and Silvanovich (1994), Bogomolov et al. (1997) was used to study the multilayers consisting of monolayers of silver nanoparticles in a dielectric matrix, and in Ponyavina et al. (2004) it was used to study the quasi-one-dimensional PhCs - the layered structures (multilayers) consisting of partially ordered monolayers of spheres with sizes comparable with the wavelength. In Kachan and Ponyavina (2002b) and Zamkovets et al. (2003b) it was found that the 1D ordering may lead to appearance of a doublet structure of the extinction spectrum and narrowing the reflection peak in the plasmon resonance region. In Ponyavina et al. (2004) it was shown that dependence of the PBG depth on the size, the refractive index of particles, and the intermonolayer distance is non-monotonic.

The method based on the QCA (Hong 1980) and the transfer-matrix method (TMM) (Katsidis and Siapkas 2002) for finding T_c and R_c of the multilayer was developed in Kachan et al. (2006). It was applied to study the light absorption by the system of monolayers of silver nanoparticles in the dielectric. It was shown that increase of absorption in a wide wavelength range can be achieved using the gradient multilayers. It should be noted that the method is simpler than the self-consistency method (Vereshchagin et al. 1991).

A model of addition (*adding method*) for coherent field was developed in Loiko and Molochko (1998), Loiko and Miskevich (2004), Loiko et al. (2005b). The analytical expressions for T_c and R_c of a thick layer of randomly arranged and partially ordered particles were derived. Under this model a layer is considered as a stack of plane-parallel elementary sublayers (monolayers). The transmission and reflection coefficients of individual monolayers were obtained in the SSA (Loiko and Molochko 1998) and the QCA (Loiko and Miskevich 2004; Loiko et al. 2005b). Oscillating dependences of T_c and R_c on the layer thickness were obtained. Multiple scattering leads to increasing the rate of oscillation decay with increasing the layer thickness (Loiko and Miskevich 2004; Loiko et al. 2005b). The results of Loiko and Molochko (1998) are in good agreement with the experimental data (Ishimaru and Kuga 1982).

2.2.3 Crystal-Like Structures

Spatially ordered particulate layers are the objects of intensive investigations due to their unique optical properties caused by the *crystal-like* structure of particle arrangement. Active research of such layers consisting of particles with sizes comparable with the wavelength of visible light began in the mid-20-th century with the study of colloidal crystals composed of the latex particles. Probably, the first systematic study of the optical properties of ordered systems was carried out in Alfrey et al. (1954). The authors observed iridescence of structures from monodisperse polystyrene particles and the visible light diffraction patterns similar to X-ray diffraction patterns by ordinary crystals. Authors of publications Krieger and O'Neill (1968) and Hiltner and Krieger (1969) proposed to use the observed diffraction effects to determine the particle sizes and lattice parameters of the colloidal crystals. Coloring of the samples was explained by the authors as a result of the Bragg diffraction by the periodically arranged monolayers of particles.

Crystal-like structure and, therefore, long-range order in the particle arrangement in ordered particulate layers cause the choice of specific approaches to their study. To date, for a rigorous theoretical description of the optical properties of such systems the most commonly used are methods based on the quantum-mechanical low-energy electron diffraction (LEED) theory (Kambe 1967, 1968; Pendry 1974) and

the Korringa–Kohn–Rostoker (KKR) method (Korringa 1947; Kohn and Rostoker 1954), generalized for photons. These methods take into account diffraction, interference and multiple scattering of waves. Let us consider the publications dedicated to the development and use of these methods.

In 1979 Ohtaka developed the theory of photon diffraction by a regular three-dimensional array of spheres (Ohtaka 1979). It is an extension of the LEED theory and the KKR method to the case of the EM fields. He used self-consistency procedure to take into account multiple scattering and the expansion of the fields and the Green's functions in terms of the vector spherical wave functions (VSWF). The equations for finding the fields and the photon energy bands were derived.

Lamb, Wood, and Ashcroft considered propagation of electromagnetic wave in a periodic array of spheres whose size is much smaller than the wavelength (Lamb et al. 1980). The theoretical description that takes into account multiple scattering effects is based on a generalization of the KKR method to the case of the EM fields. The domain of applicability of this approach was determined in Moroz (1994).

Stefanou and coauthors elaborated a method to calculate the transmittance, reflectance and photonic band structure of the three-dimensional photonic crystal, consisting of regularly arranged spheres (Stefanou et al. 1992). It is based on the consideration of the PhC as a set of plane-parallel regularly packed monolayers of particles, usage of the formalism developed by the authors earlier (Modinos 1987; Stefanou and Modinos 1991a) to find the scattering matrix of a single monolayer and the doubling method to calculate the scattering matrix of a multilayer. Calculations were made for the face-centered cubic (fcc) lattice of dielectric spheres. Later the method was named as the *layer-multiple-scattering* (LMS) method (Sainidou et al. 2005; Gantzounis and Stefanou 2006). The versions of the computer program implementing the method one can find in Stefanou et al. (1998, 2000). Its experimental verification is made in Yannopapas et al. (1997), Modinos et al. (2001a). In Modinos et al. (2001b) a theory to describe propagation of electronic, electromagnetic and elastic waves in three-dimensional periodic structures was presented. It is based on the method of Stefanou et al. (1992). In Gantzounis and Stefanou (2006) the method (Stefanou et al. 1992) was extended for layers with spheroidal particles, and in Christofi and Stefanou (2014) for layers with gyrotropic spherical particles. It was also developed and used in Yannopapas et al. (1999, 2002, 2003), Modinos et al. (2000), Psarobas et al. (2000), Almpanis et al. (2012), Yannopapas (2014).

For example, in Modinos et al. (2000) was studied the effect of imperfection of the finite-sized 3D PhC consisting of metal nanospheres on the PhC absorption coefficient. The work is an extension of a previously developed method to describe imperfection in the monolayer (Stefanou and Modinos 1993) for the case of layered system of monolayers. To simulate the disorder the *coherent potential approximation* (CPA) (Soven 1966, 1967) was used. The authors noted that it gives good results at least at a relatively low degree of disorder. The idea of the CPA is that the imperfect monolayers of actual particles are replaced by perfectly-periodic monolayers of "*effective scatterers*". The effective scatterers in different monolayers can be different. The characteristics of these scatterers depend in a complicated manner on the parameters of the monolayer. After their determination the calculations are

carried out in the framework of the previously developed methods for regular structures (Ohtaka 1979; Stefanou et al. 1992; Modinos 1987). The results for arrays of spheres ($D = 10$ nm) at different concentrations were given. The imperfection leads to the absorption increasing in the spectral region of the resonance dip. Impact of the disorder on absorption in thick PhC slab is less than in monolayer. In Almpanis et al. (2012) was studied the interaction of electromagnetic waves with a system of fractal structures from silver nanoparticles situated in a square lattice sites in the glass. Particle centers were located on the axis perpendicular to the lattice plane. Such structures can significantly enhance the local field near the boundaries of particles and thus be used as nanolenses. Their ordering into a regular lattice leads to a further enhancement of the near and far fields in certain spectral ranges. Thus, it was considered the possibility of using the hybrid nanostructures that have a combined effect of the near-field interaction and interaction caused by the long-range order. In Yannopapas (2014) was developed a "hybrid" method to model periodic structures, consisting of the "general scatterers". It is based on the discrete dipole approximation (DDA) and the LMS method (Stefanou et al. 1992). At first in the DDA is calculated the T-matrix of a single particle that is then used to find the characteristics of periodic ensemble of such particles by the LMS method (Stefanou et al. 1992). The approach was applied to study optical properties of the 2D and 3D lattices consisting of gold nanocubes.

Among other publications dedicated to the development and use of the KKR method, let us consider articles (Wang et al. 1993; Moroz 1995; Kafesaki and Economou 1999; Baryshev et al. 2007; Dorado and Depine 2009).

In Wang et al. (1993) the formalism to describe multiple scattering of electromagnetic waves in a 3D photonic crystal was developed on the basis of the KKR method for scalar waves. The photonic band structure of the "diamond" lattice of the touching spherical air voids in dielectric was calculated. It was noted that the main advantage of the theory of multiple scattering in comparison with other approaches is the direct calculation of the Green's function and the ability to consider defects and disorder of the lattice. In Moroz (1995) was developed a general theory of multiple scattering and KKR formalism for EM waves interacting with three-dimensional arrays of arbitrary shape particles. Its difference from the theory developed in Wang et al. (1993) was discussed.

In Kafesaki and Economou (1999) was developed a method for finding the transmission coefficients of regular and irregular clusters of particles (up to several hundred) and the band structures of infinite periodic lattices. The solution is based on the expansion of fields in terms of spherical wave functions. The method was used (Kafesaki et al. 2000) for calculating the transmission of acoustic waves propagating the particulate medium consisting of spherical air bubbles in water. It was shown that the gaps (which are the analogues of the PBG in the spectra of PhC) in the transmission spectra of such a medium are caused by multiple scattering and occur even at large disorder.

In Baryshev et al. (2007) were investigated "multicomponent" PhCs, consisting of inhomogeneous spheres (with radially-dependent refractive index). Their transmission spectrum and photonic band structure were calculated in frames of the KKR and the plane wave methods, respectively. It was shown that the considered type of the PhCs provides additional opportunities (as compared with the conventional PhCs) to control light fluxes. By these PhCs it is possible to "turn on and off" one stop band independently of the other ones. A good agreement between the results of calculation and experiment for opals was demonstrated.

In Dorado and Depine (2009) were studied the effects of the density of vacancies and the polydispersity of the particles on the transmission spectra of 3D photonic crystal. Calculations were made in the framework of KKR method using the averaged T-matrix, which characterizes the scattering properties of the particle, averaged over the size, position, shape, orientation, etc. At that instead of the actual lattice the perfect lattice with "average" spherical scatterers was considered. A good agreement with experimental data on the transmission and reflection of the colloidal photonic crystal was obtained. It was shown that the disorder gives a little effect on the transmittance and reflectance in the low-frequency and a strong one in the high-frequency spectral ranges. Also were considered the methods to simulate the disorder by specifying non-zero absorption index of particles and host medium. The methods yield close results and are valid for weak disorder of the PhC.

In the framework of the TMSW the 3D ordered arrays of particles were studied in works of Waterman and Pedersen (1986), Vereshchagin et al. (1991), Ponyavina and Silvanovich (1994), Bogomolov et al. (1997), Zamkovets et al. (2003b), Ponyavina et al. (2004), Kachan et al. (2006). In Vereshchagin et al. (1991), Ponyavina and Silvanovich (1994), Bogomolov et al. (1997), Zamkovets et al. (2003b), Ponyavina et al. (2004), Kachan et al. (2006) 3D particulate arrays were simulated as a periodic sequence of partially ordered monolayers. In Waterman and Pedersen (1986) scattering of electromagnetic waves by a semi-infinite periodic array of particles was considered using the T-matrix formalism. The analytical expressions for effective dielectric ε_{eff} and magnetic μ_{eff} permeabilities at $x << 1$ were obtained. The results of $\varepsilon_{eff}(x)$ and $\mu_{eff}(x)$ calculations for different lattice geometries were given. The behavior of the resonance peaks on these dependences was investigated. The expression for the reflection coefficient was obtained.

To describe interaction of waves with ordered systems are also used the *theory of dynamical diffraction* (Spry and Kosan 1986; Rundquist et al. 1989; Mittleman et al. 1999; Fedotov and Sel'kin 2011), various modifications of the *transfer matrix method* (Katsidis and Siapkas 2002; Pendry and MacKinnon 1992; Sigalas et al. 1996; Cassagne et al. 2000; Schilling et al. 2001; Rybin et al. 2009; Centurioni 2005; Troparevsky et al. 2010), the coupled dipole method (Chaumet et al. 2003), the Bragg diffraction theory (Kosobukin 2005), the scattering matrix method (Balestreri et al. 2006), and other methods (Nicorovici et al. 1995; Koenderink et al. 2005).

Since the late 1980s - early 1990s numerical methods to simulate the ordered structures are intensively developed. The impulse to the development was given by the works of John (1987) and Yablonovitch (1987), Yablonovitch and Gmitter (1989), Yablonovitch et al. (1991b, a).

So, in Yablonovitch (1987) was introduced the concept of *electromagnetic band gaps* (later these band gaps were called *photonic band gaps* (PBGs) (Yablonovitch and Gmitter 1989) or *stop gaps* (John 1987)) - the wavelength range of radiation, which cannot propagate in medium. Such band gaps arise in regular structures because of the periodic modulation of the refractive index. In Yablonovitch and Gmitter (1989) was studied the interaction of electromagnetic waves (in the microwave range) with dielectric fcc structures. The PBGs in the spectra of these structures was experimentally demonstrated. The concept of *photonic crystal* - a structure in spectrum of which there is a PBG - was introduced. It was found that a complete PBG (independent on the lighting direction) in fcc lattices can be realized with a relative refractive index of particles 3.5 and more. In Yablonovitch et al. (1991b) the structure possessing the complete PBG was prepared and studied in a microwave spectral range. It was a set of channels mechanically drilled and intersecting at a certain angle in the dielectric plate with refractive index $n \approx 3.6$. The volume fraction of voids was 78%. The calculations under the method of plane vector waves were shown that such structures have complete PBG when $n \geq 2.1$. In Yablonovitch et al. (1991a) were considered the so-called "donor" and "acceptor" modes, representing sharp peaks and dips in the region of the PBG when defects are introduced into the PhC structure. The magnitude and position of the dips and peaks depend on the size of the defects. These effects can be used in resonators and lasers with a low lasing threshold.

Numerical methods to simulate the interaction of waves with PhCs are based on Maxwell's equations. For their solution the most frequently used is the *plane-wave expansion method (plane wave method)* and its modifications (Satpathy et al. 1990; Leung and Liu 1990a, b; Economou and Zdetsis 1989; Zhang and Satpathy 1990; Ho et al. 1990; Sozuer et al. 1992; Meade et al. 1993; Busch and John 1998; van Driel and Vos 2000; Li and Zhang 2000; Galisteo-Lopez et al. 2003), the *finite element method* (Pendry and MacKinnon 1992), and the *finite difference time domain (FDTD) method* (Lavrinenko et al. 2009; Ivanov et al. 2010). The plane wave method on the one hand and the finite element and FDTD methods on the other hand are usually used to determine a photonic band structure and the transmission and reflection coefficients of the PhC, respectively.

The first investigations devoted to numerical and theoretical description of the photonic band structure of PhC were made for *scalar waves* (Satpathy et al. 1990; Leung and Liu 1990a; Economou and Zdetsis 1989; John and Rangarajan 1988), where was shown that fcc structure possesses the complete PBG when the relative refractive index of spheres is greater than $\sim 3^{1/2}$, that is inconsistent with the experimental results (Yablonovitch and Gmitter 1989).

The generalization of the plane wave method to the case of EM waves was made in Leung and Liu (1990b), Zhang and Satpathy (1990), Ho et al. (1990). In contrast to the experimental data of Yablonovitch and Gmitter (1989), the calculation results of these works show that it is impossible to obtain a complete PBG for the fcc lattice of spheres. In Ho et al. (1990) it was found that a structure from spheres arranged in a "diamond" lattice may possess such PBG. In Sozuer et al. (1992) the computational aspects of this method were considered. It was shown that discontinuous nature of the

dielectric function and the fields in a periodic array of dielectric spheres significantly limits its accuracy. The impovements of the method were proposed in Meade et al. (1993). They allow one to explore previously inaccessible to the theoretical analysis systems (e.g., lattices with the artificially introduced defects). In Busch and John (1998) under the plane wave method it was shown that the complete PBG can be obtained in lattices of voids in silicon or germanium, and that inverse opal-like structures possess such PBG if the refractive index contrast has a value of 2.8 or higher.

The formalism allowing the calculation of band structure and transmission coefficients of the ordered particulate medium is developed in Pendry and MacKinnon (1992). It is based on the solution of Maxwell's equations by the finite element method. It is a generalization of the low-energy electron diffraction theory to the case of the EM fields. Calculations are performed for a periodic array of dielectric cylinders. A good agreement with experiment for the band structure and transmission spectrum was obtained.

An overview of works on the theory of photon scattering by periodic structures was published in Pendry (1996). Various approaches for numerical solution of the problem of EM wave diffraction by such systems were considered. The main attention was paid to methods of the photonic band structure calculations.

Using numerical methods it is possible to simulate both perfect and imperfect PhCs. In the last case the most frequently used method is the *supercell technique*, under which the disorder is initially defined in a certain PhC's domain that is repeated then periodically to simulate the crystal structure. Imperfect PhCs were investigated, in particular in Li and Zhang (2000), Lavrinenko et al. (2009), Ivanov et al. (2010). So, in Li and Zhang (2000) was shown under the plane wave method that variations in size of spherical particles and in their deviations from the lattice sites greatly reduce the complete PBG in 3D PhC consisting of spherical voids in a dielectric matrix. When the magnitude of disorder is less than 2% of the lattice constant this PBG disappears even at high refractive index contrast ($n = 4$); PBG remains incomplete in this frequency domain. PBG is more sensitive to variations in the sizes of spheres than to their spatial deviations, since in the first case the filling factor varies as well. Dispersion of sphere positions and sizes was modeled using an uniform distribution. The PhC imperfection was simulated within the supercell technique. In Lavrinenko et al. (2009) in the framework of the FDTD method was studied disorder impact on photonic properties of the fcc lattice of voids in silicon. It was found that even small variations of the voids size and their deviations from the sites of perfect lattice lead to increasing the transmittance in PBG, i.e. degradation of photonic "isolating" properties of PhC. In Ivanov et al. (2010) was investigated the influence of positional disorder and polydispersity on the optical properties of 3D photonic crystal. Based on the plane wave and the FDTD methods it was shown that the disorder in the spatial positions and sizes of particles (spherical particles with $D = 200$ nm, $n = 1.6$ and 2.0 in a medium with the refractive index of 1.33 (in water) are considered) have little effect on the reflection spectrum in PBG.

In recent decades, active experimental studies of photonic crystals are carried out (Bogomolov et al. 1997; van Driel and Vos 2000; Galisteo-Lopez et al. 2003;

Rybin et al. 2008; Vasnetsov et al. 2014; Bertone et al. 1999; Koenderink et al. 2000; Mazurenko et al. 2003; Baryshev et al. 2003, 2011; Rengarajan et al. 2005; Muskens et al. 2011; Nair and Jagatap 2012).

In Bogomolov et al. (1997) was investigated the PBG transmission spectra of the 3D opal-like PhCs consisting of SiO_2 particles. When such PhCs are placed in diferent media, the relative refractive index of particles changes and, therefore, the depth and position of the PBG are changed as well. It was obtained a linear dependence of the spectral position of the PBG minimum on the refractive index n_{env} of the surrounding (host) medium when it changes from $n_{env} < n_p$ to $n_{env} > n_p$, where n_p is the particle refractive index. Similar experiments were carried out in Rybin et al. (2008), Vasnetsov et al. (2014), where the nonlinear dependences were obtained. A possible causes of the non-linearity are the heterogeneous (porous) structure of the silica particles and, as indicated by authors, dispersion of the particle parameters.

The dependence of the PBG width and the optical density of the 3D PhC slab consisting of the fcc lattice of submicron SiO_2 particles and the lattice of voids in the polymer matrix on the slab thickness was investigated in Bertone et al. (1999). The authors observed a monotonic increase of the optical density peak (corresponding to a minimum of the PBG) on the PhC thickness (the number of monolayers). The peak width decreases with increasing the number of monolayers up to a certain "critical" value, after which it practically does not change for the considered lattices.

Influence of thickness and imperfection of the PhC with the fcc lattice from polystyrene spheres on its optical properties in the PBG was studied in Galisteo-Lopez et al. (2003). It was shown that the intensity of the diffuse component of the transmitted and reflected light increases with frequency and the sample thickness. In the frequency range corresponding to the PBG edges the peaks of this intensity were occurred. The dip between these peaks was observed.

Influence of the disorder on optical properties of colloidal PhCs composed of SiO_2 particles was considered in Rengarajan et al. (2005). The radial distribution functions (RDFs) of particles in monolayers were found using the data on surface images. These functions had the shape of a sequence of sharp peaks in the region close to the coordinate origin associated with the particle center. The peaks are asymmetric: they grow sharper than fall with increasing the distance from coordinate origin. With the distance from the origin the peaks are blurred and the RDF becomes an oscillating and converging to unity function. From the measurement results was determined the threshold value $\delta = 6\%$ of particle size dispersion (polydispersity) above which there is a sharp degradation of the zero order PBG. Below this threshold the polydispersity has practically no effect on the optical properties of the photonic crystal. It was found that the most sensitive to the disordering are the magnitude and width of the reflection peak of PBG.

Influence of spatial disorder on transmittance and reflectance spectra of 3D PhC consisting of monodisperse spheres was investigated in Nair and Jagatap (2012). It was shown that with the disorder increasing the PBG disappears, and in a completely random system diffuse scattering reaches 100%.

2.2.4 Ordered Monolayers

Ohtaka in 1980 found one of the first solutions of the problem of EM wave diffraction (scattering) by the two-dimensional ensemble (monolayer) of regularly arranged homogeneous wavelength-sized particles (Ohtaka 1980). His approach is based on the extension of the low-energy electron diffraction (LEED) theory (Kambe 1967, 1968; Pendry 1974) to the case of photons. The expressions that allow one to find the transmission and reflection coefficients of the monolayer are proposed. Under this approach have been calculated reflection spectra of regular arrays of monodisperse isotropic dielectric spheres (Inoue et al. 1982). The authors identified three typical features in the obtained spectra: a broad background, anomalies in the thresholds of channels (i.e., on the edges of spectral ranges with the different number of diffraction orders) and a group of distinct peaks. The first is due to scattering by a single particle, and the other two are caused by the regularity in particle locations.

The experimental verification of the method (Ohtaka 1980) was done in Miyazaki et al. (2000), Ohtaka et al. (2000), Kondo et al. (2002), Yano et al. (2002). In addition, in Ohtaka et al. (2000) a number of theoretical methods to study photonic bands was proposed. Comparison of the calculation and experimental data on transmittance showed generally good agreement. The discrepancy between the results was associated by the authors with the imperfectness of the actual lattice, the finite size of the sample, influence of the substrate (Miyazaki et al. 2000; Ohtaka et al. 2000; Yano et al. 2002), and absorption of particles (Kondo et al. 2002), which are not taken into account in simulation. Later the method of Ohtaka (1980) was developed and used in a number of investigations of periodic monolayers of spheres (Miyazaki and Ohtaka 1998; Fujimura et al. 2000; Kurokawa et al. 2002, 2004a, b; Kondo et al. 2004; Ohtaka and Yamaguti 2004; Ohtaka et al. 2004). Furthermore, it was generalized to study regular arrays of parallel cylinders (Ohtaka et al. 1998). Let us consider in more detail some of these works.

So, in Miyazaki and Ohtaka (1998) were investigated the characteristics of the near field in the photonic bands of monolayer of periodically arranged spherical particles. It was found that the resonance peaks in the reflection spectrum coincide with the peaks of the near field, which are determined predominantly by the evanescent waves from "closed channels", i.e. from adjacent spectral regions that include higher diffraction orders. The analytical solutions for the field were obtained in a low-frequency regime (particle size parameter $x << 1$). The approach was used to describe the experimental results of Fujimura et al. (2000) where were investigated the photonic bands of monolayer of polystyrene particles by the scanning near-field optical microscopy (SNOM). In these bands the sharp changes of SNOM images were observed. A good agreement between the data of calculation and experiment was obtained for distribution of near-field intensity spectra.

In Kurokawa et al. (2002, 2004a) the method for a single monolayer (Ohtaka 1980) was extended for the case of "a monolayer on a dielectric substrate" system. It was shown that the semi-infinite substrate significantly broadens the dips in the transmission spectra, while the finite (comparable with the diameter of the spheres

and the wavelength) one leads to appearance of additional dips (Kurokawa et al. 2002). At the normal illumination the dips are shifted to lower frequencies due to the interaction with the finite-thickness substrate. The near-field intensity nearby the monolayer - substrate boundary is significantly larger than the intensity nearby the monolayer without substrate (Kurokawa et al. 2004a).

In 1987 Modinos developed a method to calculate light scattering characteristics of the monolayer of periodically arranged monodisperse metal spheres (Modinos 1987). This method, as well as the one proposed by Ohtaka (1980), is the generalization of the low-energy electron diffraction theory. However, as the author notes "*the development of the formalism is here more transparent and concise*". This paper also provides a technique to take into account disorder and polydispersity. Further development of the method was made in Stefanou and Modinos (1991a, b, 1993). Thus, in Stefanou and Modinos (1991a) symmetry of matrix elements describing multiple scattering by monolayer and the substrate effect were taken into account. The results of the calculations and comparison with experiment for a monolayer of gold nanospheres were demonstrated. In Stefanou and Modinos (1991b) was calculated the absorption of light by monolayers of metal spheres at different lighting angles, studied the effect of the substrate, the particle concentration, and the disorder in monolayers on the absorption. In Stefanou and Modinos (1993) was proposed a method of accounting the imperfection of monolayers, which is based on the coherent potencial approximation. It was applied for calculating the absorption coefficient of a monolayer of silver nanospheres. Imperfection was simulated by random filling the part of perfect lattice sites by identical particles, while the other part was remained unoccupied. The main effect of the considered type of imperfection consists in the spectral shift of the resonance peaks and their broadening.

In the framework of the TMSW the ordered monolayers was studied in the works (Kachan and Ponyavina 2002; Zamkovets et al. 2003a). In Kachan and Ponyavina (2002) was investigated influence of the type of spatial organization of monodisperse silver nanospheres in a monolayer on its direct transmittance T_c at normal illumination. Calculations were made in the framework of QCA (Hong 1980). To simulate the positional disorder the RDF was used which describes, by a set of step functions, the uniform deviation of particles from the sites of the perfect lattice. Such RDF is the sequence of rectangles situating at the distances (from coordinate origin) corresponding to the coordination circle radii (Ziman 1979). In calculations were taken into account the particles that are located within a distance of five lattice constants a (specified correlation length $l_c = 5a$). It was noted that the results for T_c, obtained with such RDFs "*are hardly dependent on the correlation length at* $l_c \geq 5a$". The authors attribute this to the strong absorption and weak lateral scattering for the considered silver particles (diameter $D = 2$ nm) near the frequency range of studied surface plasmon resonance. It was found that the T_c spectra of monolayers with triangular and square lattice with the same filling factor ($\eta = 0.6$) practically coincide in the considered wavelength range from 0.3 to 0.7 μm. A comparison with the calculation results for partially ordered monolayer and monolayer of "independent" scatterers was made. It was shown the long-wave shift of the transmittance minimum caused by surface plasmon resonance (SPR) at "transition" from monolayer

of independent scatterers to "quasi-regularly" packed and partially ordered monolayers. Using the approach similar to developed in Kachan and Ponyavina (2002) the authors described experimental data on transmittance of monolayers of silver nanoparticles in KCl ($D = 35$ nm, $\eta = 0.2$) (Zamkovets et al. 2003a). The partially ordered and quasiregular monolayers were considered. To describe the deviations of particle centers from the lattice sites the Gaussian distribution was used. It was shown that the T_c-values in the dip region caused by SPR are larger for the quasiregular monolayer as compared to the partially ordered one. With the increase in the correlation length (from $l_c = 2.74D$ to $l_c = 17.3D$) the dip minimum is slightly shifted to shorter waves and its depth is reduced. Based on this result it was concluded that the main features in the transmittance spectra are determined by the "*nearest ordering*" (i.e. nearest neighboring particles).

Among other publications devoted to theoretical and numerical investigations of ordered monolayers of spherical (and nearly spherical) particles, we emphasize the works (Skorobogatiy et al. 2005; Sun et al. 2007; Gadomskiĭ and Shalin 2007; Shalin and Moiseev 2009; Shalin 2009, 2010).

In Skorobogatiy et al. (2005) were analyzed the statistical properties of 2D PhC imperfection. A statistical model to describe PhC was developed. Authors found that at least three sets of parameters are necessary to create a minimal statistical model of 2D disorder in PhC lattices. The first set includes the size and ellipticity of particles, the second set describes the roughness of the particle surface, the third one describes the deviation of particle centers from the perfect lattice sites, which was simulated by a Gaussian distribution. The model was used to analyse photonic crystal images obtained by the scanning electron microscopy. The results showed that for the considered PhCs (with different material combinations from which these PhCs were made) there is a relatively narrow distribution of statistical parameters characterizing disorder.

Angular structure of light scattered by normally illuminated monolayer of regularly arranged dielectric spheres on a silicon substrate was calculated and analyzed in Sun et al. (2007) under the FDTD method. It was found that the intensity maxima of light transmitted through the monolayer into the substrate are focused along the lines passing through the centers of the particles. This result can be used to solve the problems of forming the nanoholes in a substrate using micrometer-sized particles.

The interaction of EM waves with a periodic monolayer of monodisperse nanospheres on the substrate and with a layered system from such monolayers was considered in the works (Gadomskiĭ and Shalin 2007; Shalin and Moiseev 2009; Shalin 2009, 2010) using the *dipole approximation* (Gadomskiĭ and Shalin 2007). The conditions were determined under which the reflectance of a structured interface (monolayer of metallic nanoparticles periodically arranged on the semi-infinite substrate) tends to zero in a wide wavelength range. Numerical results were obtained for the gold nanoparticle monolayer on glass. In Shalin and Moiseev (2009) were studied the systems with various refractive indices of particles and substrates. The comparison with the numerical calculation (by the finite-element method) of Maxwell's equations showed good agreement of the results for the systems of particles with $D = 10$ nm. For the systems with $D = 40$ nm the significant discrepancy was occurred which

is caused by neglecting the higher-order multipole fields. In Shalin (2009) the conditions were found at which for some wavelengths the transmission coefficient of sparse multilayer consisting of ordered monolayers of nanoparticles is unity. In Shalin (2010) the conditions of broadband antireflection were found for interface of glass with monolayer of regularly arranged spherical nanovoids in the subsurface region.

Consider some works on experimental investigations of ordered monolayers of spherical particles (Yamasaki and Tsutsui 1999; Rudhardt et al. 2003; Dushkina et al. 2005; Andueza and Sevilla 2007; Andueza et al. 2008, 2010, 2011, 2012).

The close-packed monolayer of SiO_2 spheres with diameter of 550 nm was examined in Yamasaki and Tsutsui (1999). The high ordering degree of sample allowed one to observe pronounced diffraction pattern at laser illumination and resonance dips in spectral transmittance. In Rudhardt et al. (2003) was proposed a fabrication method for producing interference-based electrooptic phase gratings that can be switched between diffracting and transparent states. The gratings consist of a monolayer with triangular-close-packed array of monodisperse droplets of nematic liquid crystal (LC) embedded in polymer matrix. The advantages of such gratings are small switching times and controlling electric fields ($0.1\,\mathrm{V}/\mu\mathrm{m}$), since the changes from the constructive interference to destructive one and vice versa requires a small change in refractive index of the LC droplets.

The diffraction effciency (the ratio of diffracted to incident wave intensities) of a monolayer with triangular lattice of micrometer-sized latex shperes on a glass substrate was measured in Dushkina et al. (2005) in the planar geometry (PG) and the geometry of total internal reflection (TIR). For the PG, the range of illumination angles (50°–55° relative to normal to manolayer plane) was found in which the reflection coefficient of the monolayer for zero diffraction order tends to zero. The angle at which the minimum reflectance in this range occurs is an analog of the Brewster angle. By measurement of this angle one can find the refractive index of a particulate monolayer. For TIR geometry were observed strong changes of diffraction efficiency for all orders near the critical TIR-angle (41°20′).

Andueza and Sevilla investigated, in the centimeter wavelength range, influence of packing density on the transmittance of a normally illuminated monolayer of monodisperse glass spheres with refractive index $n = 2.65$ and diameter from 4 to 8 mm, arranged in the triangular lattice (Andueza and Sevilla 2007). Analysis of the measurement and simulation results has shown that in spite of the great n-value, for which in spectra of the monolayer the resonance peaks are mainly determined by the characteristics of an individual sphere (Kondo et al. 2002; Miyazaki et al. 2000), the influence of the structure can not be neglected even at low particle concentration. In Andueza et al. (2010) a map of resonant modes of the monolayer was obtained versus particle concentration and comparison with Mie theory for single sphere and Bragg diffraction for monolayer in the $x << 1$ limit was made. Although one can expect the "convergence" of the resonances in the monolayer spectrum to the Mie resonances with concentration decreasing, the experiment showed that the position of practically all resosnances tends to the values determined by Bragg diffraction. The authors supposed that such behaviour can be caused by large n. They also found that at high packing density the positions of some resonances remain unchanged and

match the values of the Mie resonances, while the position of others ones changes smoothly with the distance between the particles (Andueza et al. 2008). Effect of the lattice type (Andueza et al. 2010), the disorder (Andueza et al. 2011), the angle of illumination (Andueza et al. 2012), and the particle permittivity (Morales et al. 2013) on the transmission spectra of the monolayer were further investigated by authors. It was found that for large and small packing densities the spectra of monolayers with triangular and square lattices are clearly different, and at moderate ones the spectra are almost the same (Andueza et al. 2010). The transmission spectrum of monolayer of spheres with $n = 2.65$ is generally weakly sensitive to structural disorder. With increasing the disorder degree the depth of resonance dips in the transmission spectrum is reduced and they become more blurred (Andueza et al. 2011). With a decrease in the packing density the Mie resonances become dominant in the photonic band structure of the monolayer. It was found that the photonic bands do not depend on the illumination angle (Andueza et al. 2012). In the framework of the FDTD method was found that the resonances in the monolayer spectra arise mainly due to the scattering modes of a separate sphere at high particle concentrations and refractive indices and due to the Bragg diffraction modes at small ones (Morales et al. 2013).

One of the key problems in theoretical consideration of optical properties of actual ordered systems is taking into accoung their imperfection. Note that, besides the described above works devoted to solution of this problem for monolayers of spheres, the arrays of cylinders are actively studied as well (Ryu et al. 1999; Kaliteevski et al. 2003; Chen et al. 2006; Meisels and Kuchar 2007; Prasad et al. 2007). In particular, in Prasad et al. (2007) was examided experimentally and numerically under the finite-element method the tansmittance of 2D PhC consisting of holes in silicon slab at the normal to the slab plane illumination. The sharp resonant peaks and dips were observed in the transmission spectra. The influence of disorder on the spectra of these resonances was investigated. The 2D RDFs of the actual samples were found for different PhC disorders. The peaks of RDF are broadened with disorder. Moreover, they are asymmetric. It was found that resonances are weakly sensitive to variations of hole sizes and roughness of boundaries, but very sensitive to deviations from spatial periodicity of the hole centers. The results obtained show the importance of accounting the spatial imperfection of crystals at theoretical description of their optical properties.

In conclusion of this section we note that many scientific groups pay great attention to study the interaction of light with the ordered particulate structures. Particularly intensively in recent decades are studied photonic crystals. This is due to the potentials of their use in photonics, optics and optoelectronics. Accordingly, there is an actual task to create the theoretical models to describe the optical properties of such systems. Most full description yield the models taking into account imperfection of actual structures, effects of diffraction, interference and multiple scattering of waves.

2.3 Theory of Multiple Scattering of Waves and Its Approximations: Application to Particulate Monolayers

2.3.1 Quasicrystalline Approximation for Determination of Direct Transmittance and Specular Reflectance

2.3.1.1 Basic Relations

Quasicrystalline approximation (QCA) of the statistical theory of multiple scattering of waves was suggested by Lax (1952) to obtain the effective field in the systems where finite size of scatterers is significant, i.e. in concentrated particulate media. Hong (1980) (as it was indicated in Sect. 2.2.2) obtained the solution of the problem of plane wave scattering by normally illuminated partially ordered monolayer of polydisperse spherical scatterers. He used the QCA to derive the equations for the direct transmission T_c and specular reflection R_c coefficients of the monolayer. In a number of problems it is reasonable to refer them as coherent components of the transmitted and reflected light. We write them as follows (Hong 1980; Loiko and Miskevich 2005a, b; Miskevich and Loiko 2011a, b):

$$T_c = |t_c|^2 = \left| 1 - \frac{\eta}{x^2} \sum_{j=1}^{N_s} (2j+1) \left(z_j + y_j \right) \right|^2, \tag{2.3.1}$$

$$R_c = |r_c|^2 = \left| -\frac{\eta}{x^2} \sum_{j=1}^{N_s} (-1)^j (2j+1) \left(z_j - y_j \right) \right|^2. \tag{2.3.2}$$

Here t_c and r_c are the amplitude direct transmission and specular reflection coefficients, respectively, η is the monolayer filling factor (the ratio of the projection areas of all particles to the area, where they are located), $x = \pi D/\lambda$ is the size parameter of particle with diameter of D, λ is the wavelength of the incident light, $N_s = x + 4.05x^{1/3} + 2$ is the number of used expansion coefficients z_j and y_j (Babenko et al. 2003). The z_j and y_j coefficients are found from the solution of the system of equations:

$$\begin{cases} z_l = b_l + \rho_0 b_l \sum\limits_{j=1}^{N_s} (A_{lj} z_j + B_{lj} y_j) \\ y_l = a_l + \rho_0 a_l \sum\limits_{j=1}^{N_s} (A_{lj} y_j + B_{lj} z_j) \end{cases}, \tag{2.3.3}$$

where a_l and b_l are the Mie coefficients, ρ_0 is the averaged over the monolayer plane numerical particle concentration in monolayer. The A_{lj} and B_{lj} coefficients are found from the solution of equations:

$$A_{lj} = \frac{2j+1}{2}[l(l+1)j(j+1)]^{-\frac{1}{2}}$$
$$\sum_{p=0,2,\ldots}^{N_s} i^{-p}(2p+1)\,[l(l+1)+j(j+1)-p(p+1)]\,\mathrm{P}_p(0)\times,$$
$$\times \begin{pmatrix} l & j & p \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l & j & p \\ 1 & -1 & 0 \end{pmatrix} \mathrm{H}_p, \quad (2.3.4)$$

$$B_{lj} = \frac{2j+1}{2}[l(l+1)j(j+1)]^{-\frac{1}{2}}$$
$$\sum_{p=0,2,\ldots}^{N_s} i^{-p}(2p+1)[(p+l-j)(p-l+j)(l+j+1+p)(l+j+1-p)]^{\frac{1}{2}}\mathrm{P}_p(0)\times$$
$$\times \begin{pmatrix} l & j & p-1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l & j & p \\ 1 & -1 & 0 \end{pmatrix} \mathrm{H}_p, \quad (2.3.5)$$

$$\mathrm{H}_p = 2\pi \int_D^{\infty} \mathrm{g}(R)\mathrm{h}_p^{(1)}(kR)R\mathrm{d}R. \quad (2.3.6)$$

Here $\mathrm{g}(R)$ is the radial distribution function (Ziman 1979), $\mathrm{h}_p^{(1)}(x)$ is the spherical Hankel function of the first kind and p-th order, $k = 2\pi/\lambda$ is the wavenumber, $\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix}$ are the Wigner $3j$-symbols which are calculated using the relations published in Varshalovich et al. (1975).

To calculate integral (2.3.6) it is convenient to write it as follows: $\mathrm{H}_p = 2\pi(\mathrm{H1}_p + \mathrm{H2}_p)$, where

$$\mathrm{H1}_p = \int_D^{\infty} \mathrm{h}_p^{(1)}(kR)R\mathrm{d}R, \quad (2.3.7)$$

$$\mathrm{H2}_p = \int_D^{\infty} (\mathrm{g}(R)-1)\,\mathrm{h}_p^{(1)}(kR)R\mathrm{d}R. \quad (2.3.8)$$

Using the recurrence relations for the spherical Bessel functions (Arfken et al. 2012), we write integral (2.3.7) as follows:

$$\mathrm{H1}_p = k^{-2}\left\{-kD\mathrm{h}_{p+1}^{(1)}(kD) + \sum_{q=0,2,\ldots}^{p}[2(p-q)+1]\frac{p!!(p-q-1)!!}{(p-1)!!(p-q)!!}\mathrm{h}_{p-q}^{(1)}(kD)\right\}. \quad (2.3.9)$$

Integral (2.3.8) is computed numerically. Let us introduce dimensionless integration variable $u = R/D$ designating the distance in the monolayer plane expressed in particle diameters. Then

$$\mathrm{H2}_p = D^2 \int_1^{\infty} (\mathrm{g}(u) - 1)\, \mathrm{h}_p^{(1)}(2xu) u \mathrm{d}u. \tag{2.3.10}$$

Integral (2.3.10) can be calculated numerically if integrand converges to zero at a finite u-value. This is provided by convergence of $\mathrm{g}(u)$ to unity. The distance where RDF becomes unity determines the correlation length l_c (Ziman 1979; Loiko et al. 1986). Thus, in Eq. (2.3.15) upper limit of integration is replaced by l_c:

$$\mathrm{H2}_p = D^2 \int_1^{l_c} (\mathrm{g}(u) - 1)\, \mathrm{h}_p^{(1)}(2xu) u \mathrm{d}u. \tag{2.3.11}$$

In the QCA the T_c and R_c coefficients are determined by the sum of waves scattered by the particles with taking into account their re-illumination and correlation in their spatial locations. The correlation is described by the radial distribution function (RDF), which characterizes the probability of any particle location in the space relative to other one.

2.3.1.2 Radial Distribution Function

There are various numerical and approximate analytical methods to calculate the RDF for partially ordered particulate monolayer (Ivanov et al. 1988; Percus and Yevick 1958; Ornstein and Zernike 1914; Fisher 1964; Lado 1968; Skryshevskii 1980; Lock and Chiu 1994). One of the most used in the QCA is the iteration method (Ivanov et al. 1988) of numerical solution of the Ornstein–Zernike integral equation (Ornstein and Zernike 1914) for hard spheres in the Percus–Yevick approximation (Percus and Yevick 1958).

The method to simulate the RDF of a near-to-regularly packed monolayer (which can be referred as a *planar crystal* (PC) with imperfect lattice) of spherical particles was proposed and developed in Miskevich and Loiko (2011a, b). Schematical view of planar crystals with perfect triangular, square, and hexagonal lattices are shown in Fig. 2.1.

To calculate the RDF of the PC with imperfect lattice, the center of any particle is selected as the coordinate origin and radii of the coordination circles (Ziman 1979; Miskevich and Loiko 2011a, b) of the perfect crystal lattice and the number of particle centers for each circle are computed. The distance dependence of the number of particle centers in a perfect lattice is a set of the infinitely narrow peaks at distances equal to the coordination circle radii (see Fig. 2.1). Accordingly, the RDF of such a lattice has the nonzero values at these distances and zero otherwise. Actual crystals typically have an imperfect lattice with coordination circles "blurred" into the "rings" with the fuzzy edges. Consequently, the peaks of the RDF are blurred as well. It was shown (Miskevich and Loiko 2011a, b) that the expression for the

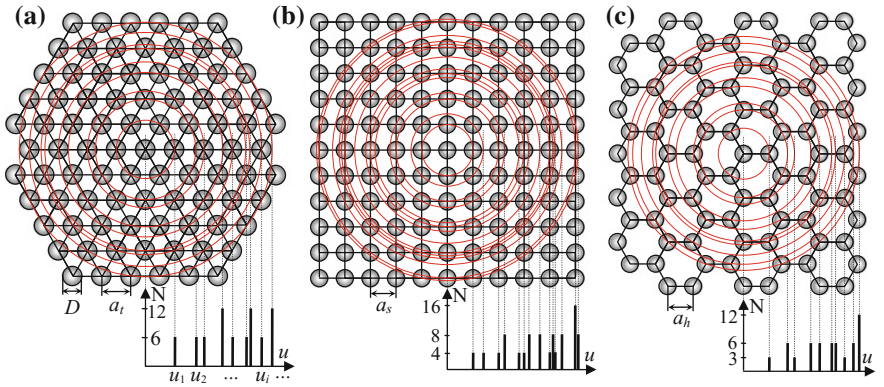

Fig. 2.1 Schematical view of perfect triangular (**a**), square (**b**), and hexagonal (**c**) lattices with constants a_t, a_s, and a_h and function N of the number of particle centers on the coordination circles with radii u_i versus dimensionless distance $u = R/D$, where R is dimension distance, D is particle diameter

RDF $g(u)$ of the near-to-regularly packed monolayer with an imperfect lattice can be written as:

$$g(u) = \rho_0^{-1} \sum_i \frac{N_i}{2\pi u_i} \frac{1}{\sqrt{2\pi}\,\sigma(u)} \exp\left(-\frac{(u-u_i)^2}{2\sigma^2(u)}\right). \tag{2.3.12}$$

Here $\sigma(u)$ is a *blurring function* that characterizes the broadening of the peaks with distance u. It is reasonable to use the linear blurring function:

$$\sigma(u) = \sigma_0 (au + b). \tag{2.3.13}$$

In Eqs. (2.3.12) and (2.3.13), $u = R/D$ is the dimensionless distance expressed in particle diameters D ($u \geq 1$), R is the distance in a monolayer plane relative to the coordinate origin; ρ_0 is the averaged numerical particle concentration in the monolayer; N_i is the number of particle centers on the coordination circle with radius R_i of a perfect crystal. The ordering degree and the scale of spatial order of the simulated crystals can be specified by changing of σ_0, a and b coefficients, respectively. Function (2.3.12) satisfies the normalization condition:

$$\frac{1}{\pi u_2^2} \int_{u_1=0}^{u_2 \to \infty} g(u) 2\pi u \, du = 1. \tag{2.3.14}$$

Equation (2.3.12) takes into account the asymmetry of the individual peaks of the RDF, which are observed in experiment (Rengarajan et al. 2005). It allows one to calculate the RDFs of the PCs in a wide range of their ordering degrees. The RDF

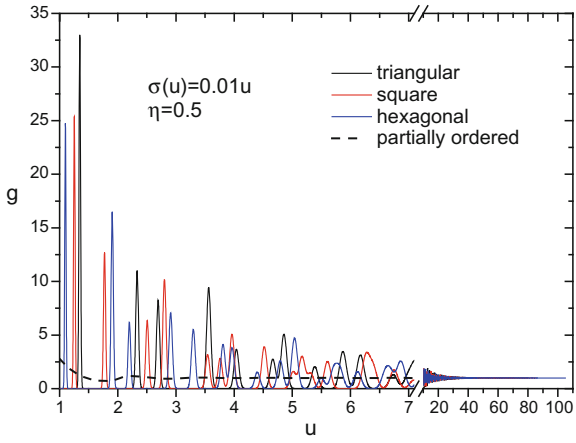

Fig. 2.2 Radial distribution functions for planar crystals with triangular, square and hexagonal imperfect lattices ($\sigma(u) = 0.01u$) and partially ordered monolayer, $\eta = 0.5$

obtained with Eqs. (2.3.12) and (2.3.13) is well adapted to utilize in the QCA. The calculation of such RDF and, consequently, transmission and reflection coefficients of the PC in the QCA, is fast and requires low amount of the computational resources.

Figure 2.2 shows the RDFs $g(u)$ describing the PCs with imperfect triangular, square and hexagonal lattices. They are calculated by Eqs. (2.3.12) and (2.3.13). The type of ordering of such PCs allows one to say that they possess the *imperfect long-range order*. Function $g(u)$ for the partially ordered monolayer, (monolayer with a *short-range order*) is shown as well. Thus, the data illustrate difference in radial functions of monolayers with short- and imperfect long-range order.

The RDF of a highly ordered PC is a sequence of narrow peaks at small $u = R/D$ values (i.e. in the region near to the coordinate origin). With the u-value increasing the peaks become wider, the function oscillates and converges to unity. Note that at $\sigma(u) = \text{const}$, Eq. (7) transforms into the known expression, which describes the Gaussian blurring of peaks (Skryshevskii 1980).

To obtain the RDF of the partially ordered monolayer of particles, the solution of the Ornstein–Zernike integral equation (Ornstein and Zernike 1914) was used. It was numerically calculated in the Percus–Yevick approximation (Percus and Yevick 1958) for a system of hard spheres by the iteration method (Ivanov et al. 1988). Note that this function is deduced from the Poisson statistics with taking into account the finite size of particles (Ivanov et al. 1988).

2.3.1.3 Direct Transmission and Specular Reflection Coefficients

Let us consider some results for direct transmission T_c and specular reflection R_c coefficients of particulate monolayers. They are calculated by the equations written in Sect. 2.3.1.1 using the RDFs shown in Fig. 2.2. The dependences of T_c and R_c coefficients and $T_c + R_c$ sum of planar crystals with triangular, square and hexagonal

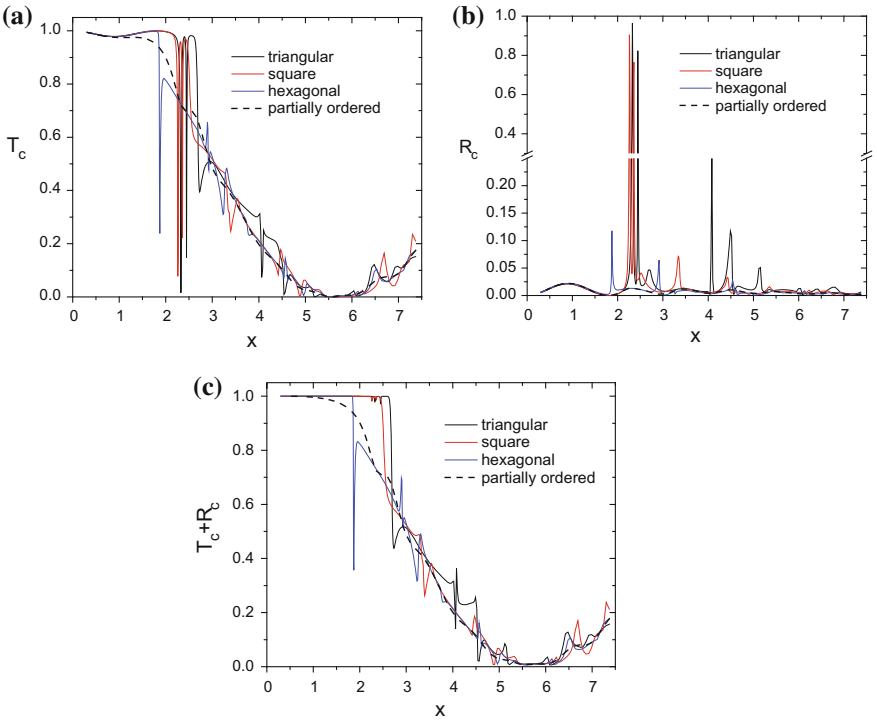

Fig. 2.3 Dependences of T_c (**a**), R_c (**b**), and the $T_c + R_c$ sum (**c**) of planar crystals with triangular, square, and hexagonal lattices ($\sigma(u) = 0.01u$) and partially ordered monolayer on the particle size parameter x. Filling factor $\eta = 0.5$ for all monolayers, particle refractive index $m = 1.4 + 5 \times 10^{-5} i$

lattices and partially ordered monolayer on the particle size parameter x are shown in Fig. 2.3.

One can see that T_c and R_c dependences of the partially ordered monolayer are smooth. The functions $T_c(x)$ and $R_c(x)$ of PC are more complicated. The sharp resonance peaks occur when particle diameters are comparable with the wavelength (see in Fig. 2.3 the range of $\sim 1.5 < x < \sim 2.5$). For the triangular and square lattices the peaks are more pronounced than for the hexagonal one and the fine structure is observed.

The $T_c + R_c$ sum of partially ordered monolayer equals to unity only for small particles ($x << 1$). With x growing (up to the value of $x \approx 5.5$) the sum monotonically decreases. For PCs the sum is unity in much wider region of size parameters than for the partially ordered monolayer. It increases with increasing the coordination number (the number of particle centers on the first coordination circle). Intervals of size parameters where $T_c + R_c = 1$ and $T_c + R_c < 1$ are separated by the resonances. If the sum $T_c + R_c = 1$ only zero diffraction order occurs. If $T_c + R_c < 1$ the

higher orders are implemented along with the zero one. This causes sharp decrease of $T_c + R_c$ sum.

Thus, long-range order provides much more interval of size parameters where only forwardly transmitted and specularly reflected light exists than short-range order. In this interval the absorption coefficient A_{ml} of monolayer consisting of absorption particles can be calculated by the simple equation: $A_{ml} = 1 - T_c - R_c$.

The results described are used in the Sect. 2.5 to calculate the absorption spectra of the layers of crystalline silicon (c-Si) particles with taking into account multiple scattering of waves.

2.3.1.4 Comparison of Theoretical and Experimental Data

The results obtained for the direct trnsmittance and specular reflectance of the PC were compared (Loiko and Miskevich 2013; Miskevich and Loiko 2014c, 2015b) with the available theoretical (Inoue et al. 1982; Miyazaki and Ohtaka 1998) and experimental (Yamasaki and Tsutsui 1999) data.

In Inoue et al. (1982), Miyazaki and Ohtaka (1998), the calculation results were given for reflection coefficients of 2D arrays of spherical particles located in the sites of perfect square and triangular lattices. They were obtained in the framework of approach, based on LEED theory and KKR method, extended for photons (Ohtaka 1980).

Figure 2.4 shows dependences of the specular reflection R_c coefficients of the normally illuminated PC with the square lattice on $Z = a/\lambda$, where a is the lattice constant and λ is the wavelength of the incident light. Solid curves show the data of Inoue et al. (1982), and dashed curves represent results of calculation in the QCA.

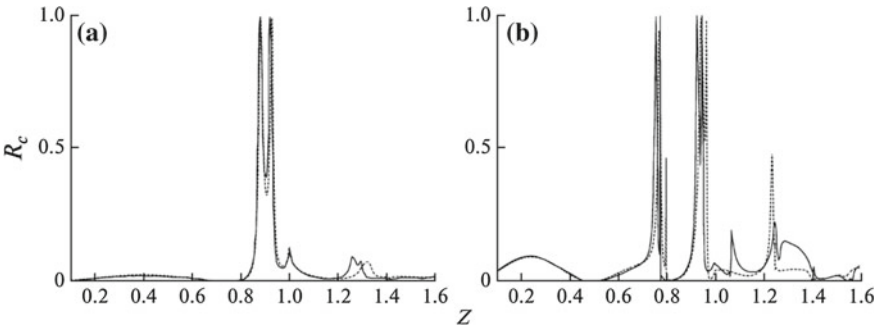

Fig. 2.4 Dependences of the specular reflection coefficients of a planar crystal with a square lattice formed from monodisperse dielectric spheres with refractive index $m = 1.6 + 0.0i$ on quantity $Z = a/\lambda$, where a is the lattice constant and λ is the wavelength of the incident wave. Normal illumination. Solid curves are data of Inoue et al. (1982), and dashed curves are results of calculation in the QCA (Loiko and Miskevich 2013). The RDF was calculated at $\sigma(u) = 0.001u$. Panel (**a**): $D/(2a) = 0.35$ ($\eta = 0.49\pi/4$); panel (**b**): (solid curve) $D/(2a) = 0.5$ ($\eta = \eta_s^{\max} = \pi/4 \approx 0.7854$) (Inoue et al. 1982), (dashed curve) $\eta = 0.78$

Figure 2.4a presents results for a monolayer with the ratio of the particle radius to the lattice constant equal to 0.35 (see Fig. 2a in Inoue et al. 1982). The filling factor of the monolayer with the perfect square lattice from monodisperse spherical particles is $\eta = \pi D^2/(4a^2)$. Correspondingly, in the QCA-calculation the value of filling factor $\eta = 0.49\pi/4$ was used (see dashed curve in Fig. 2.4a). Figure 2.4b presents dependence $R_c(Z)$ (solid curve) that was obtained in Inoue et al. (1982) for a monolayer with the ratio of the particle radius to the lattice constant equal to 0.5 (see Fig. 2d in Inoue et al. 1982). This ratio corresponds to maximal filling factor of a perfect square lattice $\eta = \eta_s^{\max} = \pi/4 \approx 0.7854$. The statistical approach that we use makes it possible to calculate the transmittance and reflectance of particulate monolayers with an imperfect lattice. The filling factor of these monolayers is $\eta < \eta^{\max}$ (at $\eta = \eta^{\max}$, a perfect lattice is realized). Therefore, to calculate the RDF of a monolayer with a lattice that is close to the perfect one a small σ_0-value, $\sigma_0 = 0.001$, was specified (Fig. 2.4a and b), and to model a lattice with the filling factor close to $\eta^{\max}$ the value $\eta = 0.78$ was chosen (Fig. 2.4b).

A little spectral shift of reflectance peak positions in Fig. 2.4 can be caused by a difference in the number of considered particles (Fig. 2.4a and b) and the value of the filling factor of the monolayer (Fig. 2.4b).

In Fig. 2.5 the data of Miyazaki and Ohtaka (1998) obtained under the approach (Ohtaka 1980), for total reflection coefficient $\mathfrak{R}$ (see Fig. 2 of Miyazaki and Ohtaka 1998) of PC with triangular lattice having maximal filling factor $\eta = \eta_t^{\max} = \pi/(2\sqrt{3}) \approx 0.9069$ is displayed by solid line. The QCA-calculated specular reflection coefficient R_c of PC with imperfect triangular lattice having filling factor $\eta = 0.9$ is shown by dashed line. The imperfection degree in this case is very low.

The data presented show that for the parameter ($\omega = \sqrt{3}a_t/(2\lambda)$, where a_t is triangular lattice constant, λ is the wavelength of incident light) lower than unity the QCA-calculation results coincide with the ones of Miyazaki and Ohtaka (1998) (except for magnitudes of resonant peaks) (Miskevich and Loiko 2014c, 2015b). The coincidence is caused by existence of only directly transmitted and specularly reflected light components in the $\omega < 1$ region. In other words, here is realized only zero diffraction order (Miyazaki and Ohtaka 1998). When $\omega > 1$, along with zero order, there are higher diffraction orders (Miyazaki and Ohtaka 1998), that results in difference of $\mathfrak{R}$ and R_c.

Thus, the comparison shows, as a whole, the good agreement of the results obtained in the framework of developed approach with the results of Inoue et al. (1982), Miyazaki and Ohtaka (1998). In the low-frequency relative to the resosnant peaks region of Z (Fig. 2.4) and ω (Fig. 2.5) the results are identical.

Let us compare the results with experimental data. In Fig. 2.6a are shown the experimental (Yamasaki and Tsutsui 1999) (Fig. 7a of Yamasaki and Tsutsui 1999) and calculated spectral dependences of direct transmission coefficient of the system consisting of monolayer of spherical SiO_2 particles on the glass plate. The calculated spectrum of direct transmission coefficient of single monolayer is displayed as well. In the experiment the monodisperse SiO_2 spheres with diameter $D = 0.55\,\mu$m formed highly ordered imperfect triangular lattice with near to maximal filling

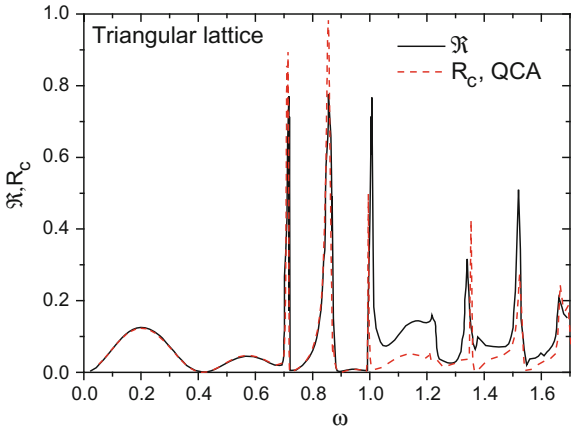

Fig. 2.5 Dependence of total reflection coefficient $\Re$ (Miyazaki and Ohtaka 1998) and specular reflection coefficient R_c of planar crystal with triangular lattice from monodisperse spherical particles with the refractive index $m = 1.6 + 0.0i$ on parameter $\omega = \sqrt{3}a_t/(2\lambda)$, where a_t is lattice constant, λ is the wavelength of incident light. Solid line: calculation results from the paper Miyazaki and Ohtaka (1998), $\eta = \eta_t^{\max} = \pi/(2\sqrt{3}) \approx 0.9069$; dashed line: calculation in the QCA, $\eta = 0.9$, $\sigma(u) = 0.001u$

factor. The substrate (glass) thickness was 1.3 mm. The transmitted-light intensity was measured at the normal illumination (Yamasaki and Tsutsui 1999).

The calculations were fulfilled in the QCA for single monolayer and in the QCA and transfer matrix method (TMM) for the system. First, in the QCA the amplitude direct transmission and specular reflection coefficients of single monolayer were found. Second, they were used in the TMM for finding the direct transmission coefficient of the "monolayer-on-substrate" system (the description of TMM is given in Sect. 2.4.2). The calculations were made at the parameters: $D = 0.55\,\mu$m, $\eta = 0.8$, substrate thickness 1.3 mm, wavelength step $\Delta\lambda = 1$ nm. The η-value was selected so that spectral position of resonant dip near the $\lambda = 0.65\,\mu$m coincides with experimental one. The sample was modeled as a system consisting of three *interfaces* (monolayer and two glass plate surfaces) and two *layers* (air between the monolayer and plate, and glass between plate surfaces). It was also supposed that whole system is surrounded by air. To simulate both spheres and plate were used the spectrum of SiO_2 refractive index (Palik 1985) shown in Fig. 2.6b. The air refractive index was set as unity.

The results of Fig. 2.6a show that, as a whole the calculated dependence describes well the measured one. The positions of transmittance maxima near $\lambda = 0.58\,\mu$m and $\lambda = 0.88\,\mu$m and minima near $\lambda = 0.65\,\mu$m and $\lambda = 1.3\,\mu$m agree well. The quantitative discrepancy can be caused by various factors among which the follows can be emphasized: (i) finite-sized receiver aperture leads to accounting not only strictly-forward component of transmitted light (which is described by theory), but the component of small-angle scattered light as well, that results in cosiderable

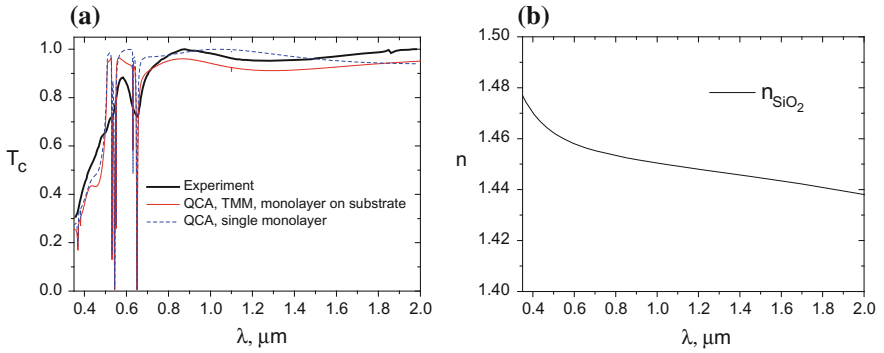

Fig. 2.6 Experimental (Yamasaki and Tsutsui 1999) and calculated in the QCA and TMM spectral dependences $T_c(\lambda)$ of system consisting of monolayer with triangular lattice from monodisperse SiO_2 spheres on the glass substrate, calculated in the QCA $T_c(\lambda)$ dependence of single monolayer (**a**) and the refractive index n_{SiO_2} spectrum of silica (SiO_2) using the data of Palik (1985) (**b**). $D = 0.55\,\mu m$, $\eta = 0.8$, $\sigma(u) = \sigma_0(2u - 1)$, $\sigma_0 = 0.001$, $l_c = 400$, substrate thickness $1.3\,\mu m$

exceeding by the measured T_c-values the calculated ones in the wavelenght ranges of $\lambda < 0.5\,\mu m$ and $\lambda > 0.7\,\mu m$; (ii) at the re-reflections between the interfaces the higher diffraction orders, which do not took into account in the computation, can contribute in the directly transmitted component (in the short-wave, relative to resonances, range, i.e. at $\lambda < 0.5\,\mu m$); (iii) the refractive index dependence of particles and substrate used in the calculation can differ from experimental one; (iv) blurring and broadening the experimentally obtained peaks and dips in the resonance region (see wavelength range 0.5–0.7 μm) may be due to lattice imperfection of actual monolayer as well as substrate influence (Kurokawa et al. 2002).

It should be noted that transmittance (and reflectance) spectra of single monolayer in air (dashed line in Fig. 2.6a) differ essentially from the ones of the "monolayer-on-substrate" system. This is caused by optical interaction between the monolayer and substrate which must be taken into account in the theoretical models. It can conditionally be divided into two components: near- and far-field interaction. In the approach described are used the methods allowing to take into account only far-field interaction: multiple reflection between the interfaces of the system (monolayer and substrate surfaces), that enables one to obtain acceptable agreement of theoretical and experimental results. Remind that this approach is based on the Hong's solution (Hong 1980) which takes into account *near* and *far* fields at multiple scattering of waves inside the monolayer and describes the average (coherent) field scattered in the strictly forward and strictly backward directions (at normal illumination) in *far-field zone* outside the monolayer. When the monolayer and substrate are in touching or on little (comarable with the wavelength) distances from each other, the near field can influence on the transmission and reflection spectra of the system (Fujimura et al. 2000; Kurokawa et al. 2002, 2004a).

The results described in this subsection can be used to solve the problem of monitoring the ordering degree of the two-dimensional lattices from analysis of the transmission and reflection coefficients.

2.3.2 *Structure Factor*

The structure factor characterizes the interference pattern of angular distribution of light, scattered by particulate medium, caused by spatial arrangement of particles.

The equation for structure factor S_2 of monolayer of monodisperse spherical particles can be written as follows (Ivanov et al. 1988):

$$S_2 = 1 + 8\eta \int_0^\infty \left[g(u) - 1\right] J_0(2xu\sin\theta)\, u\,du, \tag{2.3.15}$$

where η and $g(u)$ are the filling factor and the RDF of monolayer, $J_0(x)$ is the zero-order cylindrical Bessel function, $x = \pi D/\lambda$ is the size parameter of particle with diameter D, λ is the length of incident wave, θ is the polar scattering angle. The subscript "2" means that the 2D system is considered.

Figure 2.7 shows the dependences of structure factors of partially ordered monolayer and planar crystals on the "generalized" parameter $2x\sin\theta$ (see Eq. (2.3.15)). When S_2 of partially ordered monolayer was calculated the iteration method (Ivanov et al. 1988) of numerical solution of Ornstein–Zernike integral equation for hard spheres in the Percus–Yevick approximation was used to compute the RDF. Note that, for $\theta = 0$ the structure factor takes the form: $S_2 = (1-\eta)^3/(1+\eta)$ (Ivanov et al. 1988; Twersky 1975c). For RDF of planar crystals the method described in Sect. 2.3.1.2 was used.

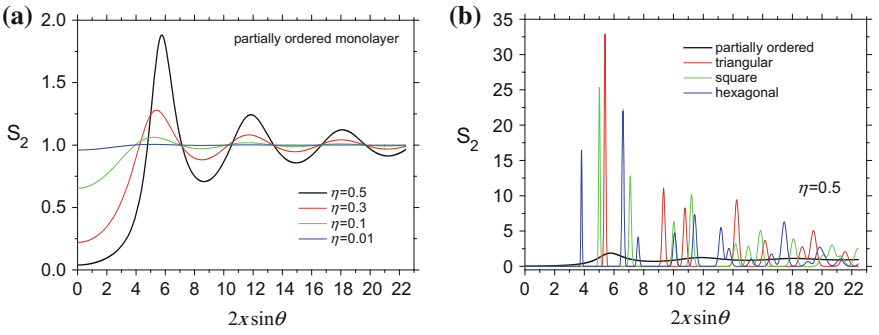

Fig. 2.7 Structure factor S_2 of partially ordered monolayer with different filling factors (**a**) and planar crystals with different lattices at $\eta = 0.5$ (**b**) versus parameter $2x\sin\theta$. The RDFs of PCs were calculated at parameters: $\sigma(u) = 0.01u$, $l_c = 100$

The maxima of S_2 correspond to constructive interference while the minima to destructive interference of scattered waves. The data obtained allows one to find parameters of the layer to realize clearly indicated maxima in the angular distribution of light. For planar crystals the number of peaks corresponds to the number of the diffraction orders.

2.3.3 Single Scattering and Interference Approximations

There are conditions in which multiple scattering of waves can be neglected. They occur, for example, in monolayers of large particles or sparse thin particulate layers. In these cases the single scattering approximation (SSA) and the interference approximation (IA) (Ivanov et al. 1988) can be used. The SSA and IA are simplest approximations in the hierarchical approach of the theory of multiple scattering of waves and they are significantly simpler than the QCA. Both SSA and IA neglect the multiple scattering of waves. Under the SSA the correlation in particle locations (i.e. spatial structure of monolayer) is neglected. Under the IA the spatial arrangement of particles is taken into account by the structure factor (see Sect. 2.3.2). The choice of SSA or IA is caused by the problem to be solved. For example, to determine the direct transittance and specular reflectance of normally illuminated monolayer the SSA can be used, while to determine the angular distribution of scattered light the IA is preferred.

Direct transmission coefficient T_c of a normally illuminated monolayer of particles can be written as follows (Bohren and Huffman 1983; van de Hulst 1957; Dick et al. 1997b, 1985, 1987a):

$$T_c = \left| 1 - \frac{2\eta}{x^2} \langle f(0) \rangle \right|^2, \tag{2.3.16}$$

where η is the monolayer filling factor, $x = \pi D/\lambda$ is the size parameter of particle with diameter D, λ is the wavelength of incident light, and $\langle f(0) \rangle$ is the configurationally averaged amplitude scattering function for zero polar scattering angle (forward-scattering amplitude) of the monolayer. Generally this function depends both on the particle parameters and the effective field acting on the particle (see Sect. 2.2), i.e., on the number of particles in the layer and their spatial arrangement.

When multiple scattering in a monolayer of monodisperse particles can be neglected (i.e. when single scattering occurs), $\langle f(0) \rangle$ is equal to the forward scattering amplitude of an isolated particle: $\langle f(0) \rangle = f(0)$, and for a spherical particle it can be calculated by Mie formulas (Bohren and Huffman 1983). Then Eq. (2.3.16) can be transformed into the form (Ivanov et al. 1988; Dick et al. 1987b):

$$T_c = 1 - Q_{ext}\eta + \frac{L}{2}(Q_{ext}\eta)^2. \tag{2.3.17}$$

Here

$$L = \frac{8\pi p(0)\Lambda}{Q_{ext}x^2}, \tag{2.3.18}$$

Q_{ext} is the extinction efficiency factor, Λ is the single-scattering albedo (ratio of the scattering to extinction efficiency factors) (Zege et al. 1991a; Mishchenko et al. 2006), $p(0)$ is the particle phase function for the zero polar scattering angle: $\theta = 0°$. It is normalized by the condition

$$2\pi \int_0^\pi p(\theta) \sin\theta \mathrm{d}\theta = 1. \tag{2.3.19}$$

Equation (2.3.17) can be easily generalized for a monolayer formed by polydisperse particles (Ivanov et al. 1988; Dick et al. 1997b):

$$T_c = 1 - \bar{Q}_{ext}\eta + \frac{\bar{L}}{2}(\bar{Q}_{ext}\eta)^2. \tag{2.3.20}$$

Here

$$\bar{Q}_{ext} = \frac{\int_0^\infty x^2 Q_{ext}(x)\mathrm{c}(x)\mathrm{d}x}{\int_0^\infty x^2 \mathrm{c}(x)\mathrm{d}x}, \tag{2.3.21}$$

$$\bar{L} = 8\frac{\left|\int_0^\infty f(0,x)\mathrm{c}(x)\mathrm{d}x\right|^2}{\left(\int_0^\infty x^2 Q_{ext}(x)\mathrm{c}(x)\mathrm{d}x\right)^2}, \tag{2.3.22}$$

$$\eta = \rho\frac{\int_0^\infty \pi r^2 \mathrm{c}(r)\mathrm{d}r}{\int_0^\infty \mathrm{c}(r)\mathrm{d}r}, \tag{2.3.23}$$

$\mathrm{c}(r)$ is the particle size distribution function, $Q_{ext}(x)$ and $f(0,x)$ are, respectively, the extinction efficiency factor and forward-scattering amplitude for the particle with size parameter x, $\rho = N/S$ is particle concentration (the ratio of the number N of particles to the monolayer area S where they are distributed).

The experimental data were obtained and their comparison with theoretical results for large latex and starch particles was made in Ivanov et al. (1988), Dick et al. (1997b). Data for latex particles are presented in Fig. 2.8. Solid line corresponds to calculation by Eq. (2.3.17). Dashed line corresponds to calculation by equation

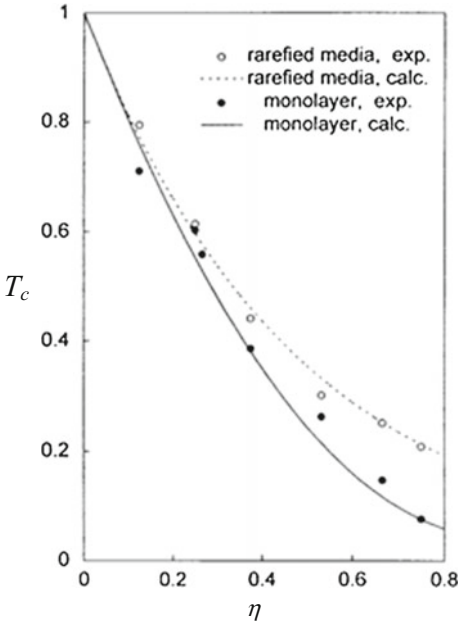

Fig. 2.8 Direct transmittance T_c of the monolayers of monodisperse latex particles versus filling factor η and the corresponding rarefied slab: experiment (dots) and calculations by the Eq. (2.3.17) (solid line), and Eq. (2.3.24) (dashed line) (Dick et al. 1997b). The relative refractive index of particles is $n = 1.144$, wavelength $\lambda = 0.65\,\mu\text{m}$, particle diameter $D = 3.75\,\mu\text{m}$

$$T_c = \exp(-Q_{ext}\eta), \tag{2.3.24}$$

which describes direct transmittance by the rarefied system. The experimentally observable dependences for monolayers are as a whole in good agreement with the results of calculation, because of the particles are large, the phase function is forward-elongated, and the role of multiple scattering is small.

We mainly consider in this chapter normal illumination of the layer. There are cases when information on transmittance at the oblique illumination plays an important or crucial role in studying the particulate system. The model to describe direct transmittance at oblique illumination in the polymer dispersed liquid crystal films is developed in Loiko and Dick (2003), Dick and Loiko (2001a, b). Transmittance at different polarization of the incident light, different characteristics of the layer and the liquid crystal droplets is close examined in these papers. The calculated results are compared with the known experimental data. The spectral behavior of transmittance of optically soft particles is studied in Dick and Loiko (2014).

Consider angular dependences of scattered light (Ivanov et al. 1988; Loiko et al. 1985; Dick et al. 1997a). At normal illumination the intensity (radiance) $I(\theta)$ of light scattered by monolayer at angle θ can be written in the IA as:

$$I(\theta) = F_0 \Lambda Q_{ext} \eta p(\theta) \mathrm{S}_2(\theta). \tag{2.3.25}$$

Here the structure factor S_2 is defined by Eq. (2.3.15) (Sect. 2.3.2), F_0 is the radiant intensity of the incident flux, Λ is single scattering albedo, Q_{ext} is particle extinction

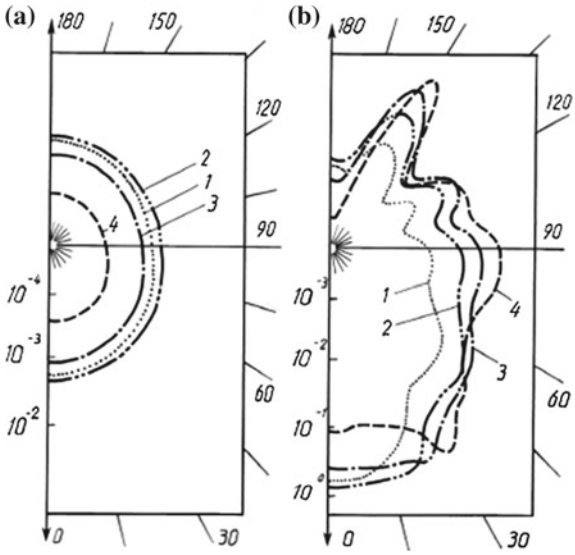

Fig. 2.9 Calculated angular distributions of intensities $I(\theta)$ of light scattered by monolayers of partially ordered nanabsorbing particles (Ivanov et al. 1988; Dick et al. 1987b). Filling factor $\eta = 0.1$ (curve 1), 0.3 (curve 2), 0.5 (curve 3), 0.7 (curve 4). Size parameter $x = 1$ (**a**) and 6 (**b**). Refractive index of particles $m = 1.4 + 0i$. The arrows indicate forward and backward directions of light propagation

efficiency factor, η is the monolayer filling factor, $p(\theta)$ is the phase function of individual particle.

The calculated by Eq. (2.3.25) angular distributions for the partially ordered monolayer are shown in Fig. 2.9 (Ivanov et al. 1988; Dick et al. 1987b) for two particle size parameters and a set of monolayer filling factors. For particles with size parameter $x = 1$ the shape of angular distribution of scattered light intensity practically does not depend on the filling factor. For $x = 6$ there are angles where intensity grows monotonically, as well as the angles where it grows nonmonotonically with η.

In Fig. 2.10 are shown the measured and calculated in the IA $I(\theta)$ dependences for monolayers with different filling factors (Dick et al. 1997a). One can see that for small scattering angles $\theta < 6°$ the intensity of scattered light first increases and then decreases with η increasing. For large θ, $I(\theta)$ dependence increases monotonically with η. The most rapid increase is observed for θ close to 9°. The different character of change in the $I(\eta)$ dependence at different θ leads to the formation of additional maximum on the angular distribution of light scattered by monolayer at large η.

In Fig. 2.11 are shown measured and calculated in the SSA angular distributions of intensity I of light scattered by the monolayer of monodisperse spherical particles of polyvinyltoluene latex in a wide range of scattering angles θ (Sarofim et al. 1968). The differences on the dependences are caused by the redistribution of energy owing to the multiple scattering and interference of waves effects.

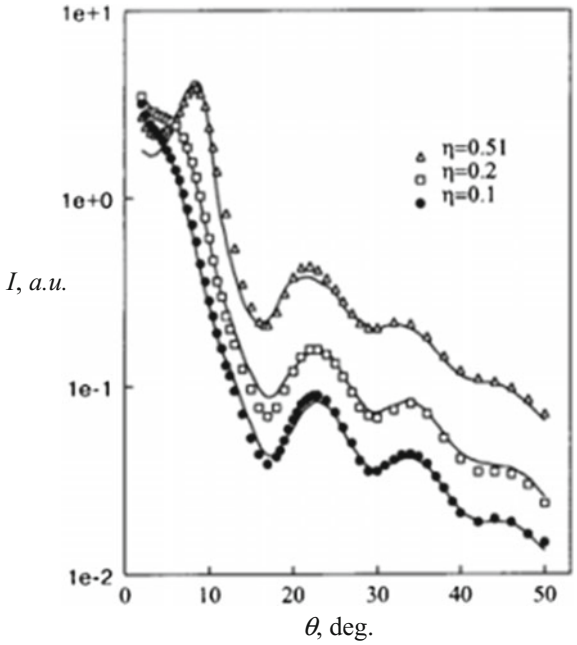

Fig. 2.10 Dependences of scattered light intensity I on polar scattering angle θ for monolayers of latex particles in water at various monolayer filling factors η (Dick et al. 1997a). Symbols correspond to experimental data. Curves represent the calculation results

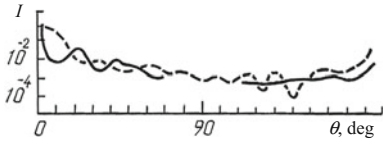

Fig. 2.11 Measured (solid line) and calculated in the SSA (dashed line) angular distributions of intensity I of light scattered by monolayer of latex particles with diameter 3.49 μm (Sarofim et al. 1968). Monolayer filling factor $\eta = 0.86$

2.3.3.1 Small-Angle Light Scattering and Direct Transmittance of Polymer Dispersed Liquid Crystal Film

Polymer dispersed liquid crystal (PDLC) films are widely used in various optic and optoelectronic devices (Zharkova and Sonin 1994; Simoni 1997; Blinov 2011; Drzaic 1988; Loiko 2013). Such a film consists of the thin polymer layer with embedded droplets of liquid crystal (LC). The optical properties of the PDLC films can be simply vary by applying the electric or magnetic fields, mechanical stretching, etc., i.e. due to externally-induced changing the properties of both individual LC droplets and their ensemble.

When the inhomogeneous anhcoring on the LC droplet – polymer surface is created, the asymmetry in small-angle scattering pattern is observed (Loiko et al. 2011, 2012, 2016a, b). The effect strongly depends on the droplet concentration.

To describe angular structure of light scattered by monolayer of LC droplets the interference approximation (Ivanov et al. 1988) was used. In this approximation for components of the intensity I of scattered light parallel (I_{vv}) and perpendicular (I_{vh}) to the polarization plane (defined by the wave vector and polarization vector of the incident wave) of the incident wave it is possible to write (Loiko et al. 2011, 2012):

$$I_{vv}(\theta_s, \varphi_s) = C\frac{\eta}{\sigma k^2}|f_{vv}(\theta_s, \varphi_s)|^2 \mathrm{S}(\theta_s), \tag{2.3.26}$$

$$I_{vh}(\theta_s, \varphi_s) = C\frac{\eta}{\sigma k^2}|f_{vh}(\theta_s, \varphi_s)|^2 \mathrm{S}(\theta_s), \tag{2.3.27}$$

Here C is the normalization constant, $f_{vv}(\theta_s, \varphi_s)$ and $f_{vh}(\theta_s, \varphi_s)$ are the vv- and vh-components of the vector amplitude scattering function in the direction of the wave vector $\mathbf{k}_s = (k\cos\theta_s, k\sin\theta_s\cos\theta_s, k\sin\theta_s\sin\theta_s)$ of the scattered wave, $\mathrm{S}(\theta_s)$ is the structure factor of the monolayer, θ_s and φ_s are the scattering angles. For a statistically isotropic ensemble of spherical or oblate shperoidal droplets, it does not depend on the φ_s angle.

Components of the amplitude scattering function in the expressions (2.3.26) and (2.3.27) are defined in terms of the elements of the amplitude scattering matrix $S_j(\mathrm{j} = 1, 2, 3, 4)$ (Loiko et al. 2012) as follows:

$$f_{vv}(\theta_s, \varphi_s) = S_2(\theta_s, \varphi_s)\cos^2(\alpha, \varphi_s) + S_1(\theta_s, \varphi_s)\sin^2(\alpha, \varphi_s) + \frac{1}{2}(S_3(\theta_s, \varphi_s) + S_4(\theta_s, \varphi_s))\sin\ 2(\alpha, \varphi_s), \tag{2.3.28}$$

$$f_{vh}(\theta_s, \varphi_s) = S_3(\theta_s, \varphi_s)\sin^2(\alpha, \varphi_s) + S_4(\theta_s, \varphi_s)\cos^2(\alpha, \varphi_s) + \frac{1}{2}(S_2(\theta_s, \varphi_s) + S_1(\theta_s, \varphi_s))\sin\ 2(\alpha, \varphi_s), \tag{2.3.29}$$

where α is the polarization angle (angle between the polarization vector of the incident wave and orientation direction of the droplet optical axes).

In Fig. 2.12 are shown the calculation results for the dependences of I_{vv} component of the intensity of light transmitted the normally illuminated monolayer of LC droplets with homogeneous surface anchoring on the scattering angle θ_s at orientation angle of the scattering plane $\varphi_s = 0$. They are obtained for the droplet radius $a = 4\,\mu\mathrm{m}$ and the refractive index of the polymer matrix $n_p = 1.532$. The normalization constant $C = 1$. Scattering plane coincides with the principal plane (it is defined by optical axis of droplet and the wave vector of the incident light) (Loiko et al. 2012). For a layer of droplets with such anchoring, the values of $I_{vv}(\theta_s)$ are identical at the same θ_s for different directions relative to the normal to the layer: $I_{vv}(\theta_s, \varphi_s = 0) = I_{vv}(-\theta_s, \varphi_s = 0)$.

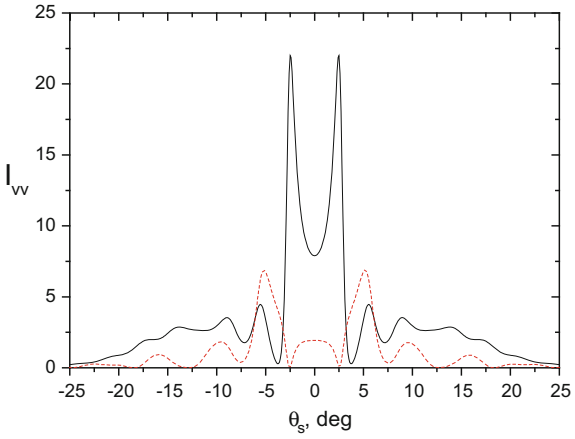

Fig. 2.12 Dependence of I_{vv}-component of the intensity of light scattered by a monolayer of monodisperse spherical oriented LC droplets with homogeneous tangential (solid line) and normal (dashed line) anchoring on the scattering angle θ_s with value of the azimuthal scattering angle $\varphi_s = 0$ (Loiko et al. 2012). No applied field. Droplet radius $a = 4\,\mu$m. $\lambda = 0.633\,\mu$m, ordinary n_o and extraordinary n_e refractive indices of LC: $n_o = 1.531$, $n_e = 1.717$, refractive index of polymer matrix $n_p = 1.532$, $\eta = 0.5$, polarization angle $\alpha = 0$

In Fig. 2.13 are shown patterns of the angular distributions of $I_{vv}(\theta_s, \varphi_s)$ and $I_{vh}(\theta_s, \varphi_s)$ components of the scattered light intensity in relative units in the angles $0 < \theta_s \leq 8°$, $0 < \varphi_s \leq 360°$ for monolayer of polydisperse droplets with uniform distribution of optical axes (directors) and homogeneous anchoring (Loiko et al. 2012).

The photograph of transmitted-light scattering pattern of PDLC film containing bipolar nematic droplets under normal illumination by the linearly polarized light is shown in Fig. 2.14.

Experimental results and theoretical model to describe angular distribution of light scattered by monolayer of nematic droplets are given in Loiko et al. (2016b). A method to simulate angular distribution of light used in this paper is based on the anomalous diffraction and interference approximations. It takes into account the director configuration within liquid crystal droplets, which is calculated using the relaxation method of the free energy minimization (Ondris-Crawford et al. 1991).

Experimental and calculated data for the I_{vv} and I_{vh} components of intensity of light scattered by the PDLC monolayer of spheroidal bipolar LC droplets versus the polar scattering angle are shown in Fig. 2.15. They are in reasonable coincidence.

For a layer of droplets with inhomogeneous anchoring, such as "tangential-normal", there is an asymmetry of the angular structure of scattered light ($I_{vv}(\theta_s) \neq I_{vv}(-\theta_s)$). This effect is most pronounced at equal parts of the tangential and

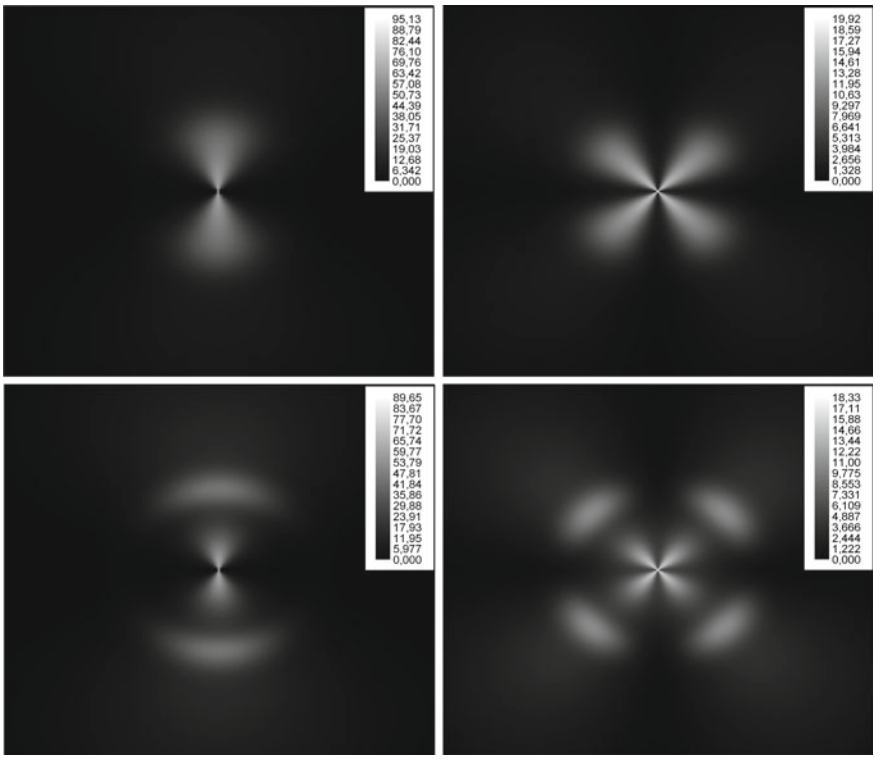

Fig. 2.13 Maps of the I_{vv} (left) and I_{vh} (right) components of intensity of light scattered in small angles by monolayer of polydisperse LC droplets with homogeneous anchoring of liquid crystal molecules at the droplet-polymer surface and uniform droplet's director configuration in a strong applied field (Loiko et al. 2012). Filling factor $\eta = 0.3$ (top), $\eta = 0.7$ (bottom). The average radius of droplets $\langle a \rangle = 2.5\,\mu\text{m}$. Variation coefficient of droplet radii is 0.2. $\lambda = 0.589\,\mu\text{m}$, ordinary n_o and extraordinary n_e refractive indices of liquid crystal are: $n_o = 1.5183$, $n_e = 1.7378$, refractive index of polymer $n_p = 1.524$, polarization angle (angle between the polarization vector of the incident wave and the average droplet director orientation) $\alpha = 0$. Droplet optical axes (directors) are distributed uniformly over the polar and azimuthal angles in the range twelve degrees. The average director of droplets is in the plane of monolayer

normal surface anchoring. Figure 2.16 shows influence of particle concentration on the $I_{vv}(\theta_s)$ dependence. Angular structure for concentrated layer possesses pronounced asymmetry determined by the shape of phase function of the individual droplet and short-range ordering of droplets in a monolayer.

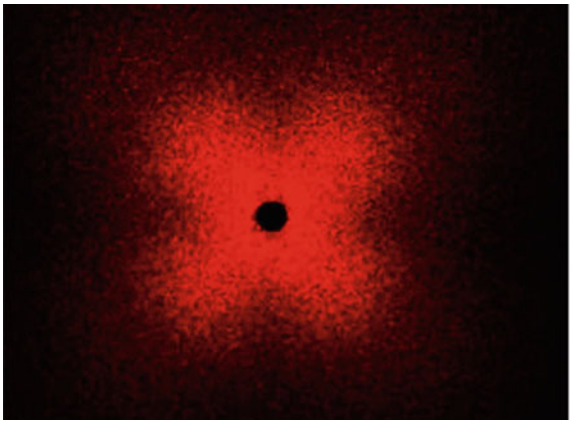

Fig. 2.14 Photograph of transmitted-light scattering pattern of polymer dispersed liquid crystal film containing bipolar nematic droplets with the average diameter 13.5 μm, and small dispersion of the droplet sizes (fraction of droplets in the range of 13.57±2.5 μm is 73%) for geometry of crossed polarizer and analyzer. Filling factor $\eta = 0.23$. Normal illumination by the linearly polarized light. $\lambda = 0.633\,\mu$m. The laser beam passing straight forward is shaded (Loiko et al. 2016b)

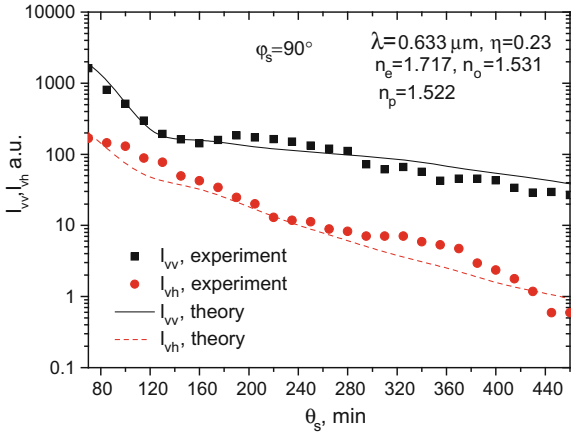

Fig. 2.15 Experimental and calculated data for the I_{vv} and I_{vh} intensities of light scattered by the PDLC monolayer of spheroidal bipolar LC droplets versus the polar scattering angle θ_s at azimuthal scattering angle $\varphi_s = 90°$. $\eta = 0.23$, ordinary n_o and extraordinary n_e refractive indices of liquid crystal are $n_e = 1.717$, $n_o = 1.531$, refractive index of polymer is 1.522. Illumination by the linearly polarized light. $\lambda = 0.633\,\mu$m (Loiko et al. 2016b)

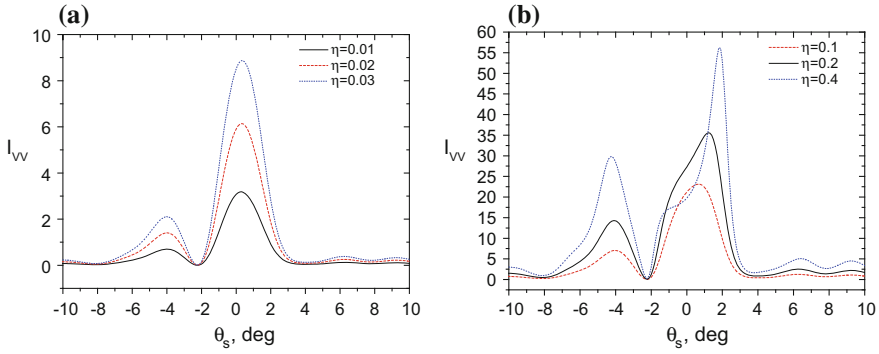

Fig. 2.16 Dependence of I_{vv}-component of the intensity of light scattered by a monolayer of spherical monodisperse oriented LC droplets with inhomogeneous anchoring at different values of filling factor $\eta = 0.01, 0.02, 0.03$ (**a**), $\eta = 0.1, 0.2, 0.4$ (**b**). No applied field. The parts with normal and tangential anchoring on the droplet-polymer surface are equal. $\lambda = 0.633\,\mu$m, $n_o = 1.531$, $n_e = 1.717$, $n_p = 1.532$, $\eta = 0.5$, $a = 5\,\mu$m, $\alpha = 0$, $\varphi_s = 0$ (Loiko et al. 2012)

2.3.4 *Direct Transmission Coefficient: Comparison of the Results Calculated in the Single Scattering and the Quasicrystalline Approximations*

Comparison of the direct transmission coefficients T_c of monolayers with a short-range ordering determined under the single scattering (SSA) and the quasicrystalline (QCA) approximations is fulfilled in Loiko et al. (1998). The regions of monolayer filling factor and particle size parameter are found where it is possible to use the simple SSA instead of the more complicated QCA. The range of filling factors ($\eta < 0.6$) is considered where the the Percus–Yevick model of the absolutely hard disks is applicable (Ziman 1979; Percus and Yevick 1958). The values of the particle size parameter x were limited by the region $x < 10.0$. For larger x, the differences in transmittance calculated in the SSA and the QCA decreases with increasing size parameter. Therefore the larger the value of x, the more justified is the use of the SSA. This is also evident from experimental results (Ivanov et al. 1988; Percus and Yevick 1958; Ornstein and Zernike 1914; Dick et al. 1997b, 1987a).

The data that enable one to see the differences in direct transmission coefficient calculated in the SSA (T_{SSA}) and the QCA (T_{QCA}) are given in Figs. 2.17 and 2.18. For relative particle refractive index $n = 1.2$ the differences are less than 1%. Generally they increase with n. For $n = 1.4$ they do not exceed 5% at $\eta < 0.4$.

In Figs. 2.17 and 2.18 the regions of small x-values are shaded. These are the regions of size parameter and filling factor where the value of the transmission coefficient calculated in the SSA is greater than unity. The SSA is not valid in these regions.

For monolayers of small ($x << 1$) scatterers it is sufficient to take into account the dipole interactions between the particles. Simple QCA- and SSA-based equations

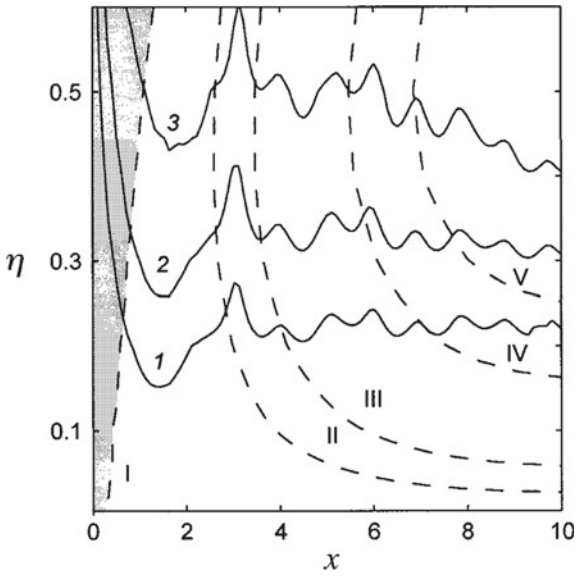

Fig. 2.17 Isolines of the $(T_{QCA}-T_{SSA})/T_{SSA}$ ratio for direct transmission coefficients of monolayer (solid lines): 0.0001 (1); 0.003 (2); 0.01 (3) and of the direct transmission coefficient T_{SSA} (dashed lines) of monolayer of nonabsorbing particles: $T_{SSA} = 1$ (I); 0.9 (II); 0.8 (III); 0.5 (IV); and 0.3 (V). $n = 1.2$ (Loiko et al. 1998)

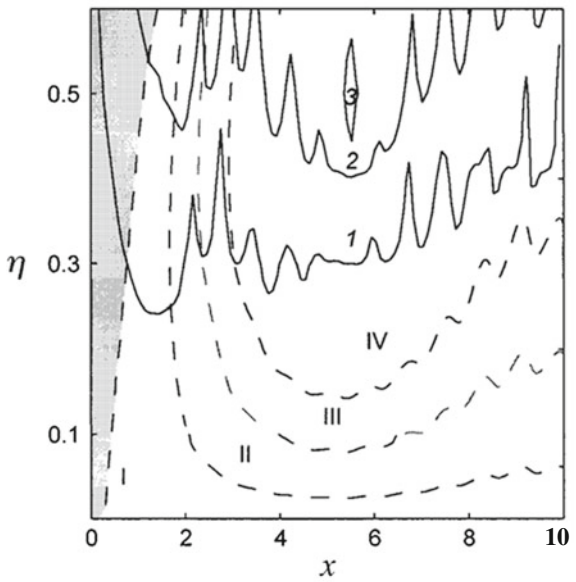

Fig. 2.18 Isolines of the $(T_{QCA}-T_{SSA})/T_{SSA}$ ratio for direct transmission coefficients of monolayer (solid lines), 0.01 (1); 0.05 (2); 0.3 (3) and of the direct transmission coefficient T_{SSA} (dashed lines) of monolayer of nonabsorbing particles $T_{SSA} = 1$ (I); 0.9 (II); 0.7 (III); 0.5 (IV). $n = 1.4$ (Loiko et al. 1998)

for direct transmission T_c and specular reflection R_c coefficients of such monolayers are obtained in Loiko and Molochko (1996).

If the absorption index (imaginary part of complex refractive index) of particles tends to zero, in the QCA (Loiko and Molochko 1996):

$$T_c = 1 - (\alpha x)^2 + O(x^4), \tag{2.3.30}$$

$$R_c = (\alpha x)^2 + O(x^4). \tag{2.3.31}$$

Here $\alpha \equiv \frac{2\nu}{1-\nu \mathrm{C}(\eta)/2}$, $\nu \equiv \eta \frac{m^2-1}{m^2+2}$, $\mathrm{C}(\eta) \equiv 1 + \int_1^\infty \frac{\mathrm{g}(t)-1}{t^2} \mathrm{d}t$, g is the radial distribution function. For $\eta < 0.7$ the approximation for $\mathrm{C}(\eta)$ by polynomials of the forth power of η is valid: $C(\eta) = 1 + 0.17\eta + 2.9\eta^2 - 8.4\eta^3 + 8.6\eta^4$.

Formulae in the SSA give the following erroneous result (Loiko and Molochko 1996):

$$T_c = 1 + (\nu x)^2 + O(x^4), \tag{2.3.32}$$

$$R_c = (2\nu x)^2 + O(x^4). \tag{2.3.33}$$

As follows from Eq. (2.3.32) the single scattering approximation cannot be used to describe optical properties of monolayers of small weakly absorbing scatterers. It gives obviously wrong result: transmittance is more than unity.

2.3.5 *Quenching of the Forwardly Transmitted Component of Light*

From Eq. (2.3.17) it follows that direct transmittance T_c is equal to zero, if:

$$L = 0.5 \tag{2.3.34}$$

and

$$\eta = 2/Q_{ext}. \tag{2.3.35}$$

It means that in these conditions the *quenching effect* for direct (coherent) transmittance is implemented. This effect was considered in Ivanov et al. (1988), Loiko et al. (2000), Dick et al. (1987a), Loiko and Konkolovich (2000, 2001a, b), Zyryanov et al. (2001), Konkolovich et al. (2000).

Equations (2.3.34) and (2.3.35) determine the parameters of individual scatterers and the filling factor of the layer, respectively, at which the quenching effect is realized. It should be noted that Eq. (2.3.35) imposes certain limits on the scattering properties of the particles. Indeed, since the particulate monolayer always has

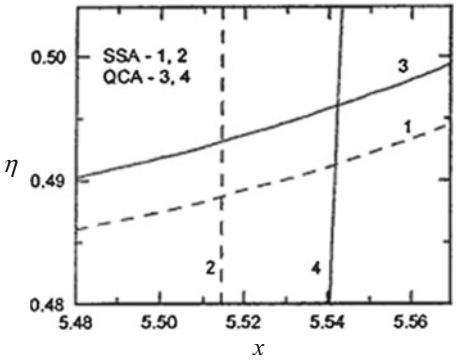

Fig. 2.19 Isolines for $L = 0.5$ (vertical) and $\eta = 2/Q_{ext}$ (tilted) calculated in the SSA (dashed lines) and the QCA (solid lines). The intersection point of the dashed (solid) lines shows the values of x_0 and η_0 calculated in the SSA (QCA) (Loiko et al. 2000)

maximal value of the filling factor $\eta_{\max}$, Eq. (2.3.35) holds only when $Q \geq 2/\eta_{\max}$ (for monolayer of monodisperse spherical particles $\eta_{\max} \approx 0.907$).

The physical meaning of Eqs. (2.3.34) and (2.3.35) is as follows. Direct transmittance is determined by the field outgoing from the layer in the forward direction, which is the sum of the field of incident wave and the fields of waves scattered in direction of the wavevector of the incident wave. Therefore the condition $T_c = 0$ can be implemented if the forward-scattered field is equal in value and opposite in phase to the field of the forwardly transmitted incident wave. Equations (2.3.34) and (2.3.35) are the mathematical representations of these conditions.

Based on Eqs. (2.3.34) and (2.3.35), it is possible to find the scatterers' parameters: the refractive index $m = n + i\kappa$, the size parameter x_0, and the values of η_0 at which the direct transmittance of the layer is zero (Ivanov et al. 1988; Loiko et al. 2000). The parameters $x = x_0$ and $\eta = \eta_0$ calculated using Eqs. (2.3.34) and (2.3.35) in the frames of the SSA (dashed lines) and QCA (solid lines) at which $T_c = 0$ are shown in Fig. 2.19 for monolayer of monodisperse nonabsorbing ($\kappa = 0$) spherical particles with $n = 1.4$. They are determined by the intersection point of the dashed (SSA-calculation) and solid (QCA-calculation) lines: the vertical one corresponds to the condition $L = 0.5$ and the tilted one corresponds to the condition $\eta = 2/Q_{ext}$.

Consider, using the SSA, typical behavior of the x_0 and η_0 for monolayer of spherical monodisperse nonabsorbing particles.

In Fig. 2.20a is shown the coupling between the particle refractive indices n and size parameters $x = x_0$ at which the quenching effect is implemented. Note that all curves except for the three lower ones have termination points determined from condition $Q_{ext} = 2/\eta_{\max} = 2.205$. It is apparent that zero transmittance can be realized at any refractive index n, but only for a few fixed values of x (we designated them as x_0). The number of x_0-values depends on n and increases with n.

The curves in Fig. 2.20a are well approximated by the expressions $x_0(n - 1) = l$, where $l = 52.2$, 5.8, 8.5, 11.6, 14.6, and so on for the curves labeled by integers 1,

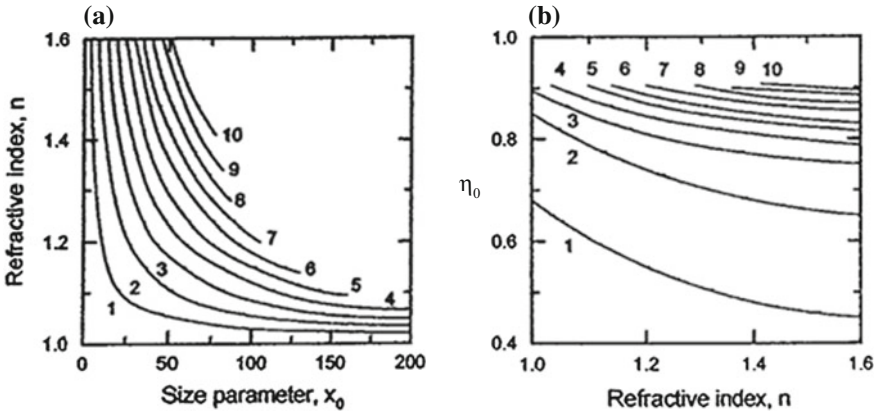

Fig. 2.20 Values of n, x_0, and η_0 at which quenching effect for direct transmittance of monolayer of monodisperse nanabsorbing spherical particles is implemented ($T_c = 0$) (Ivanov et al. 1988; Loiko et al. 2000). Simulation under the SSA. Integers near the lines correspond to numbers of maxima on dependence of the extinction efficiency factor on size parameter

2, 3, 4, 5, and so on. These integers are numbers of maxima on dependence of the extinction efficiency factor on size parameter. To satisfy the condition $T_c = 0$, as follows from the foregoing, the monolayer has to have a certain filling factor η. In this case, for each pair of n and x_0 there corresponds one filling factor $\eta = 2/Q_{ext}$ (designated as η_0).

The values of η_0 versus n are given in Fig. 2.20b. It should be noted that for each η_0 determined by curve 1, there corresponds the size parameter x_0 determined by curve 1 of Fig. 2.20a, and so on. From Fig. 2.20 it follows that as n increases, the value of η_0 decreases. The η_0 changes in a similar manner with x_0 decreasing.

The conditions for the quenching effect have been formulated using the equations obtained in the single scattering approximation: each particle is illuminated only by the incident wave and the reillumination of particles is negligibly small. The reillumination, i.e. multiple scattering of waves, can be taken into account using the QCA. Analysis shows that the direct transmittance in the QCA can be given in a form analogous to Eq. (2.3.17), but the values of L and Q_{ext} parameters become the functions of the filling factor. The intersection point of the solid lines in Fig. 2.19 shows the values of x_0 and η_0 at which quenching occurs at simulation in the QCA. It is easy to see that taking into account multiply-scattered light results in increasing the size parameter and the filling factor, but the quenching effect does not vanish. Figure 2.21 illustrates calculated in the QCA dependence of direct transmittance quenching on the filing factor and size parameter, when the monolayer consists of particles with refractive index $n = 1.4$.

The data demonstrating quenching effect experimentally are published in Loiko et al. (2000), Konkolovich et al. (2000).

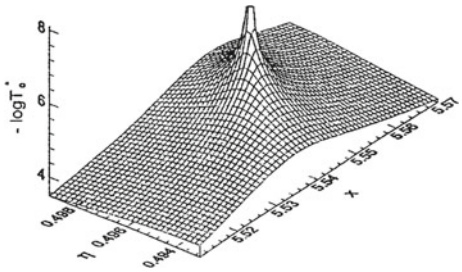

Fig. 2.21 Direct transmittance T_c as a function of monolayer filling factor η and particle size parameter x calculated in the QCA (Ivanov et al. 1988; Loiko et al. 2000). Refractive index of particles $n = 1.4$

2.3.5.1 Quenching Effect Description in the Amplitude-Phase Screen Model

For the optically soft particles in the framework of the amplitude-phase screen model the quenching effect was considered in Zege and Kokhanovsky (1991), Kokhanovsky (2001).

In the frame of this model conditions for zero transmittance of scattering medium consisting of nonabsorbing particles can be written as

$$tan\Delta = \Delta, \tag{2.3.36}$$

$$\eta = \left(1 + \frac{2(1-\cos\Delta)}{\Delta^2} - \frac{2\sin\Delta}{\Delta}\right)^{-1}. \tag{2.3.37}$$

Here $\Delta = kD(m_2 - m_1)$ is the phase shift on the particle with diameter D, m_2 is the refractive index of particles, $m_1$is the refractive index of the host medium, $k = 2\pi/\lambda$ is the wave number, λ is the length of the incident wave.

Direct transmittance T_c is zero for a discrete phase shifts Δ_l. At $\Delta_l >> 1$ they are approximately described by relation:

$$\Delta_l \approx \pi(2l+1)/2. \tag{2.3.38}$$

Here l is an integer. The values of filling factor at which $T_c = 0$ are determined by formula:

$$\eta_l = \left(1 + \frac{4}{\pi(2l+1)}\right)^{-1}, \tag{2.3.39}$$

where $l = 1, 3, 5, \ldots$.

The approach based on the amplitude-phase screen permits one to consider in details conditions of $T_c = 0$ implementation for monolayers of nonspherical, anisotropic, and structured particles.

2.3.6 Spatial Optical Noise

At illumination of the particulate layer by a wide homogeneous beam of light the intensity of light transmitted and reflected by small layer areas can depend on their locations. These fine-scale intensity fluctuations are defined as noise (Frieser 1975; Ivanov and Loiko 1983). Noise leads to loss of quality of the image recording and mapping films (photo- and holographic emulsion layers, liquid-crystal (LC) composites encapsulated in a polymer, films of image migration, and other systems of the image transfer).

Among the basic factors determining the noise nature, the following ones can be emphasized: fluctuations of the number of inhomogeneities within the limits of a pixel, fluctuations of optical properties and geometric parameters of individual inhomogeneities, statistical features of the material microstructure, and mechanisms of the process of light scattering. The sizes of inhomogeneities can vary over wide range: from tens of micrometers (for instance, large-grained emulsions) to micron and submicron levels (holographic films, LC composites encapsulated in polymers, and so on). The statistics of inhomogeneity distribution depends on technological processes of material preparation and properties of the binder.

At a low particle concentration, single scattering occurs. In this case, to describe the light scattering the statistical model of a "Poisson ensemble" of scattering centers can be used.

Dispersion materials with a high particle concentration attract considerable practical and theoretical interest. Difficulties of the theoretical study of noise properties are obvious in this case due to dense particle package inducing the short-range topological ordering and pronounced cooperative optical effects (Frieser 1975; Ivanov and Loiko 1983). As a result, the rigorous solution of the problem of predicting characteristics of the spatial optical noise should lean upon the theory of multiple scattering of waves and statistical simulation, which is unlikely to provide the solution in the complete and especially in the analytical form. Thereby, finding and the construction of adequate physically transparent approximate models is needed for the elucidation of basic features in the variation of characteristics of optical noise as functions of the concentration and optical properties of individual inhomogeneities forming a disperse material.

To characterize noise properties of the films the power spectral density (Wiener spectrum) is used (Bendat and Piersol 1971):

$$n(\boldsymbol{\nu}) = \lim_{A\to\infty} \frac{1}{A} \left\langle |TF_A\{t(\mathbf{x})\}|^2 \right\rangle. \tag{2.3.40}$$

Here $\boldsymbol{\nu}$ is the spatial frequency (Frieser 1975) vector, A is the aperture (the surface of averaging), and

$$TF_A\{t(\mathbf{x})\} = \int_A t(\mathbf{x}) \exp(-i2\pi\boldsymbol{\nu}\mathbf{x}) \mathrm{d}\mathbf{x} \tag{2.3.41}$$

is the two-dimensional Fourier transformation of $t(\mathbf{x})$. Angular brackets denote statistical averaging over realizations. The function $t(\mathbf{x})$ entering into (2.3.40) and (2.3.41) can describe various physical quantities: local spatial fluctuations of transmission or reflection coefficients and optical densities in transmission or reflection relative to their mean values. Some numerical characteristics of noise are determined from this spectrum (Frieser 1975; Ivanov and Loiko 1983).

2.3.6.1 Spatial Statistical Inhomogeneity: Fluctuation of Particle Concentration

In Sect. 2.2.1 we considered the approach based on N-particle distribution functions to statistical description of particulate media. In some practical problems, for example image processing (Frieser 1975; Ivanov and Loiko 1983; Yu 1973; Loiko and Dubovik 1989a, b, 1990; Loiko 1991; Loiko and Konkolovich 1994, 1998a, b) we should know fluctuations of particle concentration (density) in the considered space area.

Let us choose a small volume V in a particulate medium. Suppose that the density of particles in this volume is a random function of its coordinates. This factor results in fluctuation of the light flux outgoing the volume.

At low concentration of particles their spatial arrangement is described by Poisson statistics: the probability distribution function P_N to find N particles in a considered volume V is:

$$P_N = \frac{(\bar{N})^N \exp(-\bar{N})}{N!}. \tag{2.3.42}$$

Here $\bar{N}$ is the average number of particles in a volume.

If particle concentration is not small the other equations for P_N should be used. The expressions for one-, two-, and three-dimensional systems with absolutely rigid particles of finite size are written in Ivanov et al. (1988), Loiko and Dubovik (1986).

The normalized probability distribution for the 1D particulate system is (Ivanov et al. 1988):

$$P_N^n(L) = \frac{\frac{\Lambda_1^N (L - NL^*(1))^N}{N!} \exp(-\Lambda_1 (L - NL^*(1))}{\sum\limits_{k=0}^{N_{\max}} \frac{\Lambda_1^k (L - kL^*(1))^k}{k!} \exp(-\Lambda_1 (L - kL^*(1))}. \tag{2.3.43}$$

Here $\Lambda_1 = \eta_1 (1 - \eta_1)^{-1} l^{-1}$, $N_{\max} = [L/l]$, $L^*(1) = l = D$, D is the particle size, η_1 is the filling factor of the 1D particulate system. L is the size (length) of the system. The square brackets mean that the integer part is taken.

The normalized probability distribution for the 2D system is (Ivanov et al. 1988):

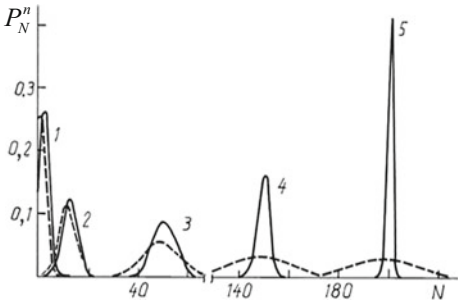

Fig. 2.22 Dependence of P_N^n as function of N. Dashed lines correspond to calculation by Eq. (2.3.42). Solid lines correspond to calculation by Eq. (2.3.44) for 2D system. The aperture is 250 times larger than the cross section of particle. Filling factor $\eta = 0.01$ (curves 1), 0.05 (curves 2), 0.2 (curves 3), 0.6 (curves 4), 0.8 (curves 5) (Ivanov et al. 1988)

$$P_N^n(S) = \frac{\frac{\Lambda_2^N (S - NS^*(1))^N}{N!} \exp(-\Lambda_2(S - NS^*(1))}{\sum_{k=0}^{N_{\max}} \frac{\Lambda_2^k (S - kS^*(1))^k}{k!} \exp(-\Lambda_2(S - kS^*(1))}. \tag{2.3.44}$$

Here $\Lambda_2 = \eta(1+\eta)^{1/2}(1-\eta)^{-3/2}s^{-1}$, $N_{\max} = \left[S\eta\left(1 - \frac{(1-\eta)^{3/2}}{(1+\eta)^{1/2}}\right)^{-1} s^{-1}\right]$, $S^* = \left(1 - \frac{(1-\eta)^{3/2}}{(1+\eta)^{1/2}}\right) s\eta^{-1}$, $s = \pi D^2/4$, D is the particle size, η is the filling factor of the 2D particulate system. S is the aperture area. The square brackets mean that the integer part of the number is taken. Expression for three-dimensional system is similar to Eq. (2.3.44) (Ivanov et al. 1988).

The calculated by Eqs. (2.3.42) and (2.3.44) dependences of P_N^n versus N are shown in Fig. 2.22. They indicate difference in statistics of points and circles.

2.3.6.2 Wiener Spectrum of Partially Ordered Monolayer

In Loiko and Konkolovich (1994, 1995, 1997, 1998a, b) the relations to describe noise properties of a monolayer of polydisperse particles are derived. They are based on the model of the stochastic amplitude-phase screen. The model of the random substitution mixture (Ziman 1979) is used to describe short-range order of particles.

The Wiener transmittance spectrum of monolayer of polydisperse particles is described by the equation (Loiko and Konkolovich 1998b)

$$n(\nu)/\langle\sigma\rangle = \eta\{(1+1/\alpha)^2(Y^2(q\sqrt{1+1/\alpha}) \times S_\nu(q\sqrt{1+1/\alpha}) + \\ + [(1+1/\alpha)(1+2/\alpha)(1+3/\alpha) - (1+1/\alpha)^2] \times Y^2(q\sqrt{(1+1/\alpha)(1+3/\alpha)})\}. \tag{2.3.45}$$

Here $Y(\mathrm{x}) = 2J_1(x)/x$, $J_1(x)$ is the first order Bessel function, $\eta = n\pi\langle R\rangle^2$ is the filling factor of the monolayer of polydisperse particles, $q = \alpha\omega/\beta = \omega\langle R\rangle$, $\omega = 2\pi\nu$, ν is the spatial frequency (Ivanov et al. 1988; Frieser 1975; Ivanov and Loiko 1983),

S_ν is the so-called spatial frequency structure factor (Loiko and Konkolovich 1998a, b), $\langle\sigma\rangle = \pi\langle R\rangle^2$, R is the particle radius, α and β are the parameters of the particle size probability density function:

$$P(R) = \frac{\beta^\alpha}{\Gamma(\alpha)} R^{\alpha-1} \exp(-\beta R), \tag{2.3.46}$$

where $\Gamma(\alpha)$ is the gamma function.

The spectrum (2.3.45) is presented in Fig. 2.23 at various filling factors of the monolayer.

Figure 2.24 illustrates the dimensionless quantity $G/\sqrt{\langle\sigma\rangle}$ as a function of filling factor η for different variation coefficients $D_R/\langle R\rangle$ of the monolayer for transmitted light. Here $G = \sqrt{n(0)}$ is the Selwyn granularity coefficient (Ivanov et al. 1988; Frieser 1975; Ivanov and Loiko 1983).

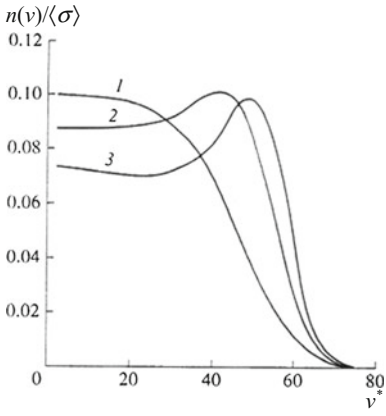

Fig. 2.23 Dimensionless Wiener transmittance spectrum at $\eta = 0.2$ (1), 0.4 (2), and 0.5 (3) (Loiko and Konkolovich 1998b). Variation coefficient $D_R/\langle R\rangle = 0.125$. Dimensionless frequency $\nu^* = \nu\pi\langle D\rangle/50$. Here the dimensions of ν and $\langle D\rangle$ are determind in mm^{-1} and μm, respectively

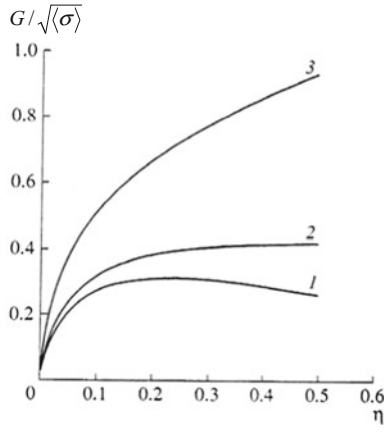

Fig. 2.24 The dependence of $G/\sqrt{\langle\sigma\rangle}$ for the transmission coefficient as a function of the filling factor η of the monolayer at different variation coefficients $D_R/\langle R\rangle = 0.125$ (curve 1), 0.25 (curve 2), and 0.5 (curve 3) (Loiko and Konkolovich 1998b)

In conclusion of this subsection note that, not only noise, but scattering in a particulate media also influences on the image quality. The peculiarities in the image transfer of the partially ordrered system caused by particle concentration are experimentally investigated in Ivanov et al. (1988), Dick et al. (1986) for layers with optically soft particles. In these publications the data for the line spread function (Ivanov et al. 1988; Zege et al. 1991a; Frieser 1975) are obtained and behavior of the modulation transfer function is discussed.

2.4 Three-Dimensional Particulate Structures

2.4.1 Adding Method for Direct Transmittance and Specular Reflectance Determination

One of the methods which allows one to obtain relatively simple solution of the problem of light scattering by the 3D particulate layer is the *adding method* (Loiko and Miskevich 2004; Loiko et al. 2005a, b; Loiko and Molochko 1998). It is based on the consideration of thick particulate layer as a set of optically interacting sublayers (monolayers) (see Fig. 2.25). Under this approach (Gurevich 1931) the analytical solution for direct transmission T and specular reflection R coefficients of a thick normally illuminated particulate layer was obtained in Loiko and Molochko (1998). It takes into account multiple reflection between the sublayers.

The adding method was used to analyze the T and R coefficients of the particulate layer and the phases of the transmitted and reflected waves (Loiko and Miskevich 2004; Loiko et al. 2005a, b; Loiko and Molochko 1998). The approach considered is applicable to layers with small and high concentration of scatterers. The equations to determine the reflection and transmission coefficients of an added layer are derived on the basis of similar characteristics of particulate monolayers which can be calculated, for example, in the SSA or QCA. The equations for amplitude coefficients of specular reflection, R, and direct transmission, T, of the 3D particulate layer can be written as follows (Loiko and Miskevich 2004; Loiko et al. 2005b; Loiko and Molochko 1998):

$$R(z, K_1, K_2) = R_\infty \frac{1 - \exp(-2Lz)}{1 - R_\infty^2 \exp(-2Lz)}, \tag{2.4.1}$$

$$T(z, K_1, K_2) = (1 - R_\infty^2) \frac{\exp(-(L + 2i)z)}{1 - R_\infty^2 \exp(-2Lz)}, \tag{2.4.2}$$

where $z \equiv \pi y/\lambda$, y is the dimensionless layer thickness, λ is the length of incident wave,

$$L^2 \equiv (K_2 - 2i)^2 - K_1^2, \quad R_\infty \equiv (K_2 - L - 2i)/K_1, \tag{2.4.3}$$

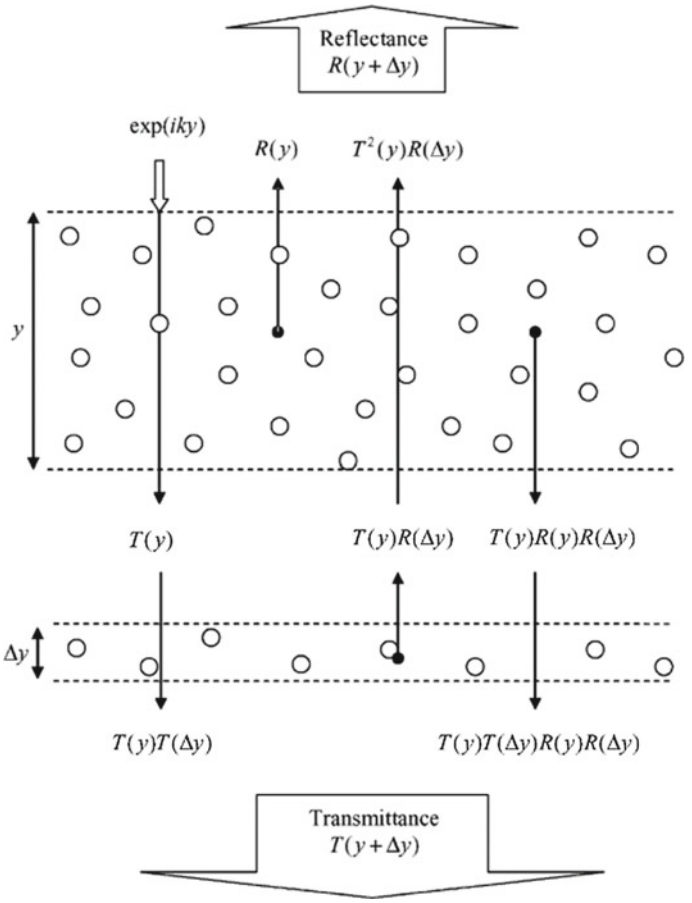

Fig. 2.25 Scheme of light propagation through a layer of thickness $y + \Delta y$ normally illuminated by a plane wave (Loiko and Miskevich 2004). The quantities $T(y)$, $R(y)$, $T(\Delta y)$, $R(\Delta y)$ are the amplitude direct transmission and specular reflection coefficients of layers with thickness y and Δy, respectively. Each thin arrow in the figure represents the sum of all possible scattering events, which result in the direct transmittance and specular reflectance of the whole layer. The formulas near the arrows indicate the amplitude reflection and transmission coefficients due to the first, second, third, etc. scattering events between layers. The dashed lines indicate the imaginary interfaces of the layer with thickness y and sublayer with thickness Δy. The gap between layers is only to indicate more clearly the multiple scattering events between the layers and to get more space to write the equations. The $R(y + \Delta y)$ and $T(y + \Delta y)$ are reflectance and transmittance of the layer of thickness $y + \Delta y$

$$K_1 = \frac{3\eta_V}{x^3} f(\pi), \quad K_2 = \frac{3\eta_V}{x^3} f(0) \tag{2.4.4}$$

$\eta_V = 2D\eta/(3y)$ is the volume filling factor of the monolayer expressed via the surface filling factor η of the monolayer, $x = kD/2$ is the size parameter of a scatterer; $k = 2\pi/\lambda$, D is the particle diameter; $f(\theta)$ is the amplitude scattering function, θ

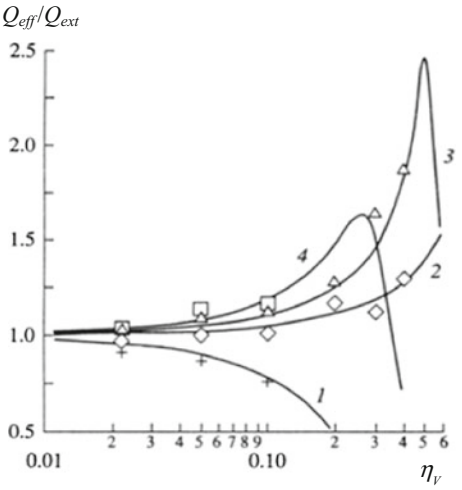

Fig. 2.26 Dependence of Q_{eff}/Q_{ext} versus volume filling factor η_V for layer of latex particles (Loiko and Molochko 1998). Q_{ext} is the extinction eficiency factor of single particle, Q_{eff} is the effective extinction efficiency factor of particle. The calculated under the adding method and the SSA data are given by lines. Experimental data of Ishimaru and Kuga (1982) are given by symbols at: $x = 3.5$, $m = 1.19$ (1); $x = 13.7$, $m = 1.19$ (2); $x = 36.2$, $m = 1.19$ (3); $x = 82.8$, $m = 1.19$ (4)

is the polar scattering angle. With growing the size of the scatterers, the thickness dependence of transmittance, T, acquires a typical exponential form:

$$|T|^2 = \exp\left(-\frac{6\eta_V}{x^3}\mathrm{Re}\, f(0)z\right) \quad . \tag{2.4.5}$$

The effective extinction efficiency factor Q_{eff} of particle in the medium is determined as

$$Q_{eff} = \frac{4D}{3\eta_V}\mathrm{Im}\, k_{eff}, \tag{2.4.6}$$

where k_{eff} is the effective wave number (Ishimaru 1978a, b).

The data for layers of monodisperse latex particles of submicron and micron sizes dispersed in water calculated under the adding method and measured in Ishimaru and Kuga (1982) are presented in Fig. 2.26. The model describes well the experimental data (Ishimaru and Kuga 1982). They correlate with experimental results on the spectral transmittance of layers of polyethylene powder in the infrared region (Zhukas et al. 2014).

The results of the calculations (by Eq. (2.4.2)) of transmission coefficient $|T|^2$ are presented in Fig. 2.27 by solid curves (Loiko and Molochko 1998). The T and R of individual monolayers (sublayers) are calculated in the SSA. It is seen that the transmission coefficient is an oscillating function of the layer thickness for small scatterers. The tilt angle of the transmission curve decreases and the oscillation

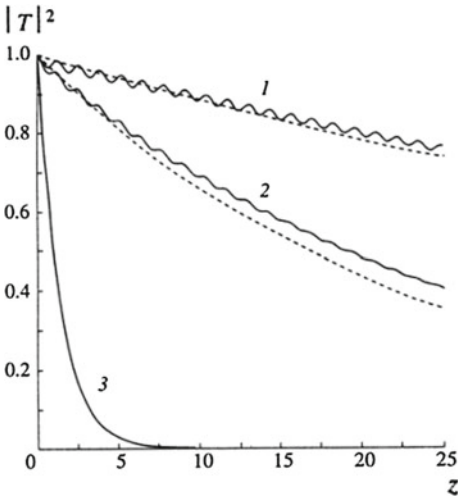

Fig. 2.27 Transmission coefficient $|T|^2$ of a layer versus layer thickness z. SSA. Solid and dashed lines were calculated from formulas (2.4.2) and (2.4.5), respectively (Loiko and Molochko 1998). (1): $K_1 = 6.1 \times 10^{-3} - 4.0 \times 10^{-1}i$, $K_2 = 6.1 \times 10^{-3} - 4.3 \times 10^{-1}i$, $(x = 0.4, m = 1.6, \eta = 0.4)$; (2): $K_1 = 2.1 \times 10^{-2} - 3.8 \times 10^{-1}i$, $K_2 = 2.1 \times 10^{-2} - 4.5 \times 10^{-1}i$, $(x = 0.6, m = 1.6, \eta = 0.4)$; (3): $K_1 = -1.1 \times 10^{-1} + 8.9 \times 10^{-3}i$, $K_2 = 3.7 \times 10^{-1} - 3.7 \times 10^{-1}i$, $(x = 2.0, m = 1.6, \eta = 0.4)$

amplitude increases as the particle sizes decrease. The shape of the curve approaches one calculated from Airy's formulas for plane-parallel plates. As the size of the scatterers grows, the magnitude of oscillations associated with interference of waves reflected from the layer boundaries decreases and the dependence acquires a typical exponential form. The calculation results obtained by Eq. (2.4.5) are shown for comparison in Fig. 2.27 by dashed lines.

Figures 2.28 and 2.29 show the dependences of $|T|^2$ and $|R|^2$ of layer on the thickness z calculated using the QCA (solid lines) and the SSA (dashed lines) (Loiko and Miskevich 2004; Loiko et al. 2005b). As follows from the data analysis, for small particle sizes $|T|^2$ and $|R|^2$ are oscillating functions of z.

With growing the volume filling factor η_V the oscillation magnitude grows. Multiple scattering leads to an increase in the velocity of damping the transmittance and reflectance oscillations with the thickness increasing. With particle sizes increasing the oscillations disapear and $T(z)$ dependence becomes exponential.

2.4.2 *Transfer Matrix Method*

Described in the previous subsection adding method can also be used to calculate the transmission and reflection coefficients of three-dimensional *ordered* particulate systems, such as photonic crystals. However, since such systems have a crystal-like

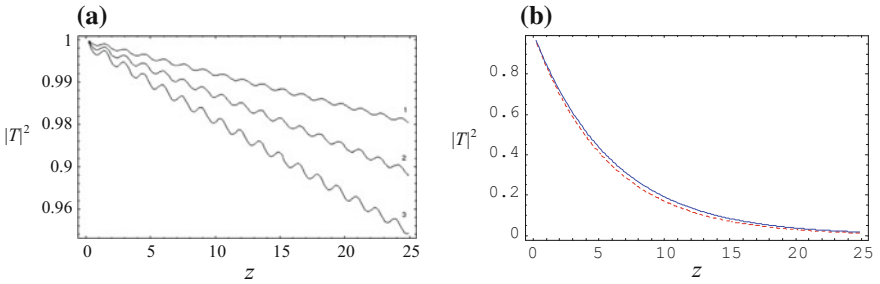

Fig. 2.28 Transmission coefficient $|T|^2$ versus z at different η_V, x and m. **a** QCA, $\eta_V = 0.3$ (curve 1), $\eta_V = 0.4$ (2), $\eta_V = 0.5$ (3), $x = 0.1$, $m = 1.1$; **b** QCA (solid curve), SSA (dashed curve), $\eta_V = 0.4$, $x = 7$, $m = 1.6$ (Loiko and Miskevich 2004)

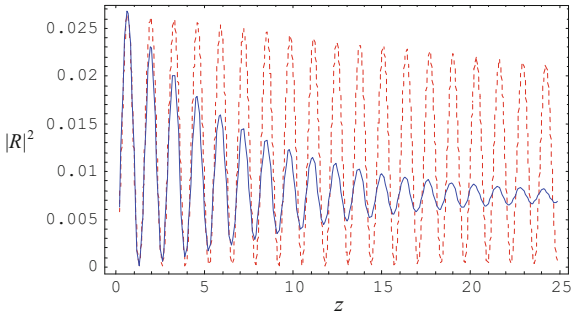

Fig. 2.29 Reflection coefficient $|R|^2$ as a function of layer thickness z. Solid and dashed lines show the results of calculation using the QCA and SSA, respectively. $\eta_V = 0.4$, $x = 0.4$, $m = 1.6$ (Loiko and Miskevich 2004)

structure with well-defined crystalline planes (particulate monolayers), it is more convenient to use the *transfer matrix method* (TMM) (Katsidis and Siapkas 2002; Centurioni 2005; Troparevsky et al. 2010).

In the framework of this method the layered system (multilayer) is considered as a sequence of *interfaces* and *layers* separated by these interfaces. Each inteface is characterized by transmission and reflection coefficients, and each layer is defined by thickness and complex refractive index. The TMM takes into account multiple reflection between the interfaces and interference of waves. Under this method one can calculate the transmission and reflection coefficients of systems with different types of interfaces, absorbing and nonabsorbing layers, and their arbitrary combinations (Katsidis and Siapkas 2002).

When TMM is used to determine the transmission and reflection coefficients of three-dimensional photonic crystal (3D PhC), the individual monolayers (crystal planes) are considered as interfaces and spaces between them as layers. The transmission and reflection coefficients of the monolayers can be determined, for example, in the QCA or SSA.

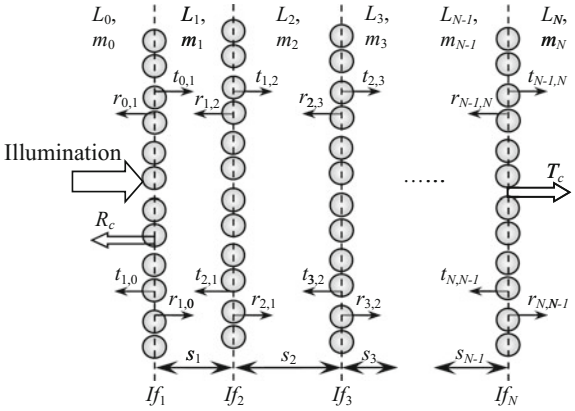

Fig. 2.30 Schematical presentation of particulate multilayer – side view along the monolayer planes designed by dashed lines. m_i and s_i are the complex refractive index and thickness of i-th layer L_i; If_i are numbers of interfaces (monolayrs); $t_{i,j}$ and $r_{i,j}$ ($t_{j,i}$ and $r_{j,i}$) are amplitude transmission and reflection coefficients for wave propagating in the illumination direction (opposite direction), $i < j$; T_c and R_c are transmission and reflection coefficients of multilayer

Let us consider the multilayer consisting of plane-parallel particulate monolayers (Fig. 2.30).

Write the basic equations of TMM for direct transmission T_c and specular reflection R_c coefficients of such a multilayer (Katsidis and Siapkas 2002):

$$T_c = |t|^2 = |1/T_{11}|^2 , \tag{2.4.7}$$

$$R_c = |r|^2 = |T_{21}/T_{11}|^2 , \tag{2.4.8}$$

where t and r are amplitude direct transmission and specular reflection coefficients of layered system, T_{ij} are elements of transfer matrix $\mathbf{T}$ of multilayer. For N-interface system (Fig. 2.30) the transfer matrix $\mathbf{T}_{0,N}$ has a form:

$$\mathbf{T}_{0,N} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} = \frac{1}{t_{0,N}} \begin{bmatrix} 1 & -r_{N,0} \\ r_{0,N} & t_{0,N}t_{N,0} - r_{0,N}r_{N,0} \end{bmatrix} . \tag{2.4.9}$$

Here $t_{0,N}$ and $r_{0,N}$ ($t_{N,0}$ and $r_{N,0}$) are amplitude transmission and reflection coefficients of multilayer for the wave propagating in the direction of wave vector of incident wave (in the opposite direction).

Transfer matrix of whole system is determined by multiplying the transfer matrices $\mathbf{T}_j$ of interfaces with propagation matrices $\mathbf{P}_j$ of layers:

$$\mathbf{T}_{0,N} = \left(\prod_{j=1}^{N-1} \mathbf{T}_j \mathbf{P}_j \right) \mathbf{T}_N . \tag{2.4.10}$$

Here

$$\mathbf{T}_j = \frac{1}{t_{j-1,j}} \begin{bmatrix} 1 & -r_{j,j-1} \\ r_{j-1,j} & t_{j-1,j}t_{j,j-1} - r_{j-1,j}r_{j,j-1} \end{bmatrix} \tag{2.4.11}$$

is the transfer matrix of j-th interface,

$$\mathbf{P}_j = \begin{bmatrix} \exp(-ik_j s_j) & 0 \\ 0 & \exp(ik_j s_j) \end{bmatrix} \tag{2.4.12}$$

is the propagation matrix of j-th layer, $t_{j-1,j}$ and $r_{j-1,j}$ ($t_{j,j-1}$and $r_{j,j-1}$) are amplitude transmission and reflection coefficients of j-th interface for the wave propagating in the direction of wave vector of incident wave (in the opposite direction); wavenumber $k_j = 2\pi m_j/\lambda$; m_j and s_j are complex refractive index and thickness (distance between the planes where centers of particles in adjacent monolayers are located) of j-th layer, λ is the wavelength in the surrounding medium.

Using the Eqs. (2.4.7)–(2.4.12) one can calculate the T_c and R_c coefficients of system consisting of dirreferent number of different monolayers in various surrounding media.

2.4.3 *Layered Particulate Structures from Spherical Dielectric Particles: Direct Transmittance and Specular Reflectance Spectra*

In this section, we consider the results for transmittance and reflectance of layered particulate structures from dielectric particles, such as 3D photonic crystals, colloidal crystals, opals, etc. They were obtained in the QCA and TMM. First, in the quasicrystalline approximation (see Sect. 2.3.1) were calculated the direct transmission and specular reflection coefficients of individual particulate monolayers. Second, these coefficients were used in the transfer matrix method (see Sect. 2.4.2) to compute the direct transmission T_c and specular reflection R_c coefficients of the layered system (multilayer) from such monolayers. The influence of the spatial distribution of particles in individual monolayers on the T_c and R_c was studied and characteristic minima in transmittance spectra of layers caused by the interference of waves were described (Loiko and Miskevich 2013; Miskevich and Loiko 2013a, 2014b).

2.4.3.1 Systems with Alumina Particles

Spectral dependences of the direct transmission T_c and specular reflection R_c coefficients of normally illuminated planar crystal with a hexagonal lattice and a partially ordered monolayer of monodisperse spherical alumina (Al_2O_3) particles are shown in Fig. 2.31a and b (Loiko and Miskevich 2013). The dependence of the $T_c + R_c$ sum

is presented in Fig. 2.31c. The value of $1–(T_c + R_c)$ gives information on the amount of light scattered at all directions exept of strictly forward and strictly backward. The filling factor of monolayers is $\eta = 0.5$. Calculations were performed using data on the spectral dependence of refractive index m of alumina published in Palik (1991). In the wavelength range of 0.3–1.0 μm, the refractive index monotonically decreases from 1.815 to 1.756. Its imaginary part in the considered spectral range is zero.

The transmittance and reflectance of monolayers were calculated in the QCA. The RDF of the planar crystal was calculated using the blurring function (see Sect. 2.3.1.2) in the form $\sigma(u) = \sigma_0 u$, where $\sigma_0 = 0.01$. To model the RDF of partially ordered monolayer, the numerical solution (Ivanov et al. 1988) of the Ornstein–Zernike integral equation in the Percus–Yevick approximation for a system of hard spheres is used. The presented results illustrate how the spatial organization of particles affects the direct transmittance and specular reflectance of the monolayer.

The spectra of partially ordered monolayers exhibit a minimum of transmission, which is caused by interference between the incident wave and waves scattered by particles. It occurs if the ratio of the wavelength to the particle size approximately corresponds to one of the main maxima of the extinction efficiency factor (Loiko et al. 2000). For a partially ordered monolayer of monodisperse spherical Al_2O_3 particles with filling factor $\eta = 0.5$ and particle diameter $D = 0.3\,\mu m$, this minimum occurs at wavelength $\lambda \approx 0.357\,\mu m$ (minimum 1' on the thick gray curve in Fig. 2.31a). Its depth and position depend on the concentration of particles and their optical properties. A similar minimum occurs in the transmittance spectra of a planar crystal. However, its position is shifted. For a PC with a hexagonal packing of Al_2O_3 particles at a considered value of the filling factor ($\eta = 0.5$), it appears at wavelength $\lambda \approx 0.339\,\mu m$ (minimum 1 on thick black curve in Fig. 2.31a). Unlike the spectrum of the partially ordered monolayer, the transmittance spectrum of the PC possesses an addition minimum (resonant dip) that is caused by the periodicity of particle arrangement (see sharp minimum 2 on thick black curve in Fig. 2.31a). In the considered case, this minimum has a fine structure. It is located in the wavelength range of 0.52–0.53 μm. This range also contains a sharp maximum of the specular reflection coefficient (see thick black curve in Fig. 2.31b). A change in the concentration of particles leads to changes in the depth and positions of minima 1 and 1', as well as in the depth, position, and fine structure of minimum 2.

In Fig. 2.31a and b, thin curves show the direct transmittance and specular reflectance dependences of a multilayer. The calculation results show that, as the number of monolayers increases, in the spectrum of the multilayer a photonic band gap is formed which is caused by the periodicity of the structure in the direction of the incident light. The thin black line in Fig. 2.31a shows the spectral dependence of the T_c coefficient of a multilayer consisting of sixty hexagonally packed planar crystals, while the thin gray line shows the corresponding dependence for a multilayer consisting of sixty partially ordered monolayers. All layers are located in air. The spectral dependences of the R_c coefficients of these systems are shown in Fig. 2.31b. These dependences are presented in the wavelength range from 0.6 to 1.0 μm, in which the indirect components of scattered light is small. As can be seen from the presented results, the spectra of the systems of partially ordered monolayers

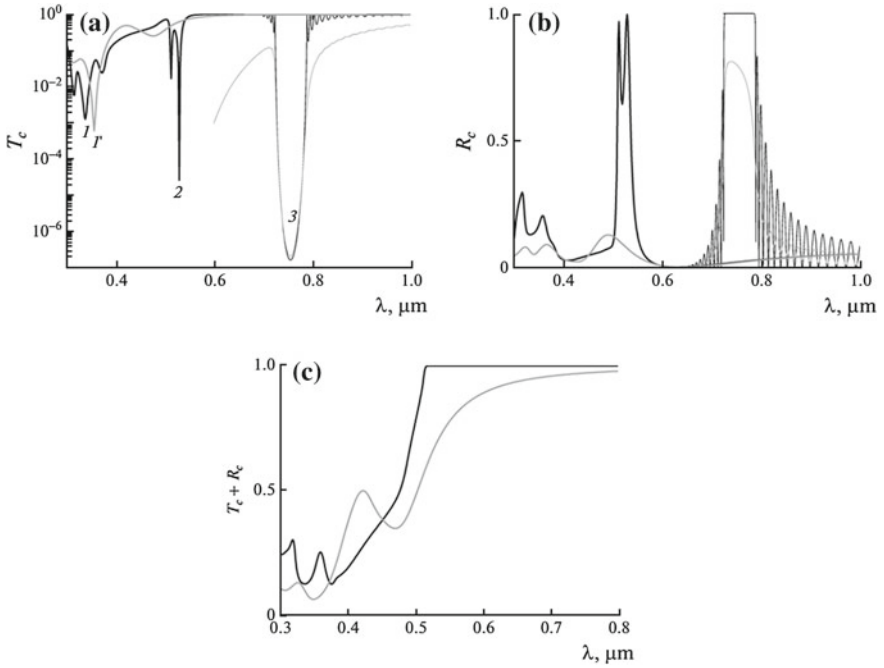

Fig. 2.31 Spectral dependences of (**a**) direct transmission T_c, (**b**) specular reflection R_c coefficients and (**c**) the $T_c + R_c$ sum of individual monolayers of alumina particles: planar crystal with a hexagonal lattice (thick black lines) and partially ordered monolayer (thick gray lines) (Loiko and Miskevich 2013). Normal illumination. The filling factor of monolayers is $\eta = 0.5$. The RDF of the PC was calculated at $\sigma(u) = 0.01u$. The diameter of particles is $D = 0.3\,\mu$m for all the layers. The digits 1, 2, and 1' designate the types of the minima in transmittance, which are characteristic for the planar crystal and the partially ordered monolayer of particles, respectively (see text). Spectral dependences of (**a**) direct transmission and (**b**) specular reflection coefficients of a layered system (a multilayer) that consists of planar crystals with a hexagonal lattice (thin black lines) and partially ordered monolayers (thin gray lines) of particles. The spacings between adjacent monolayers are $0.3\,\mu$m. The number of monolayers is 60. The $T_c(\lambda)$ and $R_c(\lambda)$ dependences of the multilayer are presented in the wavelength range from 0.6 to $1.0\,\mu$m. The digit 3 indicates a transmittance dip (photonic band gap) of the multilayer

and planar crystals have common and the different features. In particular, in both cases, the photonic band gap is formed in the wavelength range 0.73–0.78 μm. The absolute values of the photonic band gap minima differ little. However, there are also considerable differences which are caused by the spatial organization of particles in monolayers. In a system of planar crystals, the photonic band gap is more symmetric and its boundaries are sharper than for the band gap of a multilayer from partially ordered monolayers. The values of reflectance maxima $R_{c,\max}$ in the photonic band gap are also noticeably different. In a system of partially ordered monolayers, the reflection maximum in the PBG is $R_{c,\max} \approx 0.81$, whereas in a multilayer from PCs, the corresponding maximum is $R_{c,\max} \approx 1$ (Fig. 2.31b). This means, in particular,

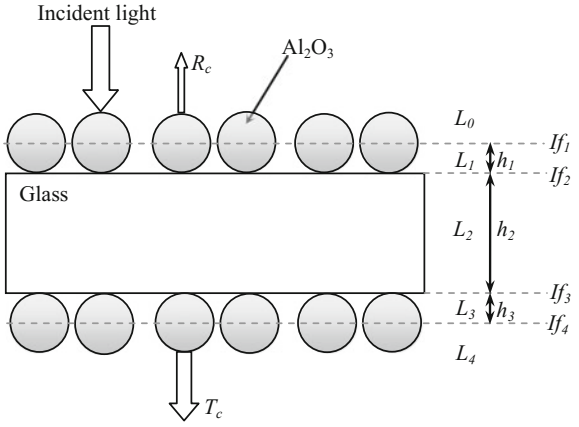

Fig. 2.32 Schematic representation of the multilayer system with monolayers of monodisperse spherical Al_2O_3 particles on the top and bottom surfaces of the glass plate (Miskevich and Loiko 2014b). h_i is the thickness of i-th layer L_i; If_i are numbers of interfaces; T_c and R_c are transmission and reflection coefficients of multilayer

that the use of systems based on PCs makes it possible to create selective reflectors and bandpass filters of better quality.

The presented results show that particulate system of nonabsorbing particles is characterized by transmittance minima of three types (Fig. 2.31). Transmission minimum 1 is determined by the concentration and optical properties of particles in the individual monolayer. It occurs in the spectra both a partially ordered monolayer (minimum 1') and a planar crystal (minimum 1). Minimum 2 is the periodicity-induced resonant dip occurred in the spectrum of a PC when sizes of particles and spacings between them are comparable with the wavelength of the incident light. In the spectrum of a short-range ordered (partially ordered) monolayer it is not observed. Transmission minimum 3 is a photonic band gap that arises in the spectrum of a multilayer when the incident wave propagates along a periodic variation of its structure.

Consider systems of ordered monolayers of Al_2O_3particles covering surfaces of a glass plate. Such layers can be used as antireflection coatings (Miskevich and Loiko 2014b). The glass plate with double-side coating (of top and bottom surfaces, see Fig. 2.32) shows better antireflection properties than single-side coated one (Miskevich and Loiko 2014b).

Spectra of system when monolayers of alumina particles have high and low ordering were simulated. Figure 2.33 shows the spectral dependences of direct transmittance and specular reflectance when the top and bottom monolayers of Al_2O_3particles are identical. Thin solid lines indicate the spectra of the system with the hexagonally packed planar crystals, the dashed lines indicate the spectra of systems with partially ordered monolayers of Al_2O_3 particles. At a double-side coating by identical monolayers of Al_2O_3 particles of diameter $D = 0.1\,\mu$m the calculated specular reflection

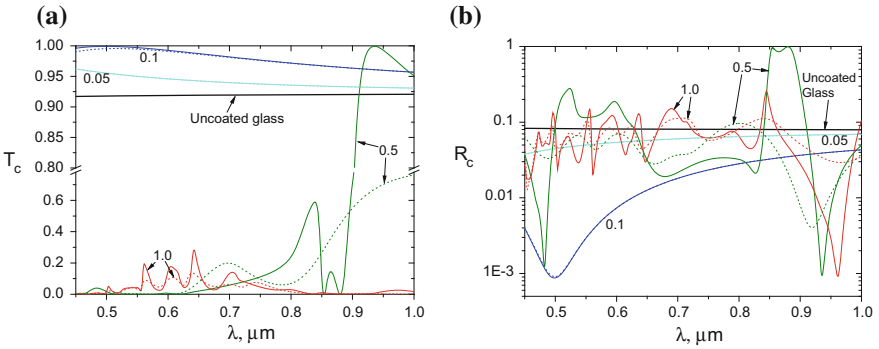

Fig. 2.33 Spectra of T_c (**a**) and R_c (**b**) of the uncoated plane-parallel glass plate (thick line) and system of identical partially ordered monolayers (dashed lines) and monolayers with hexagonal lattice of Al_2O_3 particles (thin solid lines, $\sigma(u) = 0.01u$) on the top and bottom surfaces of the glass plate at different particle sizes (Miskevich and Loiko 2014b). Thickness of the glass plate is 2 mm. $\eta = 0.5$. Diameters (in micrometers) of particles are indicated near the lines

coefficient $R_c < 0.025$ in the visible spectral range with minimum less than 0.001. If the size of Al_2O_3 particles significantly smaller than the incident light wavelength the differences in transmittance and reflectance spectra of the systems coated with PCs and partially ordered monolayers are negligible (see lines for $D = 0.05\,\mu$m and $D = 0.1\,\mu$m in Fig. 2.33). The T_c and R_c coefficients of the systems at these particle sizes are virtually identical. Only in the short-wave region, when particle sizes become comparable with the wavelength, the transmittance of the system coated with PCs becomes slightly higher than the transmittance of system coated with partially ordered monolayers (see lines for $D = 0.1\,\mu$m in Fig. 2.33). If the size of the Al_2O_3 particles is comparable with the incident light wavelength, the differences in the spectra of systems with PCs and partially ordered monolayers become significant. They are more pronounced in the region of resonant transmission dip and reflection peak caused by ordering in a PC (see lines for $D = 0.5\,\mu$m at $0.83\,\mu\text{m} < \lambda < 0.95\,\mu$m in Fig. 2.33). Thus if the creation of antireflection coatings from the Al_2O_3 particles does not required their high ordering, then for the construction of selective reflectors and transmission filters the ordering of particles in monolayers plays a crucial role.

Note that coating the plate by different monolayers can give more uniform (less selective) spectrum. This problem arises when one creates the spectrally nonselective optical elements with small values of reflection, e.g. at optimization of the display characteristics. At the considered in Miskevich and Loiko (2014b) conditions (particle diameters $D = 0.1\,\mu$m in top and $D = 0.12\,\mu$m in bottom monolayers with $\eta = 0.5$) the specular reflection coefficient in the visible spectral range is less than 0.02 with the minimum value less than 0.003.

2.4.3.2 Structures from Spherical Silica Particles: Comparison of Theoretical and Experimental Results for Transmittance Spectra of the 3D Photonic Crystal

In this subsection we compare the results obtained in the framework of the quasicrystalline approximation and the transfer matrix method (Miskevich and Loiko 2014b) with the experimental data (Bogomolov et al. 1997) for the 3D PhC – close-packed structure of spherical SiO_2 particles (artificial opal) – placed in different media (air, methanol, ethanol, cyclohexane, and toluene). The structures are considered as multilayers from planar crystals with triangular lattice. The PBGs in the spectra of these multilayers are observed. Their spectral positions depend on the refractive index of the surrounding (host) medium.

The calculated spectral dependences of direct transmittance of the layered close-packed structure of spherical SiO_2 particles in different host media, as well as the calculated and experimental dependences of the spectral position of PBG minimum of this structure on the refractive index n_{env} of the environment (surrounding medium), are presented in Fig. 2.34. For clearer presentation, the number of monolayers is chosen so that the values of PBG minima were approximately equal to the value of PBG minimum for a structure consisting of $N_{ml} = 50$ monolayers of SiO_2 particles in air (Fig. 2.34a, b). Note that the numbers N_{ml} of monolayers required to achieve the said transmission minimum differ significantly. This is due to the differences in the relative refractive index of the particles in different surrounding media. In Fig. 2.34b, the calculated values of the transmittance minimum spectral position as a function of the refractive index of the surrounding medium are indicated by the circles, the experimental values (Bogomolov et al. 1997) are indicated by the squares.

As follows from Fig. 2.34b, experimental and theoretical results are in good agreement. It means that the approach works well and yields an opportunity to determine the 3D PhC structure parameters by the PBG data. The experimental and theoretical data of work (Bogomolov et al. 1997) are shown in Fig. 2.34c. In Bogomolov et al. (1997), to simulate the individual monolayers the model of partially ordered monolayer was used. One can see that the calculation dependence possesses dramatically less agreement with the experimental data than the one shown in Fig. 2.34b which is based on the model of planar crystal with imperfect lattice. This means that the model of the system of partially ordered monolayers cannot be used to obtain the quantitative results for considered type of 3D photonic crystals.

2.4.4 Transmittance and Reflectance of Periodic, Quasiperiodic, and Aperiodic Sequences of Particulate Monolayers

Interest in study of the non-periodic structures is caused by the addition possibilities of manipulating the characteristics of transmitted and reflected light as compared to

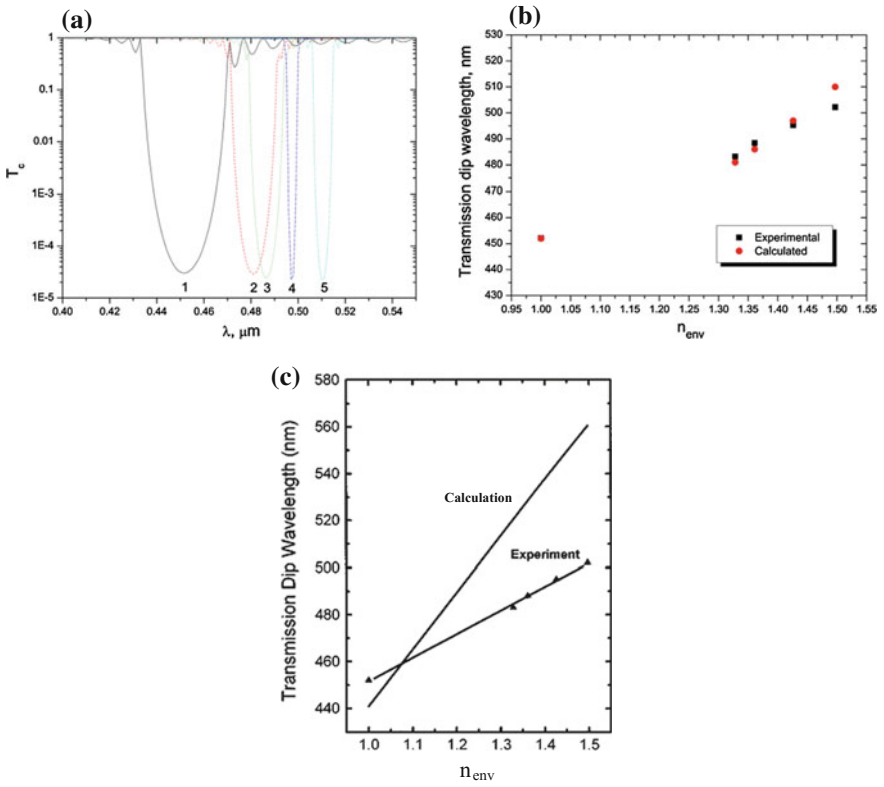

Fig. 2.34 **a** Calculated spectra of the direct transmittance (in the PBG region) of the layered regular structure consisting of monolayers with triangular lattice from spherical SiO_2 particles in different media (Miskevich and Loiko 2014b). (1) Refractive index of the environment $n_{env} = 1.0$ (air, the number of monolayers $N_{ml} = 50$), (2) $n_{env} = 1.328$ (methanol, $N_{ml} = 100$), (3) $n_{env} = 1.361$ (ethanol, $N_{ml} = 135$), (4) $n_{env} = 1.426$ (cyclohexane, $N_{ml} = 500$), (5) $n_{env} = 1.497$ (toluene, $N_{ml} = 250$). **b** Spectral dependence of the PBG minima on the refractive index of the surrounding medium: circles indicates calculation results (Miskevich and Loiko 2014b), squares indicate experimental data (Bogomolov et al. 1997). Refractive index of SiO_2 particles $n_{SiO_2} = 1.45$. $D = 0.2\,\mu m$, $\eta = 0.9$, $\sigma(u) = \sigma_0 u, \sigma_0 = 0.001$, intermonolayer spacings $s = 0.173\,\mu m$. **c** Experimental and theoretical data of Bogomolov et al. (1997)

the periodic ones (Joannopoulos et al. 2008; Levine and Steinhardt 1984; Huang et al. 2010b; Mouldi and Kanzari 2012; Rockstuhl et al. 2007; Boriskina et al. 2008). When layered structures are studied, the sequences of different homogenous layers are traditionally considered (Katsidis and Siapkas 2002; Centurioni 2005; Troparevsky et al. 2010; Born and Wolf 2002).

Consider spectra of the direct transmission T_c and specular reflection R_c coefficients of multilayers periodic, Fibonacci (quasiperiodic), and Thue-Morse (aperiodic) sequences of plane-parallel monolayers of monodisperse spherical dielectric particles (Miskevich and Loiko 2013b).

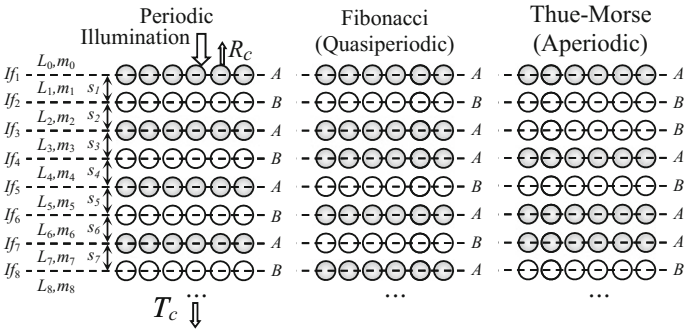

Fig. 2.35 Schematic representation of the layered structures (multilayers) consisting of the periodic, Fibonacci (quasiperiodic), and Thue-Morse (aperiodic) sequences of particulate monolayers of A and B types (side view). Normal illumination. Dashed lines depict the monolayer planes. T_c and R_c are the direct transmission and specular reflection coefficients of multilayers, respectively. If_i are numbers of interfaces (monolayers); L_i and m_i are the numbers and refractive indices of the layers (spaces between the monolayers), s_i are the spacings between the monolayers

The schematic representation of such structures is given in Fig. 2.35. The one-dimensional (1D) Fibonacci sequence L_n of symbols A and B indicating sequence constituents with different properties is built according to the inflation rule $L_n = L_{n-1}L_{n-2}$ (for $n \geq 2$, where n is the number of the sequence element), beginning from $L_0 = A$ and $L_1 = AB$: $L_0 = A, L_1 = AB, L_2 = ABA, L_3 = ABAAB, L_4 = ABAABABA, \ldots$. The 1D Thue-Morse sequence L_n of symbols A and B is built according to the rule $L_n = L_{n-1}\underline{L}_{n-1}$ (for $n \geq 1$, where $\underline{L}_{n-1}$ element is "inverted" L_{n-1} element), beginning from $L_0 = A$: $L_0 = A$, $L_1 = AB$, $L_2 = ABBA$, $L_3 = ABBABAAB$, $L_4 = ABBABAABBAABABBA$, etc.

To calculate the T_c and R_c coefficients of such multilayers, first, the amplitude direct transmission and specular reflection coefficients of the individual monolayers were computed in the QCA. Second, using these coefficients the direct transmittance and specular reflectance of the multilayer was found under the TMM.

Figure 2.36 depicts spectra of the direct transmission and specular reflection coefficients of the periodic, Fibonacci, and Thue-Morse sequences of monolayers of alumina (Al_2O_3) and silica (SiO_2) particles which are widely used in the photonics and optoelectronics. The refractive indices $n(\lambda)$ (λ is the wavelength of the incident light) of silica and alumina are taken from Palik (1985, 1991). The individual monolayers have the imperfect triangular lattice of the monodisperse spherical Al_2O_3 and SiO_2 particles with diameter $D = 0.3\,\mu m$ and the monolayer filling factor $\eta = 0.9$. The spacing s (see Fig. 2.35) between the adjacent monolayers is $0.3\,\mu m$. These parameters correspond to the three-dimensional (3D) ordered particulate structure which is similar to the structure of the colloidal crystal. The symbols A and B denote the monolayer of Al_2O_3 and SiO_2 particles, respectively.

As follows from the calculation results, with the number of monolayers increasing, the photonic band gaps appear in spectra of all considered sequences. In the spectrum

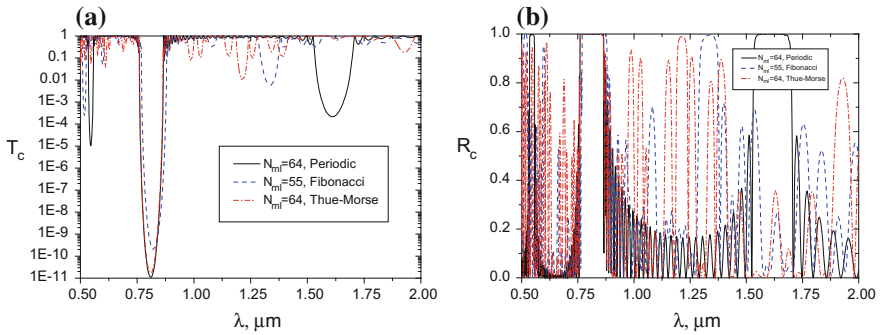

Fig. 2.36 Spectra of direct transmission T_c (**a**) and specular reflection R_c (**b**) coefficients of the multilayer consisting of periodic, Fibonacci, and Thue-Morse sequences of monolayers with a triangular lattice of monodisperse spherical Al_2O_3 (type A) and SiO_2 (type B) particles. $D = 0.3\,\mu$m, $\eta = 0.9$, $\sigma(u) = 0.001u$. The number N_{ml} of monolayers is shown in legend. $s = 0.3\,\mu$m

of the periodic sequence the one "main" and two "secondary" PBGs occur. They are caused by the periodicity of the alternating monolayers of A and B types. The "main" PBG is observed in the wavelength range from 0.76 to 0.86 μm. The spectra of multilayers with Fibonacci and Thue-Morse sequences are more complicated. The "main" PBGs of their spectra occur practically in the same spectral range as for the regular sequence, and the number of PBGs increases with the number of monolayers. This provides additional opportunities to manipulate spectrum of the multilayer in comparison with the regular one.

The PBG in the transmittance spectrum of the system consisting of two different types (A and B) of monolayers is located in the wavelength range between the PBGs of the system consisting of only A type and of the system consisting of only B type monolayers. The PBGs of the particulate structures are shifted in the short-wave range relative to those for the structures consisting of the homogeneous layers with the equivalent material volume. The PBGs observed in the spectra of sequences of the homogeneous layers typically are wider and deeper than the ones for the particulate systems. Increasing the particle concentration in the systems of monolayers and plate thicknesses in the system of homogeneous layers leads to growing the depth, width, and wavelength of the PBG minima (Miskevich and Loiko 2013b).

2.4.5 *Inverse Problem Solution for 3D Photonic Crystal: Retrieval of the Particle Refractive Index from the Photonic Band Gap Data*

The QCA- and TMM-based approach to find the direct transmission and specular reflection coefficients of ordered 3D particulate structures (see previous sections) was used to develope the method of determining (retrieving) the structure characteristics

of three-dimensional photonic crystal using the data on the PBG in its transmission spectrum (Miskevich and Loiko 2014c, 2015b).

Figure 2.37 shows the calculated spectral dependences of direct transmission coefficient $T_c(\lambda)$ of a multilayer for various refractive indices n_p and diameters D of particles, filling factors η and types of spatial particle arrangement of monolayers, spacings s between adjacent monolayers, the number of monolayers N_{ml}, and degrees of their ordering (determined by parameters of the blurring function (2.3.13)) (Miskevich and Loiko 2015b).

As seen from the figure, all considered characteristics of system influence on the PBG. But the effect is different. For example, changing in the number of monolayers leads to changing in the PBG depth, although the PBG position is not practically changed (see Fig. 2.37f). For the other considered cases are changed both the depth and position of the PBG. The above features of transmittance spectra transformation underlie the developed method of retrieval of the ordered particulate system characteristics.

The inverse problem solution was considered by an example of the particle refractive index retrieval. This problem arises, for example, when one creates colloidal crystals, synthetic opals (Bogomolov et al. 1997; Vasnetsov et al. 2014; Rojas-Ochoa et al. 2004; Galisteo-Lopez et al. 2011) and studies some biological objects (Pouya et al. 2011; Deparis et al. 2006; Campos-Fernández et al. 2011). Let characteristics of the ordered particulate structure, refractive index n_{env} of medium where the structure is situated, wavelength λ_{PBG} of the PBG minimum are known, but the refractive index of particles n_p is unknown. To find n_p we will change it in the range of expected values and solve the direct problem for each of them. We will suppose some $n_{p,\text{ret},\lambda_{\text{PBG}}}$ as a retrieved value, when at wavelength λ_{PBG} of the PBG minimum the calculated PBG minimum occurs. We designate such a method of the refractive index finding as a retrieval "by the spectral position of the PBG minimum" (way 1). If we know both the spectral position λ_{PBG} and the value T_{PBG} of the PBG minimum, then $n_{p,\text{ret},T_{\text{PBG}}}$ will be believed to be a retrieved value, if the calculated PBG minimum $T_{\text{PBG,ret}}$ and the T_{PBG} are maximally close to each other. Such a method was named as a retrieval "by the value of the PBG minimum" (way 2). It provides more accurate retrieval results (Miskevich and Loiko 2014c, 2015b). However, note that it is difficult to measure the T_{PBG} value with sufficient accuracy, when the transmittance is small.

The described techniques allow one to retrieve the refractive index of particles at the wavelength of the PBG minimum. To find refractive index at the other wavelengths we should change spectral position of the PBG. In general, this can be attained by changing the structure parameters. For example, increasing the spacing between the monolayers leads to increasing the number of PBGs. Therefore, a single transmittance measurement can provide the "multi-PBG" spectrum. However, actual structures are generally unable to change the geometrical parameters. Another technique is to change the relative refractive index of particles by filling the interparticle voids by a media with the known refractive index spectra. As a result, it is possible to obtain a number of $\lambda_{\text{PBG},i}$ and $T_{\text{PBG},i}$ (where i is number of the medium). Both the techniques were considered in Miskevich and Loiko (2015b).

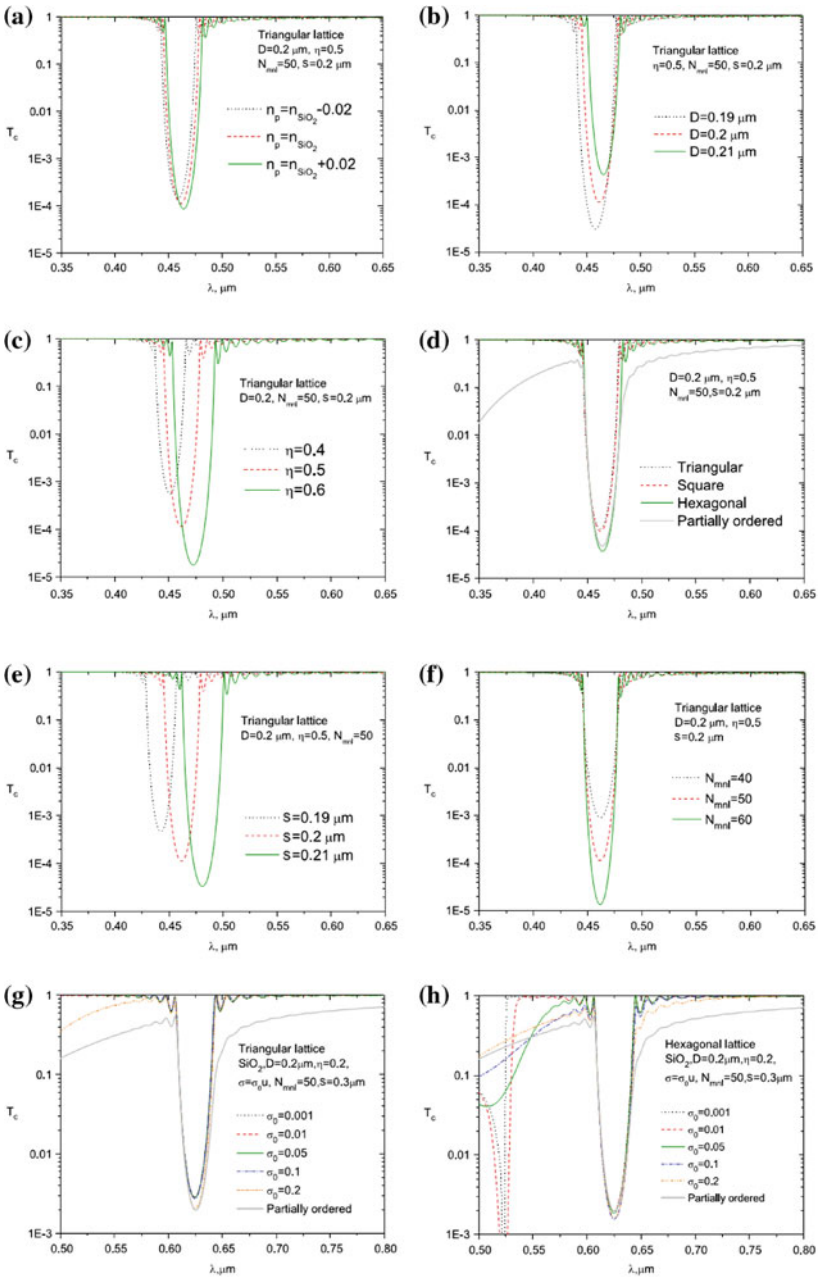

Fig. 2.37 Spectral dependences of direct transmission coefficient $T_c(\lambda)$ of multilayer at different refractive indices n_p (**a**) and diameters D (**b**) of particles; filling factors η (**c**) and types of spatial arrangement (**d**) of individual monolayers; spacings $s$between adjacent monolayers (**e**); the number of monolayers N_{ml} (**f**); and the degree of ordering of monolayers with triangular (**g**) and hexagonal (**h**) lattices (Miskevich and Loiko 2015b). **a**: $n_p = \mathrm{n}_{\mathrm{SiO}_2}$ (dashed line), $n_p = n_{\mathrm{SiO}_2} - 0.02$ (dotted line), $n_p = n_{\mathrm{SiO}_2} + 0.02$ (solid line); **b**–**h**: $n_p = n_{\mathrm{SiO}_2}$. Blurring function $\sigma(u) = 0.01u$ (**a**)–(**f**)

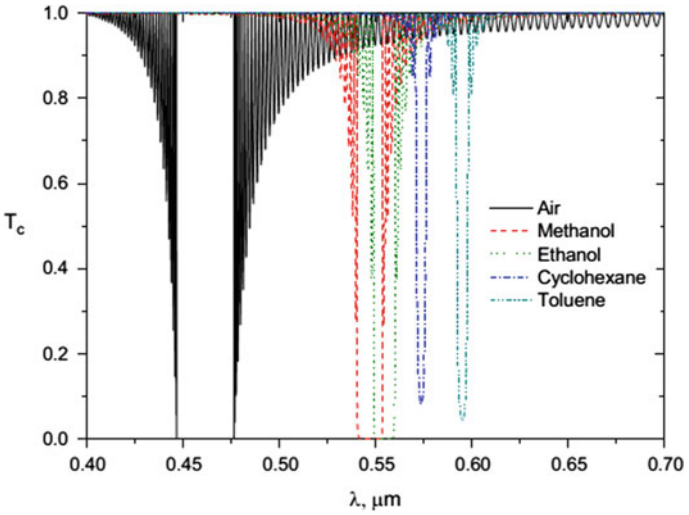

Fig. 2.38 Spectra of direct transmission T_c coefficient of multilayer in air, methanol, ethanol, cyclohexane, and toluene (Miskevich and Loiko 2015b). Individual monolayers have imperfect triangular lattice from SiO_2 particles with diameter $D = 0.2\,\mu m$ and filling factor $\eta = 0.5$. Blurring function $\sigma(u) = 0.01u$. The number of monolayers $N_{ml} = 200$, spacing between adjacent monolayers $s = 0.2\,\mu m$. Spectral position of the PBG minimum $\lambda_{PBG} = 0.4613\,\mu m$ (in air); $\lambda_{PBG} = 0.547\,\mu m$ (in methanol); $\lambda_{PBG} = 0.5546\,\mu m$ (in ethanol); $\lambda_{PBG} = 0.5738\,\mu m$ (in cyclohexane); $\lambda_{PBG} = 0.5951\,\mu m$ (in toluene)

Figure 2.38 shows the calculation results for direct transmission coefficient spectra of a multilayer in air, methanol, ethanol, cyclohexane, and toluene. Refractive index of air was taken to be unity. Refractive indices of other substances were taken from Debenham and Dew (1981), Refractive Index (2017).

Using the λ_{PBG} and T_{PBG} data (see Fig. 2.38), the refractive index of the particles was retrieved at wavelengths of the PBG minima. Actual n_p and retrieved $n_{p,ret}$ values for SiO_2 particles are shown in Fig. 2.39a. Figure 2.39b illustrates the relative error of the retrieval calculated by the equation

$$\varepsilon_{rel} = \frac{(n_{p,ret} - n_p)}{n_p} \times 100\%, \tag{2.4.13}$$

where n_p and $n_{p,ret}$ are actual and retrieved refractive indices of particles, respectively. As one can see the retrieved values of the refractive index obtained with using the simulated PBG data are in good agreement with the actual ones ($|\varepsilon_{rel,\lambda_{PBG}}| < 0.052\%$, $|\varepsilon_{rel,T_{PBG}}| < 0.004\%$). A better determination of the refractive index occurs when one uses the retrieval by the value of the PBG minimum (way 2). Note that this way is inapplicable, when one cannot provide a high enough precision of the determination of the transmittance value at wavelength of the PBG minimum.

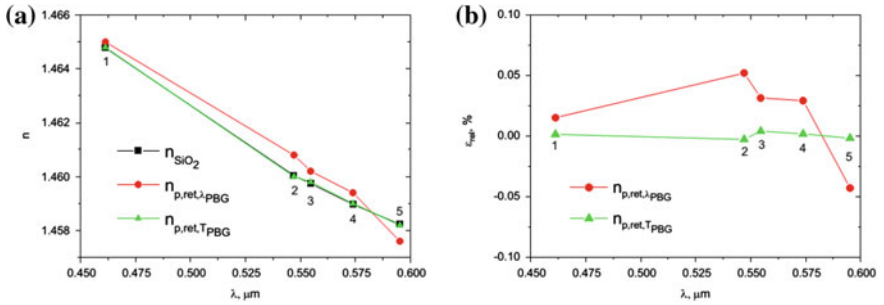

Fig. 2.39 **a** Spectral dependences of the refractive index of silica (n_{SiO_2}) and values of the refractive index of SiO_2 particles $n_{p,\mathrm{ret},\lambda_{PBG}}$ and $n_{p,\mathrm{ret},T_{PBG}}$ retrieved by the spectral position λ_{PBG} and by value T_{PBG} of the PBG minima, respectively. The system consists of 200 monolayers ($N_{ml} = 200$) with the imperfect triangular lattice, $\sigma(u) = 0.01u$, $D = 0.2\,\mu\mathrm{m}$, $\eta = 0.5$, $s = 0.2\,\mu\mathrm{m}$. Digits near the symbols indicate number of the PBG minimum in air (1); methanol (2); ethanol (3); cyclohexane (4); toluene (5). **b** Relative error of the retrieval (Miskevich and Loiko 2015b)

The method was applied (Miskevich and Loiko 2014c) to retrieve the refractive indices of the synthetic opal particles from experimental data (Vasnetsov et al. 2014) on the PBG in its transmittance spectrum. The sample was placed into various environments (liquids) to change spectral position and value of the PBG minimum. Figure 2.40a presents the experimental (Vasnetsov et al. 2014) and calculated dependences of the spectral position (λ_{PBG}) of the PBG minimum for a synthetic opal on the refractive index n_{env} of the environment. Five calculated dependences $\lambda_{PBG}(n_{env})$ are shown for different particle refractive indices n_p. The upper and lower dependences in Fig. 2.40a were obtained at $n_p = 1.4528$ and 1.33, respectively. In addition to those mentioned above, Fig. 2.40a presents two more dependences $\lambda_{PBG}(n_{env})$ obtained for structures of particles with the refractive indices $n_p = 1.4$ and 1.382 (Vasnetsov et al. 2014). The results were calculated for n_{env} from Vasnetsov et al. (2014). They are designated by symbols in Fig. 2.40a. It can be seen that the $\lambda_{PBG}(n_{env})$ dependence is linear at constant n_p. Accordingly, if the sample structure is assumed to be the same in different experiments, then it should be concluded that the effective refractive index of the particles is changed. Figure 2.40b shows the retrieved values of the effective refractive index of particles. They were obtained from the spectral positions of the PBG minima in the experimental dependence $\lambda_{PBG}(n_{env})$ (Vasnetsov et al. 2014) (Fig. 2.40a).

Figure 2.40b shows also the refractive indices n_{env} of the liquids and the spectral dependence of the SiO_2 refractive index (Palik 1985). We see that all of the retrieved values are less than n_{SiO_2}. For experiments 1–4 and 6–8, the effective refractive indices of the particles are larger and smaller than the refractive indices of the liquids, respectively. In experiments 5 and 6, they are close to the refractive indices of the liquids. These results are consistent with the estimation of the relative refractive index of the particles made in Vasnetsov et al. (2014).

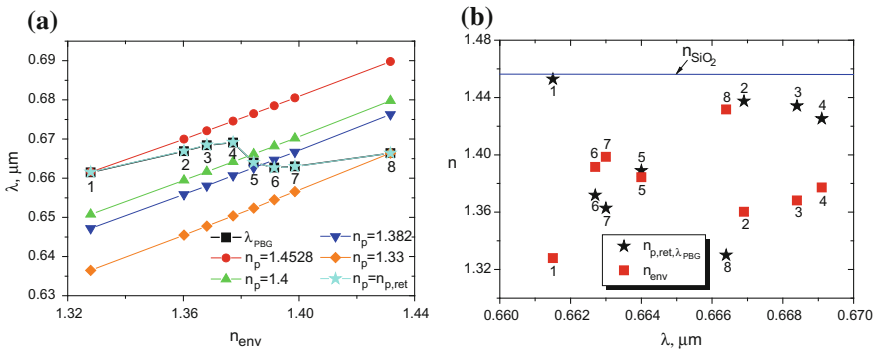

Fig. 2.40 Dependences of $\lambda_{\mathrm{PBG}}(n_{env})$, $n_{p,\mathrm{ret},\lambda_{\mathrm{PBG}}}(\lambda_{\mathrm{PBG}})$, $n_{env}(\lambda_{\mathrm{PBG}})$ and $n_{\mathrm{SiO_2}}(\lambda)$. **a** $\lambda_{\mathrm{PBG}}(n_{env})$, experiment (Vasnetsov et al. 2014) (squares), calculation (other symbols) (Miskevich and Loiko 2014c); **b** $n_{p,\mathrm{ret},\lambda_{\mathrm{PBG}}}$ (stars) (Miskevich and Loiko 2014c), n_{env} (Vasnetsov et al. 2014) (squares) and $n_{\mathrm{SiO_2}}$ spectrum (solid line). $D = 0.3\,\mu\mathrm{m}$, $\eta = 0.68$, triangular latice, $\sigma(u) = 0.01u$. $N_{ml} = 200$, $s = 0.2395\,\mu\mathrm{m}$. The digits designate experiment numbers (Miskevich and Loiko 2014c)

As can be seen from the data in Fig. 2.40b, the effective refractive index of the particles depends on the refractive index of the immersion liquid. This may be due to porous structure of particles, that is consistent with the experimental data (Bogomolov et al. 1995, 1997). When the pores are partially filled with the liquid, the effective refractive index of the particles changes. As can be seen from Fig. 2.40b, it decreases with n_{env} increasing. This gives reason to conclude that the open and closed pores are distributed nonuniformly over the particle volume. In the simplest case, a particle of the synthetic opal under consideration can be represented as a two-layered one (Babenko et al. 2003; Berdnik and Loiko 2011) consisting of a core with closed pores and a shell containing open pores. When the open pores are filled with the liquid, the effective refractive index of the shell changes. The effective refractive index of the core does not change. The effective particle refractive index retrieved within the model used is a characteristic that determines the position of the photonic band gap for the structure.

2.4.6 *Transmittance and Angular Distribution of Scattered Light: Some Experimental Data*

Consider the experimental results for thick (3D) partially ordered layers. The coherent (directly transmitted and specularly reflected) and incoherent (scattered at all other directions) fields as they are determined in Ishimaru (1978b), Ivanov et al. (1988) are investigated in Loiko et al. (1999). The coherent part characterizes the directional flux and, as a consequence, the extinction coefficient of the layer. The incoherent part determines the angular structure of the scattered light.

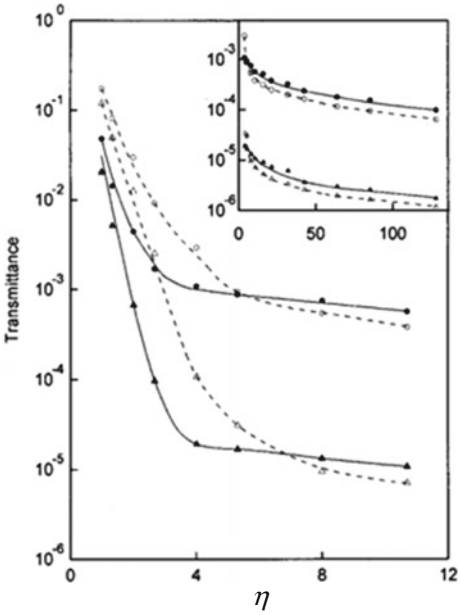

Fig. 2.41 Transmittance versus the filling factor η for the slab of latex particles with diameter $D = 3.75\,\mu\text{m}$ in water at $\lambda = 650\,\text{nm}$ for several values of volume filling factor η_V and FOV: $\eta_V < 0.006$, FOV $= 4.6°$ (open circles); $\eta_V = 0.4$, FOV $= 4.6°$ (filled circles); $\eta_V < 0.006$, FOV $= 35$ arc min (open triangles); and $\eta_V = 0.4$, FOV $= 35$ arc min (filled triangles) (Loiko et al. 1999)

At the measurements of forwardly transmitted light not only the coherent component but also a portion of the incoherent one is registered. This is because of the finite aperture of the receiver. For thin layers this contribution is small as compared with the coherent component, but generally it is not so and we are to pay attention that the instrumental transmittance T is measured at a given field of view (FOV). The results of measurements for layers with various surface η and volume η_V filling factors are shown in Fig. 2.41. Remind that surface filling factor η is the ratio of the projection area of all particles to the area over which they are distributed. For monolayers (2D systems) it is always no larger than unity. For thick (3D) layers the value of η can be greater than unity. It shows how many times the particle projections cover the surface.

As one can see from the figure, the dependences for the densely packed and the rarefied media are changed in the similar manner. In particular, at small η, the $T(\eta)$ dependences for both cases ($\eta_V = 0.4$ and $\eta_V < 0.006$) are exponential. With η increasing, the amount of scattered light increases, and the exponential dependence is disturbed. At large filling factors, the slope of curves becomes constant and, independently of the FOV, the so-called asymptotic regime is realized (Zege et al. 1991a).

The measured angular distributions of intensity of light scattered by the rarefied, $I_r(\theta)$, and densely packed, $I_d(\theta)$, layers of latex particles with different η are given in Fig. 2.42. In this figure the value of a division of vertical axis corresponds to a change in intensity by 1 order of magnitude. Note that a quantitative comparison of

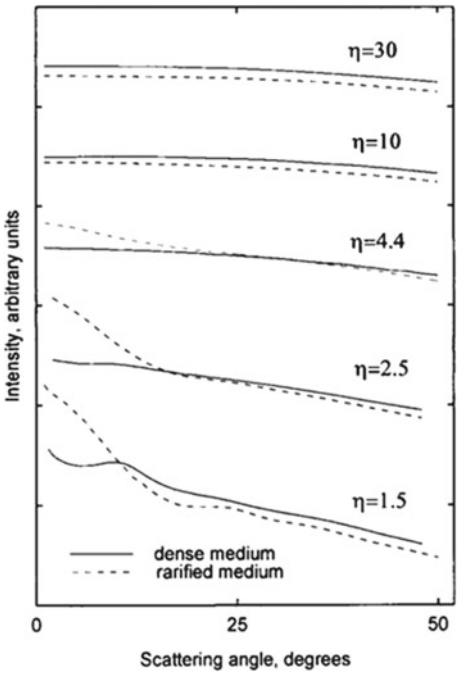

Fig. 2.42 Angular distribution of intensity of light scattered by the slab of latex particles ($D = 3.75\,\mu$m) in water for various values of η at $\lambda = 650$ nm. $\eta_V = 0.4$ (solid curves); $\eta_V < 0.006$ (dashed curves) (Loiko et al. 1999)

the curves of $I_r(\theta)$ and $I_d(\theta)$ makes sense only at the same η. At different η, only a qualitative comparison is valid.

It can see from Fig. 2.42 that, with η increasing the angular distribution of light scattered by densely packed layers changes in the same manner as the one for the rarefied layers. In particular, an increase in the contribution of multiply scattered light with η increasing leads to the smoothing of $I_d(\theta)$. For layers with $\eta \geq 10$ the angular structure becomes stable. Let us compare the $I_r(\theta)$ and $I_d(\theta)$ dependences at equal η to elucidate influence of the particle concentration on the angular distribution of scattered light. From Fig. 2.42 it follows that the effect of a change in the particle concentration on the angular distribution is the stronger, the smaller η. At $\eta = 1.5$, increasing the particle concentration leads to decreasing the intensity of light scattered in angles $\theta \leq \theta_0 = 10^\circ$ and to increasing for $\theta > \theta_0$. The shapes of the $I_r(\theta)$ and $I_d(\theta)$ dependences practically coincide at $\theta > \theta_0$.

The increase in the scattered light intensity for angles $\theta \leq \theta_0$ is due to the interference. Therefore θ_0 corresponds approximately to the minimum value of the angle at which structure factor for the 3D system (Ivanov et al. 1988; Ziman 1979) $S_3(\theta) = 1$. It can be determined by the expression: $\sin(\theta_0/2) \approx 1.5/x$.

An increase in η and, consequently, in the portion of multiply scattered light leads to the smoothing the angular distributions of the densely packed and the rarefied media. In the asymptotic regime (Zege et al. 1991a) the shapes of the angular distributions for the dense and the rarefied media are the same. For dense packing, however,

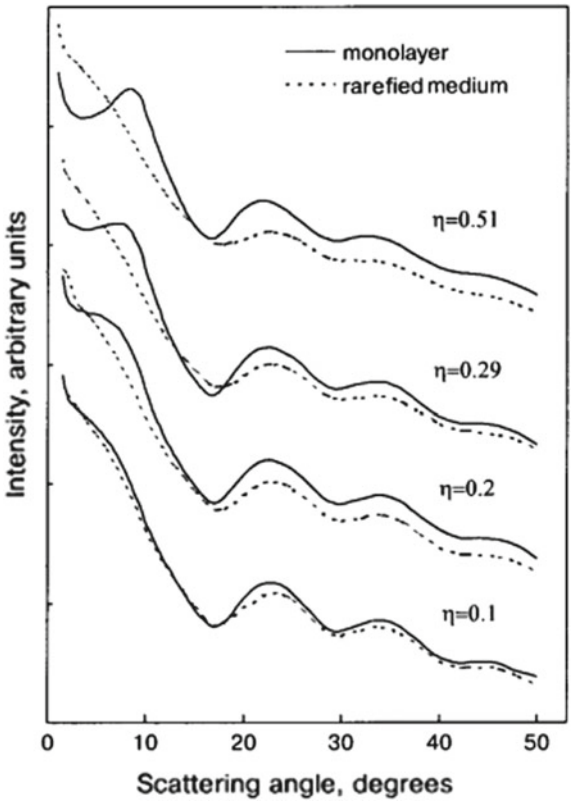

Fig. 2.43 Measured angular distributions of scattered light intensity for layers of latex particles in water at various values of filling factor η. The solid lines correspond to the monolayer, the dashed lines correspond to the layer where the distance between particles is much more than their size. Normal illumination. The incident light wavelength $\lambda = 0.650\,\mu$m. Particle diameter $D = 3.75\,\mu$m. Absolute refractive index of the particles $n = 1.580$ at wavelength $\lambda = 0.650\,\mu$m. The scale factor for the vertical axis corresponds to the change in intensity by one order of magnitude (Dick et al. 1997a)

the intensity of scattered light is higher. With increasing the number of particles the changes in transmittance for the densely packed and the rarefied media are the same.

Figure 2.43 shows experimental data on the angular distributions of scattered light intensities for rarefied media ($I_r(\theta)$) and monolayers ($I(\theta)$) of latex particles in water at various η (Dick et al. 1997a). At scattering angles $\theta < 10°$ the additional (in comparison with the rarefied layer) maxima is appeared. This can be used to determine size of the prticles (Loiko et al. 1984). Quantitative comparison of $I(\theta)$ and $I_r(\theta)$ curves of Fig. 2.43 makes sense only at equal values of η. The value of vertical axis division corresponds to the change in intensity by one order of magnitude. The shift in position of curves corresponding to various η is used only for convenience of presentation.

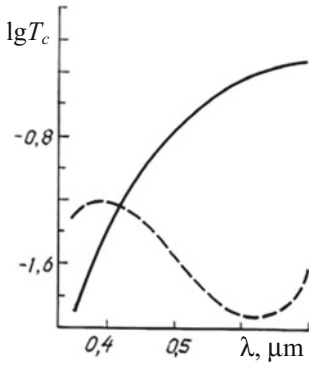

Fig. 2.44 Dependence of direct transmittance on wavelength for layers of cubic AgBr particles with the cube edge 0.49 μm in gelatin; solid line corresponds to monolayer with mean interparticle distance 0.61 μm. Dashed line corresponds to multilayer with mean interparticle distance 1.18 μm. The surface concentrations in both cases is the same (filling factor $\eta = 0.66$) (Berry 1971)

In Fig. 2.44 are shown experimental data for layers of cubic AgBr particles with the cube edge 0.49 μm in gelatin (Berry 1971). It is obvious the difference in spectra of the monolayer with mean interparticle distance 0.61 μm and the 3D layer with the same surface concentration and mean interparticle distance 1.18 μm.

2.5 Light Absorption by the Particulate Crystalline Silicon: Application to the Solar Cell Enhancement

The photovoltaic (solar) cells are the object of intensive theoretical and experimental investigations. That is caused by increasing importance of "green", renewable, and non-fossil energy sources. However, relatively low efficiency of solar cells is one of the restrictive factors for wide usage of such sources (Luque and Hegedus 2011; Sze and Ng 2007).

The significant problem in enhancement of the solar cell (SC) performance is optimization of its structure to trap more amount of the incident light. To attain this end the plasmonic and diffractive nanostructures, down-converting particles, surface texturing, nanohole patterning etc. are widely studied (Domínguez et al. 2012; Tsai et al. 2011; Sheng et al. 2011; Kocher-Oberlehner et al. 2012; Schuster et al. 2015; Abrams et al. 2011; Deinega et al. 2011; Wehrspohn and Üpping 2012; Ji and Varadan 2011; Tsakalakos 2008; Kayes et al. 2005; Vynck et al. 2012).

Currently the most-used material for solar cell production is silicon. Because it is a nondirect gap semiconductor, the probability of an electron transition from valence to conduction band due to photon absorption is small (Luque and Hegedus 2011; Sze and Ng 2007). If we deal with a homogeneous active layer of SC, we need to increase the layer thickness to increase light absorption. That can reduce the efficiency

of solar cell due to the diffusion length of the minority carriers (Saritas and McKell 1988; Toušek et al. 2003; Sharma et al. 2012) limitation. Only carriers produced in the space-charge (depletion) region and adjacent areas, that are determined by the diffusion length of the minority carriers, are separated by the electric field of the *p-n* junction and, thus, contribute in photoelectromotive force (photo-emf) generation (Luque and Hegedus 2011; Gremenok et al. 2007).

The amount and efficiency of light absorption can be increased using the particulate structure of an active layer (Miskevich and Loiko 2014a, 2015a). Such a layer consists of plane-parallel monolayers of silicon particles with diameter of the diffusion length order. That enables one to decrease reflection and simultaneously increase light absorption in comparison with the homogeneous plane-parallel silicon plate (Miskevich and Loiko 2014a, 2015a). Silicon particles of such sizes and the layers of these particles can be fabricated by chemical (Shi et al. 2012, 2013), lithographic (Staude et al. 2013) methods, plasma synthesis (Bapat et al. 2004), laser ablation in air or liquids (Barcikowski et al. 2007; Li et al. 2013), laser induced pyrolysis (Vladimirov et al. 2011) and the recently proposed method based on the laser-induced transfer of molten droplets (Zywietz et al. 2014a, b).

In crystalline silicon (c-Si) the light absorption in the long-wave part of visible and near infrared spectral regions is small while in the ultraviolet region and at the short-wave part of visible spectrum it is large. An important problem for c-Si solar cells is to enhance light absorption in the wavelength range of small imaginary part κ (absorption index) of the complex refractive index $m = n + i\kappa$ of semiconductor. Absorption by the particulate structure can be increased in conditions of the pronounced multiple scattering of waves. Such conditions occur when the particle sizes and distances between them are comparable with wavelengths of the incident light. Therefore, as one can expect, the systems of submicron and micrometer-sized particles should satisfy the mentioned conditions. Moreover, individual submicron spherical silicon particles exhibit strong peaks of scattering efficiency in the spectral range of interest due to optical resonances (also known as Mie resonances) (van de Hulst 1957; Mie 1908; Rosasco and Bennett 1978; Conwell et al. 1984; Zender and Talamantes 2006; Evlyukhin et al. 2010; Bachelard et al. 2012; Wang et al. 2015). This can be an important factor of overall absorption enhancement.

2.5.1 Spectral and Integral Absorption Coefficients of the Particulate Monolayer

To calculate absorption coefficient A_{ml} of the monolayer, the following approach is used. Initially the conditions wherein only directly transmitted and specularly reflected light exist (i.e. $T_c + R_c$ is unity) are found. To do that a monolayer of "model" nonabsorbing particles, in our case with the refractive index of crystalline silicon $n = n_{\text{c-Si}}$ and absorption index $\kappa = 0$, is considered. Spectra of direct transmission T_c and specular reflection R_c coefficients of the normally illuminated

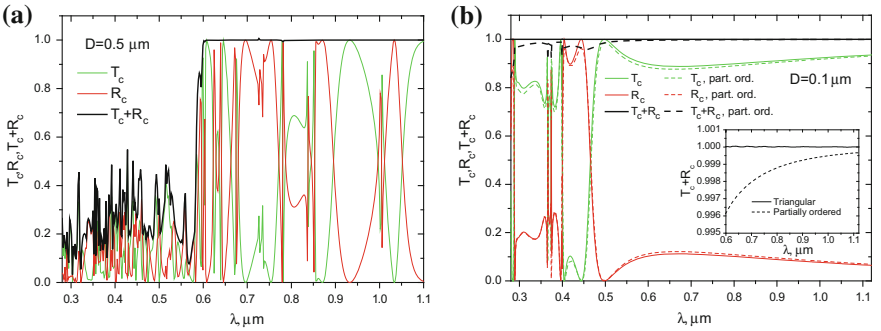

Fig. 2.45 Calculated spectra of direct transmission T_c, specular reflection R_c coefficients and $T_c + R_c$ sum of monolayer with triangular lattice (solid lines) and partially ordered monolayer (dashed lines, (**b**)) of spherical particles at particle diameters $D = 0.5\,\mu\text{m}$ (**a**) and $D = 0.1\,\mu\text{m}$ (**b**) (Miskevich and Loiko 2015a). The inset in **b** shows the large scale of $T_c + R_c$ sums in the spectral range from 0.6 to $1.12\,\mu\text{m}$. Refractive index of particles is $n = n_{c-Si}$, $\kappa = 0$. Monolayer filling factor $\eta = 0.5$. Blurring function $\sigma(u) = 0.01u$, correlation length $l_c = 85$

monolayer (planar crystal) with imperfect triangular lattice from such particles are shown in Fig. 2.45 (Miskevich and Loiko 2015a). The $T_c + R_c$ sum is also depicted in this figure. The calculations were fulfilled in the QCA. The spectral range where the value of $T_c + R_c$ is unity expands to the shorter waves with the particle size decreasing and the filling factor increasing.

Effect of particle ordering on the directly transmitted and specularly reflected light is demonstrated in Fig. 2.45b. As one can see, the $T_c + R_c$ sum of the planar crystal consisting of $0.1\,\mu\text{m}$-diameter particles is practically unity over all considered spectrum ($0.28\,\mu\text{m} \le \lambda \le 1.12\,\mu\text{m}$). The $T_c + R_c$ sum of the partially ordered monolayer is less than unity in all considered spectral range. The range of $0.6\,\mu\text{m} \le \lambda \le 1.12\,\mu\text{m}$ is presented in the inset to show the difference from unity more clear. Thus the long-range ordering provides much wider wavelength range where the $T_c + R_c$ sum is unity than the short-range ordering.

Once the ranges of wavelengths, sizes, and filling factors where the $T_c + R_c = 1$ for monolayers of nonabsorbing particles ($\kappa = 0$) were found, it is possible go to the second step. For these ranges the absorption coefficient A_{ml} of monolayer of particles with the actual complex refractive index of c-Si ($\kappa > 0$) is determined by the equation:

$$A_{ml} = 1 - T_c - R_c. \tag{2.5.1}$$

It is reasonable to compare the absorption by the particulate monolayer and by the ordinary used homogeneous plane-parallel plate under condition of the equality of the silicon volume per unit surface area in the systems. The thickness h of such "equivalent plate" is associated with the particle diameter D and the monolayer filling factor η by the relationship:

$$h = \frac{2}{3}\eta D, \tag{2.5.2}$$

Absorption coefficient A_{pl} of the homogeneous plane-parallel plate is calculated by equation:

$$A_{pl} = 1 - T_{pl} - R_{pl}, \tag{2.5.3}$$

where T_{pl} and R_{pl} are transmission and reflection coefficients of the plate. They are determined with taking into account the multiple-beam interference:

$$T_{pl} = \left|t_{pl}\right|^2 = \left|\frac{t_{01}t_{12}e^{ikh}}{1 - r_{12}r_{10}e^{2ikh}}\right|^2, \tag{2.5.4}$$

$$R_{pl} = \left|r_{pl}\right|^2 = \left|r_{01} + \frac{t_{01}t_{10}t_{12}e^{2ikh}}{1 - r_{12}r_{10}e^{2ikh}}\right|^2. \tag{2.5.5}$$

Here t_{pl} and r_{pl} are the amplitude transmission and reflection coefficients of the plate, t_{01}, t_{10} and r_{01}, r_{10} are the amplitude transmission and reflection coefficients for a wave incoming (subscript "01") and outgoing (subscript "10") the plate through the top interface, t_{12} and r_{12} are the amplitude transmission and reflection coefficients for a wave outgoing the plate (subscript "12") through the bottom interface (they are calculated by the Fresnel formulae (Born and Wolf 2002)), h is the plate thickness, $k = 2\pi m/\lambda$, m is the complex refractive index of the plate material, λ is the wavelength of the incident light.

The simulation results show that absorption coefficient A_{ml} of monolayer of c-Si particles with diameters significantly smaller than the wavelength is smaller than the absorption coefficient A_{pl} of equivalent plate. The A_{ml} coefficient can increase or decrease with filling factor increasing. It grows with particle size and becomes larger than the A_{pl} when particle diameters are comparable with the wavelength (Miskevich and Loiko 2015a).

Figure 2.46a displays the calculated in the SSA (the calculation method is described in Miskevich and Loiko 2014a) and the QCA spectra of absorption coefficients of monolayers of c-Si particles at different monolayer filling factors in the wavelength range of small absorption index of material ($0.8\,\mu\text{m} < \lambda < 1.1\,\mu\text{m}$). The data for equivalent plates are shown as well.

The peak positions of A_{ml} obtained in the SSA coincide with the ones of absorption efficiensy factor Q_{abs} of a single c-Si particle (Fig. 2.46b). The multiple scattering results in the long wave shift of the peaks and in formation of the additional ones. Thus, the absorption spectra of monolayers calculated with taking into account the multiple scattering are more complicated than the ones obtained in the SSA. They are influenced by resonances caused by individual silicon particles as well as by their spatial arrangement. In Fig. 2.46b the spectrum of extinction efficiensy factor Q_{ext} for model nonabsorbing particle ($n = n_{\text{c-Si}}$, $\kappa = 0$) with $D = 0.5\,\mu\text{m}$ is shown as well. In this case $Q_{ext} = Q_{sc}$. Comparison of the spectrum with the one of the real c-Si particle shows that, decreasing the absorption index can result in increasing as well as decreasing the amount of light scattered by individual particle. In considered case the Q_{ext}-increasing occurs in the vicinity of maxima of the resonance peaks. It is worth paying attention that, although the single particle scatters light in all directions

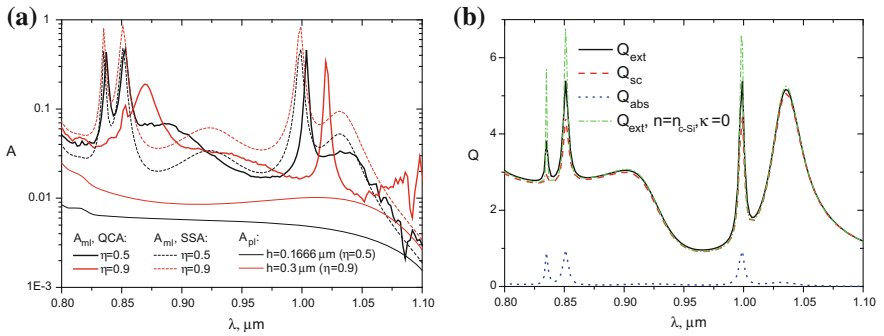

Fig. 2.46 Spectral absorption coefficients A_{ml} and A_{pl} of monolayer with triangular lattice of c-Si particles at different filling factors η and the equivalent plates with thicknesses h. $D = 0.5\,\mu\text{m}$. The filling factors of monolayers to which the plates are equivalent are indicated in parentheses in the legend. $\sigma(u) = 0.001u$, $l_c = 900$ (**a**) (Miskevich and Loiko 2015a). Spectral dependences of extinction Q_{ext}, scattering Q_{sc}, and absorption Q_{abs} efficiency factors of silicon spherical particle and Q_{ext} of the "model" nonabsorbing particle with $n = n_{\text{c-Si}}$, $\kappa = 0$ (**b**)

($0 \leq \theta \leq \pi$, θ is the polar scattering angle), the ordered monolayer of these particles scatters only in strictly forward and strictly backward directions in the considered spectral range (see Fig. 2.45a).

Calculated in the QCA spectral absorption coefficients of monolayers with filling factor $\eta = 0.5$ at different particle sizes in the wavelength range of small absorption index of silicon are presented in Fig. 2.47a (Miskevich and Loiko 2015a). The results for equivalent plates are shown as well. As can be seen from the data the number of peaks in the absorption spectra of monolayer increases with particle size. In spectral ranges of peaks the absorption efficiency significantly increases. In Fig. 2.47b are shown spectral dependences of the A_{ml}/A_{pl} ratio characterizing the difference of absorption coefficients of the monolayers and the equivalent plates. In the peaks the value of A_{ml} can be more than 100 times greater than value of A_{pl}. Therefore particulate monolayers can be used as effective light absorbers in the narrow wavelength intervals in the vicinities of the peak maxima. It is worth noting that absorption gain due to the peaks can be reduced when their spectral positions coincide with the terrestrial solar spectral irradiance dips caused by the absorption bands of atmospheric water vapor, aerosols, etc.

The integral over the solar illumination spectrum absorption coefficient of the particulate monolayer were examined as well (Miskevich and Loiko 2014a, 2015a).

The integral coefficient $\langle A \rangle$ is calculated by equation:

$$\langle A \rangle = \frac{\int_{\lambda_1}^{\lambda_2} A(\lambda) w(\lambda) \mathrm{d}\lambda}{\int_{\lambda_1}^{\lambda_2} w(\lambda) \mathrm{d}\lambda}, \tag{2.5.6}$$

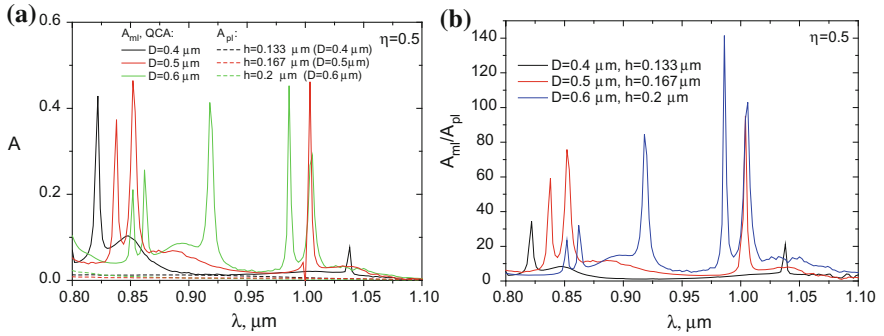

Fig. 2.47 Spectral absorption coefficients A_{ml} and A_{pl} of monolayer with triangular lattice of c-Si-particles with different diameters D and the equivalent plates with thicknesses h. The diameters of particles in monolayers to which the plates are equivalent are indicated in parentheses after the h-values (**a**) (Miskevich and Loiko 2015a). Dependences of the A_{ml}/A_{pl} ratio (**b**). $\sigma(u) = 0.001u$, $l_c = 900$. $\eta = 0.5$

where $A(\lambda)$ is a spectral absorption coefficient of a system, λ is the wavelength of incident light, $w(\lambda)$ is the power of the illumination spectrum of the source, wavelengths λ_1 and λ_2 specifies the considered spectral range.

To compare absorption by different structures, the relative absorption coefficient A_{rel} is used:

$$A_{rel} = \left(\frac{\langle A_{L_1} \rangle}{\langle A_{L_2} \rangle} - 1 \right) \times 100\%, \tag{2.5.7}$$

which characterizes difference in integral absorption coefficients of the layers under consideration. Here indices L_1 and L_2 designate the particulate system and the equivalent plane-parallel plate, respectively. The positive values of A_{rel} correspond to an increase, while the negative ones correspond to a decrease of light absorption by the particulate system in comparison with absorption by the plate.

The dependences of integral absorption coefficients of monolayer, $\langle A_{ml} \rangle$, equivalent plate, $\langle A_{pl} \rangle$, and A_{rel} on the monolayer filling factor are shown in Fig. 2.48 (Miskevich and Loiko 2015a). The calculations are made for the terrestrial solar spectral irradiance "Global tilt" ASTM G173-03 (Reference Solar Spectral Irradiance 2017) in the wavelength range from $\lambda_1 = 0.28\,\mu\text{m}$ to $\lambda_2 = 1.12\,\mu\text{m}$.

As follows from Fig. 2.48a $\langle A_{ml} \rangle$ coefficient grows with the size of particles and the monolayer filling factor. For monolayer of the 0.05 μm-diameter particles $\langle A_{ml} \rangle < \langle A_{pl} \rangle$. For other considered particle sizes $\langle A_{ml} \rangle > \langle A_{pl} \rangle$. As can be seen from Fig. 2.48b the relative absorption coefficient and, consequently, the efficiency (in comparison with the equivalent plate) of light absorption by the monolayer increase with the particle size. This result, and the other ones obtained in Miskevich and Loiko (2014a, 2015a) show that maximum of A_{rel} occurs, as one might expect, for monolayers of particles with sizes comparable with the wavelength.

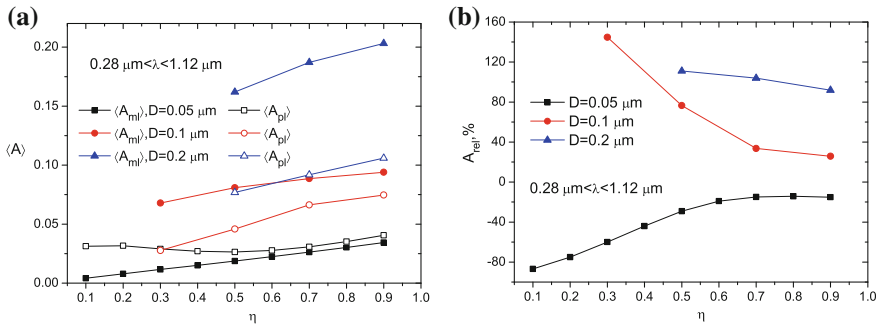

Fig. 2.48 Integral absorption coefficient $\langle A_{ml} \rangle$ of monolayer, integral absorption coefficient $\langle A_{pl} \rangle$ of the equivalent plate (**a**), and relative absorption coefficient A_{rel} (**b**) versus monolayer filling factor η (Miskevich and Loiko 2015a). Monolayers have triangular lattice, $\sigma(u) = 0.001u, l_c = 900$, $0.28\,\mu\text{m} \leq \lambda \leq 1.12\,\mu\text{m}$

2.5.2 *Absorption by Systems of Monolayers. Gradient Multilayers*

The spectral $A_{N_{ml}}$ and integral $\langle A_{N_{ml}} \rangle$ absorption coefficients of layered systems (multilayers) consisting of monolayers were also calculated and analyzed in Miskevich and Loiko (2014a, 2015a). The calculations were made under the SSA and TMM for systems of large (Miskevich and Loiko 2014a) and QCA and TMM for systems of small and wavelength-sized (Miskevich and Loiko 2015a) particles.

Figure 2.49 shows spectral absorption coefficient of the system consisting of the various number of identical monolayers of large silicon particles and the plane-parallel silicon plate with a volume equal to the particles volume of ten monolayers (Miskevich and Loiko 2014a). The absorption coefficient of a system increases with the number of monolayers. The coefficient can exceed 0.8 for large and 0.9 for moderate and small values of the absorption index of silicon. Absorption coefficient is larger for the systems of larger particles in the range of moderate and small values of the absorption index of c-Si.

Figure 2.50a shows dependences of the integral absorption coefficient $\langle A_{N_{ml}} \rangle$ of the systems of large c-Si particles with different diameters on the number N_{ml} of monolayers and the $\langle A_{pl} \rangle$ of the equivalent plane-parallel plates (Miskevich and Loiko 2014a). Dependence of $A_{rel}(N_{ml})$ is displayed in Fig. 2.50b. One can see that integral absorption coefficient of three- and more-monolayer systems is larger than the one of the equivalent plane-parallel plates for all considered particle sizes. The values of A_{rel} are larger for smaller particles. Maximum of the $A_{rel}(N_{ml})$ dependence shifts to larger N_{ml} with the particle size increasing (Fig. 2.50b). At particle diameter $D = 5\,\mu\text{m}$ the absorption by the system consisting of 6 and 7 monolayers is about 1.45 times larger than the one for the equivalent plate. However, the values of $\langle A_{N_{ml}} \rangle$ are greater for systems of larger particles (Fig. 2.50a). Pay attention that light absorption by the system increases with the particle size increasing, but gradient of

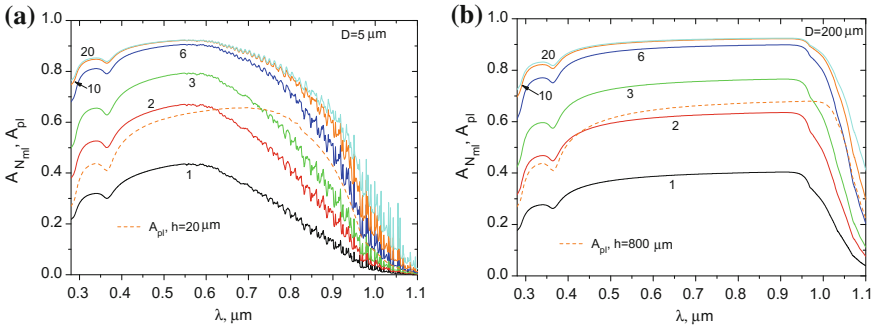

Fig. 2.49 Spectral absorption coefficient $A_{N_{ml}}$ of single monolayer of silicon particles ($N_{ml} = 1$), systems of $N_{ml} = 2$, 3, 6, 10, and 20 identical monolayers (calculations under the SSA and TMM (Miskevich and Loiko 2014a)) and A_{pl} of silicon plane-parallel plate (Miskevich and Loiko 2014a). The numbers near the curves indicate the number N_{ml} of monolayers. The filling factor of monolayers $\eta = 0.6$. Particle diameter $D = 5\,\mu\text{m}$ (**a**), $D = 200\,\mu\text{m}$ (**b**). Dashed line: spectral absorption coefficient A_{pl} of the homogeneous plane-parallel plate with a volume equal to the volume of particles in 10 monolayers (plate thickness $h = 20\,\mu\text{m}$ (**a**), $h = 800\,\mu\text{m}$ (**b**))

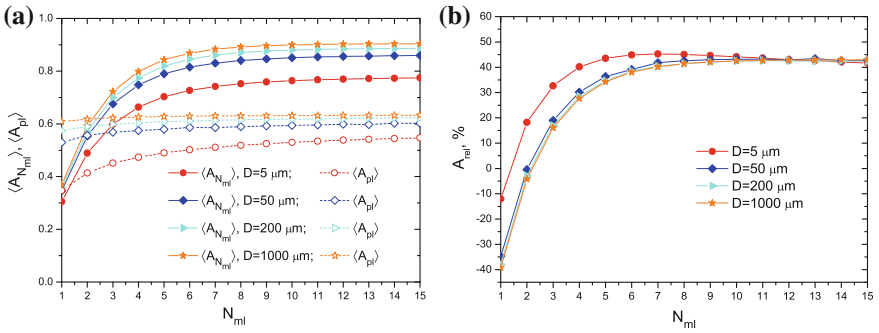

Fig. 2.50 Dependences of integral absorption coefficient $\langle A_{N_{ml}} \rangle$ of multilayer on the number N_{ml} of monolayers, $\langle A_{pl} \rangle$ of the equivalent plane-parallel plates (**a**) and relative absorption coefficient A_{rel} (**b**) (Miskevich and Loiko 2014a). Particle diameters are indicated in the legends. Values of $\langle A_{N_{ml}} \rangle$, $\langle A_{pl} \rangle$, and A_{rel} are calculated for $N_{ml} = 1, 2, \ldots, 15$. The results are indicated by symbols. Filled and empty identical symbols for each N_{ml} correspond to the same volume of material per unit surface area for the multilayer and plate, respectively

$\langle A_{N_{ml}}(D) \rangle$ function decreases. Therefore for creating the solar cells based on the layered particulate structure one should optimize size of the particles.

From the obtained results it follows that a layered particulate system consisting of three or more monolayers of large spherical silicon particles absorbs more light than the equivalent homogeneous plane-parallel silicon plate.

The systems of small and wavelength-sized particles were studied in Miskevich and Loiko (2015a). In systems of wavelength-sized particles one can obtain significant increasing the absorption efficiency (as compare to the equivalent plate) due to

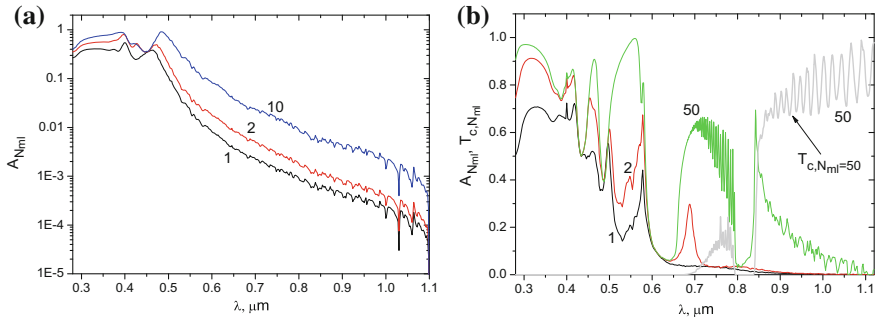

Fig. 2.51 Spectral absorption coefficient $A_{N_{ml}}$ of systems consisting of identical monolayers with triangular lattice of c-Si particles at $D = 0.1\,\mu$m, $\eta = 0.5$, $s = 0.1\,\mu$m (**a**) (Miskevich and Loiko 2015a) and at $D = 0.2\,\mu$m, $\eta = 0.7$, $s = 0.2\,\mu$m (**b**). Spectral direct transmission coefficient $T_{c,Nml}$ of multilayer consisting of 50 monolayers of 0.2 μm-diameter particles, $\eta = 0.7$, $s = 0.2\,\mu$m (**b**). QCA, $\sigma(u) = 0.001u$, $l_c = 900$. The number of monolayers N_{ml} is indicated near the lines

pronounced resonance scattering on the individual particles and multiple scattering in the particle arrays.

In Fig. 2.51 is shown spectral absorption coefficient of multilayers consisting of monolayers of the 0.1 and 0.2 μm-diameter particles. Concentration increasing can result in decreasing the absorption coefficient of the multilayer in the range of large absorption index of silicon. In the range of small absorption index the absorption coefficient of multilayer depends on particle concentration, size, and wavelength. As a whole, $A_{N_{ml}}$ grows with particle size and concentration. In the wavelength range of the photonic band gap (PBG) it sharply decreases (see $T_{c,N_{ml}}$ in the range $\sim 0.78\,\mu\text{m} \le \lambda \le \sim 0.84\,\mu$m in Fig. 2.51b). However, near the bounds of the PBG the absorption coefficient significantly grows up.

Figure 2.52a depicts integral absorption coefficients $\langle A_{N_{ml}} \rangle$ of multilayers from monolayers of the 0.1 μm-diameter particles and integral absorption coefficients $\langle A_{pl} \rangle$ of the equivalent plates. Figure 2.53a shows data for structures with the 0.2 μm-diameter particles. Figures 2.52b and 2.53b display the corresponding dependences of $A_{rel}(N_{ml})$.

As one can see from Figs. 2.52 and 2.53 the integral absorption coefficient increases with particle size. The multilayers of considered particles can absorb more light than the equivalent plates. The particulate system absorbs light more efficiently (in comparison with the equivalent plate) at lower particle concentrations and smaller number of monolayers (see Figs. 2.52b and 2.53b). Maximum relative increasing in absorption for the considered cases is about 145%.

The amount and efficiency of light absorption can be increased in *gradient multilayers* which were considered as well (Miskevich and Loiko 2015a). Such multilayers consist of monolayers with changing parameters in monolayer sequences. Figure 2.54a, b show calculated spectral and integral absorption coefficients of multilayers consisting of monolayers with different filling factors, respectively. The results for non-gradient and gradient multilayers are presented. Figure 2.54c shows

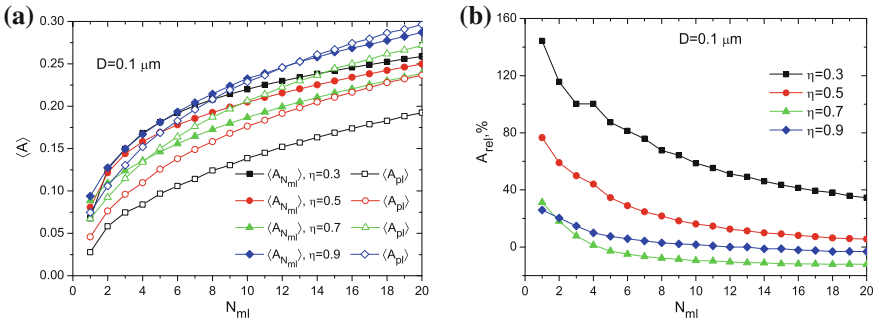

Fig. 2.52 Integral absorption coefficient $\langle A_{N_{ml}} \rangle$ of multilayer, $\langle A_{pl} \rangle$ of equivalent plate (**a**), and relative absorption coefficient A_{rel} (**b**) versus the number N_{ml} of monolayers (Miskevich and Loiko 2015a). Individual monolayers have triangular lattice, $D = 0.1\,\mu\text{m}$, $\sigma(u) = 0.001u$, $l_c = 900$. $s = 0.1\,\mu\text{m}$, $0.28\,\mu\text{m} \le \lambda \le 1.12\,\mu\text{m}$

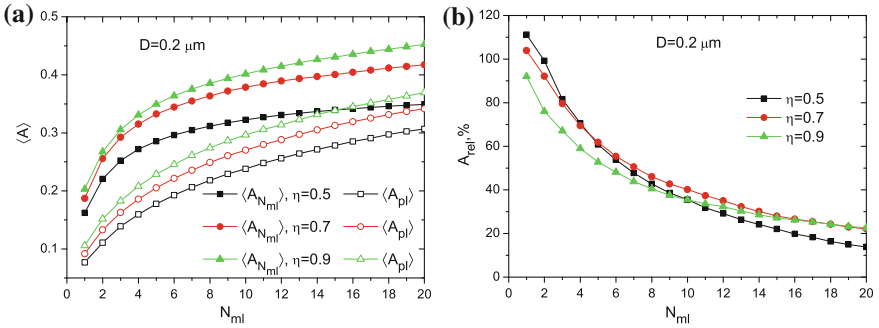

Fig. 2.53 Integral absorption coefficient $\langle A_{N_{ml}} \rangle$ of multilayer, $\langle A_{pl} \rangle$ of equivalent plate (**a**) and relative absorption coefficient A_{rel} (**b**) versus the number N_{ml} of monolayers (Miskevich and Loiko 2015a). Individual monolayers have triangular lattice, $D = 0.2\,\mu\text{m}$, $\sigma(u) = 0.001u$, $l_c = 900$. $s = 0.2\,\mu\text{m}$, $0.28\,\mu\text{m} \le \lambda \le 1.12\,\mu\text{m}$

the relative absorption coefficient $A_{rel,grad,N_{ml}}$ describing the relative absorption by gradient multilayer in comparison with the non-gradient one. It is calculated by Eq. (2.5.7), where $\langle A_{L_1} \rangle = \langle A_{grad,N_{ml}} \rangle$ is integral absorption coefficient of the gradient multilayer, $\langle A_{L_2} \rangle = \langle A_{N_{ml}} \rangle$ is integral absorption coefficient of the non-gradient multilayer consisting of identical monolayers (here calculations of $\langle A_{N_{ml}} \rangle$ are made at filling factor $\eta = 0.9$).

As can be seen from Fig. 2.54a, spectral absorption coefficient of the η-gradient multilayer is significantly larger than the one of the non-gradient multilayer (consisting of identical monolayers with $\eta = 0.9$) in the range of large absorption index of silicon. It results in significant increasing the integral absorption (see Fig. 2.54b). The integral absorption coefficient of the seven-monolayer gradient system is ~20% larger than the one of the non-gradient system (Fig. 2.54c). Thus, the larger absorption coefficient can be obtained at smaller volume of material.

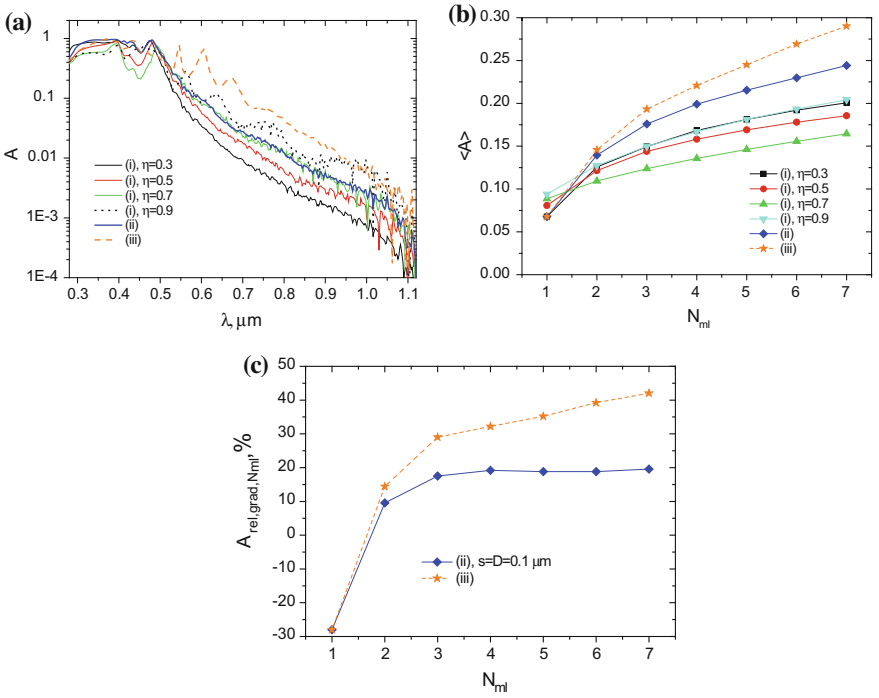

Fig. 2.54 **a** Spectral absorption coefficient of: (i) non-gradient multilayers consisting of seven identical monolayers at $\eta = 0.3$, $\eta = 0.5$, $\eta = 0.7$, $\eta = 0.9$, $D = 0.1\,\mu\text{m}$, $s = 0.1\,\mu\text{m}$; (ii) gradient multilayer consisting of seven monolayers with different filling factors: $\eta_1 = 0.3$, $\eta_2 = 0.4$, $\eta_3 = 0.5$, $\eta_4 = 0.6$, $\eta_5 = 0.7$, $\eta_6 = 0.8$, $\eta_7 = 0.9$, $D = 0.1\,\mu\text{m}$, $s = 0.1\,\mu\text{m}$; and (iii) gradient multilayer consisting of seven monolayers with different filling factors: $\eta_1 = 0.3$, $\eta_2 = 0.4$, $\eta_3 = 0.5$, $\eta_4 = 0.6$, $\eta_5 = 0.7$, $\eta_6 = 0.8$, $\eta_7 = 0.9$ and particle diameters $D_1 = 0.1$, $D_2 = 0.11$, $D_3 = 0.12$, $D_4 = 0.14$, $D_5 = 0.16$, $D_6 = 0.18$, $D_7 = 0.2\,\mu\text{m}$; $s_1 = 0.105$, $s_2 = 0.115$, $s_3 = 0.13$, $s_4 = 0.15$, $s_5 = 0.17$, $s_6 = 0.19\,\mu\text{m}$. The subscripts at η and D indicate the monolayer numbers; the subscripts at s indicate number of intermonolayer spacing between i-th and $i + 1$-th monolayer. **b** integral absorption coefficient of multilayer versus the number N_{ml} of monolayers. **c** relative absorption coefficient $A_{rel,grad,N_{ml}}$. Triangular lattice, $\sigma(u) = 0.001u$, $l_c = 900$, $0.28\,\mu\text{m} \le \lambda \le 1.12\,\mu\text{m}$ (Miskevich and Loiko 2015a)

More powerful gain can be obtained at simultaneous increase of the filling factor and diameter of particles in the monolayer sequence. This case is demonstrated in Fig. 2.54 by the dashed lines. The multilayer has following structure characteristics: $D_1 = 0.1\,\mu\text{m}$, $\eta_1 = 0.3$ (1-st monolayer), $D_2 = 0.11\,\mu\text{m}$, $\eta_2 = 0.4$ (2-nd), $D_3 = 0.12\,\mu\text{m}$, $\eta_3 = 0.5$ (3-rd), $D_4 = 0.14\,\mu\text{m}$, $\eta_4 = 0.6$ (4-th), $D_5 = 0.16\,\mu\text{m}$, $\eta_5 = 0.7$ (5-th), $D_6 = 0.18\,\mu\text{m}$, $\eta_6 = 0.8$ (6-th), and $D_7 = 0.2\,\mu\text{m}$, $\eta_7 = 0.9$ (7-th). The spacings between the monolayers were specified as the sum of particle radii in the adjacent monolayers. As can be seen from the figure, the simultaneous η- and D-growing results in more pronounced increasing the spectral and integral absorption coefficients and the value of $A_{rel,grad,N_{ml}}$. The integral absorption coefficient of this

system is more than 40% larger than the one of the non-gradient system of monolayers with $\eta = 0.9$ and $D = 0.1\,\mu$m (see Fig. 2.54c).

2.5.3 *Efficiency of Light Absorption by Layers Consisting of Silicon Particles*

Monolayer and multilayer consisting of silicon particles can be efficient light absorbers (Miskevich and Loiko 2014a, 2015a).

When particle sizes are significantly smaller than the wavelength of incident light, these systems absorb lesser than the equivalent silicon plates (Miskevich and Loiko 2015a).

Absorption coefficient of monolayer of large particles is smaller than the one of the equivalent plate. Absorption by systems of three- and more- monolayers of such particles is larger than the one of the equivalent plates. For six- and more monolayer system it can exceed 0.8 in the spectral range of large absorption index of silicon and 0.9 in the spectral range of the moderate and small values of the absorption index of silicon (Miskevich and Loiko 2014a).

Absorption by monolayer of the wavelength-sized particles can be significantly larger than the one of the equivalent plate. It is caused by strong resonant scattering by individual submicron silicon particles and strong multiple scattering in particle arrays. Absolute absorption coefficient and relative absorption coefficient of layers of such particles as a whole grow with particle size. Relative absorption coefficient attains maximum when particle sizes and distances between them are comparable with the wavelengths of incident light. In the narrow wavelength intervals (up to 10 nm) of the resonant peaks the spectral absorption coefficient of monolayer can be more than 100 times larger than the one of the equivalent plate. In the wavelength range from 0.8 μm to 1.12 μm integral absorption coefficient of monolayer can be more than 20 times higher than the one of the plate. There are spectral ranges where this ratio can be essentially larger (Miskevich and Loiko 2015a).

In multilayers consisting of monolayers of wavelength-sized particles integral absorption coefficient grows with the number of monolayers and increases in the spectral ranges near the boundaries of PBG. To increase light absorption in particulate structures the gradient multilayers can be used. The sketch of solar cell consisting of particle size- and concentration-gradient multilayer is shown in Fig. 2.55 (Miskevich and Loiko 2015a).

The solar cell consists of active layer, antireflection coating, transparent electrodes 1, transparent dielectric layers 2, and rear electrode 3. The active layer is the stack of monolayers of particles (which generally can be of any shape) of semiconductor (it can be silicon or other material). The circles sectioned by the dashed lines are the particles with the *p*-*n* junctions designated by these lines. The transparent electrodes connect the parts of particles with the same conduction type (*p* or *n*). The transparent dielectric layers separate the transparent electrodes and zones of different conduc-

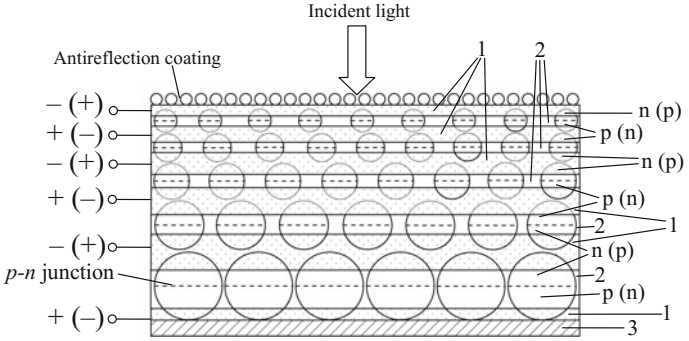

Fig. 2.55 Sketch of the solar cell with gradient layered structure of active layer (side view) (Miskevich and Loiko 2015a). The circles with the dashed lines are particles of the semiconductor material. 1: transparent electrodes, 2: transparent dielectric layers, 3: rear electrode

tion types of particles. The rear electrode 3 provides the electrical conductivity and reflection of light back into the active layer.

2.6 Light Scattering and Absorption by a Layer with High Concentration of Optically Soft Particles: Treatment Based on Radiative Transfer Theory

To describe light scattering in densely packed media, the theory of multiple scattering of waves (Ishimaru 1978a, b; Apresian and Kravtsov 1983; Twersky 1962c; Tsang et al. 2000, 2001a; Tsang and Kong 2001; Borovoi 2006) should be used (see Sect. 2.2.1). Because of the complexity of the mathematical apparatus used in this theory, complete solutions can be obtained only in a very limited number of situations. In some cases the phenomenological approach can be applied. It is simpler and widely used at the present time. This approach is based on the solution of radiative transfer equation operating with the radiation (light) intensity (radiance) (Sobolev 1956a, b; Chandrasekhar 1960; van de Hulst 1980; Barabanenkov and Finkelberg 1968; Barabanenkov 1976, 2003).

In this section we consider a model of radiation transport in a layer of particulate medium with a high concentration of optically soft particles (Loiko and Berdnik 2003; Berdnik and Loiko 2004, 2006). It describes optical interaction of particles with taking into account their spatial correlation and multiple light scattering. The model is based on the radiative transfer equation and takes into account interference effects to determine the parameters of the unit volume of a particulate media. It can be used for modeling the biological tissues, composite liquid crystal materials, polymer-dispersed liquid crystals, liquated glasses, porous glasses and structures, etc.

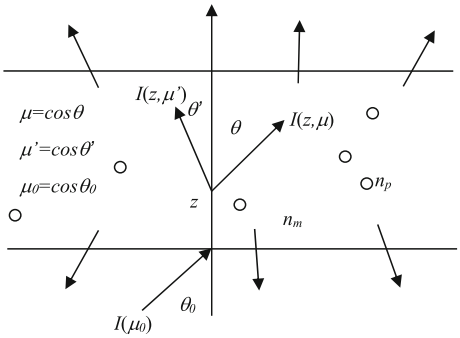

Fig. 2.56 Schematic structure of a particulate layer (Berdnik and Loiko 2006). The notations are in the text

2.6.1 Model of Radiation Transport

Let an azimuthally symmetric wide radiation beam with intensity $I(\mu_0)$ be incident at an angle θ_0 to the normal to a particulate layer (slab) with thickness determined by the planes $z = 0$ and $z = z_0$ (Fig. 2.56). The layer is a matrix with monodisperse particles of radius R_p and a relative refractive index n_p. The refractive index n_m of the matrix is equal to the refractive index of the surrounding medium and hence there are no Fresnel reflections from the boundaries. The radiation is scattered inside the layer, partially absorbed, and emerges from it.

In order to describe the propagation of radiation in the layer (Berdnik and Loiko 2006), the *radiative transfer equation* (RTE) (Sobolev 1956b; Chandrasekhar 1960; van de Hulst 1980; Minin 1988) was used which can be presented in the following form for an azimuthally-averaged intensity of scattered radiation when the layer is illuminated by a parallel radiation beam:

$$\mu \frac{\partial I(z,\mu)}{\partial z} + \varepsilon I(z,\mu) = \sigma \int_{-1}^{1} p(\mu,\mu') I(z,\mu') \mathrm{d}\mu' + \\ + I_{1n}^{+} \sigma p(\mu,\mu_0) e^{-\frac{\varepsilon z}{\mu_0}} + I_{2n}^{-} \sigma p(-\mu,\mu_0) e^{-\frac{\varepsilon(z_0-z)}{\mu_0}} . \tag{2.6.1}$$

Here $I(z,\mu)$ is the azimuth-averaged intensity of scattered radiation propagating inwards the layer at a polar angle $\theta = \arccos\mu$; σ and ε are the scattering and extinction coefficients; $\mu_0 = \cos\theta_0$; $\mathrm{p}(\mu,\mu')$ is the azimuthally-averaged phase function (also known as *redistribution function*); $p(\cos\gamma)$ is the phase function of unit volume, normalized by the condition $\int_{-1}^{1} p(\cos\gamma)\mathrm{d}\cos\gamma = 1$; $\cos\gamma = \mu\mu' + \sqrt{1-\mu^2}\sqrt{1-\mu'^2}\cos\varphi$, γ is the scattering angle; $\mu = \cos\theta$ and $\mu' = \cos\theta'$; φ is the azimuthal scattering angle; I_{1n}^{+} and I_{2n}^{-} are the intensities of the light propagating the layer in the direction of the wavevector of incident wave, at the upper and lower boundaries, respectively. The RTE is solved with the following boundary conditions:

$$I(z=0,\mu>0) = I(z=0,\mu<0), \\ I(z=z_0,\mu<0) = I(z=z_0,\mu>0). \tag{2.6.2}$$

2.6.2 Unit Volume Parameters

The simulation of the optical parameters of densely packed media is beginning from their calculation in the low-concentration limit. To calculate parameters of the unit volume in this limit, the Mie theory is used (Bohren and Huffman 1983; Babenko et al. 2003; Mie 1908). To calculate the extinction coefficient and the phase function of a medium with a high particle concentration, one must solve the problem of light diffraction (scattering) by a many body system. A rigorous solution of this problem has not been found so far, and hence various approximation methods are used. For weakly scattering particles, the interference approximation is the most convenient (Loiko and Berdnik 2006; Mishchenko 1994). According to this approach, the expressions for differential scattering coefficient $\sigma_h(\gamma)$, scattering coeficient σ_h, and the extinction coefficient ε_h of a medium consisting of identical spherical particles have the form

$$\sigma_h(\gamma) = w\sigma_{0l}\, p_l(\gamma)\mathrm{S}_3(\gamma, w), \tag{2.6.3}$$

$$\sigma_h = w\sigma_{0l}u, \tag{2.6.4}$$

$$\varepsilon_h = w(\varepsilon_{0l} - \sigma_{0l} + \sigma_{0l}u), \tag{2.6.5}$$

where

$$u = \int_0^{\pi} p_1(\gamma)\mathrm{S}_3(\gamma, w)\sin\gamma d\gamma. \tag{2.6.6}$$

Here $w \equiv \eta_V = Nv/V$ is the volume filling factor of a layer, N is the number of particles with volume v contained in volume V; $\sigma_{0l} = \Sigma_s/v$; $\varepsilon_{0l} = \alpha_{0l} + \sigma_{0l} = \Sigma_e/v$; $\alpha_0 = \Sigma_a/v$; Σ_a, Σ_s and Σ_e are the absorption, scattering and extinction cross-sections of an individual particle; and $p_l(\gamma)$ is the normalized phase function of an individual particle. Function $\mathrm{S}_3(\gamma, w)$ is the structure factor of three-dimensional particulate medium. The parameter u characterizes the degree of optical interaction of particles. For independent (single) scattering $u = 1$. The stronger the correlation in particle locations the larger the difference of the u-value from unity.

For the considered 3D layer with rigid spherical particles, the structure factor is calculated in the Percus–Yevick approximation (Ivanov et al. 1988; Percus and Yevick 1958; Berdnik and Loiko 2006, 2011):

$$\mathrm{S}_3(\gamma, w) = \left(1 - 24\omega\int_0^1 c_3(x, w)\frac{\sin yx}{yx}x^2\mathrm{d}x\right)^{-1}. \tag{2.6.7}$$

Here $x = r/2R_p$; r is the distance between two particles, $y = (8\pi R_p/\lambda)\sin(\gamma/2)$.

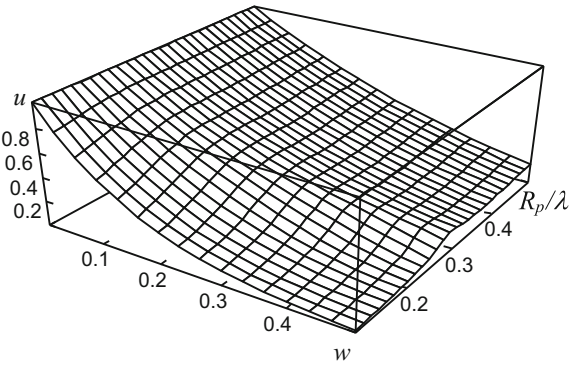

Fig. 2.57 Dependence of u on w and R_p/λ for particles with $n_p = 1.05$ (Berdnik and Loiko 2006)

$$c_3(x, w) = -a - bx - cx^3, \tag{2.6.8}$$

$$a = \frac{(1 + 2w)^2}{(1 - w)^4}, \tag{2.6.9}$$

$$b = -6w\frac{(1 + 0.5w)^2}{(1 - w)^4}, \tag{2.6.10}$$

$$c = 0.5w\frac{(1 + 2w)^2}{(1 - w)^4}. \tag{2.6.11}$$

Figure 2.57 illustrates influence of the radius and concentration of homogeneous spherical particles with $n_p = 1.05$ on u. With increasing particle concentration, u decreases monotonically. The dependence on the radius, however, is nonmonotonic. Figure 2.58 shows the dependence of u on the particle refractive index and radius for $w = 0.5$. One can see that the peak of u-value increases with decreasing the refractive index.

Figure 2.59 illustrates the change of the phase function at a change in the concentration and size of the particles. The intensity of forward-scattered radiation decreases with the volume filling factor w increasing and for quite high concentrations the phase function acquires a characteristic maximum for a nonzero scattering angle. With w increasing, the maximum is shifted towards large angles, while with R_p increasing it is shifted towards small angles.

Note that for certain values of the particle radius and concentration, the asymmetry parameter $g = \int_{-1}^{1} p(\mu)\mu \mathrm{d}\mu$ of the phase function may be zero, or even attain negative values (Fig. 2.60). The primary maximum on the phase function is formed at values of size parameter $x \geq 7/4$ (Berdnik and Loiko 2004). At large volume filling factor ($w > 0.5$) its position can be determined roughly by the expression

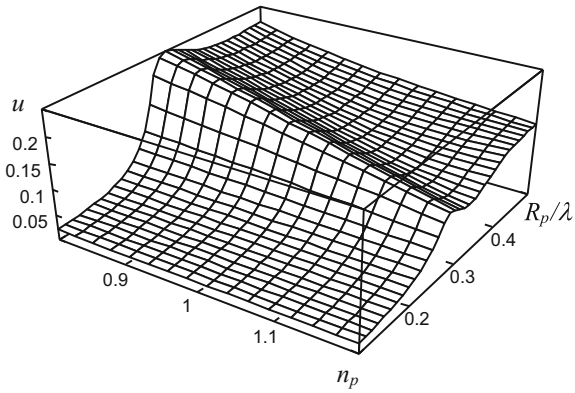

Fig. 2.58 Dependence of u on n_p and R_p/λ for $w = 0.5$ (Berdnik and Loiko 2006)

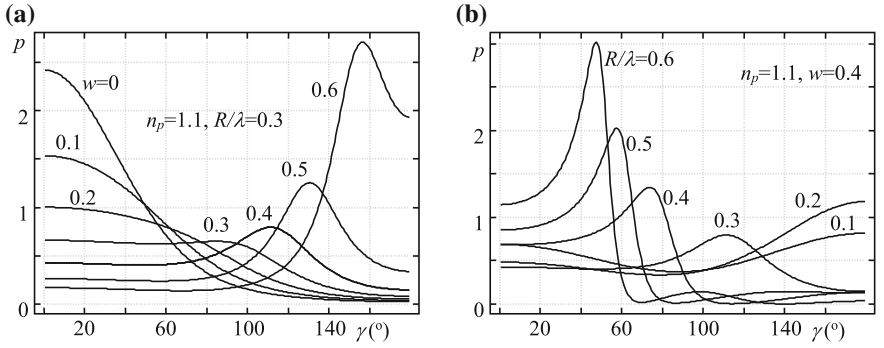

Fig. 2.59 Dependence of the phase function of a unit volume on the layer volume filling factor w (**a**) and radius R_p of particles (**b**) (Berdnik and Loiko 2006)

$$\gamma_m = 2 \arcsin \frac{7}{4x}. \tag{2.6.12}$$

The optical thickness τ_0 of the layer, the single scattering albedo Λ, and phase function $p(\gamma)$ of the unit volume are defined by the formulas

$$\tau_0 = \tau_{01}(1 - \Lambda_l(1 - u)), \tag{2.6.13}$$

$$\Lambda = \frac{\Lambda_l u}{1 - \Lambda_l(1 - u)}, \tag{2.6.14}$$

$$p(\gamma) = \frac{p_l(\gamma) S_3(\gamma)}{u}, \tag{2.6.15}$$

where τ_{0l} and Λ_l are the optical thickness of the layer and the single scattering albedo of the rarefied layer (calculated in the SSA).

In order to solve the transfer equation, we should calculate the redistribution function

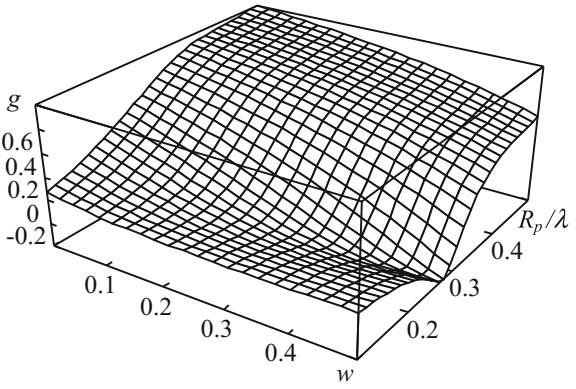

Fig. 2.60 Dependence of the asymmetry parameter g on the layer volume filling factor w and radius R_p of particles with refractive index $n_p = 1.1$ (Berdnik and Loiko 2006)

$$p(\mu, \mu') = \frac{1}{2\pi} \int_0^{2\pi} p(\mu\mu' + \sqrt{1-\mu^2}\sqrt{1-\mu'^2}\cos\varphi)d\varphi. \tag{2.6.16}$$

This function is usually calculated using the Legendre polynomial expansion (van de Hulst 1980). For the phase functions having regions with sharp peaks, however, several hundred expansion terms have to be taken into account, and the calculation of the expansion coeficients p_l becomes complicated. If the number of expansion terms used in the calculation is small, the function $p(\mu, \mu')$ acquires "ripples" appearing as a result of incorrectness of Fourier series summation with inaccurately determined coefficients. In this case, the redistribution function may acquire negative values. In order to avoid these issues, the $p(\mu, \mu')$ function is calculated using the *spline approximation method* (Berdnik and Loiko 1999; Berdnik and Mukhamedyarov 2001). The use of splines helps in reducing the ripples.

2.6.3 Angular Structure of Scattered Light

A statistically uniform layer of a scattering particulate medium is characterized by the brightness coefficients of backward, $\rho(\mu, \mu')$, and forward, $\sigma(\mu, \mu')$, diffusely scattered radiation, which are determined by the relations

$$I^-(z = 0, \mu) = \int_0^1 2\rho(\mu, \mu')\mu' I_0(\mu')\mathrm{d}\mu', \tag{2.6.17}$$

$$I^+(z = z_0, \mu) = e^{-\tau_0/\mu} I_0(\mu) + \int_0^1 2\sigma(\mu, \mu')\mu' I_0(\mu')\mathrm{d}\mu'. \tag{2.6.18}$$

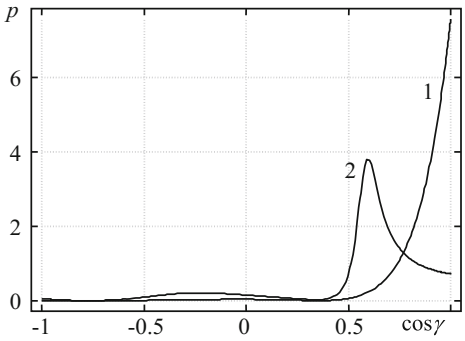

Fig. 2.61 Phase functions of a unit volume of the medium formed by particles with $R_p = 0.3\,\mu\text{m}$, $n_p = 1.1$ ($\lambda = 0.5\,\mu\text{m}$). $w = 0$ (curve 1) and $w = 0.5$ (curve 2) (Berdnik and Loiko 2006)

Here $I_0(\mu)$ is the intensity of radiation incident on the layer, $I^+(z, \mu) = I(z, \mu > 0)$ and $I^-(z, \mu) = I(z, \mu < 0)$ are the intensities of radiation outgoing the layer into forward and backward semispheres.

To determine the $\rho(\mu, \mu')$ and $\sigma(\mu, \mu')$ coefficients a calculation approach based on the layer doubling technique (van de Hulst 1980; Lenoble 1985; Plass et al. 1973; Wiscombe 1976; Hunt 1971) is used. In this method, the computations are started by choosing a layer of a quite small optical thickness τ_s so that $\tau_0 = \tau_s 2^K$, where K is an integer. For a layer with an optical thickness τ_s, the brightness coefficients are determined approximately. The methods to specify approximate values of $\rho(\mu, \mu')$ and $\sigma(\mu, \mu')$ for isotropic particulate media are developed in Tsang and Kong (1980), where τ_s is calculated in the single scattering approximation. The brightness coefficients for a layer of doubled thickness were found using the known relations derived from the balance equations at the layer boundaries (Ishimaru and Kuga 1982; Mishchenko 1994; Kuz'min et al. 2001). The initial optical thickness was assumed to be equal to $\tau_s \sim 10^{-6}$, which ensured a fairly high calculation accuracy.

Proceeding from expressions (2.6.17) and (2.6.18), we can write the expressions for the reflection (R) and transmission (T) coefficients of the particulate medium:

$$R(\mu_0) = 2\int_0^1 \rho(\mu, \mu_0)\mu r m d\mu, \tag{2.6.19}$$

$$T(\mu_0) = e^{-\frac{\tau_0}{\mu_0}} + 2\int_0^1 \sigma(\mu, \mu_0)\mu r m d\mu. \tag{2.6.20}$$

Figure 2.61 shows the phase functions of the unit volume of a medium formed by spherical particles with $R_p = 0.3\,\mu\text{m}$, $n_p = 1.1$ at $\lambda = 0.5\,\mu\text{m}$, when volume filling factor $w = 0.001$ (curve 1: practically uncorrelated ensemble) and 0.5 (curve 2: correlated ensemble).

Figures 2.62 and 2.63 show the angular dependences of the brightness coefficients of normally illuminated plane-parallel particulate layers (slabs) with different surface filling factors η. With η and, hence, the optical thickness of the layer increasing, the

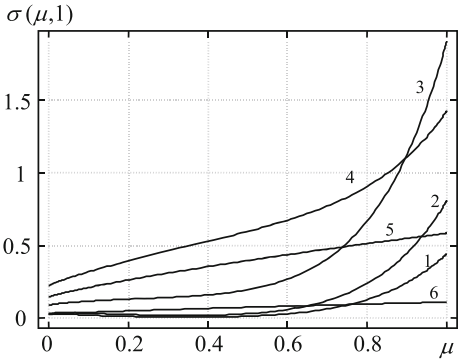

Fig. 2.62 Brightness coefficients of a layer formed by particles with $R_p = 0.3\,\mu\text{m}$, $n_p = 1.1$ ($\lambda = 0.5\,\mu\text{m}$) and $w = 0.001$ for $\eta = 0.49$ [curve (1)], 0.98 [curve (2)], 3.91 [curve (3)], 15.63 [curve (4)], 62.5 [curve (5)], and 500 [curve (6)]; $u = 0.16161$, $\tau_{0l} = 131.26$ (Berdnik and Loiko 2006)

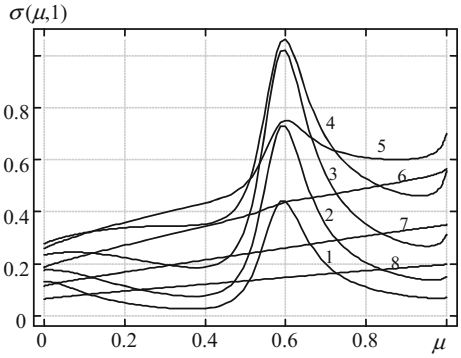

Fig. 2.63 Brightness coefficient of a layer formed by particles with $R_p = 0.3\,\mu\text{m}$, $n_p = 1.1$ ($\lambda = 0.5\,\mu\text{m}$), $w = 0.5$, $\eta = 3.91$ [curve (1)], 7.81 [curve (2)], 15.63 [curve (3)], 31.25 [curve (4)], 62.5 [curve (5)], 125 [curve (6)], 250 [curve (7)], and 500 [curve (8)]; $u = 0.16161$, $\tau_{0l} = 21.5477$ (Berdnik and Loiko 2006)

multiple scattering leads to smooth the peaks of the phase function (the peak on the angular structure observed for small η disappears), and the angular structure of scattered light does not change for quite large values of τ_0, that indicates transition to the asymptotic regime.

One can see from Fig. 2.63 that with η increasing the angular dependence of radiation scattered in forward semisphere shows a peak whose magnitude increases to its maximum value, after which it disappears for large η when the asymptotic regime is achieved. Such a behavior is a consequence of the characteristic peak in the phase function of a densely packed medium. The intensity peak in the region of small scattering angles is observed experimentally (Ivanov et al. 1988).

Angular dependences of the brightness coefficient $\sigma(\mu, \mu')$ are shown in Fig. 2.64 for various angles of illumination. With the incidence angle increasing (μ' decreasing), the characteristic peak in the angular dependence of scattered radiation is blurred. Two peaks can be formed, the distance between them increases with μ' decreasing. Note that, the peak can be formed in the strictly backward direction if the phase function of the unit volume has a peak at an angle of 90° (Berdnik and Loiko 2004, 2006).

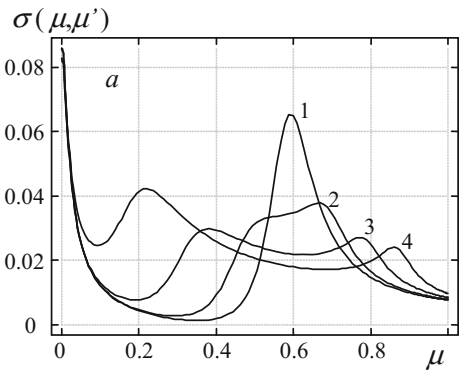

Fig. 2.64 Angular dependences of the brightness coefficient $\sigma(\mu, \mu')$ of a layer formed by particles with $R_p = 0.3\,\mu\text{m}$, $n_p = 1.1$ ($\lambda = 0.5\,\mu\text{m}$), $w = 0.5$, for $\mu' = 1$ [curve (1)], 0.99 [curve (2)], 0.96 [curve (3)], and 0.91 [curve (4)]; $u = 0.16161$, $\tau_{0l} = 131.26$, $\eta = 0.488$ (Berdnik and Loiko 2006)

2.6.4 *Light Scattering by a Layer of Correlated Particles in Liquating Glasses*

Let us pay attention to the problem of the anomalous light scattering by liquating glasses (Berdnik and Loiko 2011; Loiko and Berdnik 2006; Gurevich 1953; Kolyadin 1956; Kalmykov et al. 2000; Voishvillo 1957, 1962; Andreev et al. 1960; Shatilov 1962; Shepilov 2003). The phenomenon of anomalous forward scattering in sodium borosilicate glasses consists in the following. The power of the exponential function describing the wavelength dependence of the extinction coefficient of the glasses is much higher than four while from the theory of light scattering it follows that this power does not exceed four. Such a discrepancy is considered as an anomaly (Gurevich 1953; Kolyadin 1956; Kalmykov et al. 2000; Voishvillo 1957, 1962; Andreev et al. 1960; Shatilov 1962; Shepilov 2003). The function of angular distribution of the scattered light intensity also has an anomalous behavior. It is backward-elongated. The angular distributions of light scattered by strongly scattering glasses have maxima, whose positions shift toward smaller angles with increasing turbidity.

Let us consider, following to Loiko and Berdnik (2006), a system of spherical particles with the radius R_p and the refractive index n_p distributed in a medium with the refractive index n_m. The relative refractive index of the particles is $n = n_p/n_m$. The arrangement of the particles in space is shown schematically in Fig. 2.65. The solid-line circles correspond to the particles of the radius R_p, and the dashed-line ones show the radius of spatial interaction (the radius of the virtual "particles") of the particles. The distance between the particle centers cannot be smaller than $2R_c$. It is assumed that the spatial order of the particle arrangement in the system coincides with the order for a random arrangement of hard spheres of the radius R_c with spherical particles of the radius R_p located at their centers. The scattering of light by such a system is analyzed.

The starting point in the simulation of optical properties of densely packed media is their calculation in the limit of low concentration, where collective effects in light scattering are negligible. To calculate the characteristics of an elementary volume of

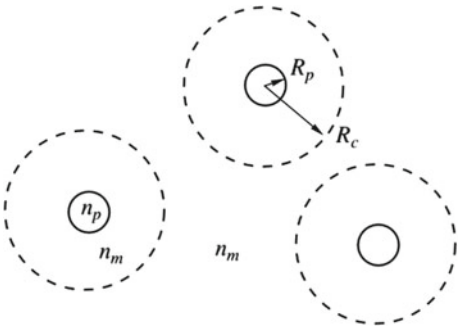

Fig. 2.65 Arrangement of scatterers (Loiko and Berdnik 2006). The notations are in the text

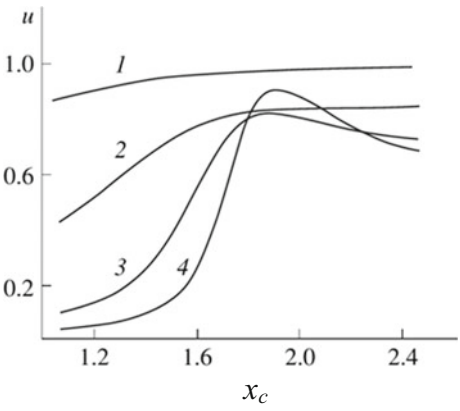

Fig. 2.66 Dependence of $u(x_c, w_c, e)$, calculated for particles with $n = 1.05$, $n_m = 1.5$, $\lambda = 0.5$ μm at $e = 0.25$ and $w_c = 0.05$ (curve 1), 0.2 (2), 0.4 (3), 0.5 (4) (Loiko and Berdnik 2006)

a medium with a low concentration of spherical scatterers, the Mie theory (Bohren and Huffman 1983; Babenko et al. 2003) is used.

To calculate the extinction coefficient and the phase function of the densely packed layers of soft particles the interference approximation (Dick and Loiko 2001a; Loiko and Berdnik 2003) is used (see Sect. 2.6.2).

The structure factor $S_3(\gamma, w_c)$ is determined as above (see Sect. 2.6.2), but instead of radius R_p we should use radius R_c.

Figure 2.66 illustrates change of the u-parameter as function of particle size parameter $x_c = 2\pi R_c/\lambda_m$ (λ_m is length of incident wave) and the volume filling factor w_c of spheres with radius R_c, when parameter $e = R_p/R_c < 1$. At low concentration, the $u(x_c)$ dependence is monotonic, as well as for $e = 1$. With the concentration increasing, a small peak arises in the dependence. It is more pronounced for larger filling factors. This peak is observed for x_c values from 1.5 to 2.

The spectral dependences of the extinction coefficient $\varepsilon_{0h}(\lambda) = \varepsilon_h(\lambda)/w_p = \sigma_{0h}(\lambda)u(R_p, w_c)$ and the phase function of the layer with nonabsorbing particles are shown in Fig. 2.67. The dependences of $\varepsilon_{0h}(\lambda)$ at certain particle sizes (see curves 2, 3, and 4) have maximum. The phase functions are backward elongated (Fig. 2.67b). With the radius R_c increasing, a peak on the phase function arises. It shifts toward

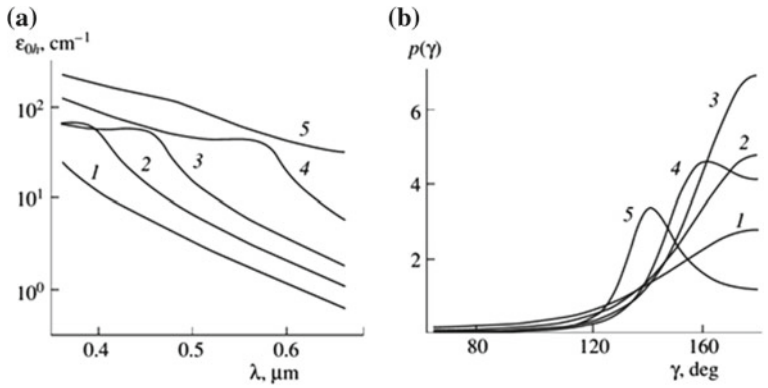

Fig. 2.67 Dependences of $\varepsilon_{0h}(\lambda)$ (**a**) and $p(\gamma)$ (**b**). $e = 0.5$ ($w_p = 0.075$); $n = 1.05$; $n_m = 1.5$; $w_c = 0.6$, R_c (μm) = 0.09 (curve 1), 0.10 (2), 0.105 (3), 0.11 (4), 0.115 (5) (**a**); $\lambda = 0.55$ μm. ε_{0h} (cm^{-1}) = 0.41(1), 1.51(2), 4.17(3), 11.93(4), 15.94(5) (**b**) (Loiko and Berdnik 2006)

smaller angles as R_c increases. Simultaneously, the extinction coefficient increases. It corresponds to the experimentally observed shift of the peak in the scattering phase function towards smaller angles with increasing scattering coefficient (Andreev et al. 1960).

The spectral dependence of the extinction coefficient $\varepsilon_{0h}(\lambda)$ in liquating glasses can be approximated as Shepilov (2003)

$$\varepsilon_{0h} = C\lambda^{-p} \tag{2.6.21}$$

where C is a constant and p is the power of the exponential. From Eq. (2.6.21) it follows that

$$p = \frac{d(\ln \varepsilon_{0h})}{d(\ln \lambda)}. \tag{2.6.22}$$

The dependence of p on the size parameter x_c and the volume filling factor w_c at $e = 0.1$ is shown in Fig. 2.68a. With w_c increase, the ranges with abnormally high values arise in the $p(x_c)$ dependence. The maxima in the $p(x_c)$ occur at the values of the x_c determined by the following equation:

$$x_c = \frac{2\pi R_c}{\lambda_m} = 0.917 + 1.506 w_c, \tag{2.6.23}$$

which is valid for w_c in the range from 0.4 to 0.63.

The dependences of p and asymmetry parameter g on w_c in this volume filling factor range at $e = 0.5$ are described by the equations:

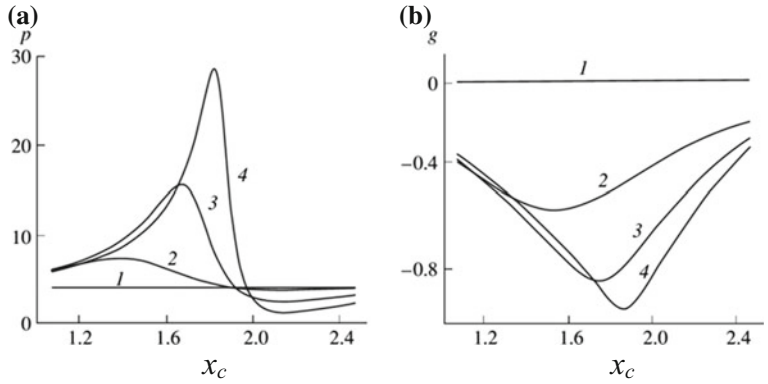

Fig. 2.68 Dependences of p (**a**) and g (**b**) on x_c at $e = 0.1$; $n = 1.05$; and $w_c = 0$ (curve 1), 0.3 (2), 0.5 (3), and 0.6 (4) (Loiko and Berdnik 2006)

$$p = -184.4 + 1243w_c - 2727w_c^2 + 2072w_c^3, \tag{2.6.24}$$

$$g = 0.207 - 2.216w_c + 0.863w_c^2. \tag{2.6.25}$$

The coefficients of dependences (2.6.23)–(2.6.25) were found by the least-squares method.

As w_c changes from 0.4 to 0.63, the p parameter changes from 9.1 to 34.4. The extremum in the size dependence of the asymmetry parameter g decreases from -0.54 to -0.86. Thus, the model presented, taking into account the correlation in spatial distribution of particles at low volume filling factors, explains the abnormally sharp change in the scattering coefficient, in particular, its abnormal behavior in sodium borosilicate glasses with a small volume filling factor.

Figure 2.68b shows the dependences of the parameter g on x_c and w_c at the same values of the medium parameters as in Fig. 2.68a. There are ranges of parameters where the backward-peaked phase functions are formed. The spectral positions of these ranges roughly coincide with the ranges of the anomalously high values of p. At some values of the medium parameters, the mean cosine of the scattering angle can be smaller than -0.9. The volume filling factor of particles in the range of abnormally high values of the p and g parameters is low, that is also in qualitative agreement with the experimental data (Kalmykov et al. 2000). At the filling factor $w_c = 0.6$ the volume filling factor of particles is $w_p = 0.075$ ($w_p = w_c e^3$).

The analysis has shown that, as the parameter $e = R_p/R_c$ increases, the maximal values of p and g decrease significantly, while the positions of these maxima change only slightly. If the value of e is small, then the values of the p and g parameters independ on the refractive index of the scattering particles and are determined by the R_c and w_c. However, at $e > 0.5$, the relative refractive index n affects significantly the p and g.

The measured anguar structures of intensities of light scattered by liquating glasses are presented in Fig. 2.69 (Voishvillo 1962).

The p-value close to nine was observed in these experiments.

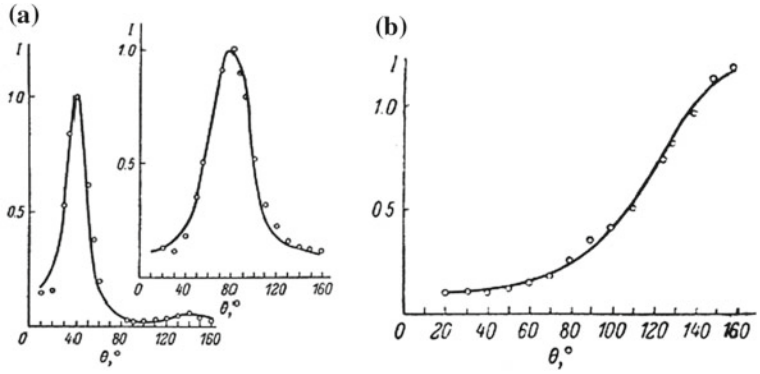

Fig. 2.69 The measured intensities I of light (I in a.u.) scattered by different types of liquating $Na_2B_2O_3SiO_2$ glasses versus polar scattering angle θ (Voishvillo 1962)

2.6.5 Light Absorption: Asymptotic Equations

Absorption coefficient of particulate layer can be determined by the data on the outgoing light. In the conditions of low particle concentration, when random spatial arrangement of particles is realized, the problem is reduced to the solution of the integro-differential radiative transfer equation (RTE) written above. Its solution is not a simple task and the set of approaches is used to overcome this difficulty.

At high optical thickness of the particulate layer an asymptotic regime is achieved. In this regime the angular and spatial variables are separated. The field is azimuthally independent and attenuates exponentially with the layer thickness increasing (Zege et al. 1991a). In this case the asymptotic equations (Sobolev 1956b; Zege et al. 1991a, b; van de Hulst 1980) can be used to describe scattering characterististics, and it is possible to find optical parameters of layer using a small number of algebraic equations. Such approach is much simpler than the solution of the RTE. In Loiko and Ruban (2000, 2004) the results of the approach were used to find light absorption by optically thick particulate layers with small weakly absorbing non-spherical silver halide microcrystals by measurements of spectral total transmittance T and reflectance R. The absorption coefficient A of a layer was determined from the relation $T + R + A = 1$.

The experimental concentration dependence of absorption coefficient $A = A(\eta_V)$ (η_V is the volume filling factor) of the layers under study is shown in Fig. 2.70. First, it increases with the volume filling factor, and then decreases starting from $\eta_V \approx 0.5$. The similar behavior of $A(\eta_V)$ of optically soft large-sized particles can be concluded from the data of (Loewinger et al. 1964; Kokhanovsky and Korolevich 1998).

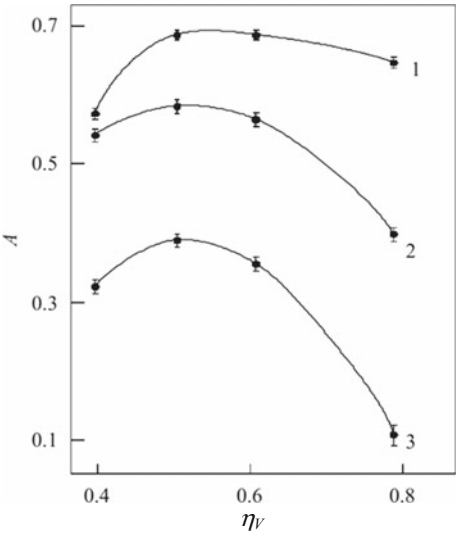

Fig. 2.70 Concentration dependence of the absorption coefficient A for a photolayer with average size of microcrystals $d = 0.11\,\mu\text{m}$ for surface concentration (mass of silver per unit area): $P =$ 20 (curve 1), 10 (curve 2), and $3\,\text{g/m}^2$ (curve 3); $\lambda = 0.436\,\mu\text{m}$ (Loiko and Ruban 2004). The measurement results are indicated by points with error bars

With the asymptotic formulas of Zege et al. (1991a), it is possible to write equation for absorption coefficient of the elementary volume (specific absorption):

$$k = (1/4)(1 - R_\infty) \lg ((T_1/T_2)/(P_1 - P_2)) . \tag{2.6.26}$$

Here R_∞ is the total reflectance of a semi-infinite layer, and T_1 and T_2 are the total transmittances determined in the asymptotic regime for two layers with surface concentrations P_1 and P_2, respectively.

The results for k are presented in Fig. 2.71. One can see that it depends nonmonotonically on the volume filling factor and has a maximum in the vicinity of $\eta_V =$ 0.5. Such a behaviour qualitatively agrees with the data of Nelson (1989) for the absorption coefficient of fractal aggregates.

In Fig. 2.72 are shown the experimental dependences of reflection coefficient R_∞ (Blewin and Brown 1961; Dubova and Khairullina 1982; Loiko 1981) of the semiinfinite layer on the volume filling factor for different particle materials.

The absorption and scattering of light by layers of closely-spaced black pigment particles are studied using the phenomenological *Kubelka–Munk theory* with the assumption of independent scattering (Gunde and Orel 2000). The absorption of radiation by concentrated biological tissues is studied in Twersky (1970b), Ivanov et al. (1987).

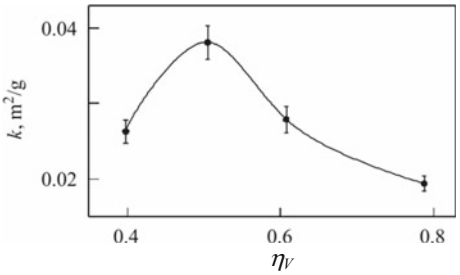

Fig. 2.71 Concentration dependence of the specific absorption k of a photolayer with average size of microcrystals $d = 0.11\ \mu$m; $\lambda = 0.436\ \mu$m (Loiko and Ruban 2000). The data obtained by measurement of T, R, R_∞ and calculation by Eq. (2.6.26) are indicated by points with the error bars

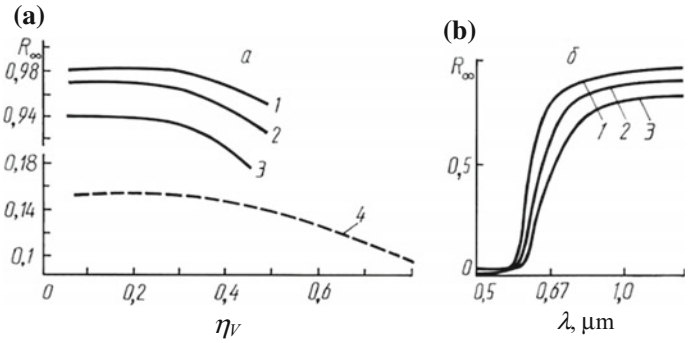

Fig. 2.72 Dependences of reflection coefficient on volume filling factor η_V (**a**) and the wavelength λ (**b**). **a** solid lines: MgO in air (Blewin and Brown 1961); dashed line: erythrocytes of human blood (Dubova and Khairullina 1982): 1: $\lambda = 0.6\ \mu$m, 2: $\lambda = 0.5\ \mu$m, 3: $\lambda = 0.4\ \mu$m, 4: $\lambda = 0.805\ \mu$m. **b** "red cadmium" pigment (Loiko 1981), 1: layer of free powder, 2: layer compressed in 1.3 times, 3: layer compressed in 1.4 times

2.7 Conclusions

The review of the investigations of wave interaction with random, partially ordered, and highly ordered particulate media is made. The main attention is devoted to the systems with short- and imperfect long-range positional order.

The recent progress in the theory of multiple scattering of waves to describe wave propagation in ordered particulate monolayers (planar crystals) is considered in details. The method to simulate spatial arrangement of particles forming the planar crystal (PC) with imperfect lattice is described. Its applicability to calculating the transmission, reflection, and absorption coefficients of PC in the quasicrystalline approximation of the theory of multiple scattering of waves is demonstrated. A number of practical applications of this method is considered. For example, it was used to simulate the antireflecion coatings on the glass, selective reflectors, multispectral

filters. It was applied also to solve the inverse scattering problem – retrieving the refractive index of particles forming the 3D photonic crystal, and to develope solar cells with effective light absorption.

A number of scattering problem solutions obtained for partially ordered particulate monolayers is described. In particular, angular distribution of light scattered by monolayer and small-angle light scattering and transmission by polymer dispersed liquid crystal film are outlined. The quenching effect for light directly transmitted the monolayer and some results for the spatial optical noise of monolayer are discussed.

The adding method and transfer matrix method (TMM) are used to describe light propagation in 3D particulate media. The last is used to study the 3D ordered structures (3D photonic crystals) from spherical alumina particles, systems of "particulate monolayers on glass substrate", periodic, quasiperiodic, and aperiodic sequences of monolayers, to solve the inverse scattering problem for 3D photonic crystal, and to calculate the absorption coefficient of system from monolayers of crystalline silicon particles.

The results based on application of the radiative transfer theory to describe light scattering and absorption by layers with high concentration of optically soft particles are considered. The data on angular structure of scattered light are presented. Features in scattering by correlated particles in liquating glasses are explained.

The review shows the significant progress made in the studying the EM wave scattering and abrorption by ordered particulate media, but the theoretical description of these phenomena is far from finishing. For example, in most of the theoretical works the systems of homogeneous isotropic spherical, cylindrical (sometimes spheroidal), or point scatterers are considered. This is caused by existing the well-developed methods to describe scattering properties of such scatterers. Of course, practically any particulate structure can be simulated numerically. But such an approach is often inconvenient to use, for example, when random or partially-ordered systems with huge quantity of particles are studied.

Below we emphasize some important, in our opinion, problems to be solved theoretically:

1. Description of angular distribution of light scattered by concentrated particulate layers with various types of ordering, including the case of oblique illumination;
2. Simulation of systems with nonspherical (noncylindrical, nonspheroidal), anisotropic, and inhomogeneous scatterers;
3. Modeling the statistically inhomogeneous concentrated paticulate media;
4. Modeling the near-field interaction between the particulate layers in multilayer and between particulate layer and homogeneous layer, for example in the "monolayer of particles on substrate" system;
5. Description of optical properties of ordered structures such as opals, colloidal crystals, etc. possessing the lattice defects other than particle deviations from lattice sites.

Acknowledgements This investigation was supported in part by the Belarusian Republican Foundation for Fundamental Research. Project No. F15IC-005 and the state research program of the Republic of Belarus "Photonics, opto- and microelectronics".

Appendix. Expansion of Fields in Terms of Scattering Orders

Let us write the main equations of the theory of multiple scattering of waves (TMSW) in the form:

$$\psi_{\mathbf{r}} = \psi^i_{\mathbf{r}} + \sum_{j=1}^{N} t_{\mathbf{r}j}\psi^e_j, \tag{A.1}$$

$$\psi^e_j = \psi^i_j + \sum_{k=1,k\neq j}^{N} t_{jk}\psi^e_k, \tag{A.2}$$

Field $\psi_{\mathbf{r}}$, i.e. the solution of Eq. (A.1), can be found by putting the (A.2) into (A.1):

$$\begin{aligned}\psi_{\mathbf{r}} &= \psi^i_{\mathbf{r}} + \sum_{j=1}^{N} t_{\mathbf{r}j}\left(\psi^i_j + \sum_{k=1,k\neq j}^{N} t_{jk}\psi^e_k\right) = \\ &= \psi^i_{\mathbf{r}} + \sum_{j=1}^{N} t_{\mathbf{r}j}\psi^i_j + \sum_{j=1}^{N}\sum_{\substack{k=1\\k\neq j}}^{N} t_{\mathbf{r}j}t_{jk}\psi^i_k + \sum_{j=1}^{N}\sum_{\substack{k=1\\k\neq j}}^{N}\sum_{\substack{l=1\\l\neq k}}^{N} t_{\mathbf{r}j}t_{jk}t_{kl}\psi^i_l \\ &+ \sum_{j=1}^{N}\sum_{\substack{k=1\\k\neq j}}^{N}\sum_{\substack{l=1\\l\neq k}}^{N}\sum_{\substack{m=1\\m\neq l}}^{N} t_{\mathbf{r}j}t_{jk}t_{kl}t_{lm}\psi^i_m + \cdots .\end{aligned} \tag{A.3}$$

The first term, $\psi^i_{\mathbf{r}}$, on the right side of (A.3) describes the field of incident wave in the observation point $\mathbf{r}$. The second and third terms represent the sum of N and $N(N-1)$ contributions of singly $t_{\mathbf{r}j}\psi^i_j$ and doubly $t_{\mathbf{r}j}t_{jk}\psi^i_k$ scattered waves, respectively. The fourth term describes the $N(N-1)^2$ contributions of triple scattering. It concludes the terms with $l = j$. Thus, this sum can be divided into two ones which describe three scattering events only by different particles, $t_{\mathbf{r}j}t_{jk}t_{kl}\psi^i_l$ $(k \neq j, l \neq k, l \neq j)$, and three scattering events at double passing by wave the same particle, $t_{\mathbf{r}j}t_{jk}t_{kj}\psi^i_j$ $(l = j \neq k)$, i.e. "forward-backward" scattering:

$$\sum_{j=1}^{N}\sum_{\substack{k=1\\k\neq j}}^{N}\sum_{\substack{l=1\\l\neq k}}^{N} t_{\mathbf{r}j}t_{jk}t_{kl}\psi^i_l = \sum_{j=1}^{N}\sum_{\substack{k=1\\k\neq j}}^{N}\sum_{\substack{l=1\\l\neq k,l\neq j}}^{N} t_{\mathbf{r}j}t_{jk}t_{kl}\psi^i_l + \sum_{j=1}^{N}\sum_{\substack{k=1\\k\neq j}}^{N} t_{\mathbf{r}j}t_{jk}t_{kj}\psi^i_j. \tag{A.4}$$

The fifth term in (A.3) describes the sum of $N(N-1)^3$ events of fourfold scattering. It can be divided into the sums describing scattering events only on different particles and at passing by wave the particles more than one time:

$$\sum_{j=1}^{N}\sum_{\substack{k=1\\k\neq j}}^{N}\sum_{\substack{l=1\\l\neq k}}^{N}\sum_{\substack{m=1\\m\neq l}}^{N} t_{\mathbf{r}j}t_{jk}t_{kl}t_{lm}\psi_m^i = \sum_{j=1}^{N}\sum_{\substack{k=1\\k\neq j}}^{N}\sum_{\substack{l=1\\l\neq k,l\neq j}}^{N}\sum_{\substack{m=1\\m\neq l,m\neq k,m\neq j}}^{N} t_{\mathbf{r}j}t_{jk}t_{kl}t_{lm}\psi_m^i + \sum_{j=1}^{N}\sum_{\substack{k=1\\k\neq j}}^{N}\sum_{\substack{l=1\\l\neq k,l\neq j}}^{N} t_{\mathbf{r}j}t_{jk}t_{kl}t_{lk}\psi_k^i +$$

$$+\sum_{j=1}^{N}\sum_{\substack{k=1\\k\neq j}}^{N}\sum_{\substack{l=1\\l\neq k,l\neq j}}^{N} t_{\mathbf{r}j}t_{jk}t_{kl}t_{lj}\psi_j^i + \sum_{j=1}^{N}\sum_{\substack{k=1\\k\neq j}}^{N}\sum_{\substack{m=1\\m\neq k,m\neq j}}^{N} t_{\mathbf{r}j}t_{jk}t_{kj}t_{jm}\psi_m^i + \sum_{j=1}^{N}\sum_{\substack{k=1\\k\neq j}}^{N} t_{\mathbf{r}j}t_{jk}t_{kj}t_{jk}\psi_k^i \qquad \text{(A.5)}$$

On the right side of (A.5) the first term describes $N(N-1)(N-2)(N-3)$ events of fourfold scattering on different particles. The second, third, and fourth ones represent the $N(N-1)(N-2)$, and fifth one describes the $N(N-1)$ events of fourfold scattering at passing by wave the particles more than one time.

References

Abrams ZR, Niv A, Zhang X (2011) Solar energy enhancement using downconverting particles: a rigorous approach. J Appl Phys 109:114905-1–9

Albada M, Lagendijk A (1985) Observation of weak localization of light in a random medium. Phys Rev Lett 55:2692–2695

Alfrey T et al (1954) Optical properties of uniform particle-size latexes. Opt Soc Am 44(8):603–609

Almpanis E et al (2012) Diffractive chains of plasmonic nanolenses: combining near-field focusing and collective enhancement mechanisms. Opt Lett 37(22):4624–4626

Andreev NS, Aver'yanov VI, Voishvillo NA (1960) Structural interpretation of the anomalous scattering of visible light in sodium borosilicate glasses. Phys Solid State (Fiz Tverd Tela) 2:1011–1021 (in Russian)

Andueza A et al (2010) Geometry influence on the transmission spectra of dielectric single layers of spheres with different compactness. J Appl Phys 107:124902-1–7

Andueza A et al (2011) Disorder effect in the transmission spectra of a noncompact single layer of dielectric spheres derived from microwave spectroscopy. Appl Opt 50(31):G91–G97

Andueza A, Sevilla J (2007) Non compact single-layers of dielectric spheres electromagnetic behaviour. Opt Quantum Electron 39:311–320

Andueza A, Echeverría R, Sevilla J (2008) Evolution of the electromagnetic modes of a single layer of dielectric spheres with compactness. J Appl Phys 104:043103-1–5

Andueza A, Morales P, Sevilla J (2012) Photonic band effect in single-layers of high refractive index spheres of different compactness. J Appl Phys 111:104902-1–7

Ao CO, Kong JA (2002) Analytical approximations in multiple scattering of electromagnetic waves by aligned dielectric spheroids. J Opt Soc Am A 19(6):1145–1156

Apresian LA, Kravtsov YuA (1983) Radiative transfer theory: statistical and wave aspects. Nauka, Moscow, 216 p (in Russian)

Arfken GV, Weber HJ, Harris FE (2012) Mathematical methods for physicists, 7th edn. Academic Press, Waltham, p 1205

Babenko VA, Astafyeva LG, Kuzmin VN (2003) Electromagnetic scattering in disperse media: inhomogeneous and anisotropic particles. Springer, Berlin, p 434

Bachelard R et al (2012) Resonances in Mie scattering by an inhomogeneous atomic cloud. Europhys Lett 97:14004-p1-p6

Balestreri A, Andreani L, Agio M (2006) Optical properties and diffraction effects in opal photonic crystals. Phys Rev E 74:036603–8

Bapat A et al (2004) Plasma synthesis of single-crystal silicon nanoparticles for novel electronic device applications. Plasma Phys Control Fusion 46:B97–B109

Barabanenkov YuN (2003) Transfer of trapped electromagnetic radiation in an ensemble of resonant mesoscopic scatterers. In: Wave scattering in complex media: from theory to applications, vol 107. NATO Science Series. pp 415–460

Barabanenkov YuN (1976) Multiple scattering of waves by ensemble of particles and the radiative transport. Sov Phys Uspekhi 18:673–689
Barabanenkov YuN, Finkelberg VM (1968) Radiation transport equation for correlated scatterers. Sov Phys JETP 26:587–591
Barcikowski S, Hahn A, Kabashin A, Chichkov B (2007) Properties of nanoparticles generated during femtosecond laser machining in air and water. Appl Phys A 87:47–55
Barrera RG, Reyes-Coronado A, García-Valenzuela A (2007) Nonlocal nature of the electrodynamic response of colloidal systems. Phys Rev B 75(18):184202-1–19
Baryshev AV, Kaplyanskii AA, Kosobukin VA, Limonov MF, Samusev KB, Usvyat DE (2003) Bragg diffraction of light in synthetic opals. Phys Solid State 45:459–471. https://doi.org/10.1134/1.1562231
Baryshev AV et al (2007) Resonant behavior and selective switching of stop bands in three dimensional photonic crystals with inhomogeneous components. Phys Rev Lett 99:063906-1–4
Baryshev AV et al (2011) Propagation of polarized light in opals: amplitude and phase anisotropy. J Exp Theor Phys 112(3):361–369
Bendat JS, Piersol AG (1971) Random data: analysis and measurement procedures. Wiley, New York
Berdnik VV, Mukhamedyarov RD (2001) Radiative transfer in plant leaves. Opt Spectrosc 90:580–591. https://doi.org/10.1134/1.1366754
Berdnik VV, Loiko VA (1999) Modelling of radiative transfer in disperse layers of a medium with a highly stretched phase function. JQSRT 61(1):49–57
Berdnik V, Loiko V (2004) Features of the angular structure of light scattered by a layer of partially ordered soft particles. J Quant Spectrosc Radiat Transf 88:111–123
Berdnik V, Loiko V (2006) Angular structure of radiation scattered by a disperse layer with a high concentration of optically soft particles. Quantum Electron 36(11):1016–1022
Berdnik VV, Loiko VA (2011) Light scattering by ensemble of nonabsorbing correlated two-layered particles: specific feature for spectral dependence of extinction coefficient. Appl Opt 50(22):4246–4251
Berry CR (1971) On the need to apply electromagnetic theoryof optical behavior of photographic emulsions. Phtogr Sci Eng 15(3):394–399
Bertone JF et al (1999) Thickness dependence of the optical properties of ordered silica-air and air-polymer photonic crystals. Phys Rev Lett 83(2):300–303
Blewin WR, Brown WJ (1961) Effect of particle on reflectance of semi-infinite diffusers. JOSA 51:129–134
Blinov LM (2011) Structure and properties of liquid crystals. Springer, Dordrecht, p 439
Bogomolov VN, Parfen'eva LS, Prokof'ev AV, Smirnov IA, Samoilovich SM, Jezowskii A, Mucha J, Miserek H (1995) Influence of periodic cluster superstructure on thermal conductivity of amorphous silica. Phys Solid State 37(11):1874
Bogomolov VN et al (1997) Photonic band gap phenomenon and optical properties of artificial opals. Phys Rev E 55(6):7619–7625
Bohren CF, Huffman DR (1983) Absorption and scattering of light by small particles. Wiley, New-York, p 530
Boriskina SV, Gopinath A, Dal Negro L (2008) Optical gap formation and localization properties of optical modes in deterministic aperiodic photonic structures. Opt Exp 16:18813–18826
Born M, Wolf E (2002) Principles of optics, 7th edn. Press, Cambridge, Univ, p 952
Borovoi AG (2006) Multiple scattering of short waves by uncorrelated and correlated scatterers. In: Kokhanovsky AA (ed) Light scattering reviews. single and multiple light scattering. Springer, Chichester, pp 181–252
Bringi VN, Varadan VV, Varadan VK (1982) The effects on pair correlation function of coherent wave attenuation in discrete random media. IEEE Trans Antennas Propag AP-30(4):805–808
Brown GS (1980) Coherent wave propagation through a sparse concentration of particles. Radio Sci 15(3):705–710

Bruning JH, Lo, YT (1971a) Multiple scattering of EM waves by spheres I - multipole expansion and ray-optical solutions. IEEE Trans Antennas Propag AP-19(3):378–390
Bruning, JH, Lo YT (1971b) Multiple scattering of EM waves by spheres II - numerical and experimental results. IEEE Trans Antennas Propag AP-19(3):391–400
Busch K, John S (1998) Photonic band gap formation in certain self-organizing systems. Phys Rev E 58(3):3896–3908
Campos-Fernández C et al (2011) Visible light reflection spectra from cuticle layered materials. Opt Mater Exp 1:85–100
Cassagne D, Reynolds A, Jouanin C (2000) Modelling of 3D photonic crystals based on opals. Opt Quantum Electron 32:923–933
Cassier M, Hazard C (2013) Multiple scattering of acoustic waves by small sound-soft obstacles in two dimensions: mathematical justification of the Foldy-Lax model. Wave Motion 50:18–28
Centurioni E (2005) Generalized matrix method for calculation of internal light energy flux in mixed coherent and incoherent multilayers. Appl Opt 44(35):7532–7539
Challa DP, Hu G, Sini M (2014) Multiple scattering of electromagnetic waves by finitely many point-like obstacles. Math Models Methods Appl Sci 24(5):863–899
Chandrasekhar S (1960) Radiative transfer. Dover, New York, p 393
Chaumet PC, Rahmani A, Bryant GW (2003) Generalization of the coupled dipole method to periodic structures. Phys Rev B 67(16):165404-1-5
Chen H et al (2006) Effect of disorder on self-collimated beam in photonic crystal. Physica E 35:64–68
Christofi A, Stefanou N (2014) Layer multiple scattering calculations for nonreciprocal photonic structures. Int J Mod Phys B 28(2):1441012-1-16
Chwolson O (1889) Grundzüge einer mathematischen Theorie der inneren Diffusion des Lichtes. Bull l'Acad Impériale Sci St Pétersbourg 33:221–256
Conwell PR, Barber PW, Rushforth CK (1984) Resonant spectra of dielectric spheres. J Opt Soc Am A 1:62–67
Cooray MFR, Ciric IR (1989) Electromagnetic wave scattering by a system of two spheroids of arbitrary orientation. IEEE Trans Antennas Propag 37(5):608–618
Davis VA, Schwartz L (1985) Electromagnetic propagation in close-packed disordered suspensions. Phys Rev B 31:5155–5165
de Vries P, van Coevorden DV, Lagendijk A (1998) Point scatterers for classical waves. Rev Mod Phys 70(2):447–466
Debenham M, Dew GD (1981) The refractive index of toluene in the visible spectral region. J Phys E Sci Instrum 14:544–545
Deinega A et al (2011) Minimizing light reflection from dielectric textured surfaces. J Opt Soc Am A 28:770–777
Deparis O et al (2006) Colorselecting reflectors inspired from biological periodic multilayer structures. Opt Exp 14:3547–3555
Dick VP, Ivanov AP, Loiko VA (1985) Density effects of a dispersive substance on the coherent component of transmitted radiation. J Appl Spectrosc 43(4):1183–1187. https://doi.org/10.1007/BF00662341
Dick VP, Ivanov AP, Loiko VA (1986) Transfer of images by dispersion layers with high concentrations of optically soft particles. J Appl Spectrosc 44(1):111–114. https://doi.org/10.1007/BF00658334
Dick VP, Ivanov AP, Loiko VA (1987a) Characteristics of the attenuation of radiation by a monolayer of discrete scatterers. J Appl Spectrosc 47(3):966–971. https://doi.org/10.1007/BF00659446
Dick VP, Ivanov AP, Loiko VA (1987b) Features of light scattering by a single-row layer of particles. J Appl Spectrosc 46:197–202. https://doi.org/10.1007/BF00665564
Dick VP, Loiko VA (2001a) Light attenuation by disperse layers with a high concentration of oriented anisotropic spherical particles. Opt Spectrosc 91:618–622. https://doi.org/10.1134/1.1412682

Dick VP, Loiko VA (2001b) Model for coherent transmittance calculation for polymer dispersed liquid crystal films. Liquid Cryst 28:1193–1198
Dick VP, Loiko VA (2014) Transmission spectra of tunable dispersion filters of the type of small particles-liquid crystal. Opt Spectrosc 117(1):111–117. https://doi.org/10.1134/S0030400X14070066
Dick VP, Loiko VA, Ivanov AP (1997a) Angular structure of radiation scattered by monolayer of particles: experimental study. Appl Opt 36(3):4235–4240. https://doi.org/10.1364/AO.36.004235
Dick VP, Loiko VA, Ivanov AP (1997b) Light transmission by a monolayer of particles: comparison of experimental data with calculation as a single-scattering approximation. Appl Opt 36(24):6119–6122
Dick VP, Ivanov AP (1999) Extinction of light in dispersive media with high particle concentrations: applicability limits of the interference approximation. J Opt Soc Am A 16(5):1034–1039
Ding KH, Tsang L (1988) Effective propagation constants of dense nontenuous media with multispecies of particles. Electromagn Waves Appl 2(8):757–777
Domínguez S et al (2012) Optimization of 1D photonic crystals to minimize the reflectance of silicon solar cells. Photonics Nanostruct Fundam Appl 10:46–53
Dorado LA, Depine RA (2009) Modeling of disorder effects and optical extinction in three-dimensional photonic crystals. Phys Rev B 79:045124-1–7
Drzaic PS (1988) Reorientation dynamics of polymer dispersed nematic liquid-crystal films. Liquid Cryst 3(11):1543–1559
Dubova GS, Khairullina AY (1982) Diffuse transmission and reflection by a thick weakly absorbing layer with close particle packing. J Appl Spectrosc. 37:1313–1316. https://doi.org/10.1007/BF00661033
Dufva TJ, Sarvas J, Sten JC-E (2008) Unified derivation of the translational addition theorems for the spherical scalar and vector wave functions. Prog Electromagn Res B 4:79–99
Dushkina N, Dushkin C, Nagayama K (2005) Diffraction from monolayer latex crystals. Annuaire de l'Universite "St K Ohridski" Fac Phys 98:23–37
Economou EN, Zdetsis A (1989) Classical wave propagation in periodic structures. Phys Rev B 40(2):1334–1337
Evlyukhin AB et al (2010) Optical response features of Si-nanoparticle arrays. Phys Rev B 82:045404-1–12
Ewald PP (1916) Zur Begründung der Kristalloptik. Ann Physik 49(4):1–38
Fedotov VG, Sel'kin AV (2011) Multiwave Bragg diffraction and interference effects in 3D photonic crystal films. Nanosyst Phys Chem Math 2:109–115 (in Russian)
Felbacq D, Tayeb G, Maystre D (1994) Scattering by a random set of parallel cylinders. J Opt Soc Am A 11(9):2526–2538
Fikioris JG, Waterman PC (1964) Multiple scattering of waves. II. "Hole Corrections" in the scalar case. J Math Phys 5:1413–1420
Fikioris JG, Waterman PC (2013) Multiple scattering of waves. III. The electromagnetic case. J Quant Spectrosc Radiat Transf 123:8–16
Fisher IZ (1964) Statistical theory of liquids. University of Chicago Press, Chicago, p 335
Foldy LL (1945) The multiple scattering of waves. I. General theory of isotropic scattering by randomly distributed scatterers. Phys Rev 67(3):107–119
Frieser H (1975) Photographic information recording. Focal, London
Fujimura T et al (2000) Near-field optical images of ordered polystyrene particle layers and their photonic band effect. J Lumin 87–89:954–956
Fuller KA, Kattawar GW (1988a) Consummate solution to the problem of classical electromagnetic scattering by an ensemble of spheres. I: linear chains. Opt Lett 13(2):90–92
Fuller KA, Kattawar GW (1988b) Consummate solution to the problem of classical electromagnetic scattering by an ensemble of spheres. II: clusters of arbitrary configuration. Opt Lett 13(12):1063–1065

Gadomskiĭ ON, Shalin AS (2007) Effect of optical blooming of a nanocrystal monolayer and the interface between two media. J Exp Theor Phys 105:761–773. https://doi.org/10.1134/S106377610710010X

Galisteo-Lopez JF et al (2003) Optical study of the pseudogap in thickness and orientation controlled artificial opals. Phys Rev B 68:115109-1–8

Galisteo-Lopez JF, Ibisate M, Sapienza R, Froufe-Prez LS, Blanco A, Lopez C (2011) Self-assembled photonic structures. Adv Mater 23:30–69

Gantzounis G, Stefanou N (2006) Layer-multiple-scattering method for photonic crystals of non-spherical particles. Phys Rev B 73(3):035115-1–10

García-Valenzuela A, Gutiérrez-Reyes E, Barrera R (2012) Multiple-scattering model for the coherent reflection and transmission of light from a disordered monolayer of particles. JOSA A 29:1161–1179

Germogenova OA (1964) Scattering of plane electromagnetic wave on two spheres. Izv AN SSSR Ser geophys 4:648–653

Gremenok VF, Tivanov MS, Zalesski VB (2007) Solar cells based on semiconductor materials. BSU, Minsk, 222 p (in Russian)

Guérin Ch-A, Mallet P, Sentenac A (2006) Effective-medium theory for finite-size aggregates. JOSA A 23:349–358

Gunde MK, Orel ZC (2000) Absorption and scattering of light by pigment particles in solar-absorbing paints. Appl Opt 39:622–628

Gurevich MM (1931) Questions of the rational classification of light-scattering substances. Trudy GOI (Trans Opt Inst Leningr) 6(57):1–18

Gurevich MM (1953) Dependence of scattering on the wavelength of the type λ^{-8}. Zh Tekh Fiz 23:986–994 (in Russian)

Hiltner PA, Krieger IM (1969) Diffraction of light by ordered suspensions. J Phys Chem 73(7):2386–2389

Ho KM, Chan CT, Soukoulis CM (1990) Existence of a photonic gap in periodic dielectric structures. Phys Rev Lett 65:3152–3155

Hong KM (1980) Multiple scattering of electromagnetic waves by a crowded monolayer of spheres: application to migration imaging films. J Opt Soc Am 70(7):821–826

https://refractiveindex.info/

Hu G, Mantile A, Sini M (2014) Direct and inverse acoustic scattering by a collection of extended and point-like scatterers. Multiscale Model Simul 12(3):996–1027

Huang K, Li P (2010b) A two-scale multiple scattering problem. Multiscale Model Simul 8(4):1511–1534

Huang K, Solna K, Zhao H (2010a) Generalized Foldy-Lax formulation. J Comput Phys 229:4544–4553

Huang H, Lin CH, Huang ZK, Lee KY, Yu CC, Kuo HC (2010b) Double photonic quasi-crystal structure effect on GaN-based vertical-injection light-emitting diodes. Jpn J Appl Phys 49:022101

Huang K, Li P, Zhao H (2013) An efficient algorithm for the generalized Foldy-Lax formulation. J Comput Phys 234:376–398

Huisman L et al (1981) Effective-medium theory of electronic states in structurally disordered metals: application to liquid Cu. Phys Rev B 24:1824–1834

Hunt GE (1971) The effect of coarse angular discretization on calculations of the radiation emerging from a model cloudy atmosphere. JQSRT 11:309–321. https://doi.org/10.1016/0022-4073(71)90016-1

Inoue M, Ohtaka K, Yanagawa S (1982) Light scattering from macroscopic spherical bodies. II. Reflectivity of light and electromagnetic localized state in a periodic monolayer of dielectric spheres. Phys Rev B 25:689–699

Ishimaru A (1978a) Wave propagation and scattering in random media, vol 1. Academic Press, New York, Single scattering and transport theory, p 255

Ishimaru A (1978b) Wave propagation and scattering in random media. In: Multiple scattering, turbulence, rough surfaces, and remote sensing, vol 2. Academic Press, New York, 340 p

Ishimaru A, Kuga Y (1982) Attenuation constant of a coherent field in a dense distribution of particles. J Opt Soc Am 72:1317–1320
Ishimaru A, Tsang L (1988) Backscattering enhancement of random discrete scatters of moderate sizes. JOSA A 5:228–236
Ivanov AP, Loiko VA (1983) Optics of photographic layer, Minsk, 304 p
Ivanov AP, Loiko VA, Dick VP (1988) Propagation of light in close-packed disperse media. Nauka i Tekhnika, Minsk, 191 p (in Russian)
Ivanov AP, Makarevich SA, Khairullina AY (1987) Radiation propagation in tissues and liquids with close particle packing. J Appl Spectrosc 47:1077–1082. https://doi.org/10.1007/BF00667708
Ivanov P et al (2010) Lattice constant tuning and disorder effects in 3D colloidal photonic crystals. J Display Technol 6(1):14–21
Javanaud C, Thomas A (1988) Multiple scattering using the Foldy-Twersky integral equation. Ultrasonics 26:341–343
Ji L, Varadan VV (2011) Fishnet metastructure for efficiency enhancement of a thin film solar cell. J Appl Phys 110:043114-1–8
Joannopoulos JD, Johnson SG, Winn JN, Meade RD (2008) Photonic crystals: molding the flow of light. University Press, Princeton
John S (1987) Strong localization of photons in certain disordered dielectric superlattices. Phys Rev Lett 58(23):2486–2489
John S, Rangarajan R (1988) Optimal structures for classical wave localization: an alternative to the Ioffe-Regel criterion. Phys Rev B 38(14):10101–10104
Kachan SM, Ponyavina AN (2002) The spatial ordering effect on spectral properties of close-packed metallic nanoparticle monolayers. Surf Sci 507–510:603–608
Kachan SM, Ponyavina AN (2002a) Spectral properties of close-packed monolayers consisting of metal nanospheres. J Phys Condens Matter 14:103–111
Kachan SM, Ponyavina AN (2002b) Spectral characteristics of confined photonic and plasmonic nanostructures. PSPIE 4705:88–94
Kachan S, Stenzel O, Ponyavina A (2006) High-absorbing gradient multilayer coatings with silver nanoparticles. Appl Phys B 84:281–287
Kafesaki M, Economou E (1999) Multiple-scattering theory for three-dimensional periodic acoustic composites. Phys Rev B 60:11993–12001
Kafesaki M, Penciu RS, Economou EN (2000) Air bubbles in water: a strongly multiple scattering medium for acoustic waves. Phys Rev Lett 84(26):6050–6053
Kaliteevski MA et al (2003) Disorder-induced modification of the attenuation of light in a two-dimensional photonic crystal with complete band gap. Phys Status Solidi(a) 195(3):612–617
Kalmykov AE, Shepilov MP, Sycheva GA (2000) Electron microscopic investigation of spatial ordering of particles formed in the course of liquid phase separation in sodium borosilicate glass. Glass Phys Chem 26:307–309. https://doi.org/10.1007/BF02738304
Kambe K (1967) Theory of low-energy electron diffraction. I. Application of the cellular method to monoatomic layers. Z Naturforschg 22a:322–330
Kambe K (1968) Theory of low-energy electron diffraction. II. Cellular method for complex monolayers and multilayers. Z Naturforschg 23a:1280–1294
Katsidis CC, Siapkas DI (2002) General transfer-matrix method for optical multilayer systems with coherent, partially coherent, and incoherent interference. Appl Opt 41(19):3978–3987
Kayes BM, Atwater HA, Lewis NS (2005) Comparison of the device physics principles of planar and radial pn junction nanorod solar cells. J Appl Phys 97:114302–114312
Kelly JR, Wu W (1993) Multiple scattering effects in polymer dispersed liquid crystals. Liq Cryst 14(6):1683–1694
Kinnan MK et al (2009) Plasmon coupling in two-dimensional arrays of silver nanoparticles: I. Effect of the dielectric medium. J Phys Chem C 113:7079–7084
Kocher-Oberlehner G et al (2012) Planar photonic solar concentrators for building-integrated photovoltaics. Sol Energy Mater Sol Cells 104:53–57

Koenderink AF et al (2000) Enhanced backscattering from photonic crystals. Phys Lett A 268:104–111
Koenderink AF, Lagendijk A, Vos WL (2005) Optical extinction due to intrinsic structural variations of photonic crystals. Phys Rev B 72:153102-1–4
Kohn W, Rostoker N (1954) Solution of the Schrödinger equation in periodic lattices with an application to metallic lithium. Phys Rev 94(5):1111–1120
Kokhanovsky AA (2001) Optics of light scattering media: problems and solutions, 2nd edn. Springer-Praxis, Chichester, p 262
Kokhanovsky AA, Korolevich AN (1998) The dependence of the diffuse reflection coefficient of blood on the concentration of red cells. J Colloid Interface Sci 208:575–577
Kolyadin AI (1956) Anomalous scattering in glass. Dokl Akad Nauk SSSR 109:64–67 (in Russian)
Kondo T et al (2002) Transmission characteristics of a two-dimensional photonic crystal array of dielectric spheres using subterahertz time domain spectroscopy. Phys Rev B 66:033111-1–4
Kondo T et al (2004) Refractive index dependence of the transmission properties for a photonic crystal array of dielectric spheres. Phys Rev B 70:235113-1–6
Konkolovich AV, Presnyakov VV, Zyryanov VYa, Loiko VA, Shabanov VF, (2000) Interference quenching of light transmitted through a monolayer film of polymer-dispersed nematic liquid crystal. J Exp Theor Phys Lett 71(12):486–488
Korringa J (1947) On the calculation of the energy of a Bloch wave in a metal. Physica (Utrecht) 13(6–7):392–400
Kosobukin VA (2005) On the theory of diffraction of light in photonic crystals with allowance for interlayer disordering. Phys Solid State 47:2035–2045. https://doi.org/10.1134/1.2131141
Krieger IM, O'Neill FM (1968) Diffraction of light by arrays of colloidal spheres. J Am Chem Soc 90(12):3114–3120
Kuga Y, Ishimaru A (1984) Retroreflectance from a dense distribution of spherical particles. J Opt Soc Am A 1(8):831–835
Kurokawa Y, Miyazaki H, Jimba Y (2002) Light scattering from a monolayer of periodically arrayed dielectric spheres on dielectric substrates. Phys Rev B 65:201102-1–4
Kurokawa Y, Miyazaki H, Jimba Y (2004a) Optical band structure and near-field intensity of a periodically arrayed monolayer of dielectric spheres on dielectric substrate of finite thickness. Phys Rev B 69:155117-1–9
Kurokawa Y, Jimba Y, Miyazaki H (2004b) Internal electric-field intensity distribution of a monolayer of periodically arrayed dielectric spheres. Phys Rev B 70:155107-1–5
Kuz'min VL, Romanov VP, Obraztsov EP (2001) Fluctuations of dielectric constant in a system of hard spheres. Opt Spectrosc 91:913–920. https://doi.org/10.1134/1.1429707
Lado F (1968) Equation of state of the hard-disk fluid from approximate integral equations. J Chem Phys 49(7):3092–3096
Lagendijk A, van Tiggelen BA (1996) Resonant multiple scattering of light. Phys Rep 270:143–215
Lamb W, Wood DM, Ashcroft NW (1980) Long-wavelength electromagnetic propagation in heterogeneous media. Phys Rev B 21(6):2248–2266
Lavrinenko AV, Wohlleben W, Leyrer RJ (2009) Influence of imperfections on the photonic insulating and guiding properties of finite Si-inverted opal crystals. Opt Exp 17(2):747–760
Lax M (1951) Multiple scattering of waves. Rev Mod Phys 23(4):287–310
Lax M (1952) Multiple scattering of waves. II. The effective field in dense systems. Phys Rev 85(4):621–629
Lenoble J (ed) (1985) Radiative transfer in scattering and absorbing atmospheres: standard computational procedures. Deepak Publishing, Hampton
Leung KM, Liu YF (1990a) Photon band structures: the plane-wave method. Phys Rev B 41(14):10188–10190
Leung KM, Liu YF (1990b) Full vector wave calculation of photonic band structures in face-centered cubic dielectric media. Phys Rev Lett 65(21):2646–2649
Levine D, Steinhardt PJ (1984) Quasicrystals: a new class of ordered structures. Phys Rev Let 53:2477

Li CQ et al (2013) Assembling of silicon nanoflowers with significantly enhanced second harmonic generation using silicon nanospheres fabricated by femtosecond laser ablation. J Phys Chem C 117:24625–24631

Li Z-Y, Zhang Z-Q (2000) Fragility of photonic band gaps in inverse-opal photonic crystals. Phys Rev B 62(3):1516–1519

Liang C, Lo YT (1967) Scattering by two spheres. Radio Sci 2:1481–1495

Liao J, Ji C (2014) Extended Foldy-Lax approximation on multiple scattering. Math Model Anal 19(1):85–98

Linton CM, Martin PA (2005) Multiple scattering by random configurations of circular cylinders: second-order corrections for the effective wavenumber. J Acoust Soc Am 117(6):3413–3423

Lloyd P, Berry MV (1967) Wave propagation through an assembly of spheres: IV. Relations between different multiple scattering theories. Proc Phys Soc 91:678–688

Lock JA, Chiu C-L (1994) Correlated light scattering by a dense distribution of condensation droplets on a window pane. Appl Opt 33(21):4663–4671

Loewinger E, Gordon A, Weinreb A, Gross J (1964) Analysis of a micromethod for transmission oximetry of whole blood. J Appl Physiol 19:1179–84

Loiko VA (1981) Reflection by disperse layer with different packing density. In: Proceedings of the academy of sciences of BSSR. Series of physics and mathematics, No. 4, pp 241–245 (in Russian)

Loiko VA (2013) Polymer films with nanosized liquid-crystal droplets: extinction, polarization, phase, and light focusing. In: Wang ZM (ed) Nanodroplets. Springer, New York, pp 195–235 Chap. 9

Loiko VA, Berdnik VV (2003) Light scattering in a disperse layer with partially ordered soft particles. In: Van Tiggelen B, Skipetrov S (eds) Proceedings of NATO ASI wave scattering in complex media: from theory to applications. Series II. Kluwer Academic Publishers, Dordrecht, pp 535–551

Loiko VA, Berdnik VV (2006) Light scattering in a layer of correlatively arranged optically soft particles. Opt Spectrosc 101(2):303–308. https://doi.org/10.1134/S0030400X06080200

Loiko VA, Dick VP (2003) Coherent transmittance of a polymer dispersed liquid crystal film in a strong field: effect of correlation and polydispersity of droplets. Opt Spectrosc 94:595–599. https://doi.org/10.1134/1.1570487

Loiko VA, Dick, VP, Ivanov AP (1984) A method to determine average particle size. SU patent No1118176, 8 June 1984

Loiko VA, Dubovik OV (1986) Program of international congress of photographic science. In: Statistics of particle distribution in a highly-concentrated photosensitive layer, 10-17 September. Cologne/Köln, p 306

Loiko VA, Ivanov AP, Dik VP (1985) Application of the radial distribution function to the analysis of light scattering in a dispersed medium. J Appl Spectrosc 42(5):571–576. https://doi.org/10.1007/BF00661410

Loiko VA, Maschke U, Zyryanov VYa, Konkolovich AV, Miskevich AA, (2011) Coherent transmission and angular structure of light scattering by monolayer films of polymer dispersed liquid crystals with inhomogeneous boundary conditions. Opt Spectrosc 111(6):866–872. https://doi.org/10.1134/S0030400X11130121

Loiko VA, Miskevich AA (2005a) Propagation of light through a monolayer of particles: analysis of phase and coherent transmittance. Opt Spectrosc 98:61–67. https://doi.org/10.1134/1.1858041

Loiko VA (1991) Fluctuations in number of irregularities and noise properties of the dispersion medium. Opt J 11:40–45 (in Russian)

Loiko VA, Dubovik OV (1989a) Fluctuations in transmission coefficient of a single layer of discrete scatterers. Opt Spectrosc 67(2):233–235

Loiko VA, Dubovik OV (1989b) Noise properties of the medium to reproduce the optical image due to fluctuations in the number of irregularities. Optoelectron Instrum Data Process 6:11–15 (in Russian)

Loiko VA, Dubovik OV (1990) Features of formation of optical noise in the particulate monolayer. Optoelectron Instrum Data Process 1:51–53 (in Russian)
Loiko VA, Konkolovich AV (1994) Concentrated recording media: theory of optical noise for monolayers. Opt Spectrosc 77(6):899–905
Loiko VA, Konkolovich AV (1995) The RMS granularity and Wiener spectrum of concentrated photolayers. Sci Appl Photo 37(2):22–231
Loiko VA, Konkolovich AV (1997) Transmittance and spatial optical noise of polymer dispersed liquid crystal layers. Mol Cryst Liq Cryst 303:41–46
Loiko VA, Konkolovich AV (1998a) Spatial optical noise of a monolayer of discrete inhomogeneities: I. Basic relations for the Wiener spectrum. Opt Spectrosc 85(4):563–567
Loiko VA, Konkolovich AV (1998b) Spatial optical noise of a monolayer of discrete inhomogeneities: II. The wiener spectrum in the model of random substitution mixture. Opt Spectrosc 85(4):568–573
Loiko VA, Konkolovich AV (2000) Interference effect of coherent transmittance quenching: theoretical study of optical modulation by surface ferroelectric liquid crystal droplets. J Phys D Appl Phys 33:2201–2210
Loiko VA, Konkolovich AV (2001a) Propagation of polarized light via polymer-dispersed liquid crystal film. J Opt B Quantum Semiclass Opt 3:155–158
Loiko VA, Konkolovich AV (2001b) Coherent transmittance quenching effect in thin films of polymer-dispersed ferroelectric liquid crystals. Opt Spectrosc 90(5):760–764
Loiko VA, Miskevich A (2004) The adding method for coherent transmittance and reflectance of a densely packed layer. J Quant Spectrosc Radiat Transf 88:125–138
Loiko VA, Miskevich AA (2005b) Light propagation through a monolayer of discrete scatterers: analysis of coherent transmission and reflection coefficients. Appl Opt 44(18):3759–3768
Loiko VA, Miskevich AA (2013) Coherent transmission and reflection spectra of ordered structures from spherical alumina particles. Opt Spectrosc 115(2):274–282
Loiko VA, Molochko VI (1995) Coherent transmission and reflection of a monolayer of discrete scatterers at oblique incidence of the light wave. Opt Spectrosc 79:304–310
Loiko VA, Molochko VI (1996) Coherent transmission and reflection by a monolayer of discrete scatterers. Part Part Syst Char 13:227–233
Loiko VA, Molochko VI (1998) The model of adding for coherent component in a concentrated layer of discrete scatterers. Opt Spectrosc 84:755–760
Loiko VA, Ruban GI (2000) Light absorption and scattering by a photolayer with closely packed particles. Opt Spectrosc 88:756–761
Loiko VA, Ruban GI (2004) Absorption by a layer of densely packed subwavelength-sized particles. J Quant Spectrosc Radiat Transf 89:271–278
Loiko VA, Zyryanov VYa, Maschke U, Konkolovich AV, Miskevich AA, (2012) Small-angle light scattering and transmittance of polymer film, containing liquid crystal droplets with inhomogeneous boundary conditions. J Quant Spectrosc Radiat Transf 113:2585–2592
Loiko VA, Ivanov AP, Dick VP (1986) Estimate of correlation radius in a disperse medium with close packing of particles. Opt Spectrosc 60(5):649–651
Loiko VA, Dick VP, Molochko VI (1998) Monolayers of discrete scatterers: comparison of the single-scattering and quasi-crystalline approximations. JOSA A 15(9):2351–2354
Loiko VA, Dick VP, Ivanov AP (1999) Passage of light through a dispersion medium with a high concentration of discrete inhomogeneities: experiment. Appl Opt 38(12):2640–2646
Loiko VA, Dick VP, Ivanov AP (2000) Features in coherent transmittance of a monolayer of particles. J Opt Soc Am A 17(11):2040–2045
Loiko VA, Konkolovich AV, Miskevich AA (2005a) Polymer dispersed liquid crystal films: adding model for coherent field. Proc SPIE 6023:41–48
Loiko VA, Miskevich AA, Konkolovich AV (2005b) Propagation of light in polymer dispersed liquid crystal films: the adding model for coherent field. Mol Cryst Liq Cryst 433:65–72

Loiko VA, Zyryanov VYa, Konkolovich AV, Miskevich AA, (2016a) Light transmission of polymer-dispersed liquid crystal layer composed of droplets with inhomogeneous surface anchoring. Opt Spectrosc 120(1):143–152

Loiko VA, Krakhalev MN, Konkolovich AV, Prishchepa OO, Miskevich AA, Zyryanov VYa (2016b) Experimental results and theoretical model to describe angular dependence of light scattering by monolayer of nematic droplets. J Quant Spectrosc Radiat Transf 178:263–268

Lommel E (1887) Die Photometrie der diffusen Zurückwerfung. Sitzber Acad Wissensch. München 17:95–124

Luque A, Hegedus S (eds) (2011) Handbook of photovoltaic science and engineering, 2nd edn. Wiley, Chichester, p 1132

Ma Y, Varadan VV, Varadan VK (1988) Scattered intensity of a wave propagating in a discrete random medium. Appl Opt 27:2469–2477

Ma Y, Varadan VK, Varadan VV (1990) Enhanced absorption due to dependent scattering. J Heat Transf 112:402–407

Mackowski D (1991) Analysis of radiative scattering for multiple sphere configurations. Proc R Soc Lond A 433:599–614

Mackowski D (1994) Calculation of total cross sections of multiple-sphere clusters. J Opt Soc Am A 11:2851–2861

Markel VA (2016) Introduction to the Maxwell-Garnett approximation. JOSA A 33(7):1244–1256. https://doi.org/10.1364/JOSAA.33.001244

Mathur NC, Yeh KC (1964) Multiple scattering of electromagnetic waves by random scatterers of finite size. J Math Phys 5(11):1619–1628

Maurel A, Martin PA, Pagneux V (2010) Effective propagation in a one-dimensional perturbed periodic structure: comparison of several approaches. Waves Random Complex Media 20(4):634–655

Maxwell Garnett JC (1904) Colours in metal glasses and in metallic films. Phil Trans R Soc Lond A 203:385–420

Mazurenko DA et al (2003) Ultrafast optical switching in three-dimensional photonic crystals. Phys Rev Lett 91(21):213903-1- 4

Meade RD et al (1993) Accurate theoretical analysis of photonic band-gap materials. Phys Rev B 48(11):8434–8437

Meisels R, Kuchar F (2007) Density-of-states and wave propagation in two-dimensional photonic crystals with positional disorder. J Opt A Pure Appl Opt 9:S396–S402

Mie G (1908) Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen. Ann Phys 25:377–445

Minin IYa, (1988) Theory of radiation transfer in the atmosphere of planets. Nauka, Moscow (in Russian)

Mishchenko MI (2008a) Multiple scattering by particles embedded in an absorbing medium. 1. Foldy-Lax equations, order-of-scattering expansion, and coherent field. Opt Exp 16(3):2288–2301

Mishchenko MI (2008b) Multiple scattering by particles embedded in an absorbing medium. 2. Radiative transfer equation. J Quant Spectrosc Radiat Transf 109:2386–2390

Mishchenko MI, Zakharova NT, Khlebtsov NG, Videen G, Wriedt T (2016) Comprehensive thematic T-matrix reference database: a 2014–2015 update. J Quant Spectrosc Radiat Transf 178:276–283. https://doi.org/10.1016/j.jqsrt.2015.11.005

Mishchenko MI (1994) Asymmetry parameters of the phase function for densely packed scattering grains. J Quant Spectrosc Radiat Transf 52:95–110

Mishchenko MI et al (2007) Multiple scattering by random particulate media: exact 3D results. Opt Exp 15(6):2822–2836

Mishchenko MI, Videen G, Babenko VA, Khlebtsov NG, Wriedt T (2004) T-matrix theory of electromagnetic scattering by particles and its applications: a comprehensive reference database. J Quant Spectrosc Radiat Transf 88:357–406

Mishchenko MI, Travis LD, Lacis AA (2006) Multiple scattering of light by particles. Press, Cambridge, Univ, p 478
Miskevich AA, Loiko VA (2011a) Two-dimensional planar photonic crystals: calculation of coherent transmittance and reflectance at normal illumination under the quasicrystalline approximation. J Quant Spectrosc Radiat Transf 112.:1082–1089
Miskevich AA, Loiko VA (2011b) Coherent transmission and reflection of a two-dimensional planar photonic crystal. JETP 113(1):1–13
Miskevich AA, Loiko VA (2013a) Periodic structures with spherical alumina particles: transmission and reflection spectra. Semicond Phys Quantum Electron Optoelectron 16(2):177–184
Miskevich AA, Loiko VA (2013b) Spectra of coherent transmittance and reflectance of periodic, Fibonacci, and Thue-Morse multilayers of dielectric particles. Nanosyst Phys Chem Math 4:778–794
Miskevich AA, Loiko VA (2014a) Light absorption by a layered structure of silicon particles as applied to the solar cells: theoretical study. J Quant Spectrosc Radiat Transf 146:355–64. https://doi.org/10.1016/j.jqsrt.2013.12.008i
Miskevich AA, Loiko VA (2014b) Layered periodic disperse structures of spherical alumina particles: coherent transmittance and reflectance spectra. J Quant Spectrosc Radiat Transf 136:58–70
Miskevich AA, Loiko VA (2014c) Method for retrieving the refractive index of ordered particles from data on the photonic band gap. JETP 119(2):211–226
Miskevich AA, Loiko VA (2015a) Solar cells based on particulate structure of active layer: investigation of light absorption by an ordered system of spherical submicron silicon particles. J Quant Spectrosc Radiat Transf 167:23–39. https://doi.org/10.1016/j.jqsrt.2015.08.003
Miskevich AA, Loiko VA (2015b) Three-dimensional ordered particulate structures: method to retrieve characteristics from photonic band gap data. J Quant Spectrosc Radiat Transf 151:260–268
Mittleman DM et al (1999) Optical properties of planar colloidal crystals: Dynamical diffraction and the scalar wave approximation. Chem Phys 111(1):345–354
Miyazaki HT et al (2000) Photonic band in two-dimensional lattices of micrometer-sized spheres mechanically arranged under a scanning electron microscope. J Appl Phys 87(10):7152–7158
Miyazaki H, Ohtaka K (1998) Near-field images of a monolayer of periodically arrayed dielectric spheres. Phys Rev B 58:6920–6937
Modinos A (1987) Scattering of electromagnetic waves by a plane of spheres-formalism. Modinos Physica A 141:575–588
Modinos A et al (2001) On wave propagation in inhomogeneous systems. Physica B 296:167–173
Modinos A, Yannopapas V, Stefanou N (2000) Scattering of electromagnetic waves by nearly periodic structures. Phys Rev B 61:8099–8107
Modinos A, Stefanou N, Yannopapas V (2001) Applications of the layer-KKR method to photonic crystals. Opt Exp 8(3):197–202
Morales P, Andueza A, Sevilla J (2013) Effect of dielectric permittivity variation in the transmission spectra of noncompact 2D-arrays of dielectric spheres. J Appl Phys 113:084906-1–4
Moroz A (1994) Inward and outward integral equations and the KKR method for photons. J Phys Condens Matter 6:171–182
Moroz A (1995) Density-of-states calculations and multiple-scattering theory for photons. Phys Rev B 51(4):2068–2081
Mouldi A, Kanzari M (2012) Design of microwave devices exploiting Fibonacci and hybrid periodic/Fibonacci one dimensional photonic crystals. Prog Electromagn Res B 40:221–240
Muskens O, Koenderink AF, Vos WL (2011) Broadband coherent backscattering spectroscopy of the interplay between order and disorder in three-dimensional opal photonic crystals. Phys Rev B 83(155101):1–9
Nair RV, Jagatap BN (2012) Engineering disorder in three-dimensional photonic crystals. Photonics Nanostruct Fundam Appl 10:581–588
Nelson J (1989) Test of a mean field theory for the optics of fractal clusters. J Mod Opt 36:1031–1057

Nicorovici NA, McPhedran RC, Da Ke B (1995) Propagation of electromagnetic waves in periodic lattices of spheres: Green's function and lattice sums. Phys Rev E 51(1):690–702
Ohtaka K (1979) Energy band of photons and low-energy photon diffraction. Phys Rev B 19(10):5057–5067
Ohtaka K (1980) Scattering theory of low-energy photon diffraction. J Phys C Solid State Phys 13:667–680
Ohtaka K et al (2000) Photonic band effects in a two-dimensional array of dielectric spheres in the millimeter-wave region. Phys Rev B 61:5267–5279
Ohtaka K, Yamaguti S (2004) Optical excitation of optically inactive photonic band modes. Photonics Nanostruct Fundam Appl 2:73–79
Ohtaka K, Ueta T, Amemiya K (1998) Calculation of photonic bands using vector cylindrical waves and reflectivity of light for an array of dielectric rods. Phys Rev B 57(4):2550–2568
Ohtaka K, Inoue J, Yamaguti S (2004) Derivation of the density of states of leaky photonic bands. Phys Rev B 70:035109-1–13
Okada Y, Kokhanovsky AA (2009) Light scattering and absorption by densely packed groups of spherical particles. J Quant Spectrosc Radiat Transf 110:902–917
Ondris-Crawford R, Boyko EP, Wagner BG, Erdmann JH, Zumer S, Doane JW (1991) Microscope textures of nematic droplets in polymer dispersed liquid crystal. J Appl Phys 69:6380–6
Ornstein LS, Zernike F (1914) Accidental deviations of density and opalescence at the critical point of a single substance. Proc Acad Sci 17:793–806
Palik ED (ed) (1985) Handbook of optical constants of solids, vol 3. Academic, San Diego, 1985–1998, vol 1, 804 p
Palik ED (ed) (1991) Handbook of optical constants of solids, vol II. Academic, San Diego, p 770
Parnell WJ, Abrahams ID (2010) Multiple point scattering to determine the effective wavenumber and effective material properties of an inhomogeneous slab. Waves Random Complex Media 20(4):678–701
Parnell WJ, Martin PA (2011) Multiple scattering of flexural waves by random configurations of inclusions in thin plates. Wave Motion 48:161–175
Parnell W, Abrahams I, Brazier-Smith P (2010) Effective properties of a composite half-space: exploring the relationship between homogenization and multiple-scattering theories. Q J Mech Appl Math 63:145–175
Pendry JB (1974) Low energy electron diffraction. Academic Press, New York, p 407
Pendry JB (1996) Calculating photonic band structure. J Phys Condens Matter 8:1085–1108
Pendry JB, MacKinnon A (1992) Calculation of photon dispersion relations. Phys Rev Lett 69(19):2772–2775
Percus JK, Yevick GJ (1958) Analysis of classical statistical mechanics by means of collective coordinates. Phys Rev 110:1–13
Peterson B, Ström S (1973) T matrix for electromagnetic scattering from an arbitrary number of scatterers and representations of E(3). Phys Rev D 8(10):3661–3678
Peterson B, Ström S (1974) T-matrix formulation of electromagnetic scattering from multilayered scatterers. Phys Rev D 10:2670–2684
Plass GN, Kattawar GW, Catchings FF (1973) Matrix operator theory of radiative transfer. 1: Rayleigh scattering. Appl Opt 12:314–329
Ponyavina AN (1998) Selection of optical radiation in scattering in partially ordered disperse media. J Appl Spectrosc 65:752–765
Ponyavina AN, Silvanovich NI (1994) Interference effects and spectral characteristics of multilayer scattering media. Opt Spektrosk 76:581–588
Ponyavina AN, Sil'vanovlch NI (1990) Coherent reflection and transmission of close-packed scatterer monolayers. J Appl Spectrosc 53(2):884–889
Ponyavina A, Kachan S, Sil'vanovich N (2004) Statistical theory of multiple scattering of waves applied to three-dimensional layered photonic crystals. JOSA B 21(10):1866–1875
Pouya C, Stavenga DG, Vukusic P (2011) Discovery of ordered and quasiordered photonic crystal structures in the scales of the beetle Eupholus magnificus. Opt Exp 19:11355–11364

Prasad T, Colvin VL, Mittleman DM (2007) The effect of structural disorder on guided resonances in photonic crystal slabs studied with terahertz time-domain spectroscopy. Opt Exp 15(25):16954–16965
Psarobas I, Stefanou N, Modinos A (2000) Scattering of elastic waves by periodic arrays of spherical bodies. Phys Rev B 62:278–291
Reference solar spectral irradiance: air mass 1.5 [Electronic resource]/American society for testing and materials (ASTM) terrestrial reference spectra for photovoltaic performance evaluation - mode of access https://rredc.nrel.gov/solar/spectra/am1.5
Rengarajan R et al (2005) Effect of disorder on the optical properties of colloidal crystals. Phys Rev E 71:016615-1–11
Rockstuhl C et al (2007) Enhanced transmission of periodic, quasiperiodic, and random nanoaperture arrays. Appl Phys Lett 91:151109-1–3
Rojas-Ochoa LF, Mendez-Alcaraz JM, Schurtenberger P, Saenz JJ, Scheffold F (2004) Photonic properties of strongly correlated colloidal liquids. Phys Rev Lett 93:073903-1–4
Rosasco GJ, Bennett HS (1978) Internal field resonance structure: Implications for optical absorption and scattering by microscopic particles. J Opt Soc Am 68:1242–1250
Roth LM (1974) Effective-medium approximation for liquid metals. Phys Rev B 9(6):2476–2484
Rudhardt D et al (2003) Phase switching of ordered arrays of liquid crystal emulsions. Appl Phys Lett 82(16):2610–2612
Rundquist PA et al (1989) Dynamical Bragg diffraction from crystalline colloidal arrays. J Chem Phys 91(8):4932–4941
Rybin MV, Samusev KB, Limonov MF (2008) On dip broadening in transmission spectra of synthetic opals. Phys Solid State 50:436–445. https://doi.org/10.1134/S1063783408030074
Rybin MV et al (2009) Fano resonance between mie and bragg scattering in photonic crystals. Phys Rev Lett 103:023901-1–4
Ryu H-Y, Hwang J-K, Lee Y-H (1999) Effect of size nonuniformities on the band gap of two-dimensional photonic crystals. Phys Rev B 59(8):5463–5469
Sainidou R et al (2005) A layer-multiple-scattering method for phononic crystals and heterostructures of such. Comput Phys Commun 166:197–240
Saritas M, McKell H (1988) Diffusion length studies in silicon by the surface photovoltage method. Solid-State Electron 31:835–842
Sarofim AF, Hottel HC, Fahimion EJ (1968) Scattering of radiation by particle layer. AIAA J 6:2262–2266
Satpathy S, Zhang Z, Salehpour MR (1990) Theory of photon bands in three-dimensional periodic dielectric structures. Phys Rev Lett 64(11):1239–1242
Schilling J et al (2001) A model system for two-dimensional and three-dimensional photonic crystals: macroporous silicon. J Opt A Pure Appl Opt 3:S121–S132
Schuster A (1905) Radiation through a foggy atmosphere. Astrophys J 21:1–22
Schuster CS et al (2015) Plasmonic and diffractive nanostructures for light trapping - an experimental comparison. Optica 2:194–200
Schwartz L, Plona TJ (1984) Ultrasonic propagation in closepacked disordered suspensions. J Appl Phys 55(11):3971–3977
Shalin AS (2009) Effect of the absolute transparency of an ordered nanocomposite. JETP Lett 90:257–262. https://doi.org/10.1134/S0021364009160073
Shalin AS (2010) Broadband blooming of a medium modified by an incorporated layer of nanocavities. JETP Lett 91:636–642. https://doi.org/10.1134/S0021364010120052
Shalin A, Moiseev S (2009) Optical properties of nanostructured layers on the surface of an underlying medium. Opt Spectrosc 106:916–925
Sharma AK et al (2012) Determination of minority carrier diffusion length from distance dependence of lateral photocurrent for side-on illumination. Sol Energy Mater Sol Cells 100:48–52
Shatilov AV (1962) Anomalous scattering as a case of scattering by a system of particles. Opt Spectrosc 13:412

Sheng X et al (2011) Optimization-based design of surface textures for thin-film Si solar cells. Opt Exp 19:A841–A850
Shepilov MP (2003) The problem of theoretically describing anomalous light scattering by liquating glasses, caused by interparticle interference. J Opt Technol 70(12):882–887. https://doi.org/10.1364/JOT.70.000882
Shi L et al (2013) Monodisperse silicon nanocavities and photonic crystals with magnetic response in the optical region. Nat Commun 4:1904
Shi L, Tuzer TU, Fenollosa R, Meseguer F (2012) A new dielectric metamaterial building block with a strong magnetic response in the sub-1.5-micrometer region: silicon colloid nanocavities. Adv Mater 24:5934–5938
Sigalas M et al (1996) Localization of electromagnetic waves in two-dimensional disordered systems. Phys Rev B 53:8340–8348
Simoni F (1997) Nonlinear properties of liquid crystals and polymer dispersed liquid crystals. Singapore, World scientific, p 278
Siqueira PR, Sarabandi K (1996) Method of moments evaluation of the two-dimensional quasi-crystalline approximation. IEEE Trans Antennas Propag 44(8):1067–1077
Skorobogatiy M, Begin G, Talneau A (2005) Statistical analysis of geometrical imperfections from the images of 2D photonic crystals. Opt Exp 13(7):2487–2502
Skryshevskii AF (1980) Structural analysis of liquids and amorphous solids. Vysshaya Shkola, Moscow, 328 p (in Russian)
Sobolev VV (1956a) A treatise on radiative transfer. Gostekhizdat, Moscow. Van Nostrand, Princeton (1963)
Sobolev VV (1956b) Radiant energy transport in athmospheres of stars and planets. State publishing house of tech.-theor. literature, Moscow, 392 p (in Russian)
Soven P (1966) Approximate calculation of electronic structure of disordered alloys-application to alpha brass. Phys Rev 151:539–550
Soven P (1967) Coherent-potential model of substitutional disordered alloys. Phys Rev 156(3):809–813
Sozuer HS, Haust JW, Inguva R (1992) Photonic bands: convergence problems mith the plane-wave method. Phys Rev B 45(24):13962–13972
Spry RJ, Kosan (DJ) Theoretical analysis of the crystalline colloidal array filter. Appl Spectrosc 40(6):782–784
Staude I et al (2013) Tailoring directional scattering through magnetic and electric resonances in subwavelength silicon nanodisks. ACS Nano 7:7824–7832
Stefanou N, Modinos A (1991a) Scattering of light from a two-dimensional array of spherical particles on a substrate. J Phys Condens Matter 3:8135–8148
Stefanou N, Modinos A (1991b) Optical properties of thin discontinuous metal films. J Phys Condens Matter 3:8149–8157
Stefanou N, Modinos A (1993) Scattering of electromagnetic waves by a disordered two-dimensional array of spheres. J Phys Condens Matter 5:8859–8868
Stefanou N, Karathanos V, Modinos A (1992) Scattering of electromagnetic waves by periodic structures. J Phys Cond Mat 4:7389–7400
Stefanou N, Yannopapas V, Modinos A (1998) Heterostructures of photonic crystals: frequency bands and transmission coefficients. Comput Phys Commun 113:49–77
Stefanou N, Yannopapas V, Modinos A (2000) MULTEM 2: a new version of the program for transmission and band-structure calculations of photonic crystals. Comput Phys Commun 132:189–196
Sun W et al (2007) Modeling light scattered from and transmitted through dielectric periodic structures on a substrate. Appl Opt 46:1150–1156
Sze SM, Ng KK (2007) Physics of semiconductor devices, 3rd edn. Hoboken, Wiley Interscience, p 815

Tishkovets VP, Jockers K (2006) Multiple scattering of light by densely packed random media of spherical particles: dense media vector radiative transfer equation. J Quant Spectrosc Radiat Transf 101:54–72
Tishkovets VP (2010) Radiation diffuse scattering by a densely packed layer of spherical particles. Radio Phys Radio Astron 1(1):69–77. https://doi.org/10.1615/RadioPhysicsRadioAstronomy.v1.i1.90
Tishkovets VP (2007) Incoherent and coherent backscattering of light by a layer of densely packed random medium. J Quant Spectrosc Radiat Transf 108:454–463
Tishkovets VP, Petrova EV, Mishchenko MI (2011) Scattering of electromagnetic waves by ensembles of particles and discrete random media. J Quant Spectrosc Radiat Transf 112:2095–2127
Toušek J, Dolhov S, Toušková J (2003) Interpretation of minority carrier diffusion length measurements in thin silicon wafers. Solar Energy Mater Solar Cells 76:205–210
Trinks W (1935) Zur Vielfachstreuung an kleinen Kugeln. Annalen der Physik 22(5):561–590
Troparevsky MC et al (2010) Transfer-matrix formalism for the calculation of optical response in multilayer systems: from coherent to incoherent interference. Opt Exp 18(24):24715–24721
Tsai MA et al (2011) Enhanced conversion efficiency of a crystalline silicon solar cell with frustum nanorod arrays. Opt Exp 19:A28–A34
Tsakalakos L (2008) Nanostructures for photovoltaics. Mater Sci Eng R 62:175–189
Tsang L et al (2000–2001) Scattering of electromagnetic waves, vol 3. Wiley, New York, vol 2, Num Simulat (2001), 705 p
Tsang L et al (1992) Inversion of snow parameters from passive microwave remote sensing measurements by a neural network trained with a multiple scattering model. IEEE Trans Geosci Remote Sens 30:1015–1024
Tsang L et al (2000) Dense media radiative transfer theory based on quasicrystalline approximation with applications to passive microwave remote sensing of snow. Radio Sci 35(3):731–749
Tsang L et al (2007) Modeling active microwave remote sensing of snow using dense media radiative transfer (DMRT) theory with multiple-scattering effects. IEEE Trans Geosci Remote Sens 45(4):990–1004
Tsang L, Ishimaru A (1984) Backscattering enhancement of random discrete scatterers. J Opt Soc Am A 1(8):836–839
Tsang L, Ishimaru A (1985a) Theory of backscattering enhancement of random discrete isotropic scatterers based on the summation of all ladder and cyclical terms. J Opt Soc Am A 2(8):1331–1338
Tsang L, Ishimaru A (1985b) Radiative wave and cyclical transfer equations for dense nontenuous media. J Opt Soc Am A 2(12):2187–2194
Tsang L, Kong JA (1980) Multiple scattering of electromagnetic waves by random distributions of discrete scatterers with coherent potential and quantum mechanical formalism. J Appl Phys 51:3465–3485
Tsang L, Kong JA (1982a) Effective propagation constants for coherent electromagnetic wave propagation in media embedded with dielectric scatters. J Appl Phys 53(11):7162–7173
Tsang L, Kong JA (1983) Scattering of electromagnetic waves from a half space of densely distributed dielectric scatterers. Radio Sci 18(6):1260–1272
Tsang L, Kong JA (1992a) Radiative wave equations for vector electromagnetic propagation in dense nontenuous media. J Electromagn Waves Appl 6(3):265–286
Tsang L, Kong JA (1992b) Scattering of electromagnetic waves from a dense medium consisting of correlated mie scatterers with size distributions and applications to dry snow. J Electromagn Waves Appl 6(3):265–286
Tsang L, Kong JA, Hahashy T (1982) Multiple scattering of acoustic waves by random distribution of discrete spherical scatterers with the quasicrystalline and Percus-Yevick approximation. J Acoust Soc Am 71:552–558
Tsang L, Mandt ChE, Ding KH (1992) Monte Carlo simulations of the extinction rate of dense media with randomly distributed dielectric spheres based on solution of Maxwell's equations. Opt Lett 17(5):314–316

Tsang L, Kong JA (2000–2001) Scattering of electromagnetic waves, vol 3. Wiley, New York, vol 3 Adv Top (2001), 413 p
Tsang L, Kong JA, Ding K-H (2000–2001) Scattering of electromagnetic waves, vol 3. Wiley, New York, vol 1, Theories and applications (2000), 426 p
Twersky V (1964) On propagation in random media of discrete scatterers. Proc Symp Appl Math Am Math Soc 16:84–116, Providence, RI
Twersky V (1952a) Multiple scattering of radiation by an arbitrary planar configuration of parallel cylinders and by two parallel cylinders. J Appl Phys 23(4):407–414
Twersky V (1952b) On a multiple scattering theory of the finite grating and the wood anomalies. J Appl Phys 23:1099–1118
Twersky V (1952c) Multiple scattering of radiation by an arbitrary configuration of parallel cylinders. J Acoust Soc Am 24:42–46
Twersky V (1962a) On scattering of waves by random distributions. II. Two-space scatterer formalism. J Math Phys 3(4):724–734
Twersky V (1962b) On scattering of waves by random distributions. I. Free-space scatterer formalism. J Math Phys 3(4):700–715
Twersky V (1962c) Multiple scattering of waves and optical phenomena. J Opt Soc Am 52(2):145–171
Twersky V (1962d) Multiple scattering by arbitrary configurations in three dimensions. J Math Phys 3(1):83–91
Twersky V (1967) Multiple scattering of electromagnetic waves by arbitrary configurations. J Math Phys 8(3):589–610
Twersky V (1970a) Interface effects in multiple scattering by large, low-refracting, absorbing particles. J Opt Soc Am 60:908–914
Twersky V (1970b) Absorption and multiple scattering by biological suspensions. J Opt Soc Am 60(8):1084–1093
Twersky V (1975a) Multiple scattering of waves by the doubly periodic planar array of obstacles. J Math Phys 16(3):633–643
Twersky V (1975b) Lattice sums and scattering coefficients for the rectangular planar array. J Math Phys 16(3):644–657
Twersky V (1975c) Transparency of pair-correlated, random distributions of small scatterers, with applications to the cornea. J Opt Soc Am 65(5):524–530
Twersky V (1978) Coherent electromagnetic waves in pair-correlated random distributions of aligned scatterers. J Math Phys 19:215–230
van de Hulst HC (1957) Light scattering by small particles. Wiley Interscience, New York (Chap. 4)
van de Hulst HC (1980) Multiple light scattering. Tables, formulas, and applications, vols 1, 2. Academic, New York
van Driel HM, Vos WL (2000) Multiple bragg wave coupling in photonic band-gap crystals. Phys Rev B 62(15):9872–9875
Varadan VK et al (1983) Multiple scattering theory for waves in discrete random media and comparison with experiments. Radio Sci 18:321–327
Varadan VV et al (1987) Effects of nonspherical statistics on EM wave propagation in discrete random media. Radio Sci 22(4):491–498
Varadan VV, Varadan VK (1980) Multiple scattering of electromagnetic waves by randomly distributed and oriented dielectric scatterers. Phys Rev D 21(2):388–394
Varadan VK, Bringi VN, Varadan VV (1979) Coherent electromagnetic wave propagation through randomly distributed dielectric scatterers. Phys Rev D 19(8):2480–2489
Varadan VK, Ma Y, Varadan VV (1984) Coherent electromagnetic wave propagation through randomly distributed and oriented pair-correlated dielectric scatterers. Radio Sci 19(6):1445–1449
Varadan VK, Ma Y, Varadan VV (1985a) A multiple scattering theory for elastic wave propagation in discrete random media. J Acoust Soc Am 77(2):375–385

Varadan VV, Ma Y, Varadan VK (1985b) Propagator model including multipole fields for discrete random media. JOSA A 2(12):2195–2201
Varadan VK, Ma Y, Varadan VV (1989) Scattering and attenuation of elastic waves in random media. PAGEOPH 131:577–603
Varshalovich DA, Moskalev AN (1988) Khersonskii VK (1975) Quantum theory of angular momentum. World Scientific, Singapore, Nauka, Leningrad
Vasnetsov MV et al (2014) Photonic bandgap examination in an immersed synthetic opal. Appl Phys B 116(3):541–548
Vereshchagin VG et al (1991) Interference effects in multilayer scattering systems. J Appl Spectrosc 54(2):184–188
Vereshchagin VG, Ponyavina AN, Sil'vanovich NI (1990) Role of cooperative effects in formation of coherent field of densely-packed monolayer. Dokl Akad Nauk BSSR 34:123–126 (in Russian)
Vladimirov A, Korovin S, Surkov A, Kelm E, Pustovoy V (2011) Synthesis of luminescent Si nanoparticles using the laser induced pyrolysis. Laser Phys 21:830–5
Voishvillo NA (1957) Influence of heat treatment on scattering properties of sodium borosilicate glass. Opt Spectrosc 2:371–376 (in Russian)
Voishvillo NA (1962) To a question of coherent scattering of light in the glass. Opt Spectrosc 12:412–418
Vynck K et al (2012) Photon management in two-dimensional disordered media. Nat Mater 11:1017–1022
Wang ZY et al (2015) Broadband optical absorption by tunable Mie resonances in silicon nanocone arrays. Sci Rep 5:7810-1-6
Wang X et al (1993) Multiple-scattering theory for electromagnetic waves. Phys Rev B 47(8):4161–4167
Wang ZL, Hu L, Ren W (1994) Multiple scattering of waves by a half-space of distributed discrete scatterers with modified T-matrix approach. J Phys D Appl Phys 27:441–446
Waterman PC (1969) New formulation of acoustic scattering. J Acoust Soc Am 45(6):1417–1429
Waterman PC (1971) Symmetry, unitarity, and geometry in electromagnetic scattering. Phys Rev D 3(4):825–839
Waterman PC, Pedersen NE (1986) Electromagnetic scattering by periodic arrays of particles. J Appl Phys 59:2609–2618
Waterman PC, Truell R (1961) Multiple scattering of waves. J Math Phys 2(4):512–537
Wehrspohn RB, Üpping J (2012) 3D photonic crystals for photon management in solar cells. J Opt 14:024003–024011
Wen B et al (1990) Dense medium radiative transfer theory: comparison with experiment and application to microwave remote sensing and polarimetry. IEEE Trans Geosci Remote Sens 28(1):46–59
West R, Gibbs D, Tsang L, Fung AK (1994) Comparison of optical scattering experiments and the quasi-crystalline approximation for dense media. JOSA A 11(6):1854–1858
Wiscombe WJ (1976) On initialization, error and flux conservation in the doubling method. JQSRT 16:637–658. https://doi.org/10.1016/0022-4073(76)90056-X
Wolf P-E, Maret G (1985) Weak localization and coherent backscattering of photons in disordered media. Phys Rev Lett 55:2696–2699
Xu Y (1995) Electromagnetic scattering by an aggregate of spheres. Appl Opt 34(21):4573–4588
Xu Y (1997) Electromagnetic scattering by an aggregate of spheres: far field. Appl Opt 36(36):9496–9508
Xu Y (1998a) Efficient evaluation of vector translation coefficients in multiparticle light-scattering theories. J Comput Phys 139:137–165
Xu Y (1998b) Electromagnetic scattering by an aggregate of spheres: asymmetry parameter. Phys Lett A 249:30–36
Xu Y, Wang RT (1998) Electromagnetic scattering by an aggregate of spheres: theoretical and experimental study of the amplitude scattering matrix. Phys Rev E 58(3):3931–3948

Yablonovitch E (1987) Inhibited spontaneous emission in solid-state physics and electronics. Phys Rev Lett 58:2059–2062
Yablonovitch E, Gmitter TJ (1989) Photonic band structure: the face-centered-cubic case. Phys Rev Lett 63(18):1950–1953
Yablonovitch E et al (1991a) Donor and acceptor modes in photonic band structure. Phys Rev Lett 67(24):3380–3383
Yablonovitch E, Gmitter TJ, Leung KM (1991b) Photonic band structure: the face-centered-cubic case employing nonspherical atoms. Phys Rev Lett 67(17):2295–2298
Yamasaki T, Tsutsui T (1999) Fabrication and optical properties of two-dimensional ordered arrays of silica microspheres. Jpn J Appl Phys 38:5916–5921
Yannopapas V (2014) Layer-multiple-scattering method for photonic structures of general scatterers based on a discrete-dipole approximation/T-matrix point-matching method. J Opt Soc Am B 31(3):631–636
Yannopapas V, Stefanou N, Modinos A (1997) Theoretical analysis of the photonic band structure of face-centred cubic colloidal crystals. J Phys Condens Matter 9:10261–10270
Yannopapas V, Modinos A, Stefanou N (1999) Optical properties of metallodielectric photonic crystals. Phys Rev B 60:5359–5365
Yannopapas V, Modinos A, Stefanou N (2002) Scattering and absorption of light by periodic and nearly periodic metallodielectric structures. Opt Quantum Electron 34:227–234
Yannopapas V, Modinos A, Stefanou N (2003) Anderson localization of light in inverted opals. Phys Rev B 68:193205-1–4
Yano S et al (2002) Optical properties of monolayer lattice and three-dimensional photonic crystals using dielectric spheres. Phys Rev B 66:075119–7
Yu FTS (1973) Introduction to diffraction. MTI Press, Cambridge, Information processing and holography
Zaccanti G, Del Bianco S, Martelli F (2003) Measurements of optical properties of high-density media. Appl Opt 42:4023–4030
Zamkovets AD, Kachan SM, Ponyavina AN (2003a) Surface plasmon resonances and light selection in metal-dielectric nanostructures of various spatial arrangement. In: Borisenko VE et al Physics, chemistry and application of nanostructures: reviews and short notes to nanomeeting 2003, Minsk, 20-23 May 2003. World Scientific, Singapore, pp 151–154
Zamkovets AD et al (2003b) Optical spectra of metal-dielectric nanocomposites with a layered subwave structure. J Appl Spectrosc 70(4):593–598
Zege EP, Kokhanovsky AA (1991) Model of an amplitude-phase screen in the optics of close-packed media. Opt Spectrosc 70(3):367–370
Zege EP, Ivanov AP, Katsev IL (1991a) Image transfer through a scattering medium. Springer, Berlin
Zege EP, Katsev IL, Kokhanovsky AA (1991b) Fenomenological model of optical properties of strongly scattered layers and its application to the optics of foams. Opt Spectrosc 71:835–840
Zender CS, Talamantes J (2006) Solar absorption by Mie resonances in cloud droplets. J Quant Spectrosc Radiat Transf 98:122–129
Zhang Z, Satpathy S (1990) Electromagnetic wave propagation in periodic structures: bloch wave solution of Maxwell's Equations. Phys Rev Lett 65(21):2650–2653
Zharkova GM, Sonin AS (1994) Liquid-crystal composites. Nauka, Novosibirsk, 214 p (in Russian)
Zhukas LA et al (2014) Size-dependent optical properties of polyethylene powders in far-IR region: on the way to universal matrix. J Quant Spectrosc Radiat Transf 147:1–7
Ziman J (1979) Models of disorder. Press, Cambridge, Univ, p 525
Zyryanov VYa, Presnyakov VV, Loiko VA (2001) High contrast light modulator based on PDNLC monolayer. Mol Cryst Liq Cryst 368:215–222. https://doi.org/10.1080/10587250108029949
Zywietz U, Reinhardt C, Evlyukhin AB, Birr T, Chichkov BN (2014a) Generation and patterning of Si nanoparticles by femtosecond laser pulses. Appl Phys A 114:45–50
Zywietz U, Evlyukhin AB, Reinhardt C, Chichkov BN (2014b) Laser printing of silicon nanoparticles with resonant optical electric and magnetic responses. Nat Commun 5:1–7. Article no. 3402

Chapter 3
Fast Stochastic Radiative Transfer Models for Trace Gas and Cloud Property Retrievals Under Cloudy Conditions

Dmitry S. Efremenko, Adrian Doicu, Diego Loyola and Thomas Trautmann

3.1 Introduction

Farman et al. (1985) revealed the annual depletion of ozone above the Antarctic. This finding stimulated intensive research on climatology and ozone photochemistry (Smirnov 1975). Clouds are an important component of the global hydrological cycle and play a major role in the Earth's climate system through their strong impact on radiation processes. The interplay of sunlight with clouds imposes major challenges for satellite remote sensing, both in terms of the spatial complexity of real clouds and the dominance of multiple scattering in radiation transport. The retrieval of trace gas products from UV/VIS spectrometers is strongly affected by the presence of clouds (see Fig. 3.1). The physics behind the influence of clouds on trace gas retrieval is well understood, and in general, there are three different contributions:

1. the albedo effect associated with the enhancement of reflectivity for cloudy scenes compared to cloud-free sky scenes;
2. the so-called shielding effect, for which that part of the trace gas column below the cloud is hidden by the clouds themselves; and
3. the increase in absorption, related to multiple scattering inside clouds which leads to enhancements of the optical path length.

D. S. Efremenko (✉) · A. Doicu · D. Loyola · T. Trautmann
Institut für Methodik der Fernerkundung (IMF), Deutsches Zentrum für Luft- und Raumfahrt (DLR),
Münchener Straße 20, 82234 Weßling, Oberpfaffenhofen, Germany
e-mail: dmitry.efremenko@dlr.de

A. Doicu
e-mail: adrian.doicu@dlr.de

D. Loyola
e-mail: diego.loyola@dlr.de

T. Trautmann
e-mail: thomas.trautmann@dlr.de

A. Kokhanovsky (ed.), *Springer Series in Light Scattering*, Springer Series in Light Scattering, https://doi.org/10.1007/978-3-319-70796-9_3

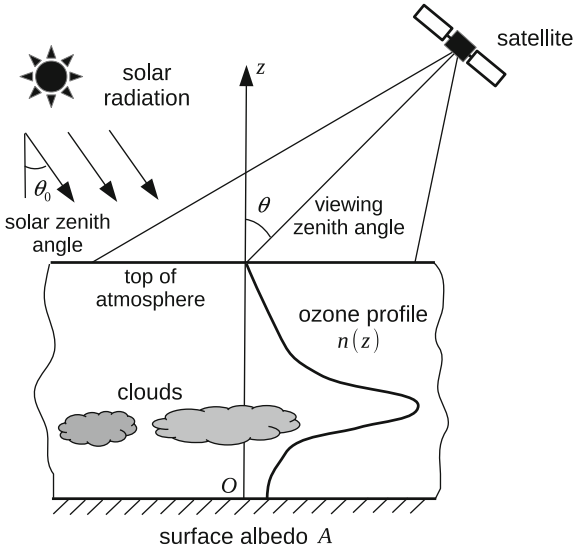

Fig. 3.1 Ozone profile retrieval in the presence of clouds

The albedo and in-cloud absorption effects increase the visibility of trace gases at and above the cloud-top, while the shielding effect normally results in an underestimation of the trace gas column. Using radiative transfer modeling several papers have quantified theoretically the influence of cloud parameters on the retrieval of trace gas columns (e.g., see Ahmad 2004; Boersma 2004; Roozendael et al. 2006 and the references therein). These studies show that the cloud fraction, cloud optical thickness (albedo), and cloud-top pressure (height) are the most important quantities for cloud correction of satellite trace gas retrievals.

Real clouds are an inhomogeneous three-dimensional scattering medium. For the new generation of satellite spectrometers with a relative high spatial resolution (such as Sentinel 5 Precursor, Sentinel 4 and Sentinel 5 Ingmann et al. 2012), it is important to account for the sub-pixel cloud inhomogeneities, or at least, to assess their effect on the radiances at the top of the atmosphere, and in particular, on the trace gas and cloud retrieval results. This assessment has to be probabilistic since the detailed structure of the clouds is unknown and only a small number of statistical properties are given. In this regard, two computational strategies can be proposed for simulating the radiance field and obtaining unknown statistical characteristics of the radiance field. The first one is based on multidimensional deterministic models and consists in the following steps:

1. simulation of sampling random realizations of an inhomogeneous cloudy field;
2. solution of radiative transfer equation for each realization, and
3. averaging of the obtained solutions over the ensemble of cloudiness realizations.

For each realization of a cloud field, a three-dimensional radiative transfer problem has to be solved by using either the Monte-Carlo method (Marchuk et al. 1980)

or multi-dimensional solvers such as SHDOM developed by Evans (1998). In both cases 3D-modeling is very time-consuming. Due to its computational complexity, this strategy is not used in operational retrieval algorithms.

In the second strategy, cloud fields are regarded as stochastic scattering media due to their internal inhomogeneity and stochastic geometry. The radiative transfer through these media is described by stochastic radiative transfer models (SRTM), in which new transport equations, relating the statistical parameters of the clouds to those of the radiance field, are derived. The goal of this approach is to obtain a relatively simple relationship between the statistical parameters of clouds and radiation that can be numerically evaluated in practice. Within the second strategy, the computations of the radiance field can be organized as follows:

1. simulation of random realizations of an inhomogeneous cloudy field,
2. obtaining the relevant statistical structure information from a set of realizations of a cloudy field, and
3. solution of the stochastic radiative transfer problem using as input the cloud statistics parameters obtained at the previous step.

In this way, time-consuming three-dimensional computations are avoided, but the SRTM is less accurate than 3D-models. However, the SRTM seems to be a good compromise between accuracy and computational costs. Although several methods for such statistical treatment were developed, stochastic models are not used in the operational retrieval algorithms due to complicated mathematical problems and the absence of reliable statistical data on cloud properties.

The majority of operational retrieval algorithms are based on the so-called external mixing model, which is the simplest deterministic model. It is the foremost application of the independent pixel approximation to homogeneous broken clouds. The external mixing procedure to calculate the radiance in an atmosphere containing partial cloudiness is to compute separately the radiances for completely cloudy and clear skies (I_{cloud} and I_{clear}, respectively), and then to express the partially cloudy radiance as a weighted linear combination of the separate radiances, the weighting provided by the effective cloud fraction $\overline{f}$:

$$I = (1 - \overline{f}) I_{\text{clear}} + \overline{f} I_{\text{cloud}}. \tag{3.1}$$

Here $\overline{f}$ is assumed not to depend on the wavelength. The clouds within each pixel are assumed to be plane-parallel and homogeneous in both horizontal and vertical directions as shown in Fig. 3.2.

The main advantage of the external mixing model is its computational efficiency, since it requires the solution of only two one-dimensional radiative transfer problems. In addition, its performance can be enhanced by reusing results from clear-sky RTM calculations for the cloud-filled scenario. Two methods have been proposed in Efremenko et al. (2014), which provide a speed-up factor of approximately 2 compared with two calls of RTM was obtained. The disadvantage of the external mixing model is that for cloudy scenes of small horizontal extent, errors due to the three-dimensional effects may be larger than 30% (Kokhanovsky 2003).

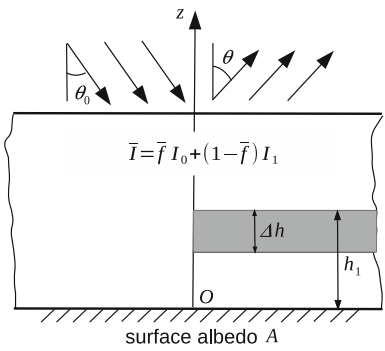

Fig. 3.2 Illustration of the external mixing model: h_1 is the cloud top height while Δh is the cloud geometric thickness

The topic addressed in this chapter is related to the ones discussed in Kassianov et al. (2012) and incorporates besides stochastic radiative transfer for Markovian cloud models also numerical aspects of efficient computations of radiance field and trace gas retrievals. The goal of our analysis is to demonstrate how the SRTM which takes into account the inhomogeneity of the scattering medium can be used in operational processors for cloud and ozone retrieval. The layout of this chapter is as follows. Section 3.2 contains a brief overview of previously developed SRTMs. In Sect. 3.3 we describe a n-order SRTM for arbitrary cloud statistics. In Sect. 3.4 a stochastic model for broken clouds is presented. Numerical aspects and the accuracy of SRTM are analyzed in Sect. 3.5. Internal mixing models are derived on the basis of stochastic models in Sect. 3.6. The impact of cloud inhomogeneities on trace gas and cloud property retrieval is studied in Sect. 3.7. The final Sect. 3.8 contains a few concluding remarks.

3.2 Overview of Stochastic Radiative Transfer Models

Techniques of computing the mean outgoing radiation field from any type of stochastic ensemble of cloud structure can be broadly categorized into two major groups. The first group is based on the Monte Carlo algorithm for the scattered medium in which some optical parameters are random functions. To minimize computational costs, the random cloud field is modeled together with the photon trajectories. The details can be found in Kargin (2000), Prigarin et al. (1998) and in references therein. This technique proved to be very efficient for forward simulations, in particular, for computing the radiance field in broken clouds. In this chapter we are focused on the second group of methods in which the analytical procedure of statistical averaging is applied directly to the radiative transfer equation. Here we are confronted with a problem of statistical closure which was first formulated and solved by Vainikko (1973a, b). A corresponding SRTM for analyzing the statistical structure of the radiance field in media with random spatial fluctuations of the optical properties was studied by Anisimov and Fukshansky (1992). It consists in a system (a hierarchy)

of stochastic equations for the statistical moments of the radiance field and for the covariances of the radiance field and the optical parameters. The technique for deriving the stochastic model is purely algebraic and applies to arbitrary statistics, while the system of stochastic equations is solved by an iterative method. Due to the emergence of the higher-order covariance terms in the stochastic equations, the problem of closing the system of equations was addressed. This problem was solved for the covariances of the radiance field and the optical parameters by assuming that the statistical fluctuations of the optical properties of the medium obey Gaussian statistics.

The radiative transfer in broken cloud layers has been cast as two coupled integro-differential equations, the so-called Levermore–Pomraning equations (Levermore et al. 1986), one for the mean radiance in the cloud, the other for the mean radiance in the clear sky. The initial derivation of the Levermore–Pomraning equations was based on the theory of alternating renewal processes (Vanderhaegen 1986; Levermore et al. 1988; Pomraning 1989). Another derivation has been provided by Adams et al. (1989) by determining a balance equation for the particles in each material, and by ensemble averaging the resulting equation over all statistical realizations. The derivation in Adams et al. (1989) is simple and applies to arbitrary binary statistical mixtures. Additional derivations have been reported by assuming that the particle trajectories are uncorrelated (Sahini 1989b) and by using nuclear reactor noise techniques (Sahini 1989a). In dealing with the Levermore–Pomraning equations we are confronted with a problem of closure, since the balance equations for the "volumetric" mean radiances involve the "interface" radiances between the cloudy and the clear skies. The standard closure is to assume that these radiances are equal, in which case, the resulting model is called the Levermore model. The Levermore model is exact for purely absorbing media with Markovian statistics, and approximate for scattering media and/or non-Markovian statistics. A more accurate model for treating non-Markovian statistics including scattering has been derived by Levermore et al. (1988) in the framework of the renewal theory. In practice, the Levermore–Pomraning equations are solved under the assumption of horizontal invariance of the statistical characteristics of the cloud field, in which case, we are faced with the solution of a system of two one-dimensional integro-differential equations. Assuming isotropic scattering and performing a Fourier cosine expansion of the ensemble-averaged radiances in the azimuthal angle, the system of equations is solved for the zeroth-order cosine mode, and so, the computed quantities are the mean radiances averaged over the azimuthal angle. The solution method is the discrete ordinate method in conjunction with an approximate iterative technique (Malvagi et al. 1993; Byrne et al. 1996). Note, that the SRTM of Anisimov and Fukshansky, as well as the SRTM based on the Levermore–Pomraning equations are formulated in terms of the total radiance, and the problem of extracting the diffuse radiance for stochastic scattering media illuminated by solar radiation was not addressed.

A Markovian cloud model involving two coupled integral equations was introduced in the atmospheric science community by Titov (1990) and Zuev and Titov (1995). The Titov integral model treats the clear sky as completely transparent, and for the standard closure assumption, it is mathematically identical to the differential Levermore model (Malvagi et al. 1993). An extension of the Titov integral model

to the solar radiative transfer problem in multilayer broken clouds and for inhomogeneous Markovian statistics has been provided by Kassianov (2003). In this work, the statistically inhomogeneous scattering medium has been represented as a set of statistically homogeneous interrelated cloud layers, each homogeneous in the vertical but inhomogeneous in the horizontal dimensions. The solution of the resulting integral equations for the mean radiance is performed by using the Monte Carlo method and the Markov chain simulation. To be used for trace gas retrieval, this model should be supplemented with the reflection from the boundary surface, and the scattering and absorption of the solar radiation by aerosols and atmospheric gases (this extension has been made using the Monte Carlo method by Titov et al. 1997 and Zhuravleva 2008). Note that Markovian statistics for a cloud appeared to be a very useful assumption, since the system of SRTM equations can be closed and the input parameters of the model (such as mean chord lengths of clouds) can be measured (e.g., Lane et al. 2002).

3.3 An *n*th Order Stochastic Radiative Transfer Model

3.3.1 Problem Formulation for the Inhomogeneous Atmosphere

We consider an atmosphere consisting of air molecules and clouds, while aerosols are not explicitly included. In the case of horizontally homogeneous aerosols, they can be included by redefining coefficients corresponding to the molecular atmosphere. The radiative transfer equation for the diffuse radiance I at the position $\mathbf{r}$ and in the direction $\boldsymbol{\Omega}$ is given by

$$\begin{aligned}\boldsymbol{\Omega}\cdot\nabla I\left(\mathbf{r},\boldsymbol{\Omega}\right) = &-\left[\sigma^0_{\mathtt{ext}}\left(\mathbf{r}\right)+\sigma^1_{\mathtt{ext}}\left(\mathbf{r}\right)\right] I\left(\mathbf{r},\boldsymbol{\Omega}\right)\\ &+\frac{1}{4\pi}\int_{4\pi}\left[\sigma^0_{\mathtt{sct}}\left(\mathbf{r}\right)P_0\left(\boldsymbol{\Omega},\boldsymbol{\Omega}'\right)+\sigma^1_{\mathtt{sct}}\left(\mathbf{r}\right)P_1\left(\boldsymbol{\Omega},\boldsymbol{\Omega}'\right)\right] I\left(\mathbf{r},\boldsymbol{\Omega}'\right)\mathrm{d}\Omega'\\ &+\frac{F_\odot}{4\pi}\left[\sigma^0_{\mathtt{sct}}\left(\mathbf{r}\right)P_0\left(\boldsymbol{\Omega},\boldsymbol{\Omega}_\odot\right)+\sigma^1_{\mathtt{sct}}\left(\mathbf{r}\right)P_1\left(\boldsymbol{\Omega},\boldsymbol{\Omega}_\odot\right)\right]\mathrm{e}^{-\tau^\odot_{\mathtt{ext}}\left(\mathbf{r}\right)},\end{aligned} \tag{3.2}$$

where $\sigma^{0,1}_{\mathtt{ext}}$, $\sigma^{0,1}_{\mathtt{sct}}$ and $P_{0,1}$ are the extinction coefficient, the scattering coefficient and the phase function corresponding to the clear sky (index 0) and to the cloud (index 1), respectively, $\tau^\odot_{\mathtt{ext}}$ is the solar optical depth, $\boldsymbol{\Omega}_\odot = (\pi-\theta_\odot, \varphi_\odot)$, with $\mu_\odot = \cos\theta_\odot > 0$, is the solar direction, and $F_\odot$ is the solar flux. In our analysis, we assume for simplicity, that the phase functions are position independent, and that the cloud single scattering albedo ω_1, defined by $\omega_1 = \sigma^1_{\mathtt{sct}}\left(\mathbf{r}\right)/\sigma^1_{\mathtt{ext}}\left(\mathbf{r}\right)$, is constant. The cloud extinction field is expressed as

$$\sigma^{1}_{\mathtt{ext}}(\mathbf{r}) = \sigma^{1\mathtt{d}}_{\mathtt{ext}}(\mathbf{r})\, f(\mathbf{r}), \tag{3.3}$$

where $\sigma^{1\mathtt{d}}_{\mathtt{ext}}$ is a deterministic function and f is the random geometric function. Consequently, the cloud scattering field is given by $\sigma^{1}_{\mathtt{sct}}(\mathbf{r}) = \sigma^{1\mathtt{d}}_{\mathtt{sct}}(\mathbf{r})\, f(\mathbf{r})$, with $\sigma^{1\mathtt{d}}_{\mathtt{sct}} = \omega_1 \sigma^{1\mathtt{d}}_{\mathtt{ext}}$. The random geometric function f reflects the statistics of the cloud field; for broken clouds, described by binary statistical mixtures, we have $f = f^2$, or more precisely, $f(\mathbf{r}) = 1$ inside the cloud and $f(\mathbf{r}) = 0$ inside the clear sky, but in general, e.g., for a bounded cascade cloud model, we have $f \neq f^2$. Note that for broken clouds, f is known as the random indicator function. The ensemble average of the cloud extinction field is $\overline{\sigma^{1}_{\mathtt{ext}}} = \sigma^{1\mathtt{d}}_{\mathtt{ext}}\overline{f}$, where here and in the following, the notation $\overline{X}$ means the ensemble average of X.

For a cloud embedded in a planar layer of infinite horizontal extent, the radiative transfer Eq. (3.2) must be supplemented with boundary conditions at the top and the bottom of the atmosphere. At the top of the atmosphere, we consider the standard homogeneous boundary condition for the downwelling radiance, i.e.,

$$I(\mathbf{r}_{\mathtt{t}}, \boldsymbol{\Omega}_{-}) = 0,$$

while at the bottom of the atmosphere we assume a Lambertian reflecting surface with surface albedo A, i.e.,

$$I(\mathbf{r}_{\mathtt{s}}, \boldsymbol{\Omega}_{+}) = \frac{A}{\pi} F_{\odot} |\mu_{\odot}| \, \mathrm{e}^{-\tau^{\odot}_{\mathtt{ext}}(\mathbf{r}_{\mathtt{s}})} + \frac{A}{\pi} \int_{2\pi} |\mu| \, I(\mathbf{r}_{\mathtt{s}}, \boldsymbol{\Omega}_{-}) \, \mathrm{d}\Omega. \tag{3.4}$$

Here, $\mathbf{r}_{\mathtt{t}}$ and $\mathbf{r}_{\mathtt{s}}$ are boundary points at the top of the atmosphere and at the underlying surface, respectively, while $\boldsymbol{\Omega}_{+}$ and $\boldsymbol{\Omega}_{-}$ denote an upwelling and a downwelling direction, respectively.

3.3.2 Statistical Averaging

As the cloud extinction coefficient $\sigma^{1}_{\mathtt{ext}}(\mathbf{r})$ is a random function, the radiance $I(\mathbf{r}, \boldsymbol{\Omega})$ is also a random function. Representing f, I and $\mathrm{e}^{-\tau^{\odot}_{\mathtt{ext}}}$ as the sum of their mean values $\overline{f}$, $\overline{I}$ and $\overline{\mathrm{e}^{-\tau^{\odot}_{\mathtt{ext}}}}$, respectively, and of their random fluctuations f', I' and $(\mathrm{e}^{-\tau^{\odot}_{\mathtt{ext}}})'$, respectively, that is,

$$f = \overline{f} + f', \quad I = \overline{I} + I', \quad \mathrm{e}^{-\tau^{\odot}_{\mathtt{ext}}} = \overline{\mathrm{e}^{-\tau^{\odot}_{\mathtt{ext}}}} + (\mathrm{e}^{-\tau^{\odot}_{\mathtt{ext}}})', \tag{3.5}$$

and by applying the procedure of statistical averaging it is possible to derive an infinite system of equations for the mean value of the radiance field and the covariances of the fluctuations of the radiance field and the geometric field. This system of equations must be truncated at a certain stage, and then closed by applying an independent hypothesis on the higher-order covariance terms. Following Anisimov and

Fukshansky (1992), inserting the Reynolds decompositions (3.5) into the radiative transfer Eq. (3.2) gives

$$\begin{aligned}
&\boldsymbol{\Omega}\cdot\nabla\overline{I}+\boldsymbol{\Omega}\cdot\nabla I' \\
&=-\sigma^{0}_{\mathtt{ext}}\left(\overline{I}+I'\right)-\sigma^{1\mathtt{d}}_{\mathtt{ext}}\left(\overline{f}\;\overline{I}+f'\overline{I}+\overline{f}\;I'+f'I'\right) \\
&+\frac{1}{4\pi}\int_{4\pi}\left[\sigma^{0}_{\mathtt{sct}}P_0\left(\overline{I}+I'\right)+\sigma^{1\mathtt{d}}_{\mathtt{sct}}P_1\left(\overline{f}\;\overline{I}+f'\overline{I}+\overline{f}\;I'+f'I'\right)\right]\mathrm{d}\Omega' \\
&+\frac{F_\odot}{4\pi}\left\{\sigma^{0}_{\mathtt{sct}}P_0\left[\overline{\mathrm{e}^{-\tau^{\odot}_{\mathtt{ext}}}}+(\mathrm{e}^{-\tau^{\odot}_{\mathtt{ext}}})'\right]\right. \\
&\left.+\sigma^{1\mathtt{d}}_{\mathtt{sct}}P_1\left[\overline{f}\,\overline{\mathrm{e}^{-\tau^{\odot}_{\mathtt{ext}}}}+f'\,\overline{\mathrm{e}^{-\tau^{\odot}_{\mathtt{ext}}}}+\overline{f}\,(\mathrm{e}^{-\tau^{\odot}_{\mathtt{ext}}})'+f'(\mathrm{e}^{-\tau^{\odot}_{\mathtt{ext}}})'\right]\right\}. \qquad (3.6)
\end{aligned}$$

Taking the ensemble average of Eq. (3.6) gives

$$\begin{aligned}
\boldsymbol{\Omega}\cdot\nabla\overline{I}&=-\sigma^{0}_{\mathtt{ext}}\overline{I}-\sigma^{1\mathtt{d}}_{\mathtt{ext}}\left(\overline{f}\;\overline{I}+\overline{f'I'}\right) \\
&+\frac{1}{4\pi}\int_{4\pi}\left[\sigma^{0}_{\mathtt{sct}}P_0\overline{I}+\sigma^{1\mathtt{d}}_{\mathtt{sct}}P_1\left(\overline{f}\;\overline{I}+\overline{f'I'}\right)\right]\mathrm{d}\Omega' \qquad (3.7) \\
&+\frac{F_\odot}{4\pi}\left\{\sigma^{0}_{\mathtt{sct}}P_0\overline{\mathrm{e}^{-\tau^{\odot}_{\mathtt{ext}}}}+\sigma^{1\mathtt{d}}_{\mathtt{sct}}P_1\left[\overline{f}\,\overline{\mathrm{e}^{-\tau^{\odot}_{\mathtt{ext}}}}+\overline{f'(\mathrm{e}^{-\tau^{\odot}_{\mathtt{ext}}})'}\right]\right\}.
\end{aligned}$$

A basic equation for computing the covariances $\overline{f'^{n}I'}$, $n\geq 1$, can be derived by taking the difference between Eqs. (3.6) and (3.7),

$$\begin{aligned}
\boldsymbol{\Omega}\cdot\nabla I'&=-\sigma^{0}_{\mathtt{ext}}I'-\sigma^{1\mathtt{d}}_{\mathtt{ext}}\left(f'\overline{I}-\overline{f'I'}+\overline{f}\;I'+f'I'\right) \\
&+\frac{1}{4\pi}\int_{4\pi}\left[\sigma^{0}_{\mathtt{sct}}P_0I'+\sigma^{1\mathtt{d}}_{\mathtt{sct}}P_1\left(f'\overline{I}-\overline{f'I'}+\overline{f}\;I'+f'I'\right)\right]\mathrm{d}\Omega' \\
&+\frac{F_\odot}{4\pi}\left\{\sigma^{0}_{\mathtt{sct}}P_0(\mathrm{e}^{-\tau^{\odot}_{\mathtt{ext}}})'\right. \\
&\left.+\sigma^{1\mathtt{d}}_{\mathtt{sct}}P_1\left[f'\,\overline{\mathrm{e}^{-\tau^{\odot}_{\mathtt{ext}}}}-\overline{f'(\mathrm{e}^{-\tau^{\odot}_{\mathtt{ext}}})'}+\overline{f}\,(\mathrm{e}^{-\tau^{\odot}_{\mathtt{ext}}})'+f'(\mathrm{e}^{-\tau^{\odot}_{\mathtt{ext}}})'\right]\right\}. \qquad (3.8)
\end{aligned}$$

Multiplying Eq. (3.8) by f' and taking the average, yields

$$\begin{aligned}
\boldsymbol{\Omega}\cdot\nabla\overline{f'I'}&=\overline{(\boldsymbol{\Omega}\cdot\nabla f')\,I'}-\sigma^{0}_{\mathtt{ext}}\overline{f'I'}-\sigma^{1\mathtt{d}}_{\mathtt{ext}}\left(\overline{f'^{2}}\;\overline{I}+\overline{f}\;\overline{f'I'}+\overline{f'^{2}I'}\right) \\
&+\frac{1}{4\pi}\int_{4\pi}\left[\sigma^{0}_{\mathtt{sct}}P_0\overline{f'I'}+\sigma^{1\mathtt{d}}_{\mathtt{sct}}P_1\left(\overline{f'^{2}}\;\overline{I}+\overline{f}\;\overline{f'I'}+\overline{f'^{2}I'}\right)\right]\mathrm{d}\Omega' \\
&+\frac{F_\odot}{4\pi}\left\{\sigma^{0}_{\mathtt{sct}}P_0\overline{f'(\mathrm{e}^{-\tau^{\odot}_{\mathtt{ext}}})'}\right. \\
&\left.+\sigma^{1\mathtt{d}}_{\mathtt{sct}}P_1\left[\overline{f'^{2}}\,\overline{\mathrm{e}^{-\tau^{\odot}_{\mathtt{ext}}}}+\overline{f}\,\overline{f'(\mathrm{e}^{-\tau^{\odot}_{\mathtt{ext}}})'}+\overline{f'^{2}(\mathrm{e}^{-\tau^{\odot}_{\mathtt{ext}}})'}\right]\right\}, \qquad (3.9)
\end{aligned}$$

while multiplying Eq. (3.8) by f'^{n}, $n\geq 2$, and taking the average, gives

$$
\begin{aligned}
&\boldsymbol{\Omega}\cdot\nabla\overline{f'^n I'}\\
&=\overline{\left(\boldsymbol{\Omega}\cdot\nabla f'^n\right) I'}-\sigma^0_{\mathtt{ext}}\overline{f'^n I'}-\sigma^{1\mathtt{d}}_{\mathtt{ext}}\left(\overline{f'^{(n+1)}}\,\overline{I}-\overline{f'^n}\,\overline{f'I'}+\overline{f}\,\overline{f'^n I'}+\overline{f'^{(n+1)}I'}\right)\\
&+\frac{1}{4\pi}\int_{4\pi}\left[\sigma^0_{\mathtt{sct}}P_0\overline{f'^n I'}+\sigma^{1\mathtt{d}}_{\mathtt{sct}}P_1\left(\overline{f'^{(n+1)}}\,\overline{I}-\overline{f'^n}\,\overline{f'I'}+\overline{f}\,\overline{f'^n I'}+\overline{f'^{(n+1)}I'}\right)\right]\mathrm{d}\Omega'\\
&+\frac{F_\odot}{4\pi}\left\{\sigma^0_{\mathtt{sct}}P_0\overline{f'^n(\mathrm{e}^{-\tau^\odot_{\mathtt{ext}}})'}\right.\\
&\left.+\sigma^{1\mathtt{d}}_{\mathtt{sct}}P_1\left[\overline{f'^{(n+1)}}\,\overline{\mathrm{e}^{-\tau^\odot_{\mathtt{ext}}}}-\overline{f'^n}\,\overline{f'(\mathrm{e}^{-\tau^\odot_{\mathtt{ext}}})'}+\overline{f}\,\overline{f'^n(\mathrm{e}^{-\tau^\odot_{\mathtt{ext}}})'}+\overline{f'^{(n+1)}(\mathrm{e}^{-\tau^\odot_{\mathtt{ext}}})'}\right]\right\}.
\end{aligned}
\tag{3.10}
$$

Equation (3.10) for computing $\overline{f'^n I'}$, involves the terms $\overline{I}$ and $\overline{f'I'}$, as well as, the higher-order covariances $\overline{(\boldsymbol{\Omega}\cdot\nabla f'^n)I'}$ and $\overline{f'^{(n+1)}I'}$. Because of the dependency on these higher-order covariance terms, we are led to an infinite system of equations for $\overline{I}$ and $\overline{f'^n I'}$, $n\geq 1$, consisting in Eqs. (3.7), (3.9), and (3.10) for $n\geq 2$. In practice, this system of equations must be truncated at some stage n, and then closed by making some assumptions on $\overline{(\boldsymbol{\Omega}\cdot\nabla f'^n)I'}$ and $\overline{f'^{(n+1)}I'}$.

The same technique can be used to derive the surface boundary conditions for $\overline{I}$ and $\overline{f'^n I'}$, $n\geq 1$. For a Lambertian reflecting surface with surface albedo A, these are given by

$$
\overline{I}_+=\frac{A}{\pi}F_\odot\,|\mu_\odot|\,\overline{\mathrm{e}^{-\tau^\odot_{\mathtt{ext}}}}+\frac{A}{\pi}\int_{2\pi}|\mu|\,\overline{I}_-\mathrm{d}\Omega,
\tag{3.11}
$$

and

$$
\overline{f'^n I'}_+=\frac{A}{\pi}F_\odot\,|\mu_\odot|\,\overline{f'^n(\mathrm{e}^{-\tau^\odot_{\mathtt{ext}}})'}+\frac{A}{\pi}\int_{2\pi}|\mu|\,\overline{f'^n I'}_-\mathrm{d}\Omega,
\tag{3.12}
$$

where $\overline{I}_+=\overline{I}(\mathbf{r}_{\mathtt{s}},\boldsymbol{\Omega}_+)$, $\overline{I}_-=\overline{I}(\mathbf{r}_{\mathtt{s}},\boldsymbol{\Omega}_-)$, $\overline{f'^n I'}_+=\overline{f'^n I'}(\mathbf{r}_{\mathtt{s}},\boldsymbol{\Omega}_+)$ and $\overline{f'^n I'}_-=\overline{f'^n I'}(\mathbf{r}_{\mathtt{s}},\boldsymbol{\Omega}_-)$.

For an nth-order stochastic model, it is convenient to define the $(n+1)$-dimensional radiance vector by

$$
\mathbf{I}=\begin{bmatrix}\overline{I}\\ \overline{f'I'}\\ \vdots\\ \overline{f'^n I'}\end{bmatrix},
$$

and to express Eqs. (3.7), (3.9), and (3.10) in matrix form as

$$\boldsymbol{\Omega}\cdot\nabla\mathbf{I}\left(\mathbf{r},\boldsymbol{\Omega}\right)=\mathbf{I}_{\nabla}\left(\mathbf{r},\boldsymbol{\Omega}\right)-\boldsymbol{\Sigma}\left(\mathbf{r}\right)\mathbf{I}\left(\mathbf{r},\boldsymbol{\Omega}\right)-\sigma_{\mathtt{ext}}^{1\mathtt{d}}\left(\mathbf{r}\right)\triangle\mathbf{I}\left(\mathbf{r},\boldsymbol{\Omega}\right)$$
$$+\frac{1}{4\pi}\int_{4\pi}\left[\mathbf{S}\left(\mathbf{r},\boldsymbol{\Omega},\boldsymbol{\Omega}'\right)\mathbf{I}\left(\mathbf{r},\boldsymbol{\Omega}'\right)+\sigma_{\mathtt{sct}}^{1\mathtt{d}}\left(\mathbf{r}\right)P_{1}\left(\boldsymbol{\Omega},\boldsymbol{\Omega}'\right)\triangle\mathbf{I}\left(\mathbf{r},\boldsymbol{\Omega}'\right)\right]\mathrm{d}\Omega'$$
$$+\frac{F_{\odot}}{4\pi}\left[\mathbf{S}\left(\mathbf{r},\boldsymbol{\Omega},\boldsymbol{\Omega}_{\odot}\right)\mathbf{I}_{\odot}\left(z\right)+\sigma_{\mathtt{sct}}^{1\mathtt{d}}\left(\mathbf{r}\right)P_{1}\left(\boldsymbol{\Omega},\boldsymbol{\Omega}_{\odot}\right)\triangle\mathbf{I}_{\odot}\left(\mathbf{r}\right)\right]. \tag{3.13}$$

Here, the vector $\mathbf{I}_{\nabla}$ involving the gradient fields $\nabla f'^{k}$, $k=1,\ldots,n$, and the solar transmission vector $\mathbf{I}_{\odot}$ are given by

$$\mathbf{I}_{\nabla}=\begin{bmatrix}0\\ \overline{\left(\boldsymbol{\Omega}\cdot\nabla f'\right)I'}\\ \vdots\\ \overline{\left(\boldsymbol{\Omega}\cdot\nabla f'^{n}\right)I'}\end{bmatrix}\text{ and }\mathbf{I}_{\odot}=\begin{bmatrix}\overline{\mathrm{e}^{-\tau_{\mathtt{ext}}^{\odot}}}\\ \overline{f'(\mathrm{e}^{-\tau_{\mathtt{ext}}^{\odot}})'}\\ \vdots\\ \overline{f'^{n}(\mathrm{e}^{-\tau_{\mathtt{ext}}^{\odot}})'}\end{bmatrix},$$

respectively, while the pseudo-phase matrix $\mathbf{S}$, depending on the product of the scattering coefficient and the phase function, and the extinction matrix $\boldsymbol{\Sigma}$ are given by

$$\mathbf{S}=\begin{bmatrix}\sigma_{\mathtt{sct}}^{0}P_{0}+\sigma_{\mathtt{sct}}^{1\mathtt{d}}\overline{f}P_{1} & \sigma_{\mathtt{sct}}^{1\mathtt{d}}P_{1} & 0 & \ldots & 0\\ \sigma_{\mathtt{sct}}^{1\mathtt{d}}\overline{f'^{2}}P_{1} & \sigma_{\mathtt{sct}}^{0}P_{0}+\sigma_{\mathtt{sct}}^{1\mathtt{d}}\overline{f}P_{1} & \sigma_{\mathtt{sct}}^{1\mathtt{d}}P_{1} & \ldots & 0\\ \vdots & \vdots & \vdots & & \vdots\\ \sigma_{\mathtt{sct}}^{1\mathtt{d}}\overline{f'^{(n+1)}}P_{1} & -\sigma_{\mathtt{sct}}^{1\mathtt{d}}\overline{f'^{n}}P_{1} & 0 & \ldots & \sigma_{\mathtt{sct}}^{0}P_{0}+\sigma_{\mathtt{sct}}^{1\mathtt{d}}\overline{f}P_{1}\end{bmatrix} \tag{3.14}$$

and

$$\boldsymbol{\Sigma}=\begin{bmatrix}\sigma_{\mathtt{ext}}^{0}+\sigma_{\mathtt{ext}}^{1\mathtt{d}}\overline{f} & \sigma_{\mathtt{ext}}^{1\mathtt{d}} & 0 & \ldots & 0\\ \sigma_{\mathtt{ext}}^{1\mathtt{d}}\overline{f'^{2}} & \sigma_{\mathtt{ext}}^{0}+\sigma_{\mathtt{ext}}^{1\mathtt{d}}\overline{f} & \sigma_{\mathtt{ext}}^{1\mathtt{d}} & \ldots & 0\\ \vdots & \vdots & \vdots & & \vdots\\ \sigma_{\mathtt{ext}}^{1\mathtt{d}}\overline{f'^{(n+1)}} & -\sigma_{\mathtt{ext}}^{1\mathtt{d}}\overline{f'^{n}} & 0 & \ldots & \sigma_{\mathtt{ext}}^{0}+\sigma_{\mathtt{ext}}^{1\mathtt{d}}\overline{f}\end{bmatrix}, \tag{3.15}$$

respectively. Finally, the higher-order covariance vectors $\triangle\mathbf{I}$ and $\triangle\mathbf{I}_{\odot}$, read as

$$\triangle\mathbf{I}=\begin{bmatrix}0\\ 0\\ \vdots\\ \overline{f'^{(n+1)}I'}\end{bmatrix}\text{ and }\triangle\mathbf{I}_{\odot}=\begin{bmatrix}0\\ 0\\ \vdots\\ \overline{f'^{(n+1)}(\mathrm{e}^{-\tau_{\mathtt{ext}}^{\odot}})'}\end{bmatrix},$$

respectively.

3.3.3 Closure Relations for the SRTM

To solve the radiative transfer Eq. (3.13), we need to employ appropriate closure relations for $\overline{(\boldsymbol{\Omega}\cdot\nabla f'^k)I'}$, $k=1,\ldots,n$, and $\overline{f'^{(n+1)}I'}$. For the covariance $\overline{f'^{(n+1)}I'}$, we have several options:

1. A standard closure procedure for an nth-order stochastic model is to assume that $\overline{f'^{(n+1)}I'}$ is negligible compared to $\overline{I}$, $\overline{f'I'},\ldots,\overline{f'^nI'}$. Thus, assuming that $\overline{f'^{(n+1)}I'}=0$, we obtain $\triangle\mathbf{I}=0$. Moreover, if

$$\overline{f'^{(n+1)}}\,\overline{\mathrm{e}^{-\tau_{\mathrm{ext}}^{\odot}}}-\overline{f'^n}\,\overline{f'(\mathrm{e}^{-\tau_{\mathrm{ext}}^{\odot}})'}+\overline{f}\,\overline{f'^n(\mathrm{e}^{-\tau_{\mathrm{ext}}^{\odot}})'}\gg\overline{f'^{(n+1)}(\mathrm{e}^{-\tau_{\mathrm{ext}}^{\odot}})'},$$

   then $\triangle\mathbf{I}_{\odot}$ can be also neglected.
2. The closure procedure proposed in Anisimov and Fukshansky (1992), in conjunction with a second-order stochastic model ($n=2$), is based on the relation between the moments of different orders, or more precisely, on the relation between the fourth-order correlation moments of any four Gaussian random variables and their second-order moments. If the statistical fluctuations of the optical properties of the medium obey Gaussian statistics, this closure procedure gives $\overline{f'^3I'}=3\overline{f'^2}\,\overline{f'I'}$. As a result, we deduce that $\mathbf{I}$ solves the stochastic equation (3.13) with $\triangle\mathbf{I}=0$, and in which, the $(3,2)$ entries of the matrices $\mathbf{S}$ and $\boldsymbol{\Sigma}$ are given by $[\mathbf{S}]_{32}=2\sigma_{\mathtt{sct}}^{\mathtt{ld}}P_1\overline{f'^2}$ and $[\boldsymbol{\Sigma}]_{32}=2\sigma_{\mathtt{ext}}^{\mathtt{ld}}\overline{f'^2}$, respectively, while the third entry of $\triangle\mathbf{I}_{\odot}$ is $[\triangle\mathbf{I}_{\odot}]_3=\overline{f'^3(\mathrm{e}^{-\tau_{\mathrm{ext}}^{\odot}})'}-3\overline{f'^2}\,\overline{f'(\mathrm{e}^{-\tau_{\mathrm{ext}}^{\odot}})'}$. If moreover, this closure procedure applies for the solar terms, i.e., $\overline{f'^3(\mathrm{e}^{-\tau_{\mathrm{ext}}^{\odot}})'}=3\overline{f'^2}\,\overline{f'(\mathrm{e}^{-\tau_{\mathrm{ext}}^{\odot}})'}$, then $\triangle\mathbf{I}_{\odot}=0$.
3. Broken clouds (binary statistical mixtures) are characterized by $f=f^2$, and a closure relation for $\overline{f'^2I'}$ readily follows. Indeed, from $f=f^2$, we find that $\overline{f^2I}=\overline{fI}$ and $\overline{f}=\overline{f}^2+\overline{f'^2}$. As a result, we infer that

$$\overline{f'^2I'}=\left(1-2\overline{f}\right)\overline{f'I'} \tag{3.16}$$

   and similarly, that

$$\overline{f'^2(\mathrm{e}^{-\tau_{\mathrm{ext}}^{\odot}})'}=\left(1-2\overline{f}\right)\overline{f'(\mathrm{e}^{-\tau_{\mathrm{ext}}^{\odot}})'}. \tag{3.17}$$

Regarding the higher-order covariances $\overline{(\boldsymbol{\Omega}\cdot\nabla f'^k)I'}$, $k=1,\ldots,n$ we make the following remarks:

1. In the zeroth-order stochastic model (3.7), the covariance vector $\mathbf{I}_{\nabla}$ does not appear, and what is required is only a closure relation for $\overline{f'I'}$. In Gabriel and Evans (1996), $\overline{f'I'}$ is neglected, while in Stephens (1988), the closure relation $\overline{f'I'}=C\overline{f}\,\overline{I}$, with $C=C(z,\mu)=\mu\tilde{C}(z)$, is applied in the framework of a two-stream model.
2. In the analysis conducted by Anisimov and Fukshansky (1992), the covariance vector $\mathbf{I}_{\nabla}$ is disregarded and a second-order stochastic model is used to analyze the

radiative transfer through overcast cloudy scenes. This model provides accurate results if $\overline{f'^2}$ and $\overline{f'^3}$ do not exceed certain critical values.
3. For broken clouds, the covariance vector $\mathbf{I}_\nabla$ is the key point in proving the equivalence with a stochastic model based on the Levermore–Pomraning equations. Therefore, the derivation of a closure relation for $\overline{(\boldsymbol{\Omega}\cdot\nabla f')I'}$ is an important task.

3.4 A Stochastic Radiative Transfer Model for Broken Clouds

As stated in Zuev and Titov (1996), Chap. 6 there are two types of cloud inhomogeneities. The first one is due to random internal structure of a cloud while the second one is referred to macro-fluctuations and described roughly in terms of "cloud" and "clear sky". The impact of the first type of inhomogeneities on the cloud field is much stronger than that due to micro-fluctuations. In this regard, for numerical analysis it is beneficial to assume a random cloud field with deterministic optical parameters inside the cloud. Below we consider stochastic radiative transfer in broken clouds.

3.4.1 A First-Order Stochastic Radiative Transfer Model

Using the closure relations (3.16) and (3.17), together with Eqs. (3.7) and (3.9) we are led to a first-order stochastic model for the two-dimensional radiance vector $\mathbf{I} = [\,\overline{I},\ \overline{f'I'}\,]^T$, that is,

$$\boldsymbol{\Omega}\cdot\nabla\mathbf{I}(\mathbf{r},\boldsymbol{\Omega}) = \mathbf{I}_\nabla(\mathbf{r},\boldsymbol{\Omega}) - \boldsymbol{\Sigma}(\mathbf{r})\,\mathbf{I}(\mathbf{r},\boldsymbol{\Omega}) + \frac{1}{4\pi}\int_{4\pi}\mathbf{S}\left(\mathbf{r},\boldsymbol{\Omega},\boldsymbol{\Omega}'\right)\mathbf{I}\left(\mathbf{r},\boldsymbol{\Omega}'\right)\mathrm{d}\Omega' + \frac{F_\odot}{4\pi}\mathbf{S}(\mathbf{r},\boldsymbol{\Omega},\boldsymbol{\Omega}_\odot)\,\mathbf{I}_\odot(\mathbf{r}), \tag{3.18}$$

where $\mathbf{I}_\nabla = [\,0,\ \overline{(\boldsymbol{\Omega}\cdot\nabla f')\,I'}\,]^T$, $\mathbf{I}_\odot = [\,\overline{\mathrm{e}^{-\tau^\odot_{\mathtt{ext}}}},\ \overline{f'(\mathrm{e}^{-\tau^\odot_{\mathtt{ext}}})'}\,]^T$,

$$\mathbf{S} = \begin{bmatrix} \sigma^0_{\mathtt{sct}}P_0 + \sigma^{1\mathtt{d}}_{\mathtt{sct}}\overline{f}\,P_1 & \sigma^{1\mathtt{d}}_{\mathtt{sct}}P_1 \\ \sigma^{1\mathtt{d}}_{\mathtt{sct}}\overline{f'^2}P_1 & \sigma^0_{\mathtt{sct}}P_0 + \sigma^{1\mathtt{d}}_{\mathtt{sct}}\left(1-\overline{f}\right)P_1 \end{bmatrix} \tag{3.19}$$

and

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma^0_{\mathtt{ext}} + \sigma^{1\mathtt{d}}_{\mathtt{ext}}\overline{f} & \sigma^{1\mathtt{d}}_{\mathtt{ext}} \\ \sigma^{1\mathtt{d}}_{\mathtt{ext}}\overline{f'^2} & \sigma^0_{\mathtt{ext}} + \sigma^{1\mathtt{d}}_{\mathtt{ext}}\left(1-\overline{f}\right) \end{bmatrix}. \tag{3.20}$$

For the first-order stochastic model (3.18), the boundary conditions (3.11) and (3.12) translate into

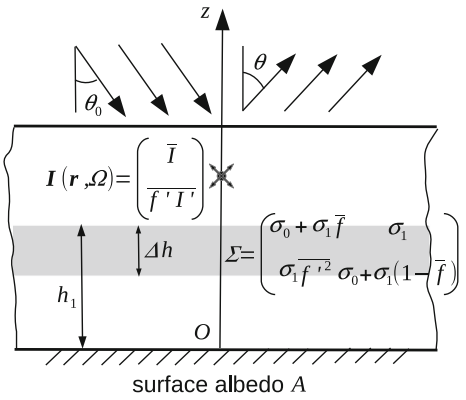

Fig. 3.3 Illustration of the first-order stochastic model for broken clouds. The model is formulated for the two-dimensional radiance vector $\mathbf{I} = [\,\overline{I},\ \overline{f'I'}\,]^T$

$$\mathbf{I}\,(\mathbf{r}_{\mathrm{t}},\,\boldsymbol{\Omega}_{-}) = 0, \tag{3.21}$$

and

$$\mathbf{I}\,(\mathbf{r}_{\mathrm{s}},\,\boldsymbol{\Omega}_{+}) = \frac{A}{\pi} F_{\odot}\,|\mu_{\odot}|\,\mathbf{I}_{\odot}\,(\mathbf{r}_{\mathrm{s}}) + \frac{A}{\pi}\int_{2\pi} |\mu|\,\mathbf{I}\,(\mathbf{r}_{\mathrm{s}},\,\boldsymbol{\Omega}_{-})\,\mathrm{d}\Omega. \tag{3.22}$$

Figure 3.3 shows a radiative transfer model corresponding to Eq. (3.18) accompanied with Eqs. (3.19) and (3.20).

Note, that Eq. (3.18) is equivalent to the matrix form representation of the Levermore–Pomraning equations, expressed in terms of the mean radiance in the cloud

$$I_1 = \frac{1}{\overline{f}}\,\overline{fI}, \tag{3.23}$$

and the mean radiance in the clear sky

$$I_0 = \frac{1}{1-\overline{f}}\,\overline{(1-f)\,I}. \tag{3.24}$$

Indeed, taking linear combinations of the two equations in (3.18) according to the transformation rules $\overline{f}\,I_1 = \overline{f}\;\overline{I} + \overline{f'I'}$ and $(1-\overline{f})I_0 = (1-\overline{f})\overline{I} - \overline{f'I'}$, and further, using $\overline{f'^2} = \overline{f} - \overline{f}^2$ and $\overline{I'\nabla f'} = \overline{I\nabla f} - \overline{I}\,\overline{\nabla f}$, we are led to a stochastic equation as in (3.18) but for the radiance vector $\mathbf{I} = [\,\overline{f}\,I_1,\ \left(1-\overline{f}\right) I_0\,]^T$, with $\mathbf{I}_{\nabla} = [\,\overline{(\boldsymbol{\Omega}\cdot\nabla f)\,I},\ -\overline{(\boldsymbol{\Omega}\cdot\nabla f)\,I}\,]^T$, $\mathbf{I}_{\odot} = [\,\overline{f\mathrm{e}^{-\tau^{\odot}_{\mathrm{ext}}}},\ \overline{(1-f)\,\mathrm{e}^{-\tau^{\odot}_{\mathrm{ext}}}}\,]^T$,

$$\mathbf{S} = \begin{bmatrix} \sigma^0_{\mathrm{sct}} P_0 + \sigma^{\mathrm{ld}}_{\mathrm{sct}} P_1 & 0 \\ 0 & \sigma^0_{\mathrm{sct}} P_0 \end{bmatrix},\ \text{and}\ \boldsymbol{\Sigma} = \begin{bmatrix} \sigma^0_{\mathrm{ext}} + \sigma^{\mathrm{ld}}_{\mathrm{ext}} & 0 \\ 0 & \sigma^0_{\mathrm{ext}} \end{bmatrix}. \tag{3.25}$$

This equation is the matrix form representation of the Levermore–Pomraning equations for binary statistical mixtures illuminated by solar radiation, and with the

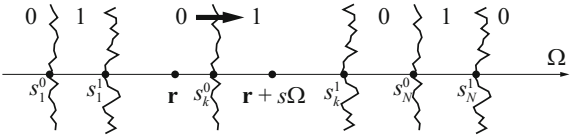

Fig. 3.4 Scheme for illustrating Eq. (3.27): $g(s) = 1$ in clouds while $g(s) = 0$ in cloud free regions

coupling terms encapsulated in the covariance vector $\mathbf{I}_\nabla$. Thus, in the stochastic model (3.18), the radiance vector comprises the mean radiance field and the covariance of the radiance field and the indicator field, while in a stochastic model based on the Levermore–Pomraning equations, the radiance vector involves the mean radiances in each material.

3.4.2 *Estimating a Covariance Vector*

The main problem which need to be solved is to find an estimate of $\overline{(\boldsymbol{\Omega} \cdot \nabla f')I'}$, which enters in the expression of $\mathbf{I}_\nabla$. Accounting of the representation

$$\overline{(\boldsymbol{\Omega} \cdot \nabla f')\, I'} = \overline{(\boldsymbol{\Omega} \cdot \nabla f)\, I} - \left(\boldsymbol{\Omega} \cdot \nabla \overline{f}\right) \overline{I}, \tag{3.26}$$

we will find an estimate for $\overline{(\boldsymbol{\Omega} \cdot \nabla f)I}$ and postpone the problem for $\overline{(\boldsymbol{\Omega} \cdot \nabla f')I'}$ to the next section. The estimate for $\overline{(\boldsymbol{\Omega} \cdot \nabla f)I}$ will be derived by using the basic closure assumption of the Levermore model, according to which, the interface radiances between the cloud and the clear sky are equal to the volumetric mean radiances. Our derivation should be regarded as an alternative to the standard derivation given by Adams et al. (1989) and Pomraning (1991).

At point $\mathbf{r}$, we trace a line along $\boldsymbol{\Omega}$, and denote by s be the spatial coordinate in the direction $\boldsymbol{\Omega}$; thus, the point $\mathbf{r} + s\boldsymbol{\Omega}$ is a point on this line, situated at a distance s from $\mathbf{r}$ in the direction $\boldsymbol{\Omega}$. Considering a realization of the indicator function f, we set

$$g(s) = f(\mathbf{r} + s\boldsymbol{\Omega}) = \begin{cases} 1, & s_k^0 \le s \le s_k^1 \\ 0, & \text{rest} \end{cases}, \tag{3.27}$$

where the s_k^i, with $i = 0, 1$ and $k = 1, \ldots, N$, are the values of s at the boundary points of the cloudy region k, and N is the number of cloudy regions (see Fig. 3.4). The significance of the points $\mathbf{r} + s_k^i \boldsymbol{\Omega}$ is given by the direction $\boldsymbol{\Omega}$. More precisely, $\mathbf{r} + s_k^i \boldsymbol{\Omega}$ are the points at the interfaces $i \to j$ in direction $\boldsymbol{\Omega}$, because at the interface point $\mathbf{r} + s_k^i \boldsymbol{\Omega}$, when moving in the direction $\boldsymbol{\Omega}$, we pass from material i to material j. Evidently, s_k^i and N are random variables. The spatial derivative of f can be expressed as

$$(\boldsymbol{\Omega} \cdot \nabla f)(\mathbf{r}) = \frac{\mathrm{d}f}{\mathrm{d}s}(\mathbf{r}) = g'(0),$$

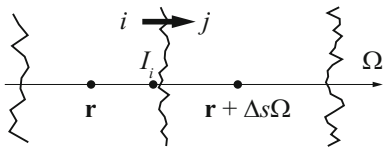

Fig. 3.5 The interface between region i and region j. The radiance at the interface $i \to j$ is assumed to be equal to the volumetric mean radiances $I_i(\mathbf{r}, \Omega)$

where g' is the derivative of g. Taking into account that

$$g'(s) = \sum_{k=1}^{N} \left[\delta\left(s - s_k^0\right) - \delta\left(s - s_k^1\right)\right], \tag{3.28}$$

and using the limit representation of the Dirac delta function

$$\delta\left(s - s^*\right) = \lim_{\triangle s \to 0_+} h_{\triangle s}\left(s, s^*\right) = \begin{cases} \frac{1}{\triangle s}, & s \in \left[s^* - \frac{\triangle s}{2}, s^* + \frac{\triangle s}{2}\right] \\ 0, & \text{rest} \end{cases}$$

we find that

$$g'(s) = \lim_{\triangle s \to 0_+} \sum_{k=1}^{N} \left[h_{\triangle s}\left(s, s_k^0\right) - h_{\triangle s}\left(s, s_k^1\right)\right]. \tag{3.29}$$

Thus, at a fixed point $\mathbf{r} + s\boldsymbol{\Omega}$, $g'(s)$ is not zero, if for sufficiently small $\triangle s$, the point $\mathbf{r} + s\boldsymbol{\Omega}$ is an interface point; $g'(s) > 0$ if $\mathbf{r} + s\boldsymbol{\Omega}$ is at the interface $0 \to 1$ in direction $\boldsymbol{\Omega}$, and $g'(s) < 0$ if $\mathbf{r} + s\boldsymbol{\Omega}$ is at the interface $1 \to 0$ in direction $\boldsymbol{\Omega}$. Therefore, for $s = 0$, we may write

$$\overline{(\boldsymbol{\Omega} \cdot \nabla f) I} = T_0 - T_1, \tag{3.30}$$

where T_i is given by

$$T_i = \lim_{\triangle s \to 0_+} \overline{I(\mathbf{r}, \Omega) \frac{1}{\triangle s}}^{i}, \tag{3.31}$$

and the bar notation "$\overline{\quad}^{i}$" means that the average is taken over that realizations, for which, the point $\mathbf{r}$ is at the interface $i \to j$ in direction $\boldsymbol{\Omega}$. For these realizations, we approximate the radiances $I(\mathbf{r}, \Omega)$ at the interfaces $i \to j$, say $\mathcal{I}_i(\mathbf{r}, \Omega)$, by the volumetric mean radiances $I_i(\mathbf{r}, \Omega)$, as given by (3.24) and (3.23), i.e.,

$$\mathcal{I}_i(\mathbf{r}, \Omega) = I_i(\mathbf{r}, \Omega). \tag{3.32}$$

as shown in Fig. 3.5.

Then, to account for the ensemble average in (3.31) we have to multiply $I_i(\mathbf{r}, \Omega)/\triangle s$ by the probability of occurrence of these events. Let us define the events A_i = "point $\mathbf{r}$ is in material i" and B_j = "point $\mathbf{r} + \triangle s\boldsymbol{\Omega}$ is in material j". For $i \neq j$, the joint probability of the events A_i and B_j, given by

$P(A_i \cap B_j) = P(B_j \mid A_i)P(A_i)$, represents the probability that "the point $\mathbf{r}$ is in material i and the point $\mathbf{r} + \triangle s\boldsymbol{\Omega}$ is in material j", or equivalently, that "the point $\mathbf{r}$ is at the interface $i \to j$ in direction $\boldsymbol{\Omega}$". Defining the transition length $\lambda_i(\mathbf{r}, \boldsymbol{\Omega})$ in material i in direction $\boldsymbol{\Omega}$ in terms of the conditional probability that "the point $\mathbf{r} + \triangle s\boldsymbol{\Omega}$ is in material j given that the point $\mathbf{r}$ is in material i", $P(B_j \mid A_i)$, by the relation

$$P\left(B_j \mid A_i\right) := \frac{\triangle s}{\lambda_i\left(\mathbf{r}, \boldsymbol{\Omega}\right)}, \tag{3.33}$$

and taking into account that $P(A_0) = 1 - \overline{f}\,(\mathbf{r})$ and $P(A_1) = \overline{f}\,(\mathbf{r})$, we obtain

$$\overline{(\boldsymbol{\Omega} \cdot \nabla f)\, I} = \frac{1 - \overline{f}}{\lambda_0} I_0 - \frac{\overline{f}}{\lambda_1} I_1. \tag{3.34}$$

For Markovian cloud statistics, the $\lambda_i(\mathbf{r}, \boldsymbol{\Omega})$ are the Markovian transition lengths; they are prescribed by the cloud dynamics and completely define the binary statistical mixture. For arbitrary cloud statistics, the $\lambda_i(\mathbf{r}, \boldsymbol{\Omega})$ can be determined numerically from $P(B_j \mid A_i)$, or more precisely, from $P(A_i \cap B_j)$ and $P(A_i)$, for an ensemble of realizations of the indicator function f. An equivalent derivation of (3.34), which is more close to the proofs of Adams et al. (1989) and Pomraning (1991), can be given by making use of a definition of the gradient operator that does not make reference to any coordinate system, that is,

$$\boldsymbol{\Omega} \cdot \nabla f = \lim_{V \to 0} \frac{1}{V} \oint_S (\boldsymbol{\Omega} \cdot \mathbf{n})\, f \, \mathrm{d}S,$$

where S is the surface bounding a volume V, and $\mathbf{n}$ is the outward normal on S.

In order to derive an estimate of $\overline{(\boldsymbol{\Omega} \cdot \nabla f) I}$ for two- and one-dimensional cloud fields, we consider a general situation. Let $\mathbf{e}_{\mathrm{u}}$ and $\mathbf{e}_{\mathrm{v}}$ be two orthonormal vectors, chosen such that the vectors $\boldsymbol{\Omega}$, $\mathbf{e}_{\mathrm{u}}$ and $\mathbf{e}_{\mathrm{v}}$ are coplanar, and that $\boldsymbol{\Omega} \cdot \mathbf{e}_{\mathrm{u}} > 0$. The representation $\boldsymbol{\Omega} = \Omega_{\mathrm{u}}\mathbf{e}_{\mathrm{u}} + \Omega_{\mathrm{v}}\mathbf{e}_{\mathrm{v}}$, yields $s\boldsymbol{\Omega} = u\mathbf{e}_{\mathrm{u}} + v\mathbf{e}_{\mathrm{v}}$, with $\Omega_{\mathrm{u}} = \boldsymbol{\Omega} \cdot \mathbf{e}_{\mathrm{u}} > 0$, $\Omega_{\mathrm{v}} = \boldsymbol{\Omega} \cdot \mathbf{e}_{\mathrm{v}}$, $u = \Omega_{\mathrm{u}} s$, and $v = \Omega_{\mathrm{v}} s$. Obviously, u and v are the spatial coordinates in the directions $\mathbf{e}_{\mathrm{u}}$ and $\mathbf{e}_{\mathrm{v}}$, respectively. Now, let us assume that f does not depend on v. As a result, we see that $f(\mathbf{r} + s\boldsymbol{\Omega}) = f(\mathbf{r} + u\mathbf{e}_{\mathrm{u}})$. Defining $h\,(u) = f(\mathbf{r} + u\mathbf{e}_{\mathrm{u}})$, and recalling (3.27), i.e., $g\,(s) = f(\mathbf{r} + s\boldsymbol{\Omega})$, we obtain $g\,(s) = h\,(u) = h(u\,(s))$, giving $g'\,(s) = \Omega_{\mathrm{u}} h'\,(u)$. By straightforward calculations, we get

$$h(u) = \begin{cases} 1, & u_k^0 \le u \le u_k^1 \\ 0, & \text{rest} \end{cases},$$

and further,

$$g'\,(s\,(u)) = \Omega_{\mathrm{u}} \sum_{k=1}^{N} \left[\delta\left(u - u_k^0\right) - \delta\left(u - u_k^1\right)\right].$$

Here, we set $u_k^i = \Omega_u s_k^i$, so that $\mathbf{r} + u_k^i \mathbf{e}_u$ are the points at the interfaces $i \to j$ in direction $\mathbf{e}_u$. Note that this result may follow directly from (3.28) and the scaling property of the Dirac delta function $\delta(ax) = (1/|a|)\delta(x)$. Employing now the previous arguments we end up with

$$\overline{(\boldsymbol{\Omega} \cdot \nabla f) I} = \Omega_u \left(\frac{1 - \overline{f}}{\lambda_{u0}} I_0 - \frac{\overline{f}}{\lambda_{u1}} I_1 \right), \tag{3.35}$$

where $\lambda_{ui}(\mathbf{r}, \mathbf{e}_u)$ is the transition length in material i in direction $\mathbf{e}_u$. An alternative proof of (3.35) can be given as follows: From $f(\mathbf{r} + s\boldsymbol{\Omega}) = f(\mathbf{r} + u\mathbf{e}_u)$, we deduce that "the probability that the point $\mathbf{r} + \triangle s \boldsymbol{\Omega}$ is in material j given that the point $\mathbf{r}$ is in material i" is equal to the probability that "the point $\mathbf{r} + \triangle u \mathbf{e}_u$ is in material j given that the point $\mathbf{r}$ is in material i". Hence, we may write

$$P\left(B_j \mid A_i\right) = \frac{\triangle s}{\lambda_i} = \frac{\triangle u}{\lambda_{ui}},$$

yielding $\lambda_i = \lambda_{ui}/\Omega_u$. This result together with (3.34) gives (3.35). It should be pointed out that the choice of the vector $\mathbf{e}_u$ with the property $\boldsymbol{\Omega} \cdot \mathbf{e}_u > 0$ can always be done. In general, for positive or negative values of $\boldsymbol{\Omega} \cdot \mathbf{e}_u$, (3.35) remains valid with $|\Omega_u|$ in place of Ω_u, and with λ_{ui} being the transition length in material i in direction $\text{sign}(\Omega_u)\mathbf{e}_u$. The result established in (3.35) can be particularized to two- and one-dimensional broken clouds as follows:

1. Two-dimensional cloud fields are homogeneous in altitude, and are characterized by the choice $f(\mathbf{r}) = f(x, y)$. For the direction $\boldsymbol{\Omega} = (\theta, \varphi)$, we consider the decomposition $\boldsymbol{\Omega} = \Omega_\rho \mathbf{e}_\rho + \Omega_z \mathbf{e}_z$, where $\mathbf{e}_\rho$ is the polar unit vector in the xy-plane, $\mathbf{e}_z$ is the Cartesian unit vector along the z-axis, $\Omega_\rho = \boldsymbol{\Omega} \cdot \mathbf{e}_\rho = \sin\theta$, and $\Omega_z = \boldsymbol{\Omega} \cdot \mathbf{e}_z = \cos\theta$. As f does not depend on z, (3.35) holds with $\Omega_u = \Omega_\rho = \sin\theta$.
2. One-dimensional cloud fields are relevant when dealing with two-dimensional media, and are described by the choice $f(\mathbf{r}) = f(x)$. For the direction $\boldsymbol{\Omega} = (\theta, \varphi)$, we consider the decomposition $\boldsymbol{\Omega} = \Omega_x \mathbf{e}_x + \Omega_{yz} \mathbf{e}_{yz}$, where $\mathbf{e}_x$ is a unit vector along the x-axis, chosen such that $\boldsymbol{\Omega} \cdot \mathbf{e}_x > 0$, $\mathbf{e}_{yz}$ is a unit vector in the yz-plane, $\Omega_x = \boldsymbol{\Omega} \cdot \mathbf{e}_x = \sin\theta\,|\cos\varphi|$, and $\Omega_{yz} = \boldsymbol{\Omega} \cdot \mathbf{e}_{yz}$. As f does not depend on y and z, (3.35) holds with $\Omega_u = \Omega_x = \sin\theta\,|\cos\varphi|$.

3.4.3 A Stochastic Model for Homogeneous Three-, Two- and One-Dimensional Broken Clouds

The solution of the boundary-value problem consisting in Eqs. (3.18), (3.21) and (3.22) is not a trivial task. To simplify our analysis, we assume that the statistics of f are independent of x and y (horizontal invariance of cloud statistics), yielding

$$\overline{f} = \overline{f}\,(z)\,,\ \overline{f'^2} = \overline{f'^2}\,(z)\,,\ \text{and}\ \lambda_i = \lambda_i\,(z,\mu)\,. \tag{3.36}$$

In addition, we suppose that the deterministic optical properties of the cloud and of the clear sky depend only on the altitude, i.e., $\sigma^{\mathtt{ld}}_{\mathtt{ext}} = \sigma^{\mathtt{ld}}_{\mathtt{ext}}\,(z)$ and $\sigma^{0}_{\mathtt{ext}} = \sigma^{0}_{\mathtt{ext}}\,(z)$. A possible choice for $\sigma^{\mathtt{ld}}_{\mathtt{ext}}$, which determines the distribution of the cloud in altitude, is

$$\sigma^{\mathtt{ld}}_{\mathtt{ext}}\,(z) = \begin{cases} \sigma_{\mathtt{c}}, & h_1 \le z \le h_2 \\ 0, & \text{rest} \end{cases}\,, \tag{3.37}$$

where h_1 and h_2 are the cloud-base and the cloud-top heights, respectively. From the above assumptions, we infer that $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}\,(z)$, $\mathbf{S} = \mathbf{S}(z, \boldsymbol{\Omega}, \boldsymbol{\Omega}')$ and $\mathbf{I}_{\odot} = \mathbf{I}_{\odot}\,(z)$. Taking into account that the boundary conditions (3.21) and (3.22) are uniform, then the mean radiative properties are independent of x and y, i.e., $\mathbf{I} = \mathbf{I}\,(z, \boldsymbol{\Omega})$, and the streaming operator $\boldsymbol{\Omega} \cdot \nabla$ in (3.18) is transformed to $\mu \mathrm{d}/\mathrm{d}z$.

The horizontal invariance assumption (3.36) becomes stronger for homogeneous cloud statistics, in which case,

$$\nabla \overline{f} = 0,\ \nabla \overline{f'^2} = 0,\ \text{and}\ \lambda_i = \lambda_i\,(\mu)\,. \tag{3.38}$$

For statistically homogeneous cloud fields, the λ_i have the physical interpretation of the mean chord lengths in material i in direction $\boldsymbol{\Omega}$, and the cloud fraction $\overline{f}$ can be expressed as

$$\overline{f} = \frac{\lambda_1\,(\mu)}{\lambda_0\,(\mu) + \lambda_1\,(\mu)}. \tag{3.39}$$

From (3.39), it is apparent that $\overline{f}$ is independent of μ, if $\lambda_0\,(\mu)$ and $\lambda_1\,(\mu)$ have the same dependency on μ. An example of a stochastic cloud model fulfilling (3.38) and parametrized by the cloud fraction $\overline{f}$ and the mean cloud aspect ratio $\gamma = l/h$, where h and l are the characteristic vertical and horizontal dimensions of the clouds, respectively, can be found in Malvagi et al. (1993). The correlation length $\lambda_{\mathtt{c}}$, associated with a homogeneous statistics is defined by

$$\frac{1}{\lambda_{\mathtt{c}}} = \frac{1}{\lambda_0} + \frac{1}{\lambda_1}. \tag{3.40}$$

Coming to the problem of estimating $\overline{(\boldsymbol{\Omega} \cdot \nabla f')I'}$, we observe that for homogeneous cloud statistics, we have $\overline{(\boldsymbol{\Omega} \cdot \nabla f')I'} = \overline{(\boldsymbol{\Omega} \cdot \nabla f)I}$, and so, from (3.34), (3.39) and (3.40), we obtain

$$\overline{(\boldsymbol{\Omega} \cdot \nabla f')\,I'} = -\frac{1}{\lambda_{\mathtt{c}}}\overline{f'I'}. \tag{3.41}$$

Homogeneous and isotropic cloud statistics are characterized by

$$\nabla \overline{f} = 0,\ \nabla \overline{f'^2} = 0,\ \text{and}\ \lambda_i = \text{constant.}, \tag{3.42}$$

and it is apparent that for two- and one-dimensional cloud fields, described by $f = f(x, y)$ and $f = f(x)$, respectively, the horizontal invariance assumption (3.36) implies the homogeneity and isotropy assumption (3.42). Therefore, for two-dimensional cloud fields, (3.35) yields

$$\overline{(\boldsymbol{\Omega} \cdot \nabla f')\, I'} = -\frac{\sin\theta}{\lambda_c} \overline{f' I'}, \tag{3.43}$$

while for one-dimensional cloud fields, (3.35) gives

$$\overline{(\boldsymbol{\Omega} \cdot \nabla f')\, I'} = -\frac{\sin\theta\, |\cos\varphi|}{\lambda_c} \overline{f' I'}. \tag{3.44}$$

The estimates (3.41), (3.43) and (3.44) are based on the closure assumption (3.32) of the Levermore model. The equivalence between the interface radiances and the volumetric radiances is exact for Markovian cloud statistics in the absence of scattering. For non-Markovian cloud statistics, this equivalence is a severe approximation (even in the absence of scattering), which can be improved by replacing the correlation length λ_c in (3.41), (3.43) and (3.44) by an effective correlation length $\lambda_{\mathtt{eff}}$ (Levermore et al. 1988). The effective correlation length is computed as $\lambda_{\mathtt{eff}} = \lambda_c / q$, where the correction factor q depends on the deterministic extinction coefficients in the cloud and in the clear sky, and the cloud statistics. In a compact notation, (3.41), (3.43) and (3.44) can be written as

$$\overline{(\boldsymbol{\Omega} \cdot \nabla f')\, I'} = -F(\boldsymbol{\Omega})\, \frac{q}{\lambda_c} \overline{f' I'}, \tag{3.45}$$

where F is an angular factor depending on the direction $\boldsymbol{\Omega} = (\theta, \varphi)$. For three-dimensional cloud fields, we have $F(\boldsymbol{\Omega}) = 1$ and $\lambda_c = \lambda_c(\mu)$, for two-dimensional cloud fields we have $F(\boldsymbol{\Omega}) = \sin\theta$ and $\lambda_c = \text{constant}$, and for one-dimensional cloud fields we have $F(\boldsymbol{\Omega}) = \sin\theta\, |\cos\varphi|$ and $\lambda_c = \text{constant}$.

The treatment of one-dimensional cloud fields is the most challenging task, because F depends not only on the zenith angle θ but also on the azimuthal angle φ. As a Fourier expansion of the function $|\cos\varphi|$ involves sine and cosine modes, a Fourier azimuthal expansion of the radiance vector $\mathbf{I}$ will also involves sine and cosine modes. In order to simplify the solution method, we make a further assumption, namely that $|\cos\varphi| \approx \cos^2\varphi$. This approximation, yielding an relative error of about 0.2 in the L_2-norm on $[0, 2\pi]$, is quite acceptable. In this context, we express $\overline{(\boldsymbol{\Omega} \cdot \nabla f')\, I'}$ as

$$\overline{(\boldsymbol{\Omega} \cdot \nabla f')\, I'} = -\sin\theta \frac{q}{\lambda_c} \overline{f' I'} + \sin\theta \left(1 - \cos^2\varphi\right) \frac{q}{\lambda_c} \overline{f' I'} \tag{3.46}$$

and rewrite the stochastic equation (3.18) as

$$\mu \frac{\mathrm{d}\mathbf{I}}{\mathrm{d}z}(z,\boldsymbol{\Omega}) = -\boldsymbol{\Sigma}_q(\mu,z)\,\mathbf{I}(z,\boldsymbol{\Omega}) + \frac{1}{4\pi}\int_{4\pi}\mathbf{S}\left(z,\boldsymbol{\Omega},\boldsymbol{\Omega}'\right)\mathbf{I}\left(\mathbf{r},\boldsymbol{\Omega}'\right)\mathrm{d}\Omega' + \frac{F_\odot}{4\pi}\mathbf{S}(z,\boldsymbol{\Omega},\boldsymbol{\Omega}_\odot)\,\mathbf{I}_\odot(z) + \sqrt{1-\mu^2}\left(1-\cos^2\varphi\right)\frac{q}{\lambda_\mathrm{c}}\mathbf{I}'(z,\boldsymbol{\Omega}) \tag{3.47}$$

where the correction vector $\mathbf{I}'$ is given by $\mathbf{I}' = [\,0,\ \overline{f'I'}\,]^T$, and the new extinction matrix $\boldsymbol{\Sigma}_q$ is given by

$$\boldsymbol{\Sigma}_q(\mu,z) = \begin{bmatrix} \sigma^0_{\mathtt{ext}}(z) + \sigma^{1\mathtt{d}}_{\mathtt{ext}}(z)\overline{f} & \sigma^{1\mathtt{d}}_{\mathtt{ext}}(z) \\ \sigma^{1\mathtt{d}}_{\mathtt{ext}}(z)\overline{f'^2} & \sigma^0_{\mathtt{ext}}(z) + \sigma^{1\mathtt{d}}_{\mathtt{ext}}(z)\left(1-\overline{f}\right) + \sqrt{1-\mu^2}\,\dfrac{q}{\lambda_\mathrm{c}} \end{bmatrix}.$$

The stochastic equation (3.47) can be used to model various scenarios:

1. A stochastic model for homogeneous three-dimensional cloud statistics is characterized by $\mathbf{I}' = 0$, $\boldsymbol{\Sigma}_q = \boldsymbol{\Sigma}_q(\mu = 0, z)$ and $\lambda_\mathrm{c} = \lambda_\mathrm{c}(\mu)$, while a stochastic model for two-dimensional cloud fields is described by $\mathbf{I}' = 0$.
2. A stochastic model, in which we neglect the covariance vector $\mathbf{I}_\nabla$ depending on the gradient of f', is characterized by $\mathbf{I}' = 0$ and $\boldsymbol{\Sigma}_q = \boldsymbol{\Sigma}_q(\mu = 1, z)$, or equivalently, by $q = 0$.
3. Homogeneous scenes are described by setting $q = 0$, and by putting $\overline{f'^2} = 0$ in the expressions of $\mathbf{S}$ and $\boldsymbol{\Sigma}_0$, as well as, $\overline{f'(\mathrm{e}^{-\tau^\odot_{\mathtt{ext}}})'} = 0$ in the expression of $\mathbf{I}_\odot$. Cloudy and clear scenes are specified by $\overline{f} = 1$ and $\overline{f} = 0$, respectively.

Equation (3.47) is an one-dimensional stochastic model, in which the stochastic properties of the clouds are encapsulated in the bulk geometrical statistics $\overline{f}$, $\overline{f'^2}$ and $\overline{\lambda_i}$, $i = 0, 1$, as well as, in the bulk statistics of the solar transmission $\overline{\mathrm{e}^{-\tau^\odot_{\mathtt{ext}}}}$ and $\overline{f'(\mathrm{e}^{-\tau^\odot_{\mathtt{ext}}})'}$. For any constructive stochastic cloud model, these statistics can be obtained by a numerical averaging method for an ensemble of realizations of the indicator function f.

For one-dimensional cloud fields, the stochastic radiative transfer Eq. (3.47) is solved by an iterative method: Starting with a zero correction vector $\mathbf{I}'^{(0)} = 0$, the radiance vector $\mathbf{I}^{(n)}$ at the iteration step n, is computed for a correction vector $\mathbf{I}'^{(n-1)}$ at the previous iteration step. Two main features influence the convergence of the iterative method:

1. As in general, $\overline{f'I'} \ll \overline{I}$, the correction vector $\mathbf{I}'$ is small compared to the radiance vector $\mathbf{I}$. In this regard, it seems that the Levermore model, formulated in terms of the radiance vector $\mathbf{I} = [\,\overline{f}I_1,\ \left(1-\overline{f}\right)I_0\,]^T$, does not exhibit this feature.
2. As a consequence of the representation (3.46), the radiance vector at the first iteration step $\mathbf{I}^{(1)}$ is the solution of a stochastic model for two-dimensional cloud fields.

The solar transmission vector deserves a word of explanation.

1. The stochastic model (3.47) is formulated in terms of the diffuse radiance and relies on the assumption that the solar transmission vector $\mathbf{I}_{\odot} = [\,\overline{\mathrm{e}^{-\tau_{\mathtt{ext}}^{\odot}}},\ \overline{f'(\mathrm{e}^{-\tau_{\mathtt{ext}}^{\odot}})'}\,]$ is known. More precisely, we suppose that we have a constructive stochastic cloud model for computing $\mathbf{I}_{\odot}$ by a numerical averaging method.
2. The direct solar beam associated to the stochastic radiative transfer equation (3.47) is given by $F_{\odot}\delta(\mu+\mu_{\odot})\delta(\varphi-\varphi_{\odot})\mathbf{I}_{\odot q}\,(z)$, where the solar transmission vector $\mathbf{I}_{\odot q}$ solves a boundary-value problem for a semi-infinite medium consisting in (I) the stochastic transfer equation

$$\mu_{\odot}\frac{\mathrm{d}\mathbf{I}_{\odot q}}{\mathrm{d}z}\,(z) = -\boldsymbol{\Sigma}_{\odot q}\,(z)\,\mathbf{I}_{\odot q}\,(z)\,, \tag{3.48}$$

with

$$\boldsymbol{\Sigma}_{\odot q} = \begin{bmatrix} \sigma_{\mathtt{ext}}^{0} + \sigma_{\mathtt{ext}}^{1\mathtt{d}}\,\overline{f} & \sigma_{\mathtt{ext}}^{1\mathtt{d}} \\ \sigma_{\mathtt{ext}}^{1\mathtt{d}}\,\overline{f'^{2}} & \sigma_{\mathtt{ext}}^{0} + \sigma_{\mathtt{ext}}^{1\mathtt{d}}\left(1-\overline{f}\right) + \sqrt{1-\mu_{\odot}^{2}}\,|\cos\varphi_{\odot}|\,\dfrac{q}{\lambda_{\mathtt{c}}} \end{bmatrix},$$

and (II) the boundary condition at the top of the atmosphere $\mathbf{I}_{\odot q}(z_{\mathtt{t}}) = [\,1,\ 0\,]^{T}$. Note that for $\mu_{\odot} = 0$, we have $\boldsymbol{\Sigma}_{\odot q} = \boldsymbol{\Sigma}_{q}$. For a stochastic model formulated in terms of the total radiance, the diffuse radiance will solve Eq. (3.47) with $\mathbf{I}_{\odot q}$ in place of $\mathbf{I}_{\odot}$. Obviously, if q is not optimal, then the errors in $\mathbf{I}_{\odot q}$ will affect the accuracy of computing $\mathbf{I}$.

Nevertheless, the q-dependent solution of the stochastic transfer Eq. (3.48) can be used to define an optimal value of the correction factor q_0 as

$$q_0 = \arg\min_{q} \left\| \left[\mathbf{I}_{\odot}\right]_1 - \left[\mathbf{I}_{\odot q}\right]_1 \right\|_{2z}, \tag{3.49}$$

where $\|\cdot\|_{2z}$ is the L_2-norm on the altitude interval of analysis, and the notation $[\mathbf{I}_{\odot}]_1$ refers to the first component of $\mathbf{I}_{\odot}$. It is clear that q_0 is a correction of the closure assumption (3.32) for the direct solar radiance (in the absence of scattering) and not for the diffuse radiance (in the presence of scattering). Thus, q_0 is a remedy of the closure assumption (3.32) for non-Markovian cloud statistics and not for scattering media. If this value of the correction factor will guarantee the optimality of the radiance vector will be checked by a numerical analysis.

For inhomogeneous cloud statistics, and in particular for vertical variations of the statistical characteristics of the cloud field as in (3.36), it is preferable to use the Levermore–Pomraning equations. Accounting of (3.34), we deduce that these equation can be expressed in matrix form as in (3.47), where $\mathbf{S}$ is given by (3.25), $\mathbf{I}' = 0$, and

$$\boldsymbol{\Sigma}_{q}\,(\mu, z) = \begin{bmatrix} \sigma_{\mathtt{ext}}^{0}(z) + \sigma_{\mathtt{ext}}^{1\mathtt{d}}(z) + \dfrac{q}{\lambda_1\,(z,\mu)} & -\dfrac{q}{\lambda_0\,(z,\mu)} \\ -\dfrac{q}{\lambda_1\,(z,\mu)} & \sigma_{\mathtt{ext}}^{0}(z) + \dfrac{q}{\lambda_0\,(z,\mu)} \end{bmatrix}.$$

Thus, the stochastic equation (3.47) in its general form describes both homogeneous and inhomogeneous cloud statistics, and from a computationally point of view, the design of an efficient solution method for such kind of transfer equations is needed.

3.5 Numerical Evaluation of the Stochastic Radiative Transfer Model

3.5.1 Discrete Ordinate Solution

Here the stochastic equation (3.47) is solved in the framework of the discrete ordinate method with matrix exponential (Doicu and Trautmann 2009a, b). Assuming $\varphi_\odot = 0$, we express the radiance vector and the pseudo-phase matrix as Fourier cosine series in the azimuthal angle φ,

$$\mathbf{I}(z, \boldsymbol{\Omega}) = \sum_{m \geq 0} \mathbf{I}_m(z, \mu) \cos m\varphi$$

and

$$\mathbf{S}\left(z, \boldsymbol{\Omega}, \boldsymbol{\Omega}'\right) = \sum_{m \geq 0} (2 - \delta_{m0})\, \mathbf{S}_m\left(z, \mu, \mu'\right) \cos\left[m\left(\varphi - \varphi'\right)\right],$$

respectively. The matrix $\mathbf{S}_m$ is given by

$$\mathbf{S}_m\left(z, \mu, \mu'\right) = \sum_{n \geq m} \boldsymbol{\Pi}_n(z)\, P_n^m(\mu)\, P_n^m\left(\mu'\right),$$

where

$$\boldsymbol{\Pi}_n = \begin{bmatrix} \sigma_{\mathtt{sct}}^0 \chi_n^0 + \sigma_{\mathtt{sct}}^{\mathtt{1d}} \overline{f} \chi_n^1 & \sigma_{\mathtt{sct}}^{\mathtt{1d}} \chi_n^1 \\ \sigma_{\mathtt{sct}}^{\mathtt{1d}} \overline{f'^2} \chi_n^1 & \sigma_{\mathtt{sct}}^0 \chi_n^0 + \sigma_{\mathtt{sct}}^{\mathtt{1d}} \left(1 - \overline{f}\right) \chi_n^1 \end{bmatrix},$$

while χ_n^0 and χ_n^1 are the expansion coefficients of the Rayleigh and the Mie phase functions in terms of the normalized Legendre polynomials, respectively.

The stochastic equation for the mth azimuthal component reads as

$$\begin{aligned} \mu \frac{\mathrm{d}\mathbf{I}_m}{\mathrm{d}z}(z, \mu) = &-\boldsymbol{\Sigma}_q(\mu, z)\, \mathbf{I}_m(z, \mu) + \frac{1}{2} \int_{-1}^{1} \mathbf{S}_m\left(z, \mu, \mu'\right) \mathbf{I}_m(z, \mu)\, \mathrm{d}\Omega' \\ &+ \frac{F_0}{4\pi} (2 - \delta_{m0})\, \mathbf{S}_m(z, \mu, -\mu_\odot)\, \mathbf{I}_\odot(z) \\ &+ \sqrt{1 - \mu^2} \frac{q}{2\lambda_{\mathrm{c}}} \left[\alpha_m \mathbf{I}'_m(z, \mu) - \beta_m \mathbf{I}'_{m-2}(z, \mu) - \gamma_m \mathbf{I}'_{m+2}(z, \mu)\right], \end{aligned} \tag{3.50}$$

where $\alpha_1 = 1/2$ and $\alpha_m = 1$ for $m \neq 1$, $\beta_0 = \beta_1 = 0$, $\beta_2 = 1$ and $\beta_m = 1/2$ for $m \geq 3$, and $\gamma_m = 1/2$ for $m \geq 0$. Let $\{\mu_k, w_k\}_{k=\overline{1,N_{\mathtt{do}}}}$ be the set of Gauss–Legendre quadrature points and weights in the interval $(0, 1)$, where $N_{\mathtt{do}}$ is the number of discrete ordinates per hemisphere. In terms of the $4N_{\mathtt{do}}$-dimensional radiance vector $\mathbf{i}_m(z) = [\mathbf{i}_m^1(z)^T, \mathbf{i}_m^2(z)^T]^T$, with

$$\mathbf{i}_m^i(z) = \left[\left[\mathbf{i}_m^i(z)\right]_1^T, \ldots, \left[\mathbf{i}_m^i(z)\right]_{N_{\mathtt{do}}}^T\right]^T, \quad i = 1, 2, \tag{3.51}$$

$[\mathbf{i}_m^1(z)]_k = \mathbf{I}_m(z, \mu_k)$ and $[\mathbf{i}_m^2(z)]_k = \mathbf{I}_m(z, -\mu_k)$ for all $k = 1, \ldots, N_{\mathtt{do}}$, the stochastic equation (3.50) can be written in the discrete ordinate space as

$$\frac{\mathrm{d}\mathbf{i}_m}{\mathrm{d}z}(z) = \mathbf{A}_m(z)\,\mathbf{i}_m(z) + \mathbf{b}_m(z) + \triangle\mathbf{b}_m(z)\,. \tag{3.52}$$

The $4N_{\mathtt{do}} \times 4N_{\mathtt{do}}$ matrix $\mathbf{A}_m$ possesses the representation

$$\mathbf{A}_m(z) = \begin{bmatrix} \mathbf{A}_m^{11}(z) & \mathbf{A}_m^{12}(z) \\ \mathbf{A}_m^{21}(z) & \mathbf{A}_m^{22}(z) \end{bmatrix},$$

where the 2×2 matrix components of the $2N_{\mathtt{do}}\times2N_{\mathtt{do}}$ block matrices $\mathbf{A}_m^{ij}, i, j = 1, 2$, are given by

$$\begin{aligned}
\left[\mathbf{A}_m^{11}(z)\right]_{kl} &= \frac{1}{2\mu_k}\left[w_l\mathbf{S}_m(z, \mu_k, \mu_l) - 2\boldsymbol{\Sigma}_q(\mu_k, z)\,\delta_{kl}\right], \\
\left[\mathbf{A}_m^{12}(z)\right]_{kl} &= \frac{1}{2\mu_k}w_l\mathbf{S}_m(z, \mu_k, -\mu_l)\,, \\
\left[\mathbf{A}_m^{21}(z)\right]_{kl} &= -\frac{1}{2\mu_k}w_l\mathbf{S}_m(z, -\mu_k, \mu_l)\,, \\
\left[\mathbf{A}_m^{22}(z)\right]_{kl} &= -\frac{1}{2\mu_k}\left[w_l\mathbf{S}_m(z, -\mu_k, -\mu_l) - 2\boldsymbol{\Sigma}_q(-\mu_k, z)\,\delta_{kl}\right]
\end{aligned} \tag{3.53}$$

for all $k, l = 1, \ldots, N_{\mathtt{do}}$. The $4N_{\mathtt{do}}$-dimensional source vector $\mathbf{b}_m$ reads as

$$\mathbf{b}_m(z) = \begin{bmatrix} \mathbf{b}_m^1(z) \\ \mathbf{b}_m^2(z) \end{bmatrix},$$

where as in (3.51), the two-dimensional vector components of the $2N_{\mathtt{do}}$-dimensional block vectors $\mathbf{b}_m^i, i = 1, 2$, are given by

$$\begin{aligned}
\left[\mathbf{b}_m^1(z)\right]_k &= \frac{1}{\mu_k}\frac{F_\odot}{4\pi}(2 - \delta_{m0})\,\mathbf{S}_m(z, \mu_k, -\mu_\odot)\,\mathbf{I}_\odot(z)\,, \\
\left[\mathbf{b}_m^2(z)\right]_k &= -\frac{1}{\mu_k}\frac{F_\odot}{4\pi}(2 - \delta_{m0})\,\mathbf{S}_m(z, -\mu_k, -\mu_\odot)\,\mathbf{I}_\odot(z)\,,
\end{aligned} \tag{3.54}$$

for all $k = 1, \ldots, N_{\text{do}}$. Finally, the $4N_{\text{do}}$-dimensional correction vector $\triangle \mathbf{b}_m$ is expressed as

$$\triangle \mathbf{b}_m(z) = \begin{bmatrix} \triangle \mathbf{b}_m^1(z) \\ \triangle \mathbf{b}_m^2(z) \end{bmatrix},$$

where

$$\left[\triangle \mathbf{b}_m^1(z)\right]_k = \frac{1}{\mu_k}\sqrt{1-\mu_k^2}\frac{q}{2\lambda_c}\left[\alpha_m \mathbf{i}_m'^1(z) - \beta_m \mathbf{i}_{m-2}'^1(z) - \gamma_m \mathbf{i}_{m+2}'^1(z)\right],$$
$$\left[\triangle \mathbf{b}_m^2(z)\right]_k = -\frac{1}{\mu_k}\sqrt{1-\mu_k^2}\frac{q}{2\lambda_c}\left[\alpha_m \mathbf{i}_m'^2(z) - \beta_m \mathbf{i}_{m-2}'^2(z) - \gamma_m \mathbf{i}_{m+2}'^2(z)\right], \quad (3.55)$$

and $\mathbf{i}_m'(z) = [\mathbf{i}_m'^1(z)^T, \mathbf{i}_m'^2(z)^T]^T$, with $[\mathbf{i}_m'^1(z)]_k = \mathbf{I}_m'(z, \mu_k)$ and $[\mathbf{i}_m'^2(z)]_k = \mathbf{I}_m'(z, -\mu_k)$ for all $k = 1, \ldots, N_{\text{do}}$.

In the discrete ordinate space, the surface boundary condition reads as

$$\left[\Im_{2N_{\text{do}}}, \mathbf{R}_m\right]\mathbf{i}_m(z_s) = \mathbf{r}_m(z_s)$$

where $\Im_n$ is the $n \times n$ identity matrix, $[\mathbf{R}_m]_{kl} = -2Aw_l\mu_l\Im_2\delta_{m0}, k, l = 1, \ldots, N_{\text{do}}$, is the reflection matrix, and $[\mathbf{r}_m]_k = (A/\pi)\, F_\odot\, |\mu_\odot|\, \mathbf{I}_\odot(z_s)\, \delta_{m0}, k = 1, \ldots, N_{\text{do}}$, is the reflection vector.

The atmosphere is discretized in homogeneous layers, and on each layer the discrete equation (3.52) is solved by using the matrix exponential formalism. The solution of the vector radiative transfer equation in the framework of the discrete ordinate method with matrix exponential has been described in Doicu and Trautmann (2009b). For the present application, the solution method is similar and has the following peculiarities:

1. The bulk geometrical statistics $\overline{f}$, $\overline{f'^2}$ and $\lambda_i, i = 0, 1$, are calculated together with the bulk statistics of the solar transmission $\overline{e^{-\tau_{\text{ext}}^{\odot}}}$ and $\overline{f'(e^{-\tau_{\text{ext}}^{\odot}})'}$ by a numerical averaging method involving ensemble and spatial averagings.
2. The discrete equation (3.52) is solved by an iterative method. The algorithm of the iterative method involves the following steps: (I) Set $\triangle \mathbf{b}_m^{(0)} = 0$. (II) For $n = 1, 2, \ldots$, compute the radiance vector $\mathbf{i}_m^{(n)}$ for a correction vector $\triangle \mathbf{b}_m^{(n-1)}$ depending on $\mathbf{i}_m'^{(n-1)}$, $\mathbf{i}_{m-2}'^{(n-1)}$ and $\mathbf{i}_{m+2}'^{(n-1)}$. (III) Stop the iteration if $\mathbf{i}_m^{(n)}$ converges.
3. The symmetry of the associated Legendre functions

$$P_n^m(-\mu_k) = (-1)^{n-m} P_n^m(\mu_k),$$

yields $\mathbf{S}_m(z, -\mu_k, -\mu_l) = \mathbf{S}_m(z, \mu_k, \mu_l)$ and $\mathbf{S}_m(z, -\mu_k, \mu_l) = \mathbf{S}_m(z, \mu_k, -\mu_l)$. This result together with $\boldsymbol{\Sigma}_q(-\mu_k, z) = \boldsymbol{\Sigma}_q(\mu_k, z)$ gives $\mathbf{A}_m^{22}(z) = -\mathbf{A}_m^{11}(z)$ and $\mathbf{A}_m^{21}(z) = -\mathbf{A}_m^{12}(z)$, and the matrix eigenvalue problem is solved by exploiting the symmetric structure of the matrix $\mathbf{A}$.
4. The handling of complex eigenvalues is based on fundamental results of the theory of matrix-exponential functions as in Doicu and Trautmann (2009b).

5. The mean radiance in an arbitrary viewing direction is computed by integrating the zeroth-order stochastic equation (3.7) in the post-processing step of the algorithm. To simplify the computational process, the value of $\overline{f'I'}$ in the viewing direction is obtained from the discrete-ordinate values of $\overline{f'I'}$ by cubic spline interpolation.
6. For highly anisotropic phase functions, the delta-M method (Wiscombe 1977) or the delta-fit method (Hu et al. 2000) can be applied.

3.5.2 *Accuracy of the Stochastic Radiative Transfer Model*

To analyze the accuracy of the stochastic model, a base line clear-sky planetary boundary-layer atmosphere with ozone trace gas absorption and Rayleigh scattering in the UV range 325–335 nm has been considered by Doicu et al. (2014b). The numerical analysis has been restricted to one-dimensional broken clouds with top height $h = 2$ km. The cloud was a liquid water-droplet "polluted-cumulus" type, with a modified Gamma size distribution

$$p\,(a) \propto a^{\alpha} \exp\left[-\frac{\alpha}{\gamma}\left(\frac{a}{a_{\mathrm{mod}}}\right)^{\gamma}\right], \tag{3.56}$$

and a droplet size range between 0.02 and 50.0 μm. The cloud droplet scattering has been computed with Mie theory at a wavelength of 330 nm. The parameters of the size distribution were $a_{\mathrm{mod}} = 3.53\,\mu\mathrm{m}$, $\alpha = 8$ and $\gamma = 2.15$, while the total numbers of expansion coefficients of the phase functions were 341. The deterministic cloud extinction field has been given by (3.37), with $\sigma_{\mathrm{c}} = 8\,\mathrm{km}^{-1}$, $h_1 = 0.5$ km and $h_2 = 1.5$ km; thus, the cloud has been placed between 0.5 and 1.5 km, and had an optical thickness of 8 (in the vertical direction). The simulations were performed for unit solar flux and a surface albedo of 0.2. We consider the one-dimensional cellular statistical model of broken clouds developed by Alexandrov et al. (2010). In this model, the atmosphere is represented by a finite lattice with N cells, that are occupied by a cloud with probability p. The continuous extension of this discrete-cell model assumes that each cell is divided in M subcells of size $\triangle x$, and allows for continuous distributions of the chord lengths and the cloud fraction (Plank 1969). The theoretical cloud fraction is independent of the sample size, and is given by

$$\overline{f}_{\mathtt{th}} = \frac{\ln\,(1-p)}{\ln p + \ln\,(1-p)}. \tag{3.57}$$

The cloud model parameters were $N = 4$, $M = 64$ (256 subcells in total) and $\triangle x = 0.025$ km. The length of the sample was $l = N_{\mathrm{x}} \triangle x = 6.4$ km, where $N_{\mathrm{x}} = NM$ is the total number of subcells. The cloud model which is parametrized by the occupation probability p, is visualized in Fig. 3.6 for $p = 0.4$ and $p = 0.8$. Note, that this model provides the exponential cloud chord distribution, what is in agreement

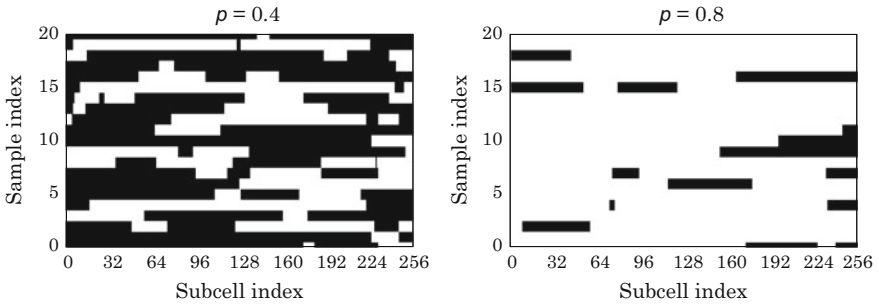

Fig. 3.6 Indicator function $f_i = f(i\triangle x), i = 1, \ldots, 256$, for the occupation probabilities $p = 0.4$ and $p = 0.8$ for 20 sample realizations: white zones correspond to the clouds while black zones correspond to the clear sky

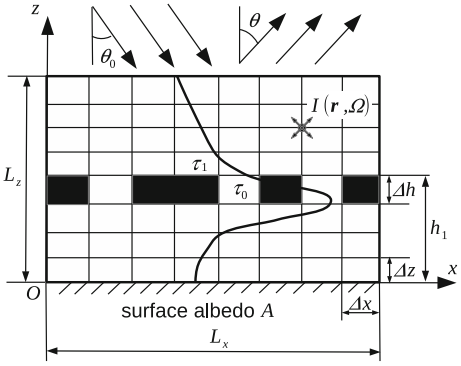

Fig. 3.7 Illustration of SHDOM model

with several studies (e.g., see Zuev and Titov 1996 and Berg and Kassianov 2008 and references therein for more information).

To estimate the accuracy of the stochastic model, the Spherical Harmonics Discrete Ordinate Method (SHDOM) developed by Evans (1998) has been chosen as a reference numerical averaging approach (see Fig. 3.7). SHDOM has been run for a two-dimensional geometry and an adaptive grid with a splitting accuracy of 10^{-4}. The spatial resolutions of the base grid were $\triangle x = \triangle z = 0.025$ km, the number of discrete ordinates was $(N_\mu = 32) \times (N_\varphi = 64)$, and the solution accuracy was 10^{-4}. The simulations were performed for periodic boundary conditions. The mean radiance at the top of the atmosphere $\overline{I}\,(h, \theta)$ has been computed by ensemble averaging over 200 realizations of the random function f, and then, by spatial averaging over the x-axis. The scattering zenith angle varied over the range $-70° < \theta < 70°$, where the positive values of θ correspond to the azimuthal angle $\varphi = 0°$, while the negative values correspond to the azimuthal angle $\varphi = 180°$. The angular resolution of the scattering zenith angle was $\triangle\theta = 1°$. In the numerical analysis the following models have been considered:

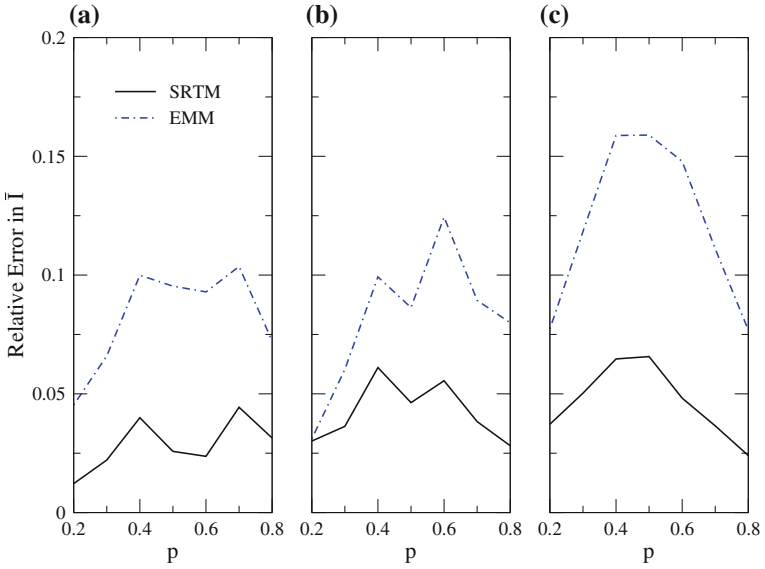

Fig. 3.8 The relative error in the mean radiances at the top of the atmosphere, averaged over the scattering zenith angle θ, as a function of the occupation probability p. The results correspond the following values of the solar zenith angle: **a** $\theta_\odot = 0°$, **b** $\theta_\odot = 30°$, and **c** $\theta_\odot = 60°$

1. a SRTM using the first iteration solution, i.e., a stochastic model based on the radiative transfer Eq. (3.47) with $\mathbf{I}' = 0$, and
2. an external mixing model (EMM), in which the mean radiance at the top of the atmosphere is computed as $\overline{I}\,(h,\theta) = \overline{f}\,I_1\,(h,\theta) + (1-\overline{f})I_0\,(h,\theta)$, where I_1 is the top radiance of a homogeneous cloudy sky, and I_0 is the top radiance of the clear sky.

Obviously, SRTM accounts for horizontal inhomogeneities in cloud optical properties and illumination of cloud sides–neither of these effects are accounted for with EMM. The relative errors in the mean radiances at the top of the atmosphere, averaged over the scattering zenith angle,

$$\varepsilon_{\mathrm{STM/EMM}} = ||\overline{I}_{\mathrm{STM/EMM}}\,(h,\theta) - \overline{I}_{\mathrm{SHDOM}}\,(h,\theta)\,||_{2\theta} / ||\overline{I}_{\mathrm{SHDOM}}\,(h,\theta)\,||_{2\theta},$$

are illustrated in Fig. 3.8. The results evidence the superiority of SRTM over EMM; the relative errors of SRTM are smaller than 6.5%, while the relative errors of EMM can reach 15%. Thus, the accuracy of the stochastic model is at least twice that of an external mixing model. More detailed numerical analysis can be found in Doicu et al. (2014b). However, the computational efficiency of the SRTM is lower, because a two-dimensional problem instead of two independent one-dimensional problems has to be solved. Further we consider several robust models derived from the SRTM.

3.6 Internal Mixing Models

3.6.1 Overview of Internal Mixing Models

The operational atmospheric retrieval algorithms are usually based on the independent column approximation, also known as the independent pixel approximation. To summarize the mathematical fundamentals of the models based on the independent column approximation, we consider a cloudy domain D and express the domain-average radiance at the top of the atmosphere $\langle I \rangle$ as

$$\langle I \rangle = \frac{1}{A} \int_S I\,(x,\, y)\, \mathrm{d}S, \tag{3.58}$$

where I is the radiance computed by a three-dimensional radiative transfer model and A is the area of the surface S bounding the domain D at the top of the atmosphere. Considering a discretization of the domain D in N discrete columns $D = \cup_{k=1}^{N} D_k$, yields $\langle I \rangle = \sum_{k=1}^{N} \nu_k \langle I_k \rangle$, with $\nu_k = A_k/A$ and $\langle I_k \rangle = (1/A_k) \int_{S_k} I\,(x,\, y) \mathrm{d}S$.

1. In the independent column approximation (Cahalan et al. 1994a, b), the horizontal interaction between the columns is neglected, and the radiances $\langle I_k \rangle$ are approximated by the plane-parallel radiances $\mathcal{I}_k$, that is, $\langle I_k \rangle \approx \mathcal{I}_k$, giving $\langle I \rangle \approx (1/N) \sum_{k=1}^{N} \mathcal{I}_k$, for a uniform column division with $\nu_{1k} = 1/N$ for all k.
2. In the independent column approximation with modified source, developed by Gabriel and Evans (1996), the direct solar beam is computed in a full three-dimensional atmosphere, and then used in an independent column diffuse radiative transfer model.
3. Other models based on the independent column approximation rely on the representation $\langle I \rangle \approx (1/N) \sum_{k=1}^{N} \mathcal{I}(\tau_k)$, where τ_k is the optical thickness of the column D_k. Assuming that τ_k are the realizations of a random variable τ, the sample mean representation of $\langle I \rangle$ translates into the integral representation

$$\langle I \rangle \approx \int_0^{\infty} \mathcal{I}(\tau)\, p(\tau) \mathrm{d}\tau, \tag{3.59}$$

   where $p(\tau)$ is a probability density function describing the variation of the optical thickness in D. In the gamma-weighted independent column approximation (e.g., Barker 1996; Barker et al. 1996; Kokhanovsky 2003), τ is assumed to follow a Gamma distribution. Otherwise, $p(\tau)$ can be approximated by a beta distribution or a lognormal distribution (Barker and Davis 2005).
4. Proceeding with (3.59), the choice $p(\tau) = \delta(\tau - \overline{\tau})$, with $\overline{\tau}$ being the mean optical thickness in D, gives $\langle I \rangle \approx \mathcal{I}(\overline{\tau})$. Several models, based on simple rescaling of the mean optical thickness $\overline{\tau}$, have attempted to improve this approximation. In the effective thickness approximation (Cahalan et al. 1994a), the representation $\langle I \rangle \approx \mathcal{I}(\eta\overline{\tau})$ has been used for marine stratocumulus layers, while in the co-packing exponent model (Gabriel et al. 1990; Davis et al. 1990), the choice

$\langle I\rangle \approx \mathcal{I}(\overline{\tau}^{\alpha})$ has been proposed for more strongly variable media. Specifically, Cahalan et al. (1994a) used self-affine fractal functions to describe the internal variability of geometrically plane-parallel clouds, while Gabriel et al. (1990) and Davis et al. (1990) used self-similar fractal sets to describe the extreme geometrical variability of their media. Moreover, in the equivalent homogeneous cloud approximation (Szczap et al. 2000a, b), the effective thickness approximation has been generalized to account for averaging-scale and solar zenith-angle effects. Computationally speaking, these methods capture in a statistical sense the unresolved spatial variability effects by performing a single one-dimensional radiative transfer computation with effective optical parameters. Similar approaches, also known as "homogenization" approaches, are the closure scheme for the fluctuations of the radiance field proposed by Stephens (1988), and the renormalization theory elaborated by Cairns et al. (2000).

5. For broken clouds described by binary statistical mixture, the probability density function in (3.59) is chosen as $p(\tau) = \overline{f}\delta(\tau - \tau_1) + (1 - \overline{f})\delta(\tau - \tau_0)$, where τ_1 and τ_0 are the optical thicknesses in the cloud and in the clear sky, respectively. In this case, we are led to the external mixing model

$$\langle I\rangle \approx \overline{f}\mathcal{I}(\tau_1) + (1 - \overline{f})\mathcal{I}(\tau_0), \tag{3.60}$$

which represents the major simplification of the independent column approximation. For an exhaustive analysis of approximate models in atmospheric three-dimensional radiative transfer we refer to Barker and Davis (2005); Davis and Marshak (2010).

Another class of fast homogenization approaches, the so-called internal mixing models have been analyzed in Doicu et al. (2014a) by considering zeroth-order stochastic models and by imposing appropriate closure relations on the first-order covariance terms.

3.6.2 Zeroth-Order Stochastic Radiative Transfer Models

The first order stochastic model based on Eq. (3.18) can be particularized to the zeroth-order stochastic models for the total radiance I_{t} and the diffuse radiance I. They read as

$$\begin{aligned}
\mu\frac{\mathrm{d}\overline{I_{\mathrm{t}}}}{\mathrm{d}z} &= -\left(\sigma_{\mathrm{ext}}^{0} + \overline{f}\sigma_{\mathrm{ext}}^{1\mathrm{d}}\right)\overline{I_{\mathrm{t}}} - \sigma_{\mathrm{ext}}^{1\mathrm{d}}\overline{f' I_{\mathrm{t}}'} \\
&\quad + \frac{1}{4\pi}\int_{4\pi}\left[\left(\sigma_{\mathrm{sct}}^{0}P_0 + \overline{f}\sigma_{\mathrm{sct}}^{1\mathrm{d}}P_1\right)\overline{I_{\mathrm{t}}} + \sigma_{\mathrm{sct}}^{1\mathrm{d}}P_1\overline{f' I_{\mathrm{t}}'}\right]\mathrm{d}\Omega' \\
\overline{I_{\mathrm{t}-}} &= F_0\delta\left(\boldsymbol{\Omega} - \boldsymbol{\Omega}_{\odot}\right), \quad z = z_{\mathrm{t}}, \\
\overline{I_{\mathrm{t}+}} &= \frac{A}{\pi}\int_{2\pi}|\mu|\,\overline{I_{\mathrm{t}-}}\mathrm{d}\Omega, \quad z = z_{\mathrm{s}},
\end{aligned} \tag{3.61}$$

and

$$\begin{aligned}
\mu\frac{\mathrm{d}\overline{I}}{\mathrm{d}z} &= -\left(\sigma^0_{\mathtt{ext}} + \overline{f}\sigma^{1\mathtt{d}}_{\mathtt{ext}}\right)\overline{I} - \sigma^{1\mathtt{d}}_{\mathtt{ext}}\overline{f'I'} \\
&\quad + \frac{1}{4\pi}\int_{4\pi}\left[\left(\sigma^0_{\mathtt{sct}}P_0 + \overline{f}\sigma^{1\mathtt{d}}_{\mathtt{sct}}P_1\right)\overline{I} + \sigma^{1\mathtt{d}}_{\mathtt{sct}}P_1\overline{f'I'}\right]\mathrm{d}\Omega' + \overline{J_\odot\left(T_\odot\right)} \qquad (3.62)\\
\overline{I}_- &= 0, \quad z = z_{\mathtt{t}}, \\
\overline{I}_+ &= \frac{A}{\pi}F_\odot\mu_\odot\overline{T_\odot} + \frac{A}{\pi}\int_{2\pi}|\mu|\,\overline{I}_-\mathrm{d}\Omega, \quad z = z_{\mathtt{s}},
\end{aligned}$$

respectively. Here, $J_\odot$ is the solar source term expressed in terms of the radiance field I and the solar transmission $T_\odot = \exp(-\tau^\odot_{\mathtt{ext}})$ by

$$J_\odot\left(\mathbf{r}, \boldsymbol{\Omega}; T_\odot\right) = \frac{F_\odot}{4\pi}\left[\sigma^0_{\mathtt{sct}}\left(\mathbf{r}\right)P_0\left(\boldsymbol{\Omega}, \boldsymbol{\Omega}_\odot\right) + f\left(\mathbf{r}\right)\sigma^{1\mathtt{d}}_{\mathtt{sct}}\left(\mathbf{r}\right)P_1\left(\boldsymbol{\Omega}, \boldsymbol{\Omega}_\odot\right)\right]T_\odot(\mathbf{r}). \qquad (3.63)$$

The radiative transfer Eqs. (3.61) and (3.62) can be solved if some closure relations for the covariance terms $\overline{f'I'_{\mathtt{t}}}$ and $\overline{f'I'}$, respectively, are imposed. We have two options:

1. The internal mixing model for the total radiance (IMMT) corresponds to the assumption
$$\overline{f'I'_{\mathtt{t}}} = -(1-\eta)\overline{f}\,\overline{I_{\mathtt{t}}} \qquad (3.64)$$
in (3.61). The coefficient $-(1-\eta)$ is given by the product of the correlation coefficient $r_{f\mathcal{I}}$ between f and $I_{\mathtt{t}}$, and the coefficients of variation c_f and c_I, where, for example, $c_f = \sqrt{\overline{f'^2}}/\overline{f}$. Therefore, the coefficient $-(1-\eta)$ can be regarded as a measure of the correlation between f and $I_{\mathtt{t}}$. In this case, the resulting mean diffuse radiance $\overline{I}$ solves a radiative transfer equation with the extinction coefficient $\sigma^0_{\mathtt{ext}} + \eta\overline{f}\sigma^{1\mathtt{d}}_{\mathtt{ext}}$, the scattering kernel $\sigma^0_{\mathtt{sct}}P_0 + \eta\overline{f}\sigma^{1\mathtt{d}}_{\mathtt{sct}}P_1$, and the solar source term
$$J^{\mathtt{IT}}_\odot = \frac{F_\odot}{4\pi}(\sigma^0_{\mathtt{sct}}P_0 + \eta\overline{f}\sigma^{1\mathtt{d}}_{\mathtt{sct}}P_1)\mathrm{e}^{-\overline{\tau^\odot_{\mathtt{ext}}}}, \quad \overline{\tau^\odot_{\mathtt{ext}}} = \frac{1}{\mu_\odot}\int_z^{z_{\mathtt{t}}}(\sigma^0_{\mathtt{ext}} + \eta\overline{f}\sigma^{1\mathtt{d}}_{\mathtt{ext}})\mathrm{d}z'. \qquad (3.65)$$
2. The internal mixing model for the diffuse radiance (IMMD) corresponds to the assumption
$$\overline{f'I'} = -\left(1-\eta\right)\overline{f}\,\overline{I}, \qquad (3.66)$$
in (3.62). Here, the mean diffuse radiance solves a radiative transfer equation with the extinction coefficient $\sigma^0_{\mathtt{ext}} + \eta\overline{f}\sigma^{1\mathtt{d}}_{\mathtt{ext}}$, the scattering kernel $\sigma^0_{\mathtt{sct}}P_0 + \eta\overline{f}\sigma^{1\mathtt{d}}_{\mathtt{sct}}P_1$, and the solar source term

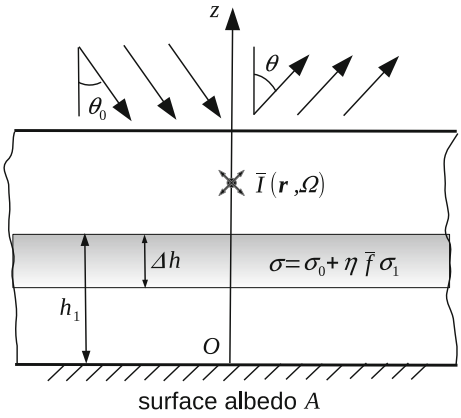

Fig. 3.9 Illustration of internal mixing models

$$J_{\odot}^{\mathrm{ID}} = \overline{J_{\odot}} = \frac{F_{\odot}}{4\pi}(\sigma_{\mathrm{sct}}^{0} P_0 + \overline{f}\,\sigma_{\mathrm{sct}}^{1\mathrm{d}} P_1)\overline{T_{\odot}} + \frac{F_{\odot}}{4\pi}\sigma_{\mathrm{sct}}^{1\mathrm{d}} P_1 \overline{f' T'_{\odot}}, \tag{3.67}$$

which is computed in a full three-dimensional atmosphere.

From (3.64) and (3.66), and for $0 \leq \eta \leq 1$, it is apparent that the internal mixing model assumes a negative and constant correlation. The assumption that the correlation does not depend on the altitude z and the direction Ω is a strong assumption, which can not be theoretically justified, although empirically it seems to be acceptable. Both mixing models require the solution of a one-dimensional radiative transfer equation with a mean extinction coefficient expressed as a linear combination of the extinction coefficients corresponding to the clear sky and to the cloud, i.e.,

$$\overline{\sigma_{\mathrm{ext}}} = (1 - \overline{f})\sigma_{\mathrm{ext}}^{0} + \overline{f}(\sigma_{\mathrm{ext}}^{0} + \eta\sigma_{\mathrm{ext}}^{1\mathrm{d}}) = \sigma_{\mathrm{ext}}^{0} + \eta\overline{f}\sigma_{\mathrm{ext}}^{1\mathrm{d}}, \tag{3.68}$$

as shown in Fig. 3.9. The difference between the two models consists in the way in which the solar source term is treated.

In the framework of a two-stream model, Stephens (1988) assumed that the correlation between f and I_{t} depends on the viewing zenith angle, that is,

$$\overline{f' I'_{\mathrm{t}}} = (1 - \eta)\mu\overline{f}\,\overline{I_{\mathrm{t}}}. \tag{3.69}$$

According to (3.69), the correlation is positive for upwelling radiation and negative for downwelling radiation. For a multi-stream model based on the discrete ordinate solution, this assumption leads to an increase of the computational time because the symmetry of the layer matrix is lost, and the order of the corresponding algebraic eigenvalue problem cannot be reduced by a factor of 2. Another angular-dependent correlation, which guarantees the symmetry of the layer matrix and which will be analyzed in the following is

$$\overline{f' I'_{\mathrm{t}}} = -(1-\eta)\,|\mu|\,\overline{f}\,\overline{I_{\mathrm{t}}}. \tag{3.70}$$

The main problem which needs to be solved is to estimate the closure coefficient η. In general, if the radiative transfer problem depends on a set of deterministic model parameters encapsulated in a vector $\mathbf{b}$, then the optimal value of η can be calculated as

$$\eta_{\mathrm{opt}}(\overline{f}, \mathbf{b}) = \arg\min_{\eta} \left\| \overline{I}(\eta\overline{f}, z_{\mathrm{t}}, \boldsymbol{\Omega}, \mathbf{b}) - \overline{I}_{\mathrm{ref}}(\overline{f}, z_{\mathrm{t}}, \boldsymbol{\Omega}, \mathbf{b}) \right\|, \tag{3.71}$$

where $\overline{I}_{\mathrm{ref}}$ is the mean radiance at the top of the atmosphere computed by a reference model, e.g., the first order stochastic model described in Sect. 3.4, $\overline{I}$ is the radiance at the top of the atmosphere computed by an internal mixing model, and $\|\cdot\|$ is the L_2-norm on the interval of variation of the viewing angle in the upper hemisphere. This approach is similar to the approach used in the equivalent homogeneous cloud approximation (Szczap et al. 2000a, b). The vector $\mathbf{b}$ may comprise the horizontal averaging scale, the solar zenith angle, the cloud parameters (optical and geometrical thicknesses, cloud-top height), the optical thickness of air molecules, the surface albedo, etc. To reduce the computational effort we may compute $\eta_{\mathrm{opt}}(\overline{f}, \mathbf{b}_0)$ for a reference state $\mathbf{b}_0$, and then use the internal mixing models with $\eta = \eta_{\mathrm{opt}}(\overline{f}, \mathbf{b}_0)$ for $\mathbf{b} \neq \mathbf{b}_0$. In view of (3.71), the internal mixing models can be regarded as a parametrization of the chosen reference model.

3.6.3 Accuracy of Internal Mixing Models

To analyze the accuracy of internal mixing models, a clear-sky planetary boundary-layer atmosphere with ozone trace gas absorption and Rayleigh scattering at a wavelength of 330 nm has been considered. The height of the atmospheric boundary layer was $z_{\mathrm{t}} = 10$ km, while the ozone profile has been taken from the model of McLinden et al. (2002). The deterministic cloud extinction field was given by (3.37), the cloud geometrical thickness was $\triangle h = h_2 - h_1 = 1$ km, the cloud single scattering albedo was $\omega_1 = 0.9$, and a Henyey–Greenstein phase function with the asymmetry parameter $g_1 = 0.85$ has been used. The simulations were performed for a unit solar flux, and the cellular stochastic model of broken clouds from Alexandrov et al. (2010) has been used.

In Figs. 3.10 and 3.11 the relative error,

$$\varepsilon(\mathbf{b}) = ||\overline{I}(z_{\mathrm{t}}, \theta, \mathbf{b}) - \overline{I}_{\mathrm{SHDOM}}(z_{\mathrm{t}}, \theta, \mathbf{b})||/||\overline{I}_{\mathrm{SHDOM}}(z_{\mathrm{t}}, \theta, \mathbf{b})||, \tag{3.72}$$

for different state vectors $\mathbf{b}$ is shown. Here, the state vector $\mathbf{b} = \left[\theta_{\odot},\ h_2,\ \tau_1,\ A\right]$ encapsulates the solar zenith angle $\theta_{\odot}$, the cloud-top height h_2, the cloud optical thickness $\tau_1 = \sigma_{\mathrm{c}} \triangle h$, and the surface albedo A. The following general conclusions can be drawn:

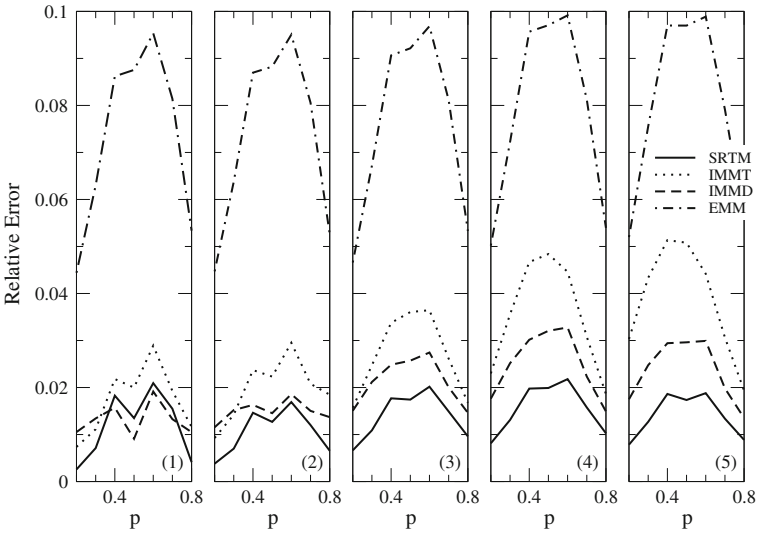

Fig. 3.10 The relative errors ε (**b**) corresponding to the vector, internal and external mixing models for a state vector **b** whose $\theta_\odot$-component takes the values: (1) $\theta_\odot = 0°$, (2) $\theta_\odot = 15°$, (3) $\theta_\odot = 30°$, (4) $\theta_\odot = 45°$, and (5) $\theta_\odot = 60°$. The remaining components of **b** are $h_2 = 5$ km, $\tau_c = 8$, and $A = 0.2$

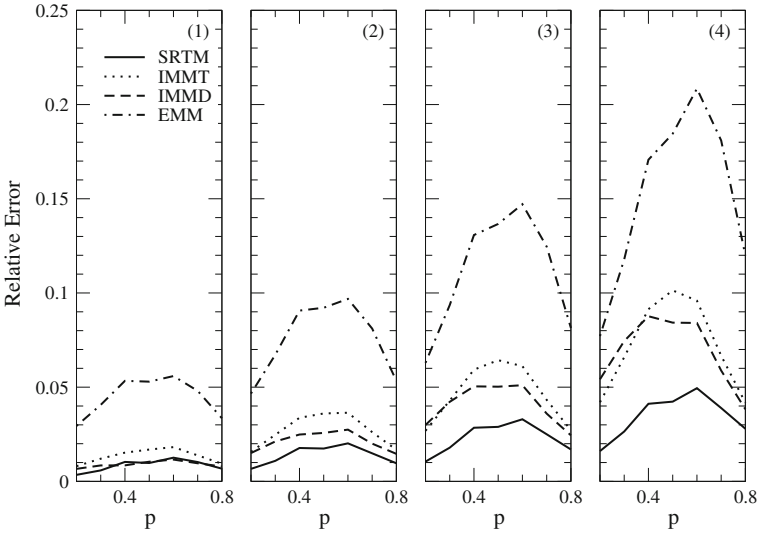

Fig. 3.11 The relative errors ε (**b**) corresponding to the vector, internal and external mixing models for a state vector **b** whose h_2-component takes the values: (1) $h_2 = 3$ km, (2) $h_2 = 5$ km, (3) $h_2 = 7$ km, and (4) $h_2 = 9$ km. The remaining components of **b** are $\theta_\odot = 30°$, $\tau_c = 8$, and $A = 0.2$

1. The first-order stochastic model is in general superior to all mixing models. The reason is that for broken clouds represented by binary statistical mixtures, the closure relation $\overline{f'^2 I'} = (1 - 2\overline{f})\,\overline{f' I'}$ holds.
2. The internal mixing models are superior to the external mixing model, because they are parameterizations of an accurate reference model (cf. (3.71)).

In particular, we have $\varepsilon_{\mathtt{STM}} \leq 5\%$, $\varepsilon_{\mathtt{IMMD}} \leq 9\%$, $\varepsilon_{\mathtt{IMMT}} \leq 10\%$, and $\varepsilon_{\mathtt{EMM}} \leq 21\%$. Regarding the computational time the following remarks are obvious:

1. The external mixing model is slower than the internal mixing models, because they require the solution of two independent plane-parallel problems. However, the performance of the external mixing model can be enhanced by using techniques described in Efremenko et al. (2014). Then the computational times of both modes are almost the same.
2. The mixing models for the diffuse radiance are slower than the mixing models for the total radiance, because they require an additional computational step consisting in the calculation of the solar transmission terms $\overline{T_\odot}$ and $\overline{f' T'_\odot}$ in a three-dimensional atmosphere by ensemble and spatial averaging.

Thus, the most accurate mixing model is internal mixing model for the diffuse radiance, and the fastest mixing model is the internal mixing model for the total radiance. In this regard, it seems promising to obtain an efficient stochastic radiative transfer model based on the method proposed by Budak et al. (2015) in which the small-angle part (Budak et al. 2010) is computed analytically while the rest (regular) part of the radiance is computed by using the two-stream approximation. As shown in Budak and Korkin (2008), this approach can be generalized for 3D computations. In the corresponding stochastic radiative transfer model, the small-angle part would be computed for each realization of the cloud field in a full three-dimensional atmosphere.

In the rest part of the chapter we analyze these models in the context of ozone retrieval problems.

3.7 Retrieval of Cloud Parameters and Ozone

3.7.1 Methodology

The retrieval of cloud parameters uses the radiance model

$$\ln R_{\mathtt{meas}}(\lambda) \approx \ln R_{\mathtt{sim}}(\lambda, \mathbf{x}) \tag{3.73}$$

where λ is the wavelength, and R is the mean radiance spectrum $\overline{I}$ normalized with respect to the solar spectrum I_0, i.e.,

$$R(\lambda) = \frac{\overline{I}(\lambda)}{I_0(\lambda)}. \tag{3.74}$$

For the sake of simplicity, the spectral corrections, also referenced as pseudo-absorbers, have been omitted in (3.73). The spectral corrections are auxiliary functions containing spectral features which are not attributed to the retrieved atmospheric species. The retrieval of trace gas columns uses the differential radiance model

$$\ln \bar{R}_{\mathtt{meas}}(\lambda) \approx \ln \bar{R}_{\mathtt{sim}}(\lambda, \mathbf{x}), \tag{3.75}$$

with

$$\ln \bar{R}_{\mathtt{sim}}(\lambda, \mathbf{x}) = \ln R_{\mathtt{sim}}(\lambda, \mathbf{x}) - P_{\mathtt{sim}}(\lambda, \mathbf{p}_{\mathtt{sim}}(\mathbf{x})) \tag{3.76}$$

and

$$\ln \bar{R}_{\mathtt{meas}}(\lambda) = \ln R_{\mathtt{meas}}(\lambda) - P_{\mathtt{meas}}(\lambda, \mathbf{p}_{\mathtt{meas}}). \tag{3.77}$$

Here, $P_{\mathtt{sim}}$ and $P_{\mathtt{meas}}$ are polynomials of low order with coefficients $\mathbf{p}_{\mathtt{sim}}$ and $\mathbf{p}_{\mathtt{meas}}$, respectively. For a state vector $\mathbf{x}$, the coefficients of the smoothing polynomials $P_{\mathtt{sim}}$ and $P_{\mathtt{meas}}$ are computed as

$$\mathbf{p}_{\mathtt{sim}}(\mathbf{x}) = \arg\min_{\mathbf{p}} \|\log R_{\mathtt{sim}}(\cdot, \mathbf{x}) - P_{\mathtt{sim}}(\cdot, \mathbf{p})\|^2,$$

and

$$\mathbf{p}_{\mathtt{meas}} = \arg\min_{\mathbf{p}} \|\log R_{\mathtt{meas}}(\cdot) - P_{\mathtt{meas}}(\cdot, \mathbf{p})\|^2, \tag{3.78}$$

respectively. In general, a smoothing polynomial is assumed to account for the low-order frequency structure due to scattering mechanisms (clouds, aerosols, and surface reflectance), so that $\ln \bar{R}$ will mainly reflect the absorption process due to gas molecules (Platt and Stutz 2008). Note, that the polynomial subtraction is somewhat similar to the background removal procedure (Seah et al. 2000; Lubenchenko et al. 2018) used in electron spectroscopy. This procedure could be avoided if we were able to take into account all factors which have impact on the spectrum (Afanas'ev et al. 2015, 2016, 2017).

As Eqs. (3.73) and (3.75) are ill-posed, some sort of regularization is required to obtain a solution with physical meaning. In the framework of Tikhonov regularization, the regularized solution $\mathbf{x}^\star$ is a minimizer of the objective function

$$\mathcal{F}(\mathbf{x}) = \frac{1}{2}\left[\left\|\mathbf{y}^\delta - \mathbf{F}(\mathbf{x})\right\|^2 + \alpha\left\|\mathbf{L}(\mathbf{x} - \mathbf{x}_{\mathrm{a}})\right\|^2\right], \tag{3.79}$$

where $\mathbf{y}^\delta = [\ln R_{\mathtt{meas}}(\lambda_i)]_{i=1,m}$ is the noisy data vector, $\mathbf{F}(\mathbf{x}) = [\ln R_{\mathtt{sim}}(\lambda_i, \mathbf{x})]_{i=1,n}$ is the forward model, $\mathbf{L}$ is the regularization matrix, and α is the regularization parameter. The minimization of the Tikhonov function can be formulated as the least squares problem

$$\min_{\mathbf{x}} \mathcal{F}(\mathbf{x}) = \frac{1}{2}\left\|\mathbf{f}(\mathbf{x})\right\|^2, \tag{3.80}$$

where the augmented vector $\mathbf{f}$ is given by

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} \mathbf{F}(\mathbf{x}) - \mathbf{y}^{\delta} \\ \sqrt{\alpha}\mathbf{L}(\mathbf{x} - \mathbf{x}_{\mathrm{a}}) \end{bmatrix}. \tag{3.81}$$

Problem (3.80) can be solved by means of Newton-type methods. Newton-type methods (NTM) are iterative methods, which employ the quadratic model

$$\mathcal{M}_k(\mathbf{p}) = \mathcal{F}(\mathbf{x}_k) + \mathbf{g}(\mathbf{x}_k)^T\mathbf{p} + \frac{1}{2}\mathbf{p}^T\mathbf{G}(\mathbf{x}_k)\mathbf{p} \tag{3.82}$$

as a reasonable approximation to the objective function in the neighborhood of the current iterate $\mathbf{x}_k$ (Doicu et al. 2010). In (3.82), k is the iteration step, while $\mathbf{g}$ and $\mathbf{G}$ are the gradient and the Hessian of $\mathcal{F}$, that is,

$$\mathbf{g}(\mathbf{x}) = \mathbf{K}_{\mathrm{f}}(\mathbf{x})^T\mathbf{f}(\mathbf{x}) \tag{3.83}$$

and

$$\mathbf{G}(\mathbf{x}) = \mathbf{K}_{\mathrm{f}}(\mathbf{x})^T\mathbf{K}_{\mathrm{f}}(\mathbf{x}) + \mathbf{Q}(\mathbf{x}), \tag{3.84}$$

respectively, where

$$\mathbf{K}_{\mathrm{f}}(\mathbf{x}) = \begin{bmatrix} \mathbf{K}(\mathbf{x}) \\ \sqrt{\alpha}\mathbf{L} \end{bmatrix} \tag{3.85}$$

is the Jacobian matrix of $\mathbf{f}(\mathbf{x})$, $\mathbf{Q}(\mathbf{x}) = \sum_{i=1}^{m}[\mathbf{f}(\mathbf{x})]_i\mathbf{G}_i(\mathbf{x})$ is the second-order derivative term, and $\mathbf{G}_i$ is the Hessian of $[\mathbf{f}]_i$.

The objective function (3.79) can be minimized by means of step-length methods. A step-length method requires the computation of a vector $\mathbf{p}_k$ called the search direction, and the calculation of a positive scalar τ_k, the step length, for which it holds that $\mathcal{F}(\mathbf{x}_{k+1}) < \mathcal{F}(\mathbf{x}_k)$, where $\mathbf{x}_{k+1} = \mathbf{x}_k + \tau_k\mathbf{p}_k$. In the Gauss–Newton method, it is assumed that the first-order term $\mathbf{K}_{\mathrm{f}}^T\mathbf{K}_{\mathrm{f}}$ in the expression of the Hessian (3.84) dominates the second-order term $\mathbf{Q}$. For small residual problems, the search direction solves the equation

$$\mathbf{K}_{\mathrm{f}}(\mathbf{x}_k)^T\mathbf{K}_{\mathrm{f}}(\mathbf{x}_k)\mathbf{p} = -\mathbf{K}_{\mathrm{f}}(\mathbf{x}_k)^T\mathbf{f}(\mathbf{x}_k). \tag{3.86}$$

3.7.2 Cloud Parameters Retrieval

In this section we analyze the accuracy of the independent pixel approximation in retrieving cloud optical thickness and design the cloud retrieval algorithm. The forward model is the radiance model (3.73). If not stated otherwise, the parameters of the simulations are: $\theta_{\odot} = 60°$, $\theta = 30°$, $A = 0.2$, and $\overline{f} = 0.5$. The measurement spectra, referred to as synthetic spectra, are the forward-model radiances corresponding to the true state and being perturbed by a measurement noise. The measurement

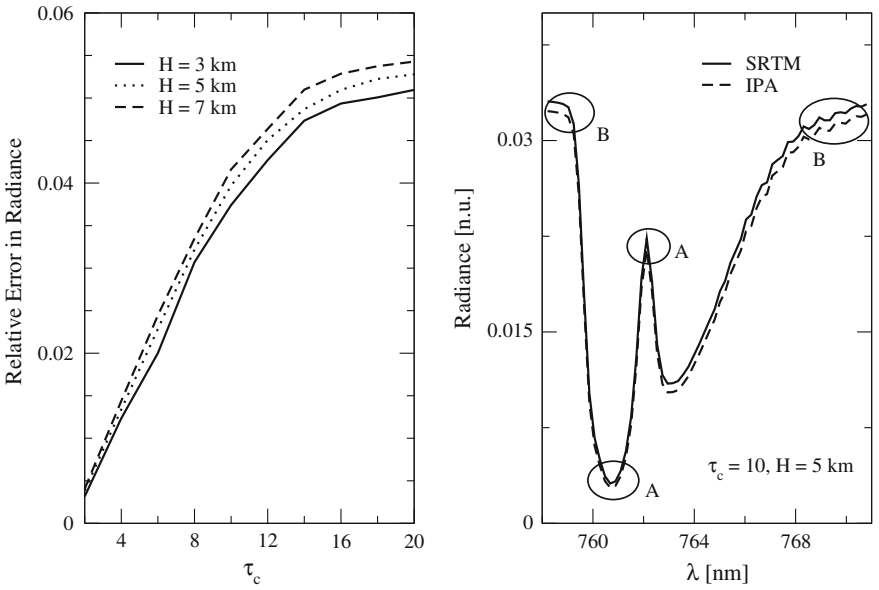

Fig. 3.12 Left: Average relative error in the radiance, computed with the independent pixel approximation, as a function of τ_c and for different values of H. Right: Radiances computed with the stochastic radiative transfer model (SRTM) and the independent pixel approximation (IPA). The spectral region A contains information about the cloud top height, while the spectral region B contains information about the cloud optical thickness

noise is Gaussian, and the signal-to-noise ratio is $SNR = 1000$. The regularization parameter is of the form $\alpha = \sigma^p$, where p is a free parameter to be optimized,

$$\sigma = \frac{\|\overline{I}(\lambda)\|}{\sqrt{m}SNR} \approx 2.4 \cdot 10^{-6}$$

is the noise standard deviation, and $m = 62$ is the number of spectral points.

In a first step we take the radiance model for stochastic clouds as a reference and estimate the relative errors in the radiances computed with the independent pixel approximation. The results are illustrated in Fig. 3.12. The plots in the left panel show that the relative error in the radiance (averaged over the spectral interval) increases with the cloud optical thickness τ_c and the cloud top height H, and attains a maximum level of about 0.05. The plots in the right panel show that the deviations between the radiance curves are more pronounced in the spectral region B (containing information about τ_c) than in the spectral region A (containing information about H). These errors in the data space are forward model errors due to stochastic clouds (simply called stochastic clouds errors here), and in view of Fig. 3.12, mainly affects the retrieval of cloud optical thickness.

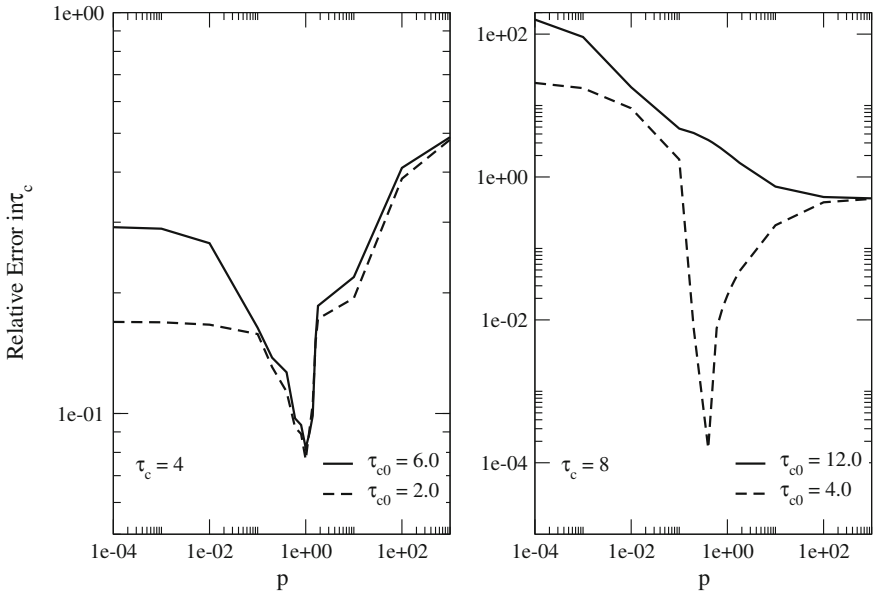

Fig. 3.13 Relative errors in the cloud optical thickness. In the left panel the true cloud optical thickness is $\tau_c = 4.0$ and two initial guesses, namely $\tau_{c0} = 6.0$ and $\tau_{c0} = 2.0$, have been considered. In right panel, the case with $\tau_c = 8.0$, $\tau_{c0} = 12.0$ and $\tau_{c0} = 4.0$ is illustrated

As a next step we analyze the accuracy of the independent pixel approximation in retrieving the cloud optical thickness. The synthetic spectra are computed with the stochastic cloud model, and the retrieval is performed by using the independent pixel approximation and the Gauss–Newton method. In Fig. 3.13 we plot the relative errors in the retrieved cloud optical thickness. The cloud top height is kept constant and is chosen as $H = 4$ km. The true cloud optical thickness are $\tau_c = 4.0$ and $\tau_c = 8.0$, and for each τ_c, two initial guesses τ_0 (smaller and larger than τ_c) are considered. For $\tau_c = 4$, the error curves have a global minimum, and the best retrieval results are characterized by a relative error of about 0.06. For $\tau_c = 8$ and $\tau_{c0} < \tau_c$, it is possible to attain an error of about $2 \cdot 10^{-3}$ provided that the optimal regularization parameter α_{opt} is appropriately chosen. For $\tau_{c0} > \tau_c$, the relative errors decrease but not beyond 0.2. Also note the result $\alpha_{\mathrm{opt}}(\tau_c = 4.0) < \alpha_{\mathrm{opt}}(\tau_c = 8.0)$, which is agreement with the relative errors plotted in Fig. 3.12. Thus, the retrieval based on the independent pixel approximation is considerably influenced by stochastic clouds errors: the retrieval errors are either too large, or the retrieval is influenced by the choice of the initial guess. Because of this inaccuracy, the cloud retrieval algorithm will rely on the stochastic cloud model.

The cloud retrieval algorithm uses the lookup table for the mean radiances of stochastic clouds. The retrieved parameters are the cloud optical thickness and the cloud-top height. The cloud fraction is assumed to be retrieved by an independent algorithm such as OCRA (Loyola et al. 2007, and is considered as an input parameter of the code.

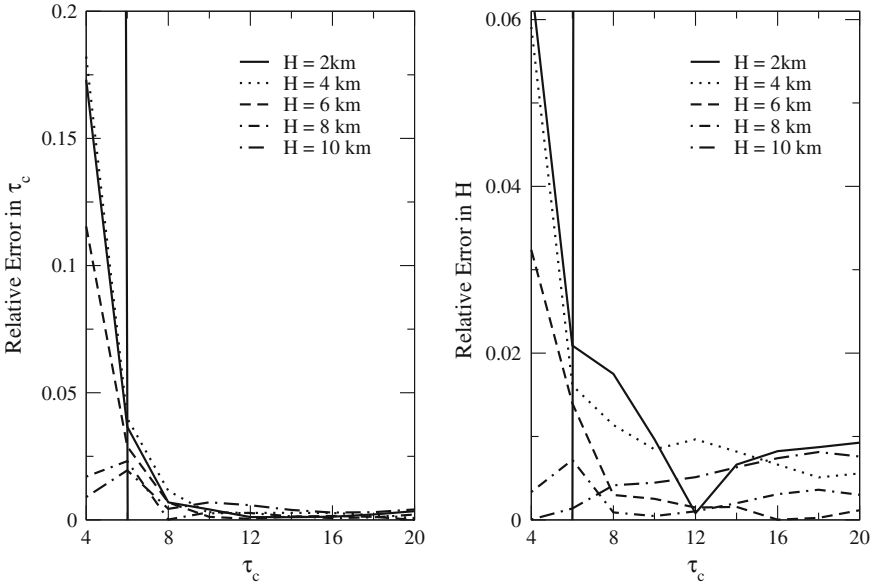

Fig. 3.14 Relative errors in the retrieved cloud optical thickness (left) and cloud top height (right). The vertical lines define the regions with large errors

First, we perform the inversion by using the Gauss–Newton method, and generate synthetic spectra with the stochastic cloud model. The retrieval results of the cloud optical thickness and cloud-top height are shown in Fig. 3.14. The true cloud optical thickness τ_c ranges between 4 and 20 in steps of 2, and the true cloud-top height H ranges between 2 and 10 km in steps of 2 km. Because the forward model errors consist only of interpolation errors (associated to the lookup table for the stochastic clouds radiances) and the noise errors are small, a low value of the regularization parameter is chosen, namely $\alpha = 10^{-6}(p \approx 0.9)$. For $\tau_c \geq 6$ and $H \geq 6$, the relative errors in τ_c are smaller than of about 0.03, while the relative errors in H are below 0.02. The solution domain

$$S_0 = \{(\tau_c, H) | \, 2 \leq \tau_c \leq 6, \;\; 2 \leq H \leq 6\,(\text{km})\}$$

is characterized by large retrieval errors. To understand the source of these errors we plot in Fig. 3.15 the Tikhonov function (3.79) in the cases ($\tau_c = 6$, $H = 6$ km) and ($\tau_c = 4$, $H = 4$ km). The results show that the objective function has only one minimum in the first case and several minima in the second case. The conclusion which can be reached is that the Gauss–Newton method, as a local optimization algorithm, is not able to find the global minimum in the solution domain S_0. In this case global optimization algorithms rather than the Gauss–Newton method should be used. Analysis of global optimization methods for cloud parameter retrieval is given in Efremenko et al. (2016).

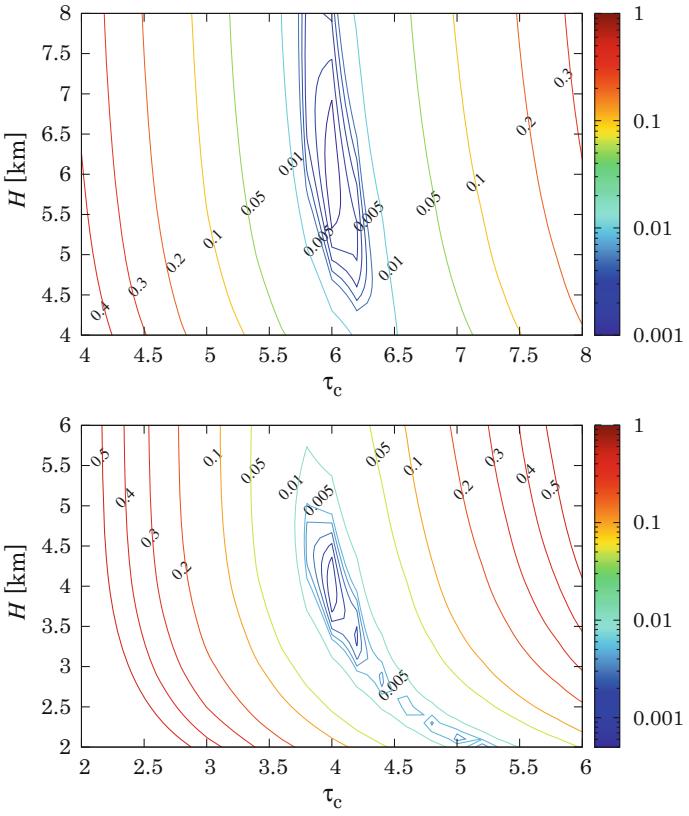

Fig. 3.15 Tikhonov functions in the cases ($\tau_c = 6$, $H = 6$ km) (upper panel) and ($\tau_c = 4$, $H = 4$ km) (lower panel)

3.7.3 Ozone Retrievals

In this section we consider the retrieval of atmospheric ozone from UV measurements. Before proceeding we estimate the forward model errors due to stochastic clouds for the radiance model (3.73) and the differential radiance model (3.75). The results in Fig. 3.16 show that the deviations between the radiance curves corresponding to the differential radiance model are much smaller than those corresponding to the radiance model; they are localized in the peaks of the curves as indicated by the ellipses in Fig. 3.16. In this respect, the goal of our simulations is to analyze the capability of the differential radiance model equipped with the independent pixel approximation to eliminate the stochastic cloud errors. Thus, the retrieval algorithm is based on the differential radiance model with the independent pixel approximation, whereas the inversion algorithm is the Gauss–Newton method. The total and partial columns are retrieved in the spectral interval between 315 and 335 nm. It is well

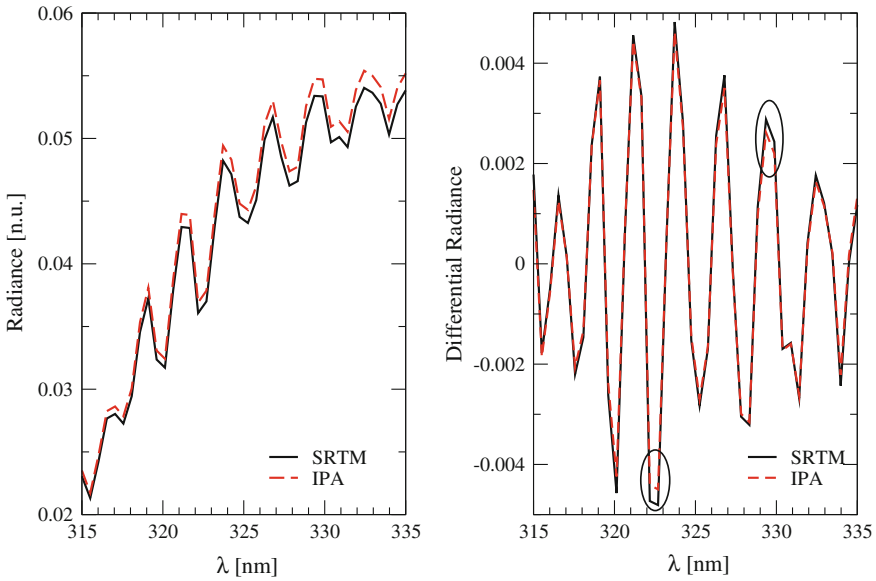

Fig. 3.16 Radiances computed with the stochastic radiative transfer model (SRTM) and the independent pixel approximation (IPA) in the framework of a radiance model (left) and a differential radiance model (right)

known that for a nadir viewing geometry, the errors in the retrieved total column are small, but the errors in the retrieved partial columns are large (the problem is severely ill-posed).

The relative errors in ozone total column and partial columns (averaged over the altitude) are shown in Fig. 3.17. The synthetic spectra are computed with: (1) the independent pixel approximation and $\tau_c = 6.0$, (2) the stochastic cloud model and $\tau_c = 6.0$, and (3) the cloud stochastic model and $\tau_c = 4.0$. The cloud-top height is assumed to be known, and is taken as $H = 4$ km. The retrievals are performed with $\tau_c = 6.0$. In this regard, case (1) considers only the noise error, case (2) includes both the noise error and the forward model error due to stochastic clouds, while case (3) includes in addition to case (2) uncertainties in the cloud optical thickness. The last scenario corresponds to an incorrect retrieval of the cloud optical thickness, for example, by a radiance model with the independent pixel approximation. The results can be summarized as follows:

1. For the total column retrieval, the smallest relative errors are 0.007 in case (1), 0.010 in case (2), and 0.016 in case (3). The interval of variation of the exponent p corresponding to the optimal value of the regularization parameter $\alpha_{\mathtt{opt}} = \sigma^p$ is large ($0.2 \leq p \leq 2.0$), and is the same in all cases.

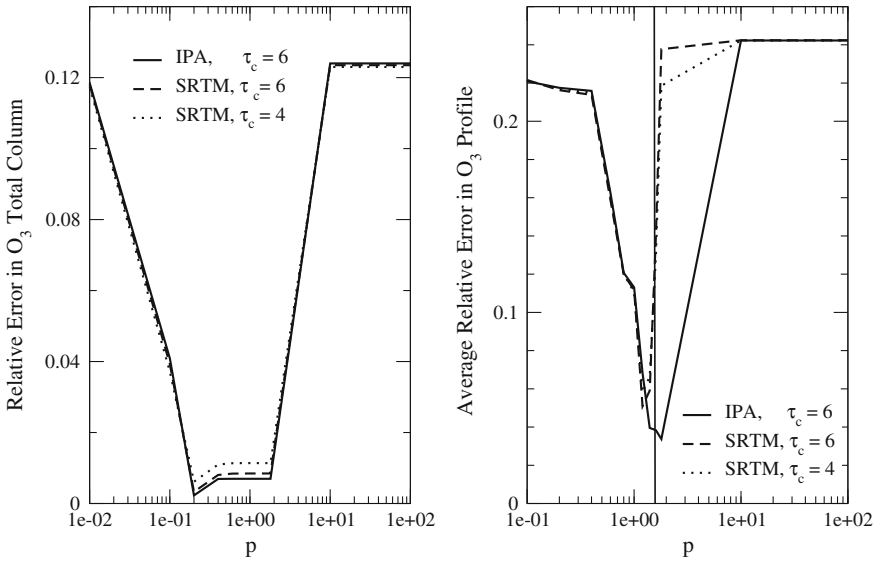

Fig. 3.17 Relative errors in O_3 total columns (left), and O_3 partial columns (right). The synthetic spectra are computed with the independent pixel approximation (IPA) and the stochastic radiative transfer model (SRTM)

2. For the partial columns retrieval, the smallest relative errors are 0.033 in case (1), 0.051 in case (2), and 0.060 in case (3). The optimal value of α corresponds to $p = 1.8$ in case (1), and to $p = 1.2$ in cases (2) and (3). Thus, $\alpha_{\mathtt{opt}}$ increases with the error level ($\alpha_{\mathtt{opt}}$(case (1)) $< \alpha_{\mathtt{opt}}$(case (2))), and it must be precisely calculated.

These results show that the differential radiance model with the independent pixel approximation can only partially eliminate the stochastic cloud errors in trace gas retrievals. The retrieval errors in ozone total column remain small, whereas the relative errors in ozone partial columns are amplified and may become significant. The retrieval of partial columns by using a linearized forward model for stochastic clouds may reduce the relative errors from 0.051 to 0.033, while an accurate knowledge of the cloud optical thickness may reduce the relative errors by 0.01.

The most important conclusion of our numerical analysis is related to the selection of the optimal value of the regularization parameter. In the framework of Tikhonov regularization, the selection of $\alpha_{\mathtt{opt}}$ by an a priori parameter choice method relies on the analysis of the relative errors curves. The question which arises here is which cloud model should be used for generating the synthetic spectra. For total column retrievals, the optimal regularization parameter may vary in a large interval, and does not depend on the choice of the cloud model. For partial columns retrieval, the situation is different. If the independent pixel approximation is used to generate the synthetic spectra, the regularization parameter is not optimal and the relative errors are, for example, of about 0.237 in the case $p = 1.8$, and 0.139 in the case

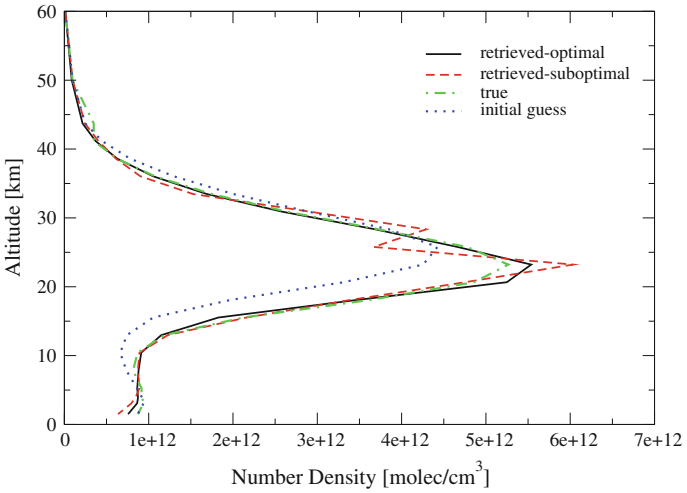

Fig. 3.18 Retrieved profiles corresponding to an optimal ($p = 1.2$) and a suboptimal ($p = 1.6$) value of the regularization parameter. The optimal value is predicted by the stochastic cloud model, while the suboptimal value is predicted by the independent pixel approximation

$p = 1.6$ (the vertical line in Fig. 3.17). If the stochastic cloud model is used to generate the synthetic spectra, the regularization parameter is optimal ($p = 1.2$) and the relative errors are of about 0.051. Essentially, the use of the independent pixel approximation for regularization parameter selection leads to an underestimation of the regularization parameter and so, to an underregularized profile as shown in Fig. 3.18.

3.8 Conclusions

In this chapter, we have reviewed the stochastic radiative transfer models based on the analytical procedure of statistical averaging of the radiative transfer equation. By representing the radiance and the geometric fields as the sum of their mean values and their random fluctuations, we derived an nth-order stochastic model for the solar radiation problem and arbitrary statistics. The stochastic model is expressed in matrix form, and is equipped with appropriate closure relations for the higher-order covariance terms. For broken clouds, the nth-order stochastic model reduces to the first-order stochastic model for a two-dimensional radiance vector, whose entries are the mean radiance field and the covariance of the radiance and the indicator fields. Using the Spherical Harmonics Discrete Ordinate Method (SHDOM) (Evans 1998) as a reference, we found that the stochastic cloud model is superior to the independent pixel approximation; the relative errors in the domain-averaged radiance are below 2% for the stochastic cloud model, and of about 10% for the independent pixel approximation.

The internal mixing models have been introduced by considering zeroth-order stochastic models and by imposing closure relations for the covariance of the fluctuations of the radiance field and the indicator field. The closure relations involve constant and angular-dependent correlations with a correlation coefficient which minimizes the discrepancy between the radiance fields computed by an internal mixing model and a reference model. For this reason, the internal mixing models can be regarded as parameterizations of the reference model. An important aspect of our analysis is the efficient computation of the correlation coefficient for the internal mixing models by taking the first-order stochastic model as reference. This approach is robust and provides the means to compute lookup tables for the correlation coefficient under realistic conditions.

An algorithm for accurate retrieving ozone and clouds parameters has been designed. Because the independent pixel approximation fails to predict accurate radiances, the retrieval algorithm uses a stochastic radiative transfer model. The retrieved parameters are the cloud optical thickness and the cloud-top height. The retrieval accuracy is of about 1%. The use of a lookup table for stochastic clouds radiances in conjunction with multi-linear interpolation is robust and does not introduce performance bottlenecks in the retrieval algorithm. For ozone retrievals we have shown that the stochastic cloud errors can only be partially eliminated by a differential radiance model based on the independent pixel approximation. The errors in the retrieved ozone total column are not significant, but the errors in the retrieved ozone partial columns may become large. The recommendation is to compute the optimal value of the regularization parameter in the framework of Tikhonov regularization by means of a stochastic cloud model. The SRTM and retrieval algorithms presented in this chapter will allow the accurate retrieval of cloud and trace gas properties from the next generation of Copernicus atmospheric composition Sentinel missions.

References

Adams ML, Larsen EW, Pomraning GC (1989) Benchmark results for particle transport in a binary Markov statistical medium. J Quant Spectrosc Radiat Transf 42(4):253–266

Afanas'ev VP, Golovina OY, Gryazev AS, Efremenko DS, Kaplya PS (2015) Photoelectron spectra of finite-thickness layers. J Vac Sci Technol B 33(3):03D101

Afanas'ev VP, Efremenko DS, Kaplya PS (2016) Analytical and numerical methods for computing electron partial intensities in the case of multilayer systems. J Electron Spectrosc Relat Phenom 210:16–29

Afanas'ev VP, Gryazev AS, Efremenko DS, Kaplya PS (2017) Differential inverse inelastic mean free path and differential surface excitation probability retrieval from electron energy loss spectra. Vacuum 136:146–155

Ahmad Z (2004) Spectral properties of backscattered UV radiation in cloudy atmospheres. J Geophys Res 109(D1)

Alexandrov MD, Marshak A, Ackerman AS (2010) Cellular statistical models of broken cloud fields. Part I: theory. J Atmos Sci 67(7):2125–2151

Anisimov O, Fukshansky L (1992) Stochastic radiation in macroheterogeneous random optical media. J Quant Spectrosc Radiat Transf 48(2):169–186

Barker HW (1996) A parametrization for computing grid-averaged solar fluxes for inhomogeneous marine boundary layer clouds: I. Methodology and homogeneous biases. J Atmos Sci 53(16):2289–2303

Barker HW, Davis AB (2005) Approximation methods in atmospheric 3D radiative transfer. Part 2: unresolved variability and climate applications. In: Marshak A, Davis AB (eds) 3d radiative transfer in cloudy atmospheres, vol 6. Springer, Berlin, pp 343–383

Barker HW, Wielicki BA, Parker L (1996) A parametrization for computing grid-averaged solar fluxes for inhomogeneous marine boundary layer clouds: II. Validation using satellite data. J Atmos Sci 53(16):2304–2316

Berg LK, Kassianov EI (2008) Temporal variability of fair-weather cumulus statistics at the ACRF SGP site. J Clim 21(13):3344–3358

Boersma KF (2004) Error analysis for tropospheric NO 2 retrieval from space. J Geophys Res 109(D4)

Budak VP, Korkin SV (2008) On the solution of a vectorial radiative transfer equation in an arbitrary three-dimensional turbid medium with anisotropic scattering. J Quant Spectrosc Radiat Transfr 109(2):220–234

Budak VP, Klyuykov DA, Korkin SV (2010) Convergence acceleration of radiative transfer equation solution at strongly anisotropic scattering. In: Kokhanovsky AA (ed) Light scattering reviews, vol 5. Springer, Berlin, pp 147–203

Budak VP, Kaloshin GA, Shagalov OV, Zheltov VS (2015) Numerical modeling of the radiative transfer in a turbid medium using the synthetic iteration. Opt Express 23(15):A829

Byrne RN, Somerville RCJ, Subasilar B (1996) Broken-cloud enhancement of solar radiation absorption. J Atmos Sci 53(6):878–886

Cahalan RF, Ridgway W, Wiscombe WJ, Bell TL, Snider JB (1994a) The albedo of fractal stratocumulus clouds. J Atmos Sci 51(16):2434–2455

Cahalan RF, Ridgway W, Wiscombe WJ, Gollmer S, Harshvardhan S, Gollmer S (1994b) Independent pixel and Monte Carlo estimates of stratocumulus albedo. J Atmos Sci 51(51):3776–3790

Cairns B, Lacis AA, Carlson BE (2000) Absorption within inhomogeneous clouds and its parameterization in general circulation models. J Atmos Sci 57:700–714

Davis A, Gabriel PM, Lovejoy SM, Schertzer D, Austin GL (1990) Discrete angle radiative transfer III: numerical results and meteorological applications. J Geophys Res: Atmos 95(D8):11729–11742

Davis AB, Marshak A (2010) Solar radiation transport in the cloudy atmosphere: a 3D perspective on observations and climate impacts. Rep Prog Phys 73(2):026801

Doicu A, Trautmann T (2009a) Discrete-ordinate method with matrix exponential for a pseudo-spherical atmosphere: scalar case. J Quant Spectrosc Radiat Transf 110(1–2):146–158

Doicu A, Trautmann T (2009b) Discrete-ordinate method with matrix exponential for a pseudo-spherical atmosphere: vector case. J Quant Spectrosc Radiat Transf 110(1–2):159–172

Doicu A, Trautmann T, Schreier F (2010) Numerical regularization for atmospheric inverse problems. Springer, Berlin

Doicu A, Efremenko DS, Loyola D, Trautmann T (2014a) Approximate models for broken clouds in stochastic radiative transfer theory. J Quant Spectrosc Radiat Transf 145:74–87

Doicu A, Efremenko DS, Loyola D, Trautmann T (2014b) Discrete ordinate method with matrix exponential for stochastic radiative transfer in broken clouds. J Quant Spectrosc Radiat Transf 138:1–16

Efremenko DS, Loyola D, Spurr RJD, Doicu A (2014) Acceleration of radiative transfer model calculations for the retrieval of trace gases under cloudy conditions. J Quant Spectrosc Radiat Transf 135:58–65

Efremenko DS, Schüssler O, Doicu A, Loyola D (2016) A stochastic cloud model for cloud and ozone retrievals from UV measurements. J Quant Spectrosc Radiat Transf 184:167–179

Evans KF (1998) The spherical harmonic discrete ordinate method for three-dimensional atmospheric radiative transfer. J Atmos Sci 55(3):429–446

Farman JC, Gardiner BG, Shanklin JD (1985) Large losses of total ozone in Antarctica reveal seasonal ClOx/NOx interaction. Nature 315(6016):207–210

Gabriel PM, Evans KF (1996) Simple radiative transfer methods for calculating domain-averaged solar fluxes in inhomogeneous clouds. J Atmos Sci 53(6):858–877

Gabriel PM, Lovejoy SM, Davis A, Schertzer D, Austin GL (1990) Discrete angle radiative transfer II: renormalization approach for homogeneous and fractal clouds. J Geophys Res: Atmos 95(D8):11717–11728

Hu Y-X, Wielicki B, Lin B, Gibson G, Tsay S-C, Stamnes K, Wong T (2000) δ-fit: a fast and accurate treatment of particle scattering phase functions with weighted singular-value decomposition least-squares fitting. J Quant Spectrosc Radiat Transf 65(4):681–690

Ingmann P, Veihelmann B, Langen J, Lamarre D, Stark H, Bazalgette Courreges-Lacoste G (2012) Requirements for the GMES atmosphere service and ESA's implementation concept: Sentinels-4/-5 and -5p. Remote Sens Environ 120:58–69

Kargin BA (2000) Statistical modeling of stochastic problems of the atmosphere and ocean optics. In: Matvienko GG, Panchenko MV (eds) Seventh international symposium on atmospheric and ocean optics. SPIE-International society for optical engineering

Kassianov E (2003) Stochastic radiative transfer in multilayer broken clouds. Part I: Markovian approach. J Quant Spectrosc Radiat Transf 77(4):373–393

Kassianov E, Lane-Veron DE, Berg LK, Ovchinnikov M, Kollias P (2012) Markovian approach and its applications in a cloudy atmosphere, vol 7. Light scattering reviews. Springer Science + Business Media, New York, pp 69–107

Kokhanovsky AA (2003) The influence of horizontal inhomogeneity on radiative characteristics of clouds: an asymptotic case study. IEEE Trans Geosci Remote Sens 41(4):817–825

Lane DE, Goris K, Somerville RCJ (2002) Radiative transfer through broken clouds: observations and model validation. J Clim 15(20):2921–2933

Levermore CD, Pomraning GC, Sanzo DL, Wong J (1986) Linear transport theory in a random medium. J Math Phys 27(10):2526–2536

Levermore CD, Wong J, Pomraning GC (1988) Renewal theory for transport processes in binary statistical mixtures. J Math Phys 29:995–1004

Loyola D, Thomas W, Livschitz Y, Ruppert T, Albert P, Hollmann R (2007) Cloud properties derived from GOME/ERS-2 backscatter data for trace gas retrieval. IEEE Trans Geosci Remote Sens 45(9):2747–2758

Lubenchenko AV, Batrakov AA, Pavolotsky AB, Lubenchenko OI, Ivanov DA (2018) XPS study of multilayer multicomponent films. Appl Surf Sci 427(Part A):711–721

Malvagi F, Byrne N, Pomraning GC, Somerville RCJ (1993) Stochastic radiative transfer in a partially cloudy atmosphere. J Atmos Sci 50(14):2146–2158

Marchuk GI, Mikhailov GA, Nazaraliev MA, Darbinjan RA, Kargin BA, Elepov BS (1980) The Monte Carlo methods in atmospheric optics, vol 12. Springer series in optical sciences. Springer, Berlin

McLinden CA, McConnell JC, Griffioen E, McElroy CT (2002) A vector radiative-transfer model for the Odin/OSIRIS project. Can J Phys 80:375–393

Plank VG (1969) The size distribution of cumulus clouds in representative Florida populations. J Appl Meteorol 8(1):46–67

Platt U, Stutz J (2008) Differential optical absorption spectroscopy: principles and applications. Springer, Berlin

Pomraning GC (1989) Statistics, renewal theory and particle transport. J Quant Spectrosc Radiat Transf 42(4):279–293

Pomraning GC (1991) A model for interface intensities in stochastic particle transport. J Quant Spectrosc Radiat Transf 46:221–236

Prigarin SM, Kargin BA, Oppel UG (1998) Random fields of broken clouds and their associated direct solar radiation, scattered transmission and albedo. Pure Appl Opt 7(6):1389–1402

Roozendael VM, Loyola D, Spurr R, Balis D, Lambert J-C, Livschitz Y, Valks P, Ruppert T, Kenter P, Fayt C, Zehner C (2006) Ten years of GOME/ERS-2 total ozone data: the new GOME data processor (GDP) version 4: I. Algorithm description. J Geophys Res: Atmos 111(D14311(1–21))

Sahini DC (1989a) An application of reactor noise techniques to neutron transport problems in a random medium. Ann Nucl Energy 16:397–408

Sahini DC (1989b) Equivalence of generic equation method and the phenomeno-logical model for linear transport problem in a two-state random scattering medium. J Math Phys 30:1554–1559

Seah MP, Gilmore IS, Spencer SJ (2000) Background subtraction: II. General behaviour of REELS and the Tougaard universal cross section in the removal of backgrounds in AES and XPS. Surf Sci 461(1–3):1–15

Smirnov BM (1975) Ekologicheskie problemy atmosfery zemli. Uspekhi Fizicheskikh Nauk 117(10):313–332 (in Russian)

Stephens GL (1988) Radiative transfer through arbitrarily shaped optical media. Part II: group theory and simple closures. J Atmos Sci 45(12):1818–1848

Szczap F, Isaka H, Saute M, Guillemet B, Iotukhovski A (2000a) Effective radiative properties of bounded cascade absorbing clouds: definition of an effective single-scattering albedo. J Geophys Res: Atmos 105(D16):20635–20648

Szczap F, Isaka H, Saute M, Guillemet B, Iotukhovski A (2000b) Effective radiative properties of bounded cascade absorbing clouds: definition of the equivalent homogeneous cloud approximation. J Geophys Res: Atmos 105(D16):20617–20633

Titov GA (1990) Statistical description of radiative transfer in clouds. J Atmos Sci 47(1):24–38

Titov GA, Zhuravleva TB, Zuev VE (1997) Mean radiation fluxes in the near-IR spectral range: algorithms for calculation. J Geophys Res: Atmos 102(D2):1819–1832

Vainikko GM (1973a) Correlation of direct solar radiance in the broken cloudiness. Statistical investigations of broken cloudiness. Meteorological investigations. Trudy MGK SSSR, pp 65–74 (in Russian)

Vainikko GM (1973b) The equations for mean radiance in the broken clouds. Statistical investigations of broken cloudiness. Meteorological investigations. Trudy MGK SSSR, pp 28–37 (in Russian)

Vanderhaegen D (1986) Radiative transfer in statistically heterogeneous mixtures. J Quant Spectrosc Radiat Transf 36(6):557–561

Wiscombe WJ (1977) The delta-M method: rapid yet accurate radiative flux calculations for strongly asymmetric phase functions. J Atmos Sci 34(9):1408–1422

Zhuravleva TB (2008) Simulation of solar radiative transfer under different atmospheric conditions. Part I. The deterministic atmosphere. Atmos Ocean Opt 21(2):81–95

Zuev VE, Titov GA (1995) Radiative transfer in cloud fields with random geometry. J Atmos Sci 52(2):176–190

Zuev VE, Titov GA (1996) Atmosphere optics and climate. Institute of Atmospheric Optics, Tomsk (in Russian)

Chapter 4
Neural Networks and Support Vector Machines and Their Application to Aerosol and Cloud Remote Sensing: A Review

Antonio Di Noia and Otto P. Hasekamp

4.1 Introduction

According to the Intergovernmental Panel on Climate Change (IPCC) fifth Assessment Report (Boucher et al. 2013), aerosols and clouds are among the least known components of the climate system. Satellite remote sensing can be of great help in improving our knowledge of aerosol and cloud properties at global scale. In the last decades, several types of satellite instruments have been launched that provide useful information on aerosols and clouds. Multispectral imagers such as the Moderate Resolution Imaging Spectroradiometer (MODIS) are a standard source of information regarding global cloud cover and aerosol optical thickness. Passive multiangle instruments such as the Multi-angle Imaging Spectroradiometer (MISR) onboard the NASA satellite Terra, the Polarization and Directionality of the Earth's Reflectances (POLDER) onboard the satellites ADEOS and PARASOL, and in the future the Multi-channel Multi-viewing and Multi-polarization Imager (3MI) onboard the ESA Sentinel-5 satellite, provide useful information for the determination of the microphysical properties of aerosol and cloud particles. Active instruments such as the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) provide information on the vertical variability of aerosols and clouds.

The retrieval of aerosol and cloud microphysical properties from satellite measurements is a complex task, that often requires running computationally expensive forward models. This, in some cases, makes it difficult to process the enormous amount of data provided by the most recent satellite instruments in order to provide daily global estimates of aerosol and cloud properties. A well known approach to

A. Di Noia (✉) · O. P. Hasekamp
SRON Netherlands Institute for Space Research, Sorbonnelaan 2,
3584 CA Utrecht, The Netherlands
e-mail: a.di.noia@sron.nl

O. P. Hasekamp
e-mail: o.p.hasekamp@sron.nl

A. Kokhanovsky (ed.), *Springer Series in Light Scattering*, Springer Series
in Light Scattering, https://doi.org/10.1007/978-3-319-70796-9_4

reduce the computational complexity of retrieval algorithms is represented by look-up tables (LUTs). The use of LUTs allows to bypass the need for on-line radiative transfer computations by storing the results of such computations for a number of combinations of the parameters to be retrieved, after which a retrieval is performed by choosing the set of retrieved parameters corresponding to a stored simulation that is "closest" to a satellite measurement according to some predefined metric. An alternative to the LUT-based approach is represented by machine learning algorithms, such as artificial neural networks and support vector machines. Machine learning algorithms perform tasks of classification, function approximation and regression, and therefore can be used in the processing chain of a remote sensing product whenever one of these tasks must be performed. The main advantage of machine learning algorithms is their speed. While they require a preliminary training phase that can be time-consuming, their application once the training phase is complete is essentially instantaneous. For this reason, their use can be attractive in situations where a large amount of data must be processed in a limited amount of time, such as, for example, the generation of a daily, global remote sensing product. In the context of aerosol and cloud remote sensing, tasks that can be carried out using machine learning methods are:

- approximation of forward model calculations,
- detection and classification of clouds and aerosols in satellite imagery,
- retrieval of aerosol and cloud properties through statistical regressions from radiometric measurements.

As will be clear later, maching learning algorithms can not only be used as standalone algorithms, alternative to those based on LUTs or on full radiative transfer computations, but can be also used in combination with these. For example, if a machine learning method is used to approximate a forward model, this can be then inverted using one of the traditional iterative retrieval schemes. If, instead, a machine learning algorithm is used as an inverse method itself, its result can be further refined by using it as a first guess in an iterative scheme based on radiative transfer calculations.

The goal of this paper is to describe the theoretical justification behind the use of machine learning algorithms in the aforementioned tasks and to present a selection of situations in which these have been applied to cloud and aerosol remote sensing. This paper is structured as follows. In Sect. 4.2 we present the problem of extracting information from remote sensing data as a statistical estimation problem. Since, in the following, we will both discuss classification problems and retrieval problems, we tried to treat these in a unified way, in order to emphasize that, while a classification problem seems completely different from a retrieval problem at a first glance, the ultimate goal is the same, i.e. to associate to a set of satellite observations a "decision" about a set of parameters describing the atmospheric state. In Sect. 4.3 we give a theoretical description of artificial neural networks, making a distinction between supervised and unsupervised neural networks. Furthermore, we discuss two important properties of supervised neural networks: (1) their capability of approximating any continuous function to an arbitrary accuracy, and (2) their capability

of approximating the posterior expectation of their output variable given their input variables. Furthermore, we discuss the concept of generalization, i.e. the conditions under which neural networks can be used on observations outside the training set. In Sect. 4.4 we discuss the concept of support vector machine, which is another machine learning method based on a different learning paradigm with respect to neural networks. In Sect. 4.5 we review a broad set of literature papers describing applications of neural networks and support vector machines to various aerosol and cloud remote sensing problems, such as approximation of radiative transfer models, cloud detection and classification and retrieval of aerosol and cloud properties from remote sensing data.

4.2 Extraction of Information from Remote Sensing Data

The goal of a remote sensing system is to use a set of measurements to make inferences about the observed system. Such inferences may take the form of quantitative estimates of one or multiple physical quantities (e.g., the temperature of the ocean surface, the total amount of ozone that is present in an atmospheric region) or they may consist of "qualitative" statements (e.g., whether a certain pixel is free from clouds or not, whether it is covered by vegetation, snow, bare soil etc.). In the latter case it is customary to speak about a "classification" problem, whereas in the former case the term "inversion" or "retrieval" is usually adopted in the remote sensing community. Although these two classes of problems may seem different, they can be treated in a unified way as statistical decision problems. A remote sensing instrument measures a set of quantities $\mathbf{m} \in \mathbb{R}^m$, that are related to an unknown "state" $\mathbf{s} \in \mathbb{R}^n$ through some functional relationship $\mathbf{F}$. In the case of retrieval problems $\mathbf{s}$ can take continuous values, whereas in classification problems it can only take values in a finite, countable subset of $\mathbb{R}^n$. The goal of a remote sensing system is to find an estimate for $\mathbf{s}$.

The problem of estimating $\mathbf{s}$ from $\mathbf{m}$ is ill-posed if $\mathbf{m}$ does not contain sufficient information to determine $\mathbf{s}$ with useful accuracy. This may happen because $\mathbf{F}$ may be poorly invertible and because the presence of measurement errors may make it impossible to find an exact solution for the problem even if $\mathbf{F}$ is invertible. Nevertheless, several methods exist to build approximate solutions. In formal terms, in order to construct an estimator of $\mathbf{s}$ from $\mathbf{m}$, a decision rule d that maps the measurement vector $\mathbf{m}$ into an estimate $\hat{\mathbf{s}}$ of the state vector is sought. Let $\mathcal{D}$ be the space of all the possible decision rules. Methods for choosing $d \in \mathcal{D}$ can be developed once $\mathcal{D}$ is equipped with an order relationship, with respect to which two decision functions can be compared (Robert 2007). The order relationship essentially specifies a "quality criterion" that allows us to decide which estimate $\mathbf{d} = d(\mathbf{m})$ is the "best" estimate of $\mathbf{s}$, thereby providing us with some "objectivity" in the choice of the estimate. The choice of the estimate $\hat{\mathbf{s}}$ takes thus the form of an optimization problem. It is important to emphasize, though, that the objectivity in the chosen estimate only holds for a given quality criterion, but the choice of the quality criterion itself is

always subjective. With respect to this latter point several possibilities exist, and in this section we will review some of them.

4.2.1 Constrained Least Squares Methods

Let us suppose that we know the relationship $\mathbf{F}$ that relates $\mathbf{m}$ to $\mathbf{s}$. Under the hypothesis of additive noise we can write:

$$\mathbf{m} = \mathbf{F}(\mathbf{s}) + \boldsymbol{\nu} \tag{4.1}$$

where $\boldsymbol{\nu}$ is a random variable describing the measurement noise. In the following, we will assume that the noise can be described as a Gaussian random variable with zero mean and covariance matrix $\mathbf{S}_\nu$. Constrained least-squares methods search for an estimate of the state $\mathbf{s}$ by minimizing a cost function of the following form

$$C(\mathbf{s}) = \|\mathbf{m} - \mathbf{F}(\mathbf{s})\|_\nu^2 + \sum_{i=1}^{N} \gamma_i c_i(\mathbf{s}) \tag{4.2}$$

where $\|\mathbf{m} - F(\mathbf{s})\|_\nu^2 = [\mathbf{m} - \mathbf{F}(\mathbf{s})]\mathbf{S}_\nu^{-1}[\mathbf{m} - \mathbf{F}(\mathbf{s})]$, the $c_i(\mathbf{m})$ are a number of constraints on $\mathbf{s}$, that depend on the problem at hand, and the γ_i are real numbers that specify the strengths of the constraints. Thus, the cost function written in Eq. (4.2) is the sum of two terms. The first term assigns a penalty to solutions that do not fit the measurement vector $\mathbf{m}$ within the measurement accuracy, whereas the constraints that compose the second term assign a penalty to solutions that violate certain, problem-dependent, "regularity" conditions. For example, if the variable $\mathbf{s}$ is an atmospheric profile, we may require that the profile be "smooth" with respect to a vertical coordinate. In this case a penalty term on $\mathbf{s}$ will be given by the value of its first derivative with respect to that vertical coordinate. Same applies, for example, if $\mathbf{s}$ is a spectrally dependent quantity (e.g., surface albedo, aerosol refractive index) and we want to ensure that its estimate is smooth with respect to wavelength. In other cases, we may want to penalize solutions that differ from a prior value $\mathbf{s}_a$, in which case the constraint may take the form $\|\mathbf{s} - \mathbf{s}_a\|^2$, and so on. The approach we just described is known as Phillips–Tikhonov regularization (Phillips 1962; Tikhonov and Arsenin 1977). As discussed in Tikhonov and Arsenin (1977) and – in a context more specifical to remote sensing problems – in Dubovik and King (2000), the regularization parameters γ_i are trade-off parameters between the relative weights given to the two terms in the cost function (4.2). Examples of the application of the Phillips–Tikhonov approach to satellite aerosol retrievals are shown in Dubovik et al. (2011) and Hasekamp et al. (2011). From the above discussion, it is probably clear that the Phillips–Tikhonov approach is more relevant in the context of retrieval problems than it is for classification problems, as it is inherently targeted at selecting suitable solutions among a continuum of possible solutions of an estimation problem.

Such an issue is not present in classification problems, where the set of possible solutions is, in itself, discrete and finite. In any case, the unconstrained least-squares approach can be applied, in principle, also to classification problems, by creating a look-up table (LUT) of values of $\mathbf{F}(\mathbf{s})$ associated to each possible value of $\mathbf{s}$ and by choosing as estimate of $\mathbf{s}$ the value of $\mathbf{s}$ for which the first term in (4.2) is minimum.

4.2.2 Statistical Methods

The existence of random errors in the measurement warrants that the relationship between $\mathbf{m}$ and $\mathbf{s}$ is not completely deterministic, and therefore it is legitimate to express such relationship in terms of a probability density $f_{M|S}(\mathbf{m}|\mathbf{s})$, that is, the conditional probability density of the measurement vector $\mathbf{m}$ for a given value of the state vector $\mathbf{s}$. This fact is exploited in the statistical approaches to the estimation of the state vector. These approaches can be broadly divided in two classes:

1. Frequentist approaches
2. Bayesian approaches

The main difference between these two classes of statistical methods arises from the way they interpret the state vector $\mathbf{s}$. In the frequentist view, $\mathbf{s}$ is an unknown, deterministic parameter, and the only source of randomness in the estimation problem is due to the measurement process. In the Bayesian view, instead the state vector $\mathbf{s}$ is not just seen as an unknown deterministic quantity, but is seen as a random variable in itself, with its own probability density function, called the *a priori* probability density $f_S(\mathbf{s})$, which quantifies our prior beliefs about $\mathbf{s}$ before performing the measurement $\mathbf{m}$, and an *a posteriori* probability density $f_{S|M}(\mathbf{s}|\mathbf{m})$, which represents our uncertainty about $\mathbf{s}$ after the measurement has been performed. The a posteriori probability density is related to the a priori density and to $f_{M|S}(\mathbf{m}|\mathbf{s})$ through the Bayes rule:

$$f_{S|M}(\mathbf{s}|\mathbf{m}) = \frac{f_{M|S}(\mathbf{m}|\mathbf{s}) f_S(\mathbf{s})}{f_M(\mathbf{m})} \tag{4.3}$$

where $f_M(\mathbf{m})$ is the marginal probability density of the measurement vector, i.e. the joint probability $f_{MS}(\mathbf{m}, \mathbf{s})$ integrated over all the possible values of $\mathbf{s}$.

One of the most famous frequentist approaches to the estimation of $\mathbf{s}$ from the measurement $\mathbf{m}$ is the maximum likelihood (ML) approach, described in Fisher (1922) and other works by the same author. In this approach, the optimal estimate of $\mathbf{s}$ is the estimate that maximizes the likelihood function

$$L(\mathbf{s}|\mathbf{m}) = f_{M|S}(\mathbf{m}|\mathbf{s}) \tag{4.4}$$

The meaning of maximizing this function is, according to Fisher (1922) and other works by the same author, that of estimating the state that is most likely to have caused a given set of observations. If the observations that constitute $\mathbf{m}$ are statistically

independent and identically distributed (i.e. they follow the same probability distribution with the same parameters), then ML estimates possess some important properties:

1. *Consistency*. The ML estimator $\hat{\mathbf{s}}_m$ built out of m independent and identically distributed observations, converges in probability to the true value of $\mathbf{s}$ for $m \to \infty$, where m is the dimension of the measurement vector.
2. *Asymptotic normality and unbiasedness*. The probability distribution of $\hat{\mathbf{s}}_m$ converges for $m \to \infty$ to a Gaussian distribution with mean value equal to the true value of $\mathbf{s}$ and covariance matrix equal to the inverse of the Fisher information matrix $\mathcal{I}$, defined by

$$\mathcal{I}_{ij}(\mathbf{s}) = \mathrm{cov}\left[\frac{\partial f_{M|S}(\mathbf{m}|\mathbf{s})}{\partial s_i}\frac{\partial f_{M|S}(\mathbf{m}|\mathbf{s})}{\partial s_j}\right] \tag{4.5}$$

3. *Efficiency*. For $m \to \infty$, the variance of a ML estimator is equal to the Cramér-Rao lower bound (CRLB), that is, the minimum possible variance for a statistical estimator. The CRLB for the variance of an estimator is equal to the reciprocal of the Fisher information for that estimator.

It is easy to see that, in the assumption of Gaussian measurement noise, $f_{M|S}(\mathbf{m}|\mathbf{s})$ is a Gaussian density with expectation equal to $\mathbf{F}(\mathbf{s})$ and covariance matrix equal to a $\mathbf{S}_v$, and that the application of the ML principle leads to the unconstrained least-squares solution.

It must be recognized that, in satellite remote sensing using passive instruments, a measurement vector made of a large number of independent, identically distributed measurements, is almost never available. A typical measurement vector $\mathbf{m}$ is usually made of a single observation of different quantities (e.g. radiances at several wavelengths or viewing angles) which are usually statistically dependent, rather than of multiple observations of the same quantity (or set of quantities) under identical measurement conditions. As a consequence, the maximum likelihood approach is usually not adopted, at least not in its "asymptotic regime".

As said above, the Bayesian approach treats the state vector itself as a random variable. Finding an optimal estimate in the Bayesian framework involves defining a *loss function* $L(\mathbf{s}, \mathbf{d})$, that quantifies the penalty we associate with estimating the true state $\mathbf{s}$ with a generic guess $\mathbf{d}$, and an *a posteriori* conditional expected loss as

$$C_L(\mathbf{d}|\mathbf{m}) = \mathrm{E}_{\mathbf{s}}\left[L(\mathbf{s}, d(\mathbf{m}))|\mathbf{m}\right] \tag{4.6}$$

where the operator $\mathrm{E}_{\mathbf{s}}$ is the expected value over $\mathbf{s}$ conditioned to the observed $\mathbf{m}$. Once L and C_L are defined, the optimal decision function d is the function that minimizes C_L given $\mathbf{m}$. In general, different choices for L give rise to different decision rules. We will now show the consequences deriving from the two simplest choices for the loss function: the 0/1 loss function and the quadratic error loss function.

- 0/1 loss function.

The 0/1 loss function is defined differently depending on whether $\mathbf{s}$ has discrete or continuous values. In the discrete case (classification), the 0/1 loss function is defined as

$$L(\mathbf{s}, \mathbf{d}) = \begin{cases} 0 & \text{if } \mathbf{d} = \mathbf{s} \\ 1 & \text{if } \mathbf{d} \neq \mathbf{s} \end{cases} \tag{4.7}$$

That is, no penalty is associated to a correct decision and a unitary penalty is associated to an incorrect decision. Let us now suppose that $\mathbf{s}$ can take values in a set of n possibilities, that we enumerate as $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$, and let $P(\mathbf{s}_i|\mathbf{m})$ the *a posteriori* probability that the correct decision is $\mathbf{s}_i$, $i = 1, \ldots, n$. The conditional expected loss associated with the decision $\mathbf{d}$ is then

$$C_L(\mathbf{d}|\mathbf{m}) = \sum_{i=1}^{n} L(\mathbf{s}_i, \mathbf{d}) P(\mathbf{s}_i|\mathbf{m}) \tag{4.8}$$

Given that L is 1 if $\mathbf{d}$ is incorrect and 0 otherwise, it is easy to see that

$$C_L(\mathbf{d}|\mathbf{m}) = \sum_{i=1}^{n} P(\mathbf{s}_i|\mathbf{m}) - P(\mathbf{d}|\mathbf{m}) = 1 - P(\mathbf{d}|\mathbf{m}) \tag{4.9}$$

Thus, the problem of minimizing $C_L(\mathbf{d}|\mathbf{m})$ with respect to $\mathbf{d}$ is equivalent to that of maximizing $P(\mathbf{d}|\mathbf{m})$. In other words, the choice of the 0/1 loss function leads to the maximum a posteriori (MAP) classifier as the optimum classifier.

The same reasoning can be extended to a retrieval problem, for which $\mathbf{s}$ is a continuous variable. For this case, the 0/1 loss function is defined as

$$L(\mathbf{s}, \mathbf{d}) = \begin{cases} 0 & \text{if } \|\mathbf{d} - \mathbf{s}\| < \varepsilon \\ 1 & \text{if } \|\mathbf{d} - \mathbf{s}\| \geq \varepsilon \end{cases} \tag{4.10}$$

That is, no penalty is associated to a solution that differs from the true one for less than a threshold ε and a unitary penalty is associated to a solution that exceeds such threshold. Let $\mathcal{S} \subseteq \mathbb{R}^n$ be the domain of variation of $\mathbf{s}$, and let $f_{S|M}(\mathbf{s}|\mathbf{m})$ be the a posteriori probability density of the state $\mathbf{s}$ given the measurements $\mathbf{m}$. Now the expected loss C_L becomes an integral:

$$C_L(\mathbf{d}|\mathbf{m}) = \int_{\mathcal{S}} L(\mathbf{s}, \mathbf{d}|\mathbf{m}) f_{S|M}(\mathbf{s}|\mathbf{m}) \mathrm{d}\mathbf{s} \tag{4.11}$$

Let us now denote by $\mathcal{B}_\varepsilon(\mathbf{d})$ the hypersphere with radius ε centred in $\mathbf{d}$. Given that L is 0 in $\mathcal{B}_\varepsilon(\mathbf{d})$ and 1 elsewhere, and that $f_{S|M}$ is a probability density, we can

write

$$C_L(\mathbf{d}|\mathbf{m}) = \int_{\mathcal{S}} f_{S|M}(\mathbf{s}|\mathbf{m})\mathrm{d}\mathbf{s} - \int_{\mathcal{B}_\varepsilon(\mathbf{d})} f_{S|M}(\mathbf{s}|\mathbf{m})\mathrm{d}\mathbf{s} = 1 - \int_{\mathcal{B}_\varepsilon(\mathbf{d})} f_{S|M}(\mathbf{s}|\mathbf{m})\mathrm{d}\mathbf{s} \tag{4.12}$$

Therefore, the decision rule that minimizes C_L is

$$\hat{\mathbf{s}} = \arg\max_{\mathbf{d}} \int_{\mathcal{B}_\varepsilon(\mathbf{d})} f_{S|M}(\mathbf{s}|\mathbf{m})\mathrm{d}\mathbf{s} \tag{4.13}$$

It is worth noting that

$$\lim_{\varepsilon\to 0} \int_{\mathcal{B}_\varepsilon(\mathbf{d})} f_{S|M}(\mathbf{s}|\mathbf{m})\mathrm{d}\mathbf{s} = f_{S|M}(\mathbf{d}|\mathbf{m}) \tag{4.14}$$

Therefore, in the limit for $\varepsilon \to 0$ the maximization problem yields once again the usual MAP retrieval (Rodgers 2000) as the optimal one. In literature the MAP retrieval is often derived by directly maximizing the a posteriori probability density of the state vector. It should be kept in mind, though, that the point value of the probability density for a continuous random variable does not represent a probability, and care must be exercised when saying that its maximum represents the "most probable value" of the random variable at hand, as this carries the risk of overinterpreting the concept of optimality of a MAP estimate. The procedure we have just presented maximizes the probability that the true state vector lies in a hypersphere of radius ε around the estimate, and obtains the MAP estimate in the limit for $\varepsilon \to 0$. In other words, in the MAP method the "mode" of the a posteriori density is taken as estimate for $\mathbf{s}$. In the case of a multimodal density, this procedure leads to selecting the highest mode.

- Squared error loss function.

The squared error loss function is defined, for both the discrete and the continuous case, as

$$L(\mathbf{s}, \mathbf{d}) = k\|\mathbf{s} - \mathbf{d}\|^2 \tag{4.15}$$

with $k > 0$. In this case, the conditional expected loss is

$$C_L(\mathbf{d}|\mathbf{m}) = \mathrm{E}_S\left[k\|\mathbf{s} - \mathbf{d}\|^2 \middle| \mathbf{m}\right] \tag{4.16}$$

By using the linearity of the expectation operator, we have that

$$C_L(\mathbf{d}|\mathbf{m}) = k\mathrm{E}_S\left[\|\mathbf{s}\|^2|\mathbf{m}\right] + k\|\hat{\mathbf{s}}\|^2 - k\mathbf{d}^T\mathrm{E}_S\left[\mathbf{s}|\mathbf{m}\right] \tag{4.17}$$

This expression is minimized if $\hat{\mathbf{s}} = \mathrm{E}_S\left[\mathbf{s}|\mathbf{m}\right]$ is chosen as estimate of $\mathbf{s}$. Therefore, minimizing the expected loss in the case of the squared error loss function leads to the choice of the posterior mean (PM) as estimate of $\mathbf{s}$. It can be shown that, if

the a posteriori distribution is Gaussian, the PM estimate is the same as the MAP estimate.

Once the problem of estimating the state vector **s** from the measurements **m** has been set up according, e.g. to one of the methods described above, the forward model **F** is used to compute explicitly the cost function that has to be minimized with respect to **s**. Methods for carrying out the minimization are described in Rodgers (2000).

4.3 Artificial Neural Networks

Scientific research on artificial neural networks – the beginning of which approximately dates back to the 1940s – has been originally motivated by an attempt to develop machines capable of imitating the computational structure of the brain of vertebrates. Such attempts were based on the observation of the fact that the brain can be seen as a computer based on an alternative architecture with respect to the traditional von Neumann paradigm, which forms the basis of most of the modern computers (Hertz et al. 1991). While a von Neumann machine is essentially a serial machine, capable of executing a sequence of elementary instructions provided by the user, the brain is a massively parallel processor, composed by a large number of elementary units (*neurons*), which exchange information through connection channels (*synapses*). The brain is thus a *network* of neurons. The interest in the computational structure of the brain was motivated by its ability to perform certain complex operations (e.g. object recognition, motor control) at speeds that – at the time – exceeded enormously those that could be achieved using a traditional computer (Haykin 1999).

The first mathematical model of an artificial neuron derives from a work by McCulloch and Pitts (1943). Their original model was only addressing Boolean computations. In particular, the McCulloch-Pitts neuron consists of n excitatory inputs and m inhibitory inputs. The output of the neuron is 1 if no inhibitory inputs are 1 and at least one excitatory input is 1, and zero otherwise. The McCulloch-Pitts neuron owes its importance to the fact that any Boolean function can be realized through combinations of such neurons. However it must be noted that these neurons have no free parameters. An extension of the McCulloch-Pitts model was proposed by Rosenblatt (1958), who introduced the following differences: (i) no inhibitory inputs are used, (ii) the excitatory inputs are multiplied by real-valued weights, (iii) the weights can be updated by means of a learning rule. Rosenblatt's neuron model, called the *perceptron*, is closer to the artificial neuronal model that is used in most applications today, the only difference being that the output of a Rosenblatt's perceptron is still discrete, whereas the models used today also allow for continuous outputs.

The mathematical model of a neuron that is still used in most of today's artificial neural network models consists of three elements:

1. A set of synaptic weights $\{w_i\}$ that are multiplied by input signals $\{x_i\}$ of the neuron.

2. A linear combiner, that computes a sum of the input signals weighted by the respective synaptic connections. The result of this sum is usually called *action potential* of the neuron.
3. An activation function φ, that is applied to the action potential to give the actual output y of the neuron.

Thus, the mathematical equation of a neuron based on this model can be written as

$$y = \varphi\left(\sum_{i=1}^{n} w_i x_i + b\right) \tag{4.18}$$

where b is a *bias* term that is usually included in order not to constrain the action potential to be zero if all the inputs to the neuron are zero. The activation function φ can take several forms. Some of the most common choices are:

- The threshold function: $\varphi(v) = 1$ if $v \geq 0$ and 0 otherwise.
- The logistic function: $\varphi(v) = 1/(1 + \exp(-v))$
- The hyperbolic tangent function: $\varphi(v) = \tanh(v)$.

It may be noted that the bias can be also interpreted as an additional synaptic weight w_0 multiplying an input x_0 that is always clamped at value 1.

It is worthwhile to remark that the mathematical model of Eq. (4.18) is not the only possible model for a neuron. Later in this paper we will mention neural network types, such as Radial Basis Functions (RBF) and Self Organizing Maps (SOM), in which the mathematical processing applied by a neuron is different from Eq. (4.18). In general, however, a neuron applies a quite simple mathematical transformation to the data it receives as inputs, and such transformation contains some free parameters, such as the weights w_i and the bias b in the model of Eq. (4.18). The process of adjusting these parameters according to the inputs coming from the external environment is what we call *training*, and is the essential ingredient through which we induce the neural network to acquire the behaviour we desire. Several classes of training methods exist (Haykin 1999), but we will only discuss the two principal ones:

1. Supervised training
2. Unsupervised training

In supervised training, the neural network is provided with a set of realizations of the input vector $\mathbf{x}$ and of a target vector $\mathbf{t}$, that represents the response we would like to obtain from the neural network when the input $\mathbf{x}$ is applied. The free parameters of the neuron are updated by trying to minimize the misfit between the desired response $\mathbf{t}$ and the actual response $\mathbf{y} = \mathbf{g}(\mathbf{x}, \mathbf{w})$, where $\mathbf{g}$ is the function computed by the neural network and $\mathbf{w}$ is a vector containing all the free parameters of the network.

In unsupervised training, the neural network is not provided with a target vector $\mathbf{t}$, but the goal is to partition the input space into subregions of "similar" input vectors, in such a way that similar input vectors produce similar responses in the neural network, while dissimilar input vector produce dissimilar responses. This is achieved through a so-called "competitive" learning process, an example of which will be illustrated

in Sect. 4.3.4. In this process the neurons compete among each other to achieve the maximum output in response to a given input and, as the training proceeds, the "winning neuron" corresponding to a certain input $\mathbf{x}$ becomes more and more likely to be the winning neuron also for inputs that are close to $\mathbf{x}$ in the signal space. To summarize, what matters in unsupervised networks is not the particular value of the network response to a certain input, but the differences between the responses to different inputs.

4.3.1 Multilayer Perceptron

The multilayer perceptron (Werbos 1974) is probably the most popular type of supervised neural network. Neurons in this network are arranged in layers:

- An input layer, where the input signal is applied.
- One or more hidden layers of neurons obeying Eq. (4.18), with nonlinear activation function φ.
- An output layer of neurons still obeying Eq. (4.18), whose activation function can be either linear or nonlinear. These neurons compute the final output of the neural network (e.g., the vector of retrieved quantities, in one of the applications of our interest).

The layers are connected in a feed-forward way, that means that the output of the ith layer forms the input for the $i+1$th layer and there are no feedback connections, neither from layer to layer nor at neuron level.

Since the neurons in a multilayer perceptron follow the model given in Eq. (4.18), the output of the neuron k in the layer l of a multilayer perceptron can be written as

$$y_k^{(l)} = \varphi_l \left(\sum_{j=0}^{N_{l-1}} w_{jk}^{(l)} y_j^{(l-1)} \right) \tag{4.19}$$

where N_l is the number of neurons in layer l, we have made the (reasonable) assumption that all the neurons in a layer l have the same activation function φ_l, and we have used the convention of including the bias $b_k^{(l)}$ in the weights, so that $w_{0k}^{(l)} = b_k^{(l)}$ and $y_0^{(l-1)} = 1$ for all l. For $l = 1$, N_{l-1} is the dimension of the input vector, and $y_j^{(l-1)}$ is simply the input x_j.

A property that may be of interest is that, as long as the derivatives of φ_l can be computed analytically, the derivatives of the output of a neuron with respect to any of its inputs can be computed analytically as well. For example, for the first derivative we have

$$\frac{\partial y_k^{(l)}}{\partial y_j^{(l-1)}} = w_{jk}^{(l)} \varphi_l' \left(\sum_{j=0}^{N_{l-1}} w_{jk}^{(l)} y_j^{(l-1)} \right) \tag{4.20}$$

The implication of this property is that, if a multilayer perceptron is used as approximate forward model, the approximate Jacobi matrix of the forward model can be obtained through a simple analytical computation, namely by implementing Eq. (4.20). If, instead, the neural network is used as retrieval algorithm, the gain matrix of the retrieval can be computed similarly. It should be noted that the activation functions used in practice usually do have analytically computable derivatives. For example, if $\varphi(v) = \tanh(v)$ then $\varphi'(v) = (1 - \varphi^2(v))$, whereas if $\varphi(v) = (1 + \exp(-v))^{-1}$ then $\varphi'(v) = \varphi(v)(1 - \varphi(v))$.

The most popular training algorithm for the multilayer perceptron is the "error backpropagation". In its simplest version, due to Rumelhart et al. (1986), for each input-target pair $(\mathbf{x}_n, \mathbf{t}_n)$ in the training set, the following error function is defined

$$E_n = \frac{1}{2}\|\mathbf{g}(\mathbf{x}_n, \mathbf{w}) - \mathbf{t}_n\|^2 \tag{4.21}$$

and is iteratively minimized with respect to $\mathbf{w}$ by using a steepest-descent approach. This means that the weight vector is initialized at a random value $\mathbf{w}_0$, and is iteratively updated according to the equation

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \eta \frac{\partial E_n}{\partial \mathbf{w}}\bigg|_{\mathbf{w}=\mathbf{w}_n} \tag{4.22}$$

where η is a user-defined parameter called "learning rate". The process is continued until a user-defined convergence criterion is met. The learning rate must be chosen based on a trade-off between the speed and the stability of the learning process. Lower learning rates ensure higher stability (i.e. a steady decrease of the training error) at the cost of a slower learning, whereas higher learning rates accelerate the learning process but increase the risk of divergence of the learning process. Bös and Amari (1999) show that decreasing the learning rate during the training speeds up the descent of the algorithm towards better minima of the error hypersurface. This technique is called "learning rate annealing".

The error backpropagation algorithm can be also applied in "batch mode", that is, by defining and minimizing a cost function that encompasses all the training cases, such as

$$E = \frac{1}{2N}\sum_{n=1}^{N}\|\mathbf{g}(\mathbf{x}_n, \mathbf{w}) - \mathbf{t}_n\|^2 \tag{4.23}$$

where N is the number of training cases. The update of the weight vector is still carried out using a steepest-descent scheme, as in Eq. (4.22) but replacing E_k with E and by updating the network weights after the presentation of the entire training set and not after each case. The role of the learning rate is similar as with the standard (or online) backpropagation; namely, it serves as a step size for the update of the weight vector. However, the optimal value of the learning rate in a batch learning algorithm does not necessarily decrease while the training proceeds, but depends on the curvature of the error hypersurface around the current weight vector $\mathbf{w}_i$.

Sophisticated speed-up methods, such as Levenberg-Marquardt training (Hagan and Menhaj 1994) and scaled conjugate gradient (Møller 1993), try to infer the optimal step size by estimating second order information over $E(\mathbf{w})$, so that the learning rate is not specified by the user. Other, more empirical, speed-up techniques, keep the concept of learning rate, but make an elaborate decision of whether to increase it or decrease it based on conditions on the change in sign in the derivative of the error surface. Some of these techniques employ different learning rates for different weights. An overview of these methods can be found in Riedmiller (1994).

4.3.1.1 Universal Approximation Theorems

The use of the multilayer perceptron as a tool for function approximation is justified by its *universal approximation* property, that has been demonstrated independently by Cybenko (1989), Funahashi (1989), and Hornik et al. (1989). While the three works mentioned above state the universal approximation theorem in slightly different forms, the essence of the theorem may be described as follows. The family of functions $\mathcal{F}$ generated by Eq. (4.19) with a single hidden hidden layer is dense in the space $C(\mathcal{X})$ of continuous functions over $\mathcal{X}$ – where $\mathcal{X}$ is any compact subset of $\mathbb{R}^n$ – provided that the φ_l of the neurons in the hidden layer are *sigmoidal*, where a sigmoidal function is roughly defined as a function $\varphi(v)$ that is infinitely differentiable, strictly monotonic and has horizontal asymptotes for $v \to \pm\infty$. The validity of the universal approximation theorems, originally stated for single-output functions, can be simply extended to multiple-output functions (Hornik et al. 1989). Hornik et al. (1990) extended the universal approximation property of a multilayer perceptron with sigmoidal activation functions to the derivatives of the function being approximated.

We recall that by saying that $\mathcal{F}$ is dense in $C(\mathcal{X})$ we mean that for any $F \in C(\mathcal{X})$ and for every $\varepsilon > 0$, a $\Phi \in \mathcal{F}$ exists such that the distance between Φ and F is less than ε. In the context of the universal approximation theorems, the distance between two functions in $C(\mathcal{X})$ is defined as $\sup_{\mathbf{x}\in\mathcal{X}} |\Phi(\mathbf{x}) - F(\mathbf{x})|$. It is straightforward to realize that the logistic function and the hyperbolic tangent are examples of sigmoidal functions, which explains why they are commonly used as activation functions for multilayer perceptrons.

The requirement that the activation functions φ_l be sigmoidal has been relaxed in later formulations of the universal approximation theorems. Hornik (1991) showed that $\mathcal{F}$ is dense on $C(\mathcal{X})$ and on $L^p(\mathbb{R}^n)$ (space of functions f such that $|f(\mathbf{x})|^p$ is integrable on $\mathbb{R}^n$) just provided that the φ_l are continuous, bounded and nonconstant. Leshno et al. (1993) proved that the only requirement on the φ_l for the universal approximation property to hold is that they be nonpolynomial, thereby removing the requirement for continuity, the necessity of which had been previously considered by Hornik (1991) as an "open problem".

To summarize, universal approximation theorems state that any continuous function on a compact subset of $\mathbb{R}^n$ can in principle be approximated to an arbitrary

accuracy by a function such as those computed by a multilayer perceptron. Hornik et al. (1989) make this point particularly clear by writing (verbatim):

> Any lack of success in applications must arise from inadequate learning, insufficient numbers of hidden units or the lack of a deterministic relationship between input and target.

The following points are remarkable about the universal approximation theorems:

- The basis for their validity does not derive from neural network theory. They are just a consequence of the mathematical properties of the functions realized by multilayer perceptrons, regardless of the fact that these functions are used to build neural networks. The consequence of this will become clearer later in this paper, when support vector machines are discussed.
- They state that a multilayer perceptron with a single hidden layer can in principle approximate any continuous function on a compact set, but they do not indicate how to build such a function, i.e. they are merely existence theorems. This explains why, in real applications of the multilayer perceptron, more than a hidden layer is often used to achieve good approximation properties, despite the results of these theorems.

4.3.2 Radial Basis Function Networks

Radial basis function (RBF) networks are a supervised neural network architecture that is alternative to multilayer perceptrons. Architecturally, a RBF network differs from that of a multilayer perceptron in the following aspects:

1. Usually there is only one hidden layer.
2. Neurons in the hidden layer have a localized activation function, i.e. a function that is significantly nonzero on a bounded subset of the input space (e.g., a Gaussian).
3. The free parameters of the network are: (i) the weights of the synaptic connection from the hidden to the output layer; (ii) the positions of the centres of the activation functions; and (iii) parameters specifying the extension of the area of the input space where the activation functions are nonzero.

The activation function of a hidden neuron in a RBF network has the form

$$y = \varphi(-\|\mathbf{x} - \mathbf{c}\|^2, \sigma) \tag{4.24}$$

where $\mathbf{x}$ is the vector of input signals, $\mathbf{c}$ is the center associated to the neuron and σ is a parameter that controls the width of the function. Examples of activation functions include the Gaussian function

$$\varphi(-\|\mathbf{x} - \mathbf{c}\|^2, \sigma) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{c}\|^2}{2\sigma^2}\right) \tag{4.25}$$

and the inverse multiquadric function

$$\varphi(-\|\mathbf{x}-\mathbf{c}\|^2,\sigma)=\left(\|\mathbf{x}-\mathbf{c}\|^2+\sigma^2\right)^{-1/2} \tag{4.26}$$

Considering that RBF networks usually have a hidden layer only, the output of the whole network in response to an input signal $\mathbf{x}$ can be written as

$$y_k=\sum_{j=0}^{N_h} w_{jk}\varphi_j(\mathbf{x}) \tag{4.27}$$

In this expression, y_k is the response of the kth output neuron, N_h is the number of hidden neurons, $\varphi_0\equiv 1$ and $\phi_j(\mathbf{x})=\varphi(-\|\mathbf{x}-\mathbf{c}_j\|^2,\sigma_j)$ for $j=1,\ldots,N_h$, $\mathbf{c}_j$ and σ_j being the center and the width parameter of the jth hidden neuron respectively.

Universal approximation theorems for RBF networks, similar to those we have seen for multilayer perceptrons, have been proved by Park and Sandberg (1991, 1993). Training a RBF network involves determining the values of the weights $\{w_{jk}\}$ as well as of the centers $\mathbf{c}_j$ and the widths σ_j. In principle, these parameters can be chosen through the usual least-squares procedure. This, however, often makes the training process computationally demanding. For this reason, alternative training methods exist where first the centers and the widths of the RBFs are chosen either randomly or through some unsupervised clustering technique, and then the weights of the output layer are determined from the training set through a simple pseudo-inverse multiplication (details in Haykin 1999).

4.3.3 *Statistical Interpretation of Supervised Neural Networks*

A fundamental result is derived by Bishop (1995a) from the asymptotic properties of the sum-of-squares error function for large training sets. Let us consider a training set made of N input-target vector pairs $(\mathbf{x}_n,\mathbf{t}_n)$, and the sum-of-squares error function

$$E=\frac{1}{2N}\sum_{n=1}^{N}\|\mathbf{g}(\mathbf{x}_n,\mathbf{w})-\mathbf{t}_n\|^2 \tag{4.28}$$

where $\mathbf{g}$ is the vector-valued function realized by the neural network and $\mathbf{w}$ is the vector of neural network parameters. The main argument made by Bishop (1995a) is that, as the number of training cases increases, the sum written in Eq. (4.28) – that is essentially an arithmetical average – approaches the expectation of $\frac{1}{2}\|\mathbf{g}(\mathbf{x},\mathbf{w})-\mathbf{t}\|^2$. More formally

$$\frac{1}{2N}\sum_{n=1}^{N}\|\mathbf{g}(\mathbf{x}_n,\mathbf{w})-\mathbf{t}_n\|^2\xrightarrow{P}\frac{1}{2}\iint\|\mathbf{g}(\mathbf{x},\mathbf{w})-\mathbf{t}\|^2 f_{x,t}(\mathbf{x},\mathbf{t})\mathrm{d}\mathbf{t}\mathrm{d}\mathbf{x} \tag{4.29}$$

where $\xrightarrow{P}$ denotes convergence in probability, $f_{x,t}$ is the joint probability density function of $\mathbf{x}$ and $\mathbf{t}$. At this point, if the conditional expectation of $\mathbf{t}$ given $\mathbf{x}$ is denoted by $\langle \mathbf{t}|\mathbf{x} \rangle$, and the following manipulation is performed

$$\mathbf{g}(\mathbf{x}, \mathbf{w}) - \mathbf{t} = \{\mathbf{g}(\mathbf{x}, \mathbf{w}) - \langle \mathbf{t}|\mathbf{x} \rangle\} + \{\langle \mathbf{t}|\mathbf{x} \rangle - \mathbf{t}\} \tag{4.30}$$

then, by substituting (4.30) back into (4.29) and after a number of algebraic manipulations, it can be shown that minimizing with respect to $\mathbf{w}$ the error function E in the limit for $N \to \infty$ is equivalent to minimizing

$$E_\infty = \frac{1}{2} \int \| \mathbf{g}(\mathbf{x}, \mathbf{w}) - \langle \mathbf{t}|\mathbf{x} \rangle \|^2 f_x(\mathbf{x}) \mathrm{d}\mathbf{x} \tag{4.31}$$

where f_x is the marginal probability density function for $\mathbf{x}$. Given that the function being integrated is nonnegative, the value of $\mathbf{w}$ that minimizes the integral is the value that makes the integral zero, i.e. the $\mathbf{w}_o$ such that

$$\mathbf{g}(\mathbf{x}, \mathbf{w}_o) = \langle \mathbf{t}|\mathbf{x} \rangle \tag{4.32}$$

In other words, a neural network trained in such a way to optimize the error function of Eq. (4.28) tends to return – in the limit for large training sets – an approximation of the conditional expectation of the target vector $\mathbf{t}$ given the input vector $\mathbf{x}$, i.e. the regression of $\mathbf{t}$ on $\mathbf{x}$.

It is worthwhile to mention that the generality of the result stated above goes beyond multilayer perceptrons and even beyond neural networks at all. All that is required for it to hold is that a regression algorithm is trained using the error function (4.28). Supervised neural networks are just a practically interesting subset of mathematical models satisfying this property.

The result mentioned above is extremely important for our purposes, as it elucidates – perhaps better than the universal approximation theorems – the reasons why neural networks can be used for solving inverse problems. In fact, as discussed in Sect. 4.2, an inverse problem is not simply the problem of finding the inverse of a given function, but is instead a statistical decision problem, where an estimate of an unknown state $\mathbf{s}$ must be constructed out of a vector of observations $\mathbf{m}$. The property (4.32) indicates that a neural network trained to retrieve $\mathbf{s}$ from $\mathbf{m}$ so as to minimize the error function (4.28) will tend to approximate the posterior-mean estimator of the state vector $\mathbf{s}$ given the observation vector $\mathbf{m}$. As noted by Bishop (1995a), in order to achieve a good approximation it is important to have large training sets as well as a large number of free parameters, so that the family of functions $\mathbf{g}(\mathbf{x}, \mathbf{w})$ is more likely to contain members that make the right-hand of Eq. (4.31) small enough.

4.3.4 Self-organizing Maps

The self-organizing map (SOM), introduced by Kohonen (1982), is probably the most popular type of unsupervised neural network. It consists of a lattice of neurons. The jth neuron in the network is characterized by:

1. A weight vector $\mathbf{w}_j$
2. A *neighbourhood function* h_{jk}, where k is the index of another neuron.

The neighbourhood function is a decreasing function of a topological distance d_{jk} between two neurons j and k. For example, if the lattice is one-dimensional, d_{jk} can be simply defined as $|j - k|$, whereas for a multidimensional lattice one can simply set $d_{jk} = \|\mathbf{r}_j - \mathbf{r}_k\|$, where $\mathbf{r}_j$ and $\mathbf{r}_k$ are the vectors of the indices defining the neurons j and k in the lattice dimensions. A commonly used neighbourhood function is

$$h_{jk} = \exp\left(-\frac{d_{jk}^2}{2\sigma^2}\right) \tag{4.33}$$

where σ is, as usual, a parameter controlling the width of the function.

Prior to the training, the weights of the SOM are initialized to random values. The neural network is then trained as follows (Haykin 1999). For each input vector $\mathbf{x}$ in the training set:

1. The distance between $\mathbf{x}$ and each weight vector $\mathbf{w}_j$ is computed, and the neuron $i(\mathbf{x})$ with the smallest distance is chosen as the "winner":

$$i(\mathbf{x}) = \arg\min_j \|\mathbf{x} - \mathbf{w}_j\| \tag{4.34}$$

   This is called the "competitive phase" of the learning process.
2. Each weight vector $\mathbf{w}_j$ is updated according to

$$\mathbf{w}_j \leftarrow \mathbf{w}_j + \eta h_{ji(\mathbf{x})}(\mathbf{x} - \mathbf{w}_j) \tag{4.35}$$

   where η is a small positive number. This is called the "cooperative phase".
3. The width of the neighbourhood function around the winning neuron is decreased.

Equations (4.34) and (4.35) are crucial for understanding the meaning of unsupervised training. Each time an input signal $\mathbf{x}$ is presented to the neural network, the weight vector of the neuron $i(\mathbf{x})$ that is closest (i.e. most similar) to $\mathbf{x}$ is brought even closer to the input signal. The weights of the other neurons, instead, are updated in a way that depends on how close they are to the winning neuron $i(\mathbf{x})$ topologically, i.e. in terms of their relative positions within the lattice. The closer a neuron is to $i(\mathbf{x})$ the higher the value of the neighbourhood function, and thus the higher the update. On the contrary, for neurons that are very far from the winner the neighbourhood function will have a very small value, and thus a very small update (virtually, no update at all) will take place. It should be clear, at this point, what the asymptotic

effect of this process is: namely, that each neuron, or group of neurons, will focus on a different subregion of the input space, and therefore inputs that are "similar" in the signal space will produce similar activation patterns in the neural network response, whereas inputs that are very dissimilar in the signal space will tend to activate different regions of the neural network lattice.

The behaviour described above makes SOMs useful tools for data classification and clustering, and the applicability of this neural network model to remote sensing problems is clear. It is important to recognize that the output of a SOM is not a classification, but only a clustering of the input space. The actual classification – that is, in remote sensing, the final decision about the "state" of the observed scene (cloud, land, vegetation, ocean etc.) must be performed at a later stage. Several ways exist to do this, and the most suitable one depends on the problem at hand. Some examples are discussed in Sects. 4.5.5 and 4.5.6. For the time being we note that from Eq. (4.35) it is probably clear that, as the training proceeds, the weight vector of a neuron that specializes on a given subregion of the input space gets closer and closer to a "representative" input for that subregion. For example, if the input space we are clustering is a space of multispectral reflectances, the weight vector of a neuron specializing on recognizing vegetation will tend to resemble a typical vegetation reflectance spectrum, that is, a "code word" for the subset of vegetation spectra. This can be useful in the process of mapping SOM clusters to physically meaningful classes.

4.3.5 Learning and Generalization

As we have seen in the previous sections, the parameters of an artificial neural network are tuned by optimizing some kind of cost function on a training set. It is important to realize, though, that minimizing a cost function on the training set is not the final goal of the design of a neural network. The final goal is to achieve a correct behaviour on observations that are not in the training set (what we call "generalization"), while the use of a training set in order to specify the neural network model is just a means for achieving this goal. In light of this, it is clear that designing a neural network in order to perform a certain task only makes sense if the available training set can be seen as a statistically representative sample of an underlying process following a certain probabilistic law. To be more precise, let us suppose that we train our neural network by minimizing the cost function

$$E_N(\mathbf{w}) = \frac{1}{2N} \sum_{i=1}^{N} \|\mathbf{g}(\mathbf{x}_n, \mathbf{w}) - \mathbf{t}_n\|^2 \tag{4.36}$$

over a training set of pairs $(\mathbf{x}_i, \mathbf{t}_i)$ with $i = 1, \ldots, N$. What we would really want from our neural network is not just an optimal behaviour over this particular set of data, but a reasonably good behaviour over whatever data pair $(\mathbf{x}, \mathbf{t})$ from the joint

probability distribution of $\mathbf{x}$ and $\mathbf{t}$. In other words, we would like to minimize the following, theoretical cost function

$$E_\infty(\mathbf{w}) = \iint \|\mathbf{g}(\mathbf{x}, \mathbf{w}) - \mathbf{t}\|^2 f_{xt}(\mathbf{x}, \mathbf{t}) d\mathbf{x} d\mathbf{t} \tag{4.37}$$

which is unknown to us. However, as long as $E(\mathbf{w})$ can be seen as a "good" estimate of $E_\infty(\mathbf{w})$, using a training set in order to optimize the cost function (4.37) is theoretically justifiable. As we have seen in Sect. 4.3.3, $E_N(\mathbf{w})$ converges to $E_\infty(\mathbf{w})$ for $N \to \infty$ as long as the training data are independent and identically distributed with probability density $f_{x,t}(\mathbf{x}, \mathbf{t})$. The *principle of empirical risk minimization* states that, as long as $E_N(\mathbf{w})$ converges to $E_\infty(\mathbf{w})$ in probability, also the set of parameters $\mathbf{w}$ that minimizes E_N converges in probability to the $\mathbf{w}$ that minimizes E_∞ (Haykin 1999).

Another issue connected to generalization is that of architecture selection, that is, of selecting the number of neurons and layers in a neural network. It is intuitively clear that network architectures with many layers and neurons have a higher "representation power" than networks with few layers and neurons, i.e. they span "larger" function spaces, and as such, they are capable of approximating more complex functions. However, they also have more potential for "overfitting" the training data, i.e. to fit such data with non-smooth functions, in such a way that very small errors are achieved in the training set but much larger errors are observed when processing unseen data. Limiting the complexity of the architecture can be seen as a regularization procedure, because it tends to reduce the space of possible fitting functions to a class of "smooth" functions. Vapnik (1998) derived explicit upper bounds for the generalization error of a learning algorithm, and showed that such bounds increase with a function of the network complexity called the Vapnik-Chervonenkis (VC) dimension, which we will not discuss in detail here. The idea of limiting the risk of overfitting by limiting the network complexity is called *structural risk minimization* principle. It must be kept in mind, though, that this is not the only possible way to design networks that possess good generalization ability. Other regularization methods include the limitation of the amplitude of the network weights (Bartlett 1997) and adding noise to the input data during the training process, which, as shown by Bishop (1995b), has the same effect as applying a Tikhonov regularization to the training cost function.

4.4 Support Vector Machines

The support vector machine (SVM) is a supervised learning model alternative to neural networks. It has been originally proposed in an context of binary classification and single-output regression, but it can be extended to multi-class classification and multiple-output regression, although the theoretical details of how this can be done

will not be discussed in this paper. The interested reader can refer to Lee et al. (2004a), Tuia et al. (2011), and references therein.

4.4.1 Classification SVM

The basic idea behind the design of SVMs for classification is that of a binary linear classifier, designed so as to maximize a so-called "classification margin". In order to define the concept of classification margin, let us start by recalling that a binary classifier can be seen as a decision rule that maps an observation vector $\mathbf{x}$ into a decision $y \in \{-1, 1\}$ after applying a linear operator on $\mathbf{x}$. An example of decision rule is

$$y = \text{sign}(\mathbf{w}^T\mathbf{x} + b) = \begin{cases} 1 & \text{if } \mathbf{w}^T\mathbf{x} + b \geq 0, \\ 0 & \text{if } \mathbf{w}^T\mathbf{x} + b < 0 \end{cases} \tag{4.38}$$

This decision rule partitions the input space in two regions, separated by the hyperplane with equation $\mathbf{w}^T\mathbf{x} + b = 0$. This hyperplane is called a "decision surface": any $\mathbf{x}$ located above this hyperplane will be classified as 1, any vector located below this hyperplane will be classified as -1. If we have a training set $\{(\mathbf{x}_k, y_k)\}_{k=1}^N$, where N is the number of examples, and we indicate with $\mathbf{x}^+$ and $\mathbf{x}^-$ the vectors belonging to the class 1 and -1 that are closest to the separation hyperplane, we call "classification margin" the projection of the difference vector $\mathbf{x}^+ - \mathbf{x}^-$ on the normal to the separation hyperplane. Mathematically, the margin is thus defined as

$$M = \frac{\mathbf{w}^T(\mathbf{x}^+ - \mathbf{x}^-)}{\|\mathbf{w}\|} \tag{4.39}$$

Intuitively, maximizing the classification margin means maximizing the distance between the data points that are most difficult to classify. It can be proved (see, e.g., Kecman 2001; Schölkopf and Smola 2002, for details) that the problem of finding $\mathbf{w}$ and b to maximize M is equivalent to the following constrained optimization problem:

$$\begin{aligned} &\underset{\mathbf{w},b}{\text{minimize}} && \frac{1}{2}\mathbf{w}^T\mathbf{w} \\ &\text{subject to} && y_i[\mathbf{w}^T\mathbf{x}_i + b] \geq 1 \quad i = 1, \ldots, N \end{aligned} \tag{4.40}$$

This is a quadratic optimization problem. An important property of this class of problems is that the function being optimized is convex, and therefore there are no local minima. The Lagrangian function for this problem is:

$$L(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_{i=1}^{N} \alpha_i \{y_i[\mathbf{w}^T\mathbf{x}_i + b] - 1\} \tag{4.41}$$

where $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_N]^T$ is a vector of Lagrange multipliers and where $\mathbf{w}^T\mathbf{w}$ has been multiplied by 1/2 for notational convenience. For $\mathbf{w}$ and b to be a solution of the problem (4.40), the first-order Karush-Kuhn-Tucker (KKT) conditions must be imposed (Nocedal and Wright 1999):

$$\partial L/\partial \mathbf{w} = 0 \tag{4.42}$$

$$\partial L/\partial b = 0 \tag{4.43}$$

$$\alpha_i\{y_i[\mathbf{w}^T\mathbf{x}_i + b] - 1\} = 0 \qquad i = 1, \ldots, N \tag{4.44}$$

These are necessary conditions, but the convexity of the cost function makes them also sufficient. Conditions (4.42) and (4.43) imply

$$\mathbf{w} = \sum_{i=1}^{N} \alpha_i \mathbf{x}_i \tag{4.45}$$

$$\sum_{i=1}^{N} \alpha_i y_i = 0 \tag{4.46}$$

which, substituted back into the problem (4.40), allow to formulate the dual problem

$$\begin{aligned} \underset{\alpha}{\text{maximize}} \quad & \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} y_i y_j \alpha_i \alpha_j \mathbf{x}_i^T \mathbf{x}_j \\ \text{subject to:} \quad & \sum_{i=1}^{N} \alpha_i y_i = 0 \\ & \sum_{i=1}^{N} \alpha_i = 0 \\ & \alpha_i \geq 0 \qquad i = 1, \ldots, N \end{aligned} \tag{4.47}$$

The only unknowns in the dual problem are the Lagrange multipliers α_i. Once the problem is solved, using any of the available techniques for quadratic optimization, the weights of the optimal separating hyperplane can be recovered by using Eq. (4.45).

It is important to note that, in the optimal solution of problem (4.47), the only nonzero Lagrange multipliers will be those relative to the *active* constraints, i.e. the constraints that are satisfied with strict equality. In practice, these are the constraints relative to the data points $\mathbf{x}_i$ whose distance from the optimal separating hyperplane is minimum. Such points are called "support vectors". As a consequence, the summations needed to recover the optimal hyperplane weights are not really to be performed over the entire training set, but only over the support vectors.

Once the weights $\mathbf{w}$ have been determined, the bias b of the separating hyperplane can be computed as

$$b = \frac{1}{N_{sv}} \left[\sum_{s=1}^{N_{sv}} \left(\frac{1}{y_s} - \mathbf{w}^T \mathbf{x}_s \right) \right] \tag{4.48}$$

where N_{sv} is the number of support vectors and the summation runs over the support vectors only.

The algorithm described above needs to be modified when a nonlinear problem is dealt with. In a nonlinear problem, the classes cannot be perfectly separated by a hyperplane, i.e. a hyperplane is guaranteed to misclassify some of the data points. In the framework of SVMs, the strategy followed to address nonlinear problems consists of two elements, that will be now explained:

1. Mapping vectors from the input space to a higher-dimensional feature space through a nonlinear function.
2. Relaxing the constraints in Problem (4.40) to allow for misclassification of some data points.

Mapping input vectors to a higher-dimensional space is justified by a result known as Cover's theorem on separability of patterns (Cover 1965). Such theorem can be summarized by saying that a classification problem is more likely to become linearly separable when nonlinearly cast into a higher-dimensional space than in a lower-dimensional space (Haykin 1999). While this may seem counterintuitive at first sight, it is not difficult to formulate simple examples in which this line of reasoning works. If, for example, the true separating hypersurface of a nonlinear problem has equation $(2x_1 + x_2)^2 = 0$, it is evident that the separating surface becomes a hyperplane if the problem is cast in terms of three transformed variables $z_1 = x_1^2$, $z_2 = x_2^2$ and $z_3 = x_1 x_2$. The example we have just shown is, of course, an ad hoc construction, but the reader may think of many other situations where the same line of reasoning applies. Thus, in formal terms, in a SVM for nonlinear classification the problems (4.40) or (4.47) are not formulated in terms of the original variables $\mathbf{x}_i \in \mathbb{R}^n$ but in terms of transformed variables $\mathbf{z}_i = \boldsymbol{\Phi}(\mathbf{x}_i) \in \mathbb{R}^m$ with $m > n$, where $\boldsymbol{\Phi}$ is a nonlinear mapping from the original input space to a higher dimensional feature space.

Since the mapping to a higher-dimensional space may still be not enough to make the problem completely linear, the constraint that all the vectors in a given class lie "above" or "below" the separating hyperplane is relaxed in the following way:

1. N so-called "slack variables" ξ_i are defined, and the constraints in the problem (4.40) are changed to

$$y_i[\mathbf{w}^T \mathbf{x}_i + b] \geq 1 - \xi_i \qquad i = 1, \ldots, N$$

2. The cost function in the problem 4.47 is updated as

$$\frac{1}{2}\mathbf{w}^T \mathbf{w} + C \left(\sum_{i=1}^{N} \xi_i \right)^k$$

with $k \in \mathbb{Z},\ k > 0$, and $C \in \mathbb{R},\ C \geq 0$ a trade-off parameter chosen by the user. Choosing large values of C drives the algorithm towards large-margin separating hyperplanes, at the cost of more misclassified training points. Small values of C, instead, push towards hyperplanes that misclassify as few training points as possible, at the cost of a lower margin between the classes.

To summarize, the problem solved by a SVM classifier becomes (for $k = 1$):

$$\begin{aligned} \underset{\mathbf{w},b}{\text{minimize}} \quad & \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{N}\xi_i \\ \text{subject to} \quad & y_i[\mathbf{w}^T\boldsymbol{\Phi}(\mathbf{x}_i) + b] \geq 1 - \xi_i \quad i = 1, \ldots, N \\ & \xi_i \geq 0 \qquad\qquad\qquad\qquad i = 1, \ldots, N \end{aligned} \tag{4.49}$$

and it can be shown that the corresponding dual problem is (Kecman 2001):

$$\begin{aligned} \underset{\alpha}{\text{maximize}} \quad & \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} y_i y_j \alpha_i \alpha_j \boldsymbol{\Phi}^T(\mathbf{x}_i)\boldsymbol{\Phi}(\mathbf{x}_j) \\ \text{subject to:} \quad & \sum_{i=1}^{N}\alpha_i y_i = 0 \\ & \sum_{i=1}^{N}\alpha_i = 0 \\ & 0 \leq \alpha_i \leq C \qquad i = 1, \ldots, N \end{aligned} \tag{4.50}$$

After solving the dual problem (4.50), the parameters of the separating hyperplane can be recovered as

$$\mathbf{w} = \sum_{s=1}^{N_{sv}} \alpha_s y_s \boldsymbol{\Phi}(\mathbf{x}_s) \tag{4.51}$$

$$b = \frac{1}{N_{sv}}\left[\sum_{s=1}^{N_{sv}}\left(\frac{1}{y_s} - \mathbf{w}^T\boldsymbol{\Phi}(\mathbf{x}_s)\right)\right] \tag{4.52}$$

where we have emphasized that the Lagrange multipliers α_i are only nonzero for the support vectors, and thus the sums can be computed by just considering such vectors instead of the entire training set.

The equation of the separating hyperplane in the transformed space is $\mathbf{w}^T\boldsymbol{\Phi}(\mathbf{x}) + b = 0$, where $\mathbf{w}$ and b are given by Eqs. (4.51) and (4.52). Substituting Eq. (4.51) in the expression we see that the separating hypersurface has the following equation:

$$\sum_{i=1}^{N_{sv}} y_i \alpha_i \boldsymbol{\Phi}^T(\mathbf{x}_i)\boldsymbol{\Phi}(\mathbf{x}) + b = 0 \tag{4.53}$$

This equation is important because it introduces us to the last central concept in SVM theory: that of kernel. A kernel function is a function of $\mathbf{x}$ defined by

$$K(\mathbf{x}_i, \mathbf{x}_j) = \boldsymbol{\Phi}^T(\mathbf{x}_i)\boldsymbol{\Phi}(\mathbf{x}_j) \tag{4.54}$$

for some vector-valued function $\boldsymbol{\Phi}$. Only functions that can be written as inner products such as in Eq. (4.54) are admissible as kernels for a SVM. The conditions for a function $K(\mathbf{x}_i, \mathbf{x}_j)$ to be an admissible kernel are stated in a theorem due to Mercer (1909), which we will not discuss. Some of the most common kernel functions are the polynomial function

$$K(\mathbf{x}_i, \mathbf{x}_j) = (1 + \mathbf{x}_i^T \mathbf{x}_j)^d \tag{4.55}$$

with d chosen by the user; the Gaussian function

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{1}{2}(\mathbf{x}_i - \mathbf{x}_j)^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \mathbf{x}_j)\right) \tag{4.56}$$

and the hyperbolic tangent

$$K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\mathbf{x}_i^T \mathbf{x}_j + b) \tag{4.57}$$

for some values of b.

The importance of the kernel concept lies in the fact that its introduction allows to compute the inner products $\boldsymbol{\Phi}^T(\mathbf{x}_i)\boldsymbol{\Phi}(\mathbf{x}_j)$ in the input space rather than in the transformed space, which has a higher dimension. It can be also noted that a Gaussian kernel gives rise to an input-output function that is exactly analogous to that of a RBF neural network, whereas a hyperbolic tangent kernel gives rise to the input-output function of a multilayer perceptron. Therefore, all the results concerning the universal approximation properties of neural networks are also inherited by SVMs with kernels chosen accordingly. Finally, we note that SVMs with Gaussian or hyperbolic tangent kernels can be seen as neural networks trained with a different method, in which the number of hidden layers is always fixed to 1 and the number of hidden neurons is not chosen by the user but is determined by the algorithm based on the number of support vectors found in the training set (we remind that the support vectors are those training data that correspond to nonzero Lagrange multipliers in the optimization procedure). This feature is often appreciated by developers who consider it as a less subjective approach to the design of machine learning algorithms. However, it should be remembered that the number of support vectors is influenced by the choice of the parameter C in solving the problem (4.49), which is just as subjective as the choice of the number of hidden neurons in a neural network. Another attractive feature of

SVM training is that the optimization problem (4.49) is convex, and therefore no local minima exist in the problem of determining the weights of a SVM for a given C. However, this does not guarantee that the global minimum found by training of a SVM will be better (i.e. lead to a more accurate classifier) than any local minimum found by training a neural network as the cost functions being optimized are different (see, e.g., Taravat et al. 2015, for a practical example).

4.4.2 Regression SVM

In regression problems the output variable y is not binary, but can be any real number. It has been seen that a SVM essentially computes an input-output function having the form

$$f(\mathbf{x}, \mathbf{w}) = \mathbf{w}^T \boldsymbol{\Phi}(\mathbf{x}) + b \tag{4.58}$$

that is, it applies a linear operation to a vector of nonlinear functions of the input $\mathbf{x}$. The goal of SVM regression is to simultaneously minimize an error function and the squared norm $\|\mathbf{w}\|^2$. Minimizing the error function means maximizing the quality of the fit between the regression function and the training data, whereas minimizing $\|\mathbf{w}\|^2$ is useful in order to avoid overfitting of the training data. While many regression algorithm minimize the squared error, in SVM regression a different type of error function is often preferred, proposed by Vapnik (1995) in his seminal work on SVM and called ε-insensitive error function:

$$|y - f(\mathbf{x}, \mathbf{w})| = \begin{cases} 0 & \text{if } |y - f(\mathbf{x}, \mathbf{w})| \leq \varepsilon, \\ |y - f(\mathbf{x}, \mathbf{w})| - \varepsilon & \text{otherwise.} \end{cases} \tag{4.59}$$

The user-defined parameter ε defines a tolerance region, often called "ε-tube", within which the misfit between the regression function and the data is counted as zero.

Reasoning in a similar way to what has been done for classification, it can be shown that the problem of training a SVM for regression is (Kecman 2001)

$$\begin{aligned} \underset{\mathbf{w},\boldsymbol{\xi},\boldsymbol{\xi}^*}{\text{minimize}} \quad & \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\left(\sum_{i=1}^{N}\xi_i + \sum_{i=1}^{N}\xi_i^*\right) & & \\ \text{subject to} \quad & y_i - \mathbf{w}^T\boldsymbol{\Phi}(\mathbf{x}_i) - b \leq \varepsilon + \xi_i & & i = 1, \ldots, N \\ & y_i - \mathbf{w}^T\boldsymbol{\Phi}(\mathbf{x}_i) - b \leq \varepsilon + \xi_i^* & & i = 1, \ldots, N \\ & \xi_i \geq 0 & & i = 1, \ldots, N \\ & \xi_i^* \geq 0 & & i = 1, \ldots, N \end{aligned} \tag{4.60}$$

where $\boldsymbol{\xi}$ and $\boldsymbol{\xi}^*$ are vectors of slack variables. After writing down the Lagrangian function for this problem and applying the KKT conditions, the dual problem can be formulated as

$$\begin{aligned}
\underset{\boldsymbol{\alpha},\boldsymbol{\alpha}^*}{\text{maximize}} \quad & -\varepsilon\sum_{i=1}^{N}(\alpha_i^*+\alpha_i)+\sum_{i=1}^{N}(\alpha_i^*-\alpha_i)y_i-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i^*-\alpha_i)(\alpha_j^*-\alpha_j)K(\mathbf{x}_i,\mathbf{x}_j) \\
\text{subject to} \quad & \sum_{i=1}^{N}\alpha_i^*=\sum_{i=1}^{N}\alpha_i \\
& 0\le\alpha_i^*\le C \qquad i=1,\ldots,N \\
& 0\le\alpha_i\le C \qquad i=1,\ldots,N
\end{aligned} \tag{4.61}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}^*$ are vectors of Lagrange multipliers. Please note that we have used the kernel notation $K(\mathbf{x}_i,\mathbf{x}_j)=\boldsymbol{\Phi}^T(\mathbf{x}_i)\boldsymbol{\Phi}(\mathbf{x}_j)$.

Once again, usually the dual problem (4.61) is solved, and then the SVM parameters are derived from the Lagrange multipliers as

$$\mathbf{w}=\sum_{i=1}^{N_{sv}}(\alpha_i^*-\alpha_i)\boldsymbol{\Phi}(\mathbf{x}_i) \tag{4.62}$$

$$b=\frac{1}{N_{sv}}\left[\sum_{i=1}^{N_{sv}}(y_i-\mathbf{w}^T\boldsymbol{\Phi}(\mathbf{x}_i))\right] \tag{4.63}$$

and once again we stress that these summations only run over the support vectors.

4.5 Applications

4.5.1 Forward Modeling

The universal approximation properties of neural networks, discussed in Sect. 4.3.1.1 justifies – in principle – their use in building approximate forward models. Indeed, a forward model can certainly be seen as a function that maps a set of atmospheric and surface properties into a set of radiative quantities (e.g. top-of-atmosphere radiance or full Stokes vector). It seems justified to assume that the function is smooth enough that the assumptions underlying the universal approximation theorems for neural networks are satisfied. The main advantage of approximating a forward model through a neural network lies in the computation speed of the approximate model. This can be important when the model is used in an iterative retrieval algorithm, as it has the potential of speeding up each iteration by a factor up to 10^6 (Chevallier et al. 1998).

As an example of use of neural networks as approximate forward model, here we show a very simple experiment, in which we approximate the I, Q and U elements of the Stokes vector at the top the atmosphere for multiple scattering from a set of aerosol parameters. Our input vector consists of the aerosol optical thickness at 550 nm for fine and coarse mode, the viewing zenith angle, the surface albedo (assumed constant with wavelength) and the surface pressure, whereas our output vector consists of the coefficients for the Fourier azimuthal expansion of the multiple scattering Stokes vector at 440, 490 and 670 nm. We used 8 coefficients per wavelength for I and Q, 7 coefficients per wavelength for U. In this simple experiment we kept the aerosol microphysical properties fixed. We trained a multilayer perceptron neural network on about 5000 radiative transfer simulations and tested it on about 4000 independent simulations, generated with data drawn from the same statistical population as the training data. Figure 4.1 shows plots of the neural network estimated versus forward model simulated Fourier coefficients of the Stokes parameter for multiple scattering at all wavelengths. An almost perfect correspondence between the forward model and the neural network estimates can be observed.

Figure 4.1 shows plots of neural network estimated Fourier coefficients of I, Q and U versus the values simulated by the radiative transfer model. An almost perfect agreement between the true and the estimated values can be shown. It is important, though, to evaluate the effect of replacing the forward model Fourier coefficients with the neural network estimates in the explicit computation of the actual Stokes parameters through the formulae (de Haan et al. 1987)

$$I(\varphi) = \sum_{m=0}^{N-1} I_m \cos(m\varphi) \tag{4.64}$$

$$Q(\varphi) = \sum_{m=0}^{N-1} Q_m \cos(m\varphi) \tag{4.65}$$

$$U(\varphi) = \sum_{m=1}^{N-1} U_m \sin(m\varphi) \tag{4.66}$$

where I_m, Q_m and U_m are the Fourier coefficients and φ is the relative azimuth angle. In our experiment we chose $N = 8$. In fact the Stokes parameters – or, as an alternative, intensity and degree of linear polarization – as a function of the scattering angle are the quantity of ultimate interest in retrieval algorithms. In Fig. 4.2 scatter plots are shown of the Stokes parameter I and Q estimated by the neural network against those simulated by the radiative transfer model at 440, 490 and 670 nm for a relative azimuth angle of 0° (principal plane). U is not shown because it is zero at this particular angle, as shown by Eq. (4.66). In Fig. 4.3 examples are shown of estimated degrees and angles of linear polarization at 490 nm for relative azimuth angles of 60° and 150°. The degree and angle of linear polarization δ_L and α_L are defined respectively as

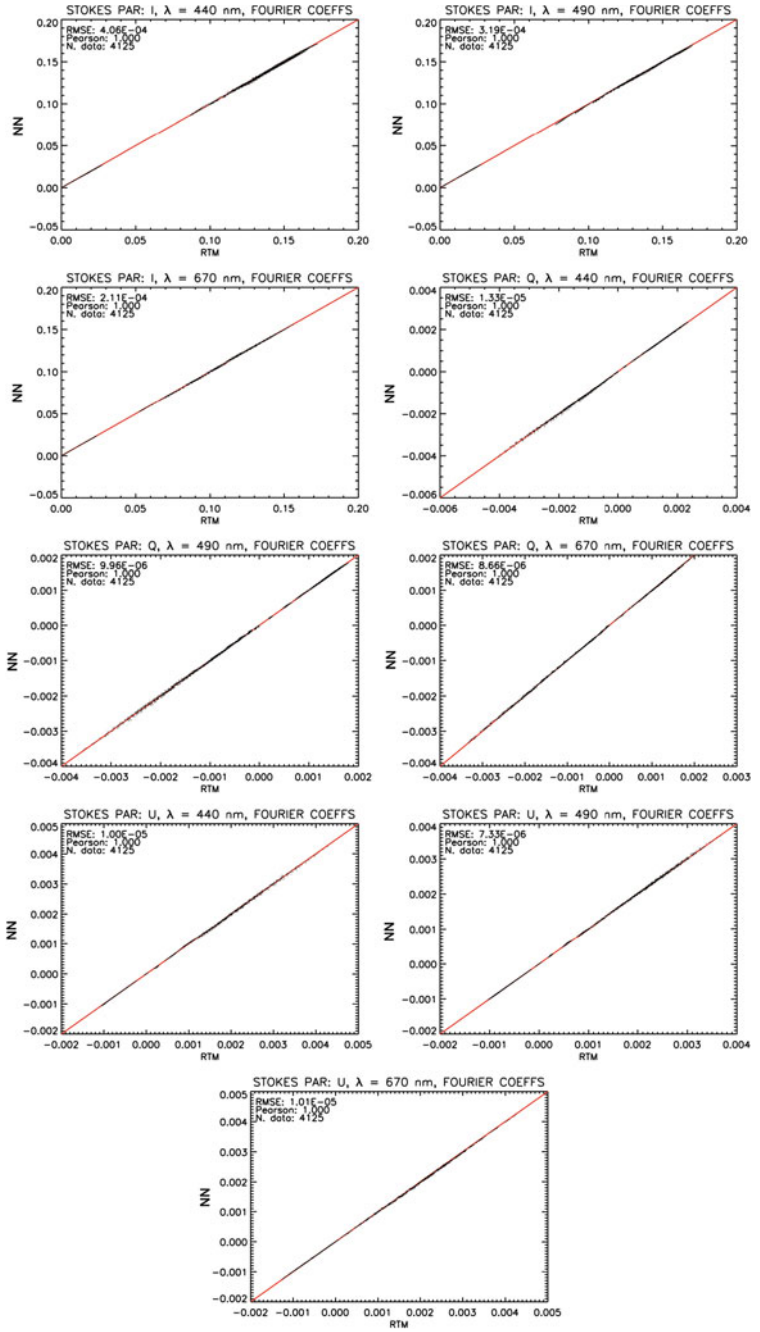

Fig. 4.1 Plots of neural network versus forward model coefficients of the azimuthal Fourier expansion of the multiple scattering Stokes vector at 440, 490 and 670 nm

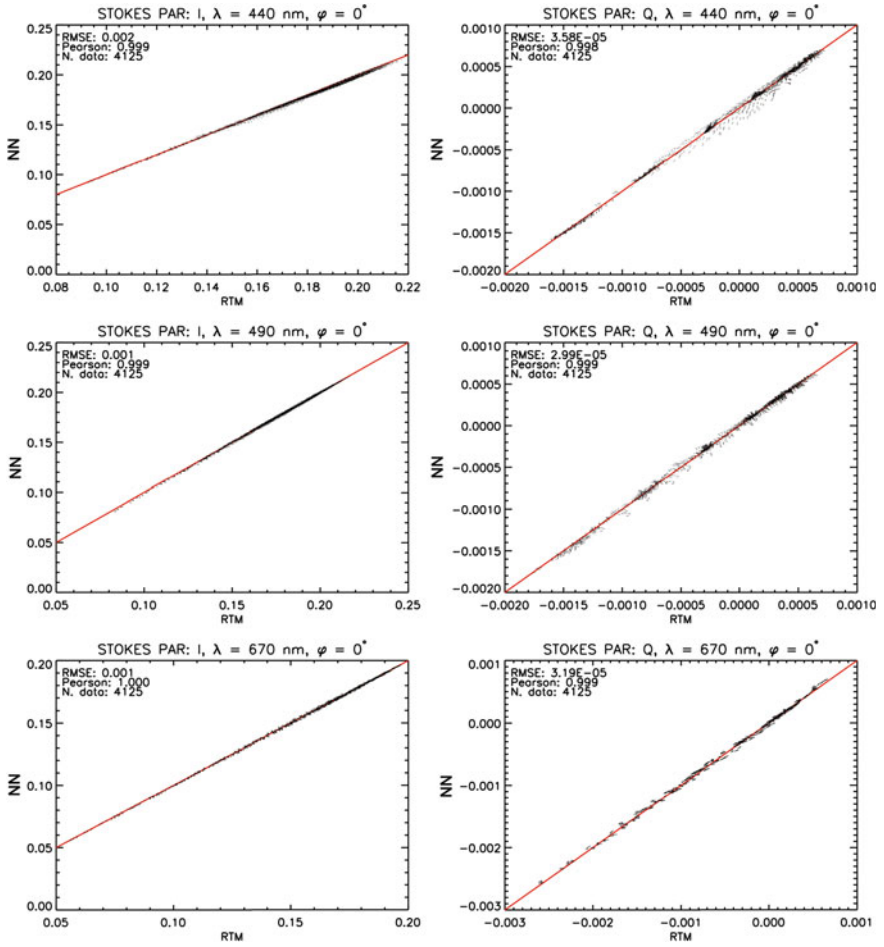

Fig. 4.2 Plots of neural network versus forward model simulations of the Stokes parameters I (left) and Q (right) at 440 (top), 490 (middle) and 670 nm (bottom) for a relative azimuth angle of 0°

$$\delta_L(\varphi) = \frac{\sqrt{Q^2(\varphi) + U^2(\varphi)}}{I(\varphi)} \tag{4.67}$$

$$\alpha_L(\varphi) = \frac{1}{2} \arctan \frac{U(\varphi)}{Q(\varphi)} \tag{4.68}$$

The root-mean-square error in the estimated degree of linear polarization appears to be in the order of 10^{-4}.

What we have shown is a very simple example of use of neural networks to perform approximated radiative transfer calculations. The approximation of a full radiative transfer model is a much more complex task, as it involves accounting

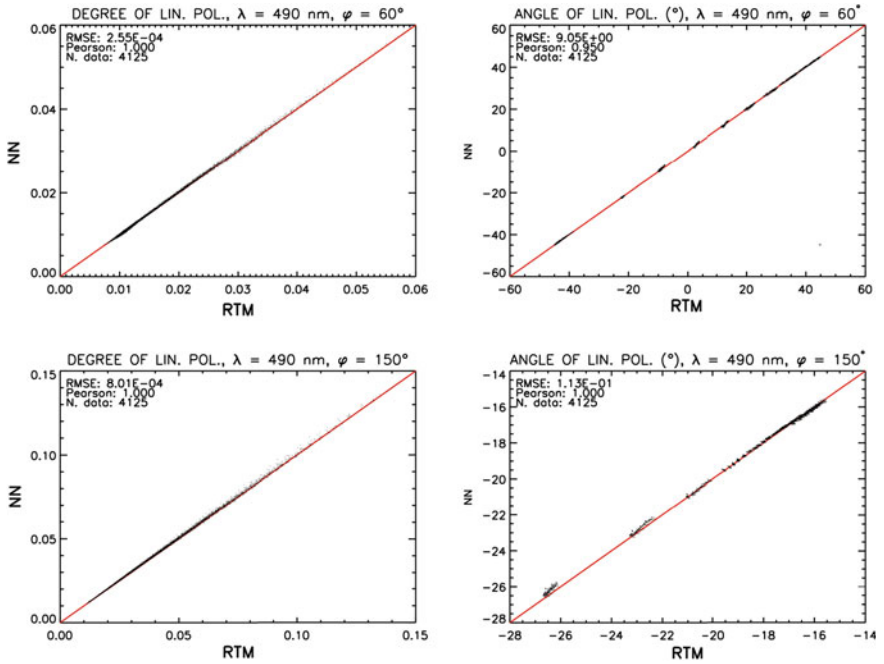

Fig. 4.3 Plots of neural network versus forward model simulations of the degree (left) and angle (right) of linear polarization at 490 nm for a relative angles of 60° (top) and 150° (bottom)

for the variability of a large number of parameters in the input vector of the neural network. Such parameters involve, of course, the aerosol microphysical properties as well as the surface properties. If atmospheric absorption must be taken into account, also the profiles of the most important absorbing gases have to be varied. While this fact has no conceptual consequences on the process of developing a neural network for approximate radiative transfer calculations, it creates practical difficulties, as the dimension of the training set needed to approximate a multivariate function to a certain accuracy grows roughly exponentially with the number of the independent variables that form the input space. This is one of the aspects of the so-called "curse of dimensionality" problem (Bellman 1961), and is easy to understand intuitively. For example, a set of 5000 randomly chosen points may constitute a sufficiently dense sampling of a 6-dimensional space, but may be an extremely sparse sampling of a 10-dimensional space. This forms an important practical limitation for the use of neural networks in the approximation of forward models with many input variables, as it leads to an increase in both computational demands and training times that is difficult to cope with, unless very powerful computers are used.

Another example of application of neural networks to simple radiative transfer problems is shown in Berdnik and Gallyamova (2012), where neural networks are used to approximate the transmittance of a plane layer containing a disperse medium.

The authors start from randomly chosen values of microphysical parameters (mean radius, ratio between second and first moment of the particle size distribution, complex refractive index and optical thickness), convert them into scattering phase function, single scattering albedo and asymmetry parameter using the Mie theory, and then train a number of RBF neural networks to simulate the reflectance and the transmittance of the layer at 12 incident radiation angles. In particular, a separate neural network is trained for each incident angle and radiation parameter (transmittance or reflectance).

Forward models that are particularly affected by speed issues are three-dimensional (3D) radiative transfer models, which are important in order to simulate the effect of inhomogeneous clouds on satellite measurements. While in one-dimensional, plane-parallel models, clouds are treated as horizontally homogeneous layers, in 3D models horizontal inhomogeneity is taken into account, and the horizontal transport of electromagnetic radiation is modeled. 3D models solve the radiative transfer equation using a Monte Carlo approach, where the electromagnetic radiation is represented as a discrete collection of "photons" which travel in a cloudy atmosphere undergoing random absorption and scattering processes (Marchuk et al. 1980). The computation of the radiation field in the atmosphere is stochastic, and the electromagnetic trajectories of a large number of "photons" must be computed in order to achieve good accuracies. Experiments aimed at approximating 3D radiative transfer computations using neural networks are described in Faure et al. (2001a, b). Faure et al. (2001b) trained multilayer perceptron networks with single hidden layer in order to model adjacency effects in inhomogeneous cloud scenes. Using a 3D radiative transfer model, the authors generated synthetic inhomogeneous cloud scenes with pixel sizes of 50 m. Then, they trained a number of neural networks to estimate the radiative properties of each pixel (reflectance and transmittance, treated separately) from those of a number of neighbouring pixels. In order to describe the effect of neighbouring pixels in the input vector they tried three alternatives: (1) simply using the reflectance (or the transmittance) of the adjacent pixels together with those of the target pixel; (2) using the differences in the considered radiative property between the target pixel and the neighbouring pixels; (3) using the difference between the target pixel and the radiative property averaged on all the pixels located between the target pixel and each neighbouring pixel. They report RMS errors smaller than 0.01 in the estimated reflectances, and around 0.035 in the transmittances. Faure et al. (2001a) described a similar experiment, but they tried to improve the performances of their neural networks by using multiple hidden layers. They also tested the effect of the assumed cloud model on the results of their neural networks by training them on synthetic scenes generated using a bounded cascade inhomogeneous model and testing them on scenes generated using two different models (white noise inhomogeneous, bounded cascade with fractional cloud cover). On test scenes generated used the same model as that used to generate the training set, their neural network approach outperformed the Independent Pixel Approximation (IPA) and its nonlocal version (NIPA, Marshak et al. 1998) in the approximation of the full 3D cloud radiative properties, whereas on scenes generated using the other two model a degradation in the neural network performance was observed. The conclusion reached by the authors is

that more complete training sets, encompassing multiple cloud generation models, may be used to improve the generalization of the neural network scheme.

If an approximate forward model based on neural networks is meant to be used in an iterative algorithm such as those discussed in Sect. 4.2, it is crucial that not only the radiative quantity of interest is simulated with sufficient accuracy, but that also its derivatives with respect to the quantities forming the state vector are. This is indeed crucial in order to ensure that the updates of the state vector performed during an iterative retrieval point in the right direction. While the universal approximation theorems stated in Sect. 4.3.1.1 indicate that the space of functions computed by neural networks can also approximate the derivatives of a continuous function to an arbitrary accuracy, the design of neural networks that actually possess this property may be difficult. This problem is discussed, for example, by Aires et al. (1999), where a strategy called "weight smoothing" is proposed in order to stabilize the Jacobi matrix of the neural network. Such strategy consists of updating the cost function for the neural network training with the introduction of a regularization matrix. Application of this method to forward model simulations in the thermal infrared has given encouraging results. Krasnopolsky (2007), instead, proposed a solution based on neural network ensembles, i.e., in training multiple neural networks and averaging their outputs so to reduce random error, with a beneficial impact on the accuracy of the neural network derivatives.

4.5.2 Inverse Scattering Problems

A first application of neural networks to inverse problems that are relevant to aerosol and cloud studies is presented in Ishimaru et al. (1990), where the mean radius and standard deviation of a log-normal size distribution are retrieved from simulated backscattered intensity measurements at three wavelengths (0.56, 1.06 and 2.12 μm) using a multilayer perceptron neural network. The experiment simply consisted in inverting the Fredholm integral equation that relates the particle size distribution to the backscattered intensity in an optically thin medium. The reported retrieval errors are smaller than 10% except for small values of the size distribution parameters. It is not clear, though, whether such results were obtained on noisy or noise-free data.

Ulanowski et al. (1998) and Wang et al. (1999) addressed the retrieval of the radius and refractive index of a single spherical particle from simulated intensities at a single wavelength and multiple scattering angles, using a RBF neural network. In both works a hierarchical approach is proposed. First, a neural network is trained generating a training set that covers the whole parameter range of interest for the experiment; then 10 networks are trained covering overlapping subsets of the radius range. The output of the first network would determine which of the 10 secondary networks would be used for the final retrieval. The performance of the algorithm was seen to be dependent on the type of preprocessing applied to the intensities (scaling and weighting by an angular factor). In the best case, which involved linear scaling and weighting by the fourth power of the sine of the scattering angle, mean errors of

the order of 10^{-5} and maximum errors of the order of 10^{-3} were reported for both the radius and the refractive index of the sphere (Ulanowski et al. 1998).

A hierarchical approach similar to the one described above was also used by Berdnik et al. (2004) in the retrieval of orientation and shape parameters of spheroidal particles. Instead of using RBF networks, the authors designed a three-level hierarchy of multilayer perceptrons with linear activation functions. The retrieved quantities are the radius of a sphere having the same volume as the considered spheroid (called the "equivolume sphere radius"), the shape parameter (ratio between the difference and the sum of the semiaxes) and an angle describing the orientation of the particle. The reported standard errors are of 0.003 μm for the equivolume sphere radius, 0.02 for the shape parameter and 6° for the orientation angle. Berdnik and Loiko (2009) performed an experiment similar to that carried out by Ulanowski et al. (1998), namely, to retrieve radius and refractive index of a sphere from multiangular intensity measurements at scattering angles between 10° and 60°. In this work, the authors did not use the hierarchical approach but solved the problem with a more classical multilayer perceptron with two hidden layers. They validated their results on noise-corrupted simulated data, reporting radius errors between 0.6 and 0.8% and refractive index errors between 3.7 and 8.1% (depending on the noise level). A similar experiment is presented in Berdnik and Loiko (2016), where, in addition, the performances of different neural network schemes (multilayer perceptron, RBF and two hierarchical approaches) are compared and their sensitivity to multiplicative intensity noise is assessed.

4.5.3 Cloud Detection and Classification

The simplest task in cloud remote sensing that a neural network can perform is cloud detection. In this task, we are simply interested in labeling pixels (or groups of pixels) in an image as "clear" or "cloudy".

A first example of application of neural networks to cloud detection in Advanced Very High Resolution Radiometer (AVHRR) imagery over ocean is presented in Yhann and Simpson (1995). They trained a number of multilayer perceptrons over pixels selected randomly, using combinations TOA reflectances for the visible channels and brightness temperatures for the thermal channels as inputs. As for the output, the authors tested two approaches. The first approach was to use one output unit taking values between 0 and 1, and to identify pixels with output higher than 0.5 as cloudy. The second approach was to use two output units taking values between 0 and 1 – one identifying cloud pixels and the other identifying ocean pixels – and to assign a pixel to the class whose corresponding output unit had the highest value.

Jang et al. (2006) developed a neural-network-based cloud detection scheme for SPOT VEGETATION images. Their neural network – a classical multilayer perceptron – receives as input all the four bands of the instrument and yields a single-valued output that is 0 for a cloud pixel, 1 for vegetation and 2 for water. The reported accuracies are between 97 and 98%.

Cloud classification goes a step beyond cloud detection, in the sense that not only a pixel or a region in an image is declared as clear or cloudy, but further inferences are made about cloud properties (e.g. cloud phase, cloud height).

A simple example of use of neural network for cloud classification is presented in Key et al. (1989). The authors of this study combined images from the AVHRR instrument onboard the NOAA-7 satellite and from the Nimbus-7 Scanning Multichannel Microwave Radiometer (SMMR) over the Arctic, to classify clouds and the surface underneath. They defined four surface classes (snow-free land, snow or ice covered land, open water and sea ice) and three broad cloud classes (low, middle and high cloud). This led them to further specialize the cloud classification into eight classes (low/high clouds over land/water/ice, middle clouds over water/ice). They trained a multilayer perceptron with a single hidden layer over areas manually selected over merged AVHRR/SMMR imagery. The inputs for the neural network were measurements from the two instrument channels (reflectances for AVHRR, brightness temperatures for SMMR), whereas the output was a vector of 12 units (one per class), each of which had the meaning of a "class membership indicator" between 0 and 1. In order to test the results, they compared the results of the neural network to a manually classified image. The reported correct classification rates were of almost 100% for the class "water", around 82% for the class "land", 69% for ice and 59.5% for snow. The correct classification accuracies for cloud classes were highly variable (from around 32% for "low clouds over water"and "medium clouds over ice" to 78% for "high clouds over ice"). It should be recognized, though, that the manual validation approach has clear limitations, especially on classes that are difficult to separate visually in an image, such as clouds over snow and ice.

A different approach to cloud classification using neural networks is presented in Lee et al. (1990). This approach is not based on radiometric quantities but on textural information, that is, information about the spatial variability of the radiance (or simply the digital number) within an image. Textural features within an image may be defined in different ways (Haralick et al. 1973). In Lee et al. (1990) the textural information content of an image is quantified by the Gray Level Difference Vector (GLDV), which is essentially a vector of statistical quantities derived from the probability distribution of the grey level differences between a certain pixel and a number of neighbouring pixels. This approach has been applied to imagery from the Landsat Multispectral Scanner (MSS). The GLDV was used as input vector for a multilayer perceptron with two hidden layers and a three-element output layer consisting of class membership indicators for stratocumulus, cumulus and cirrus. Comparisons with two traditional classification methods (nearest neighbour and discriminant analysis) showed a superior performance of the neural network classifier, which reached an overall accuracy of 93%, with 96% accuracy for cirrus, 92% for stratocumulus and 90% for cumulus. These results were obtained using single-channel visible imagery.

A neural network based cloud classification scheme combining radiometric and textural information is presented in Miller and Emery (1997). The algorithm has been named Cloud Automated Neural Network (CANN), and has been developed for AVHRR imagery. Prior to the cloud classification, a binary decision is made to identify each pixel as clear or cloudy based on the Normalized Difference

Vegetation Index (NDVI), on the channel 1 (0.63 μm) reflectance and on the channel 4 (10.8 μm) brightness temperature. Specifically, a pixel is considered to be cloud or water if it has NDVI less than zero and cloud or land if its NDVI is larger than zero. A reflectance threshold of 0.16 in channel 1 is used to distinguish between cloud and water, whereas a brightness temperature threshold of 295 K in channel 4 is used to distinguish between cloud and land. For the pixels identified as cloudy, a number of features (cloud fraction, cloud albedo, cloud top temperature and effective droplet radius) are estimated from radiance and brightness temperatures in a number of AVHRR channels, and are combined with six textural parameters (cloud connectivity, contrast, homogeneity, low/middle/high etage fraction) estimated by dividing AVHRR images into subregions. These features are used as inputs for the neural classifier, that distinguishes between nine classes (no cloud, stratus, stratocumulus, cumulus, cumulus congestus, altostratus/altocumulus/nimbostratus, cumulonimbus, cirrostratus, cirrus). The neural network has been trained on 57 images and tested on 5 independent images, with overall classification accuracies ranging from 73 to 87%.

Lee et al. (2004a, b) discuss the an extension of SVM classification to problems with more than two classes and present an application of their method to cloud detection in MODIS imagery. The classes they defined are cloud-free scenes, water clouds and ice clouds. In order to carry out the classification the authors tried four combinations of observables: (1) MODIS channel 2 radiance and the base-10 logarithm of the ratio between channel 5 and channel 6 radiances; (2) like (1) plus ratio between channel 1 and channel 2 radiances, brightness temperature in channel 31 and brightness temperature differences between channels 32 and 29; (3) just radiances in 12 channels (1–7, 27–29, 31 and 32); (4) logarithm of (3). The SVM were trained with synthetic data generated for a number of realistic atmospheric temperature and humidity profiles over ocean. A comparison with the MODIS algorithm over real images showed misclassification rates lower than 1% for the combinations of observables from (2) to (4), whereas the misclassification rate for combination (1), with only two observables, was around 4%. Another application of SVM is described by Mazzoni et al. (2007), who developed a method to classify MISR pixels into clouds, aerosol, land, water and ice, trained on real data, which achieved correct classification rates ranging from 70 to 80% depending on the considered class, but with occasionally high misclassification rates between ice and clouds, perhaps due to the intrinsic difficulty in distinguishing between these two surface types in the radiance domain at the four MISR wavelengths.

4.5.4 Retrieval of Cloud Properties

In this section we will review some applications of neural networks aimed at providing quantitative estimates of cloud properties, such as the size of cloud particles, cloud optical thickness (COT) and cloud albedo.

The first application of neural networks to cloud property retrieval from satellite data is, to our knowledge, due to Faure et al. (2001c). This study aimed at assessing the potential of neural network for retrieving cloud optical properties from instruments such as MODIS or the Global Imager (GLI) onboard ADEOS-2. The authors used a Monte Carlo radiative transfer model to simulate top-of-the atmosphere reflectances at visible and near-infrared wavelengths for realistic inhomogeneous cloud scenes. Then they designed two neural network schemes to retrieve the mean optical thickness at 0.64 μm, the particle effective radius, the fractional cloud cover and a parameter describing the inhomogeneity in the cloud cover. The first neural network only used the reflectance at four wavelengths (0.64, 1.6, 2.2 and 3.7 μm) averaged over a pixel of 0.8×0.8 km as input vector, whereas a second neural network also used, for each pixel, the reflectance spectra at the same wavelengths in 8 adjacent pixels. The errors in the retrieved optical thickness and effective radius were generally around 3–4% and 1–3% respectively, whereas the relative errors in the fractional cloud cover and in the cloud heterogeneity were more variable, ranging between 2 and 20% for the former and between 12 and 26% for the latter. According to the authors, including ancillary information on adjacent pixels did not result in a clear improvement in the performance of the neural network but gave mixed results. Cornet et al. (2004) refined the method by Faure et al. (2001c) by considering 6 cloud parameters to be retrieved (the cloud inhomogeneity parameter was split as COT and effective radius inhomogeneity, and the cloud top temperature was included in the retrieved state vector). Furthermore, while the neural networks presented by Faure et al. (2001c) were designed to work at a limited set of solar and viewing angles, Cornet et al. (2004) presented a way to process measurements at arbitrary solar and incidence angles by interpolating them to the angles at which the neural networks were trained. The interpolation was performed through a dedicated neural network mapping the radiances at an arbitrary solar-viewing angle combination, together with the distance between the considered combination and the nearest combination for which a network was trained and the surface albedo to the radiance at the desired solar and viewing angles. After this interpolation step, the neural-network-based retrieval was similar to that described by Faure et al. (2001c). The retrieval method was tested on three different types of clouds (flat-top bounded cascade, flat-top and non-flat-top Gaussian process cloud). Depending of the considered cloud type, the RMS retrieval errors approximately ranged between 1 and 2 for COT (with true values distributed between 0 and 30), between 1.5 and 1.9 μm for effective radius (true values between almost 0 to 25 μm), between 0.95 and 2 for COT inhomogeneity (true values between 0 and 15), between 0.6 and 1.5 for effective radius inhomogeneity (true values between and 15) and between 0.08 and 0.12 for fractional cloud cover. The RMS errors for cloud top temperature were around 0.2 K for flat-top clouds and 0.3 K for non-flat-top clouds. The approach we have just described has been applied by Cornet et al. (2005) to MODIS data, with limited adaptations in the input vector due to the unavailability of some of the required input variables in MODIS data. In order to test the sensitivity of their retrieval scheme to the assumptions made in the generation of the training datasets, the authors trained three different neural networks, one assuming a homogeneous cloud model, one assuming the bounded-cascade model and one

assuming a Gaussian model. By comparing retrievals from these three networks to the standard MODIS product the authors concluded that their retrieval scheme is not very sensitive to the cloud model used to generate the training set. In absolute terms, a good agreement between the retrieved COTs and those of the MODIS products was found, whereas the differences between the effective radii were larger, but consistent with previous studies about differences between homogeneous and inhomogeneous cloud effects (the MODIS product is based on a homogeneous cloud assumption).

Faure et al. (2002) developed a method similar to that described above for retrievals from high spatial resolution instruments. According to the authors, cloud retrievals from high-resolution imagers require a different approach with respect to retrievals from moderate-resolution imagers, as the 3D radiative transfer effects at high spatial resolution are different. The authors modified their retrieval method accordingly, by redefining the input vector for the neural network. The choice made by the authors was to include, for each pixel, the reflectance of that pixel and the reflectance difference between a number of pairs of contiguous pixels around that pixel. The retrieved parameters in this new scheme were the cloud optical thickness and the cloud effective radius, and two wavelengths were used instead of four (all the combinations between the four wavelengths listed before were tried). They tested their retrieval scheme on two different cloud types: inhomogeneous clouds with uniform effective radius, and inhomogeneous clouds with horizontally variable effective radius. The retrieved COTs and effective radii generally showed a good agreement with the true values.

Loyola et al. (2007) proposed the application of neural networks to the estimation of cloud properties from the Global Ozone Monitoring Experiment (GOME) onboard the ESA ERS-1 satellite. Their retrieval scheme, named Retrieval of Cloud Information using Neural Networks (ROCINN), uses high-spectral resolution reflectances in the oxygen A-band (758–772 nm), plus the cloud fraction estimated from the Polarization Measurement Devices (PMDs), to retrieve cloud-top height and cloud-top albedo. It was trained on synthetic spectra simulated using the Linearized Discrete Ordinate Radiative Transfer (LIDORT) forward model (Spurr et al. 2001) with line-by-line calculations of the oxygen absorption in the A-band, and neglecting molecular scattering, aerosol scattering and absorption, diffuse surface reflection and oxygen absorption within and below clouds. To validate their method, the authors compared the neural network retrievals to co-located Spinning Enhanced Visible and Infrared Imager (SEVIRI) retrievals for five months in 2004. The authors observed a bias of -0.64 km and a standard deviation of 1.15 km of the ROCINN cloud-top height versus SEVIRI. The comparison between the cloud-top albedos could only be qualitative, because of the inherent differences between the two instruments, but nevertheless revealed a good agreement, with a slight mean overestimation of 0.09 and a standard deviation of 0.1. Global comparisons between the cloud parameters retrieved by ROCINN and data from the International Satellite Cloud Climatology Project (ISCCP) are presented in Loyola et al. (2010). A negative bias in the cloud-top height is reported, probably due to the absence of cirrus clouds in ROCINN retrievals.

Another case of application of neural network to cloud retrievals from hyperspectral instruments is discussed in Saponaro et al. (2013), where a neural network

algorithm for fractional cloud cover estimation from Ozone Monitoring Instrument (OMI) data is presented. The input vector for the neural network proposed by the authors OMI reflectance spectra in the 349–504 nm channel, compressed through singular value decomposition (SVD), plus OMI small-pixel radiance at 388 nm (that can be seen as an indicator of the scene heterogeneity), climatological surface albedo and solar zenith angle. The cloud fraction is the only output of the neural network. The authors generated a training set by matching a number of OMI spectra with cloud fractions estimated from coincident MODIS measurements, and trained their neural network using the standard back-propagation and the Extreme Learning Machine (ELM) algorithm (Huang et al. 2005), an alternative training method. The cloud fractions retrieved by the neural network were in reasonable agreement with MODIS retrievals over ocean, whereas over land an overestimation of the cloud fraction was observed over bright surfaces.

An application of neural networks to the retrieval of the optical properties of cirrus clouds from SEVIRI is presented by Kox et al. (2014). The goal of the authors was to use SEVIRI data in order to retrieve cirrus optical thicknesses and top altitudes from SEVIRI thermal channels during day and night. In order to accomplish this, they generated a training set by matching retrievals of the cirrus properties from the Cloud and Aerosol Lidar with Orthogonal Polarization (CALIOP) lidar with coincident SEVIRI observations, and trained a multilayer perceptron network to map the SEVIRI observations to the CALIOP cirrus retrievals. The SEVIRI measurement they used in the network input vector were the brightness temperatures at 7.3, 9.7, 12 and 13.4 μm, the brightness temperature differences between 6.2 and 7.4 μm, 8.7 and 12 μm and 10.8 and 12 μm. In addition the authors included latitude, viewing zenith angle and a land/sea mask as ancillary data. They trained the data on a very large dataset, with 8 millions input-output pairs, and tested it on 1 million independent CALIOP data and on a set of retrievals performed using an airborne High Spectral Resolution Lidar (HSRL), reporting a standard error of 0.25 for cirrus optical thickness and 750 m for the top altitude.

4.5.5 Aerosol Classification

The impact aerosols exert on atmospheric radiation does not only depend on the concentration of aerosol particles in the atmosphere, but also on their size and their absorbing properties. Distinguishing between large and small particles or between strongly-absorbing and little- or non-absorbing aerosols is an important task, not only when the aerosol properties are the main object of study but also when aerosols are seen as a nuisance, such as in the atmospheric correction of land or ocean imagery. Also in cases where the available information is insufficient to determine all the aerosol properties (or most of them), it may be still possible to make a qualitative inference on whether the aerosols inside a scene are absorbing or not, or on whether they mainly consist of small or large particles. The recognition of this fact has led, over time, to the notion of “aerosol typing”, that is, to the custom of summarizing

the salient radiative properties of aerosols by means of a "type". A popular typing scheme is, for instance, that by Shettle and Fenn (1979), where aerosols are classified as "rural", "marine", "urban" and "tropospheric", but also other subdivisions into types have been proposed (e.g., Kahn et al. 2001; Holzer-Popp et al. 2008; Russell et al. 2010, 2013).

An interesting application of neural networks to aerosol classification has been proposed by Niang et al. (2003), and is based on self-organizing-maps. The authors applied an extension of the SOM network seen in Sect. 4.3.4, called Probabilistic SOM (PRSOM), to the clustering of unpolarized reflectances from ADEOS-POLDER at five wavelengths (443, 490, 565, 670, 865 nm) and a single viewing zenith angle over the Cape Verde region. Their algorithm defined a map of 20×20 neurons, thereby defining 400 different clusters. Each cluster is uniquely characterized by the weights of the corresponding neuron, that can be seen as a "representative reflectance spectrum" (called "reference vector" by the authors) for that cluster. The 400 clusters were then mapped into four physically meaning classes (marine aerosols, clouds and two types of dust aerosols) by introducing "expert knowledge" established as follows. Thresholds on POLDER AOTs at 865 nm were used to define three aerosol classes and a cloud coverage test was used to define the "cloud" class. Then, for each pixel of the POLDER image the reference vector of the winning SOM neuron was compared to the result of the "expert classification", and the reference vectors that were always associated with the same class were mapped to that class, whereas reference vectors associated to multiple classes were mapped to the class with the largest number of correspondences. The results reported by Niang et al. (2003) indicate that the SOM-based classifier was able to correctly identify almost all the cloudy pixels identified by the "expert" classification and to correctly identify marine aerosols and thick, homogeneous Sahara dust clouds in more than 90% of cases.

4.5.6 Aerosol Retrievals

As opposed to aerosol classification, where the goal is to just make a qualitative inference about the type of aerosol that is present in a certain scene, aerosol retrieval schemes aim at a quantitative estimation of the aerosol load and, in some cases, of the aerosol microphysical properties such as the particle dimension and shape and their complex refractive index. The first example of application of neural networks to aerosol retrievals is, to our awareness, due to Han et al. (2006), who investigated the application of multilayer perceptron networks to retrieve the aerosol optical thickness (AOT) from Multiangular Imaging Spectroradiometer (MISR) data. A number of neural networks, one called "global" plus several others called "region-specific", were trained using MISR radiometric measurements and AOTs derived from such measurements with another algorithm. The global neural network was trained on data collected over the entire globe whereas the region-specific ones were trained over single passes of MISR over certain regions. Validation over four 16-day cycles of

MISR data showed that the root-mean-square errors of the neural network schemes, as well as those of other data-driven approaches, were similar to those of the operational AOT product (approximately ranging from 0.065 to 0.1).

The concept devised by Han et al. (2006) for MISR was further developed by Vucetic et al. (2008) for AOT retrievals from Moderate Resolution Imaging Spectroradiometer (MODIS) data. Rather than using a single neural network, the authors of this study decided to train an ensemble of 10 neural networks using MODIS radiances and radiometric uncertainties at 7 wavelengths, plus solar, viewing and scattering angles and a number of surface type and cloud-contamination indicators, as inputs, and producing the AOT as output. In general, the goal of using an ensemble of networks is that of reducing the random errors arising a single network by averaging the outputs of different networks. Based on a comparison with the MODIS C005 product (Levy et al. 2007) on a set of AERONET test data, the neural network retrievals proved more accurate than the C005 retrievals.

In a new neural network scheme for AOT retrievals from MODIS, Radosavljevic et al. (2010) introduced a different way of constructing the ensemble of neural networks and a method of combining the outputs of the ensemble members alternative to simple averaging. Instead of training all the ensemble members on the same dataset, the authors proposed to train a subset of the ensemble members on cases of "low" AOTs and a subset on cases of "high" AOTs, following an approach that reminds the hierarchical one we discussed in Sect. 4.5.2) in the context of inverse scattering problems. Regarding the combination of the output of the ensemble members, the authors tested the possibility of performing a weighted average between the ensemble members with weights decided by another neural network. To this end, they explored two alternatives: (i) a third neural network trained by using the outputs of the ensemble members as inputs and providing the target AOT value as output; (ii) a "gating network", conceived as a classification network labeling the AOT of the scene as "high" or "low" and distributed the weights between the high-AOT and to the low-AOT ensemble networks according to this decision. These two new ensemble-based schemes were compared to the classical single-network scheme and to the ensemble scheme based on simple averaging, and an increase in the correlation coefficient between the retrieved AOTs and those measured by AERONET was observed. The results found by Vucetic et al. (2008) regarding the comparisons between neural network retrievals and C005 retrievals were confirmed by Radosavljevic et al. (2010) and by a later study by Ristovski et al. (2012), who also tried to assess the credibility of the uncertainties calculated on test datasets through a bootstrap approach.

A completely different neural-network-based approach to aerosol retrieval is proposed in Niang et al. (2006) and applied to Sea-Viewing Wide Field-of-View Sensor (SeaWiFS) imagery over the Mediterranean basin. Such approach is based on unsupervised neural networks, and is similar to the classification scheme by Niang et al. (2003) discussed in Sect. 4.5.5, but with the difference that the expert knowledge used to interpret the output consists of a set of synthetic reflectance spectra computed using a radiative transfer model from different combinations of AOT at 865 nm and aerosol Ångström exponent. Comparisons between the retrieved AOTs and those measured

by the Lampedusa AERONET station showed RMS errors around 0.035-0.04, comparable or smaller than those of the SeaWiFS official product.

A neural network algorithm proposed by Taylor et al. (2014) aims at using multispectral MODIS reflectances to estimate not only the AOT but also other aerosol parameters such as complex refractive indices, single scattering albedo (SSA) and size distribution. The approach the authors used to accomplish this task was not to directly estimate these quantities from MODIS reflectances, but to use MODIS AOTs at 470, 550 and 660 nm as input vectors for their neural network scheme. The authors also tested the use of coincident columnar water vapour retrieved by MODIS and absorption AOT at 550 nm from OMI as additional inputs. They applied their approach to the Saharan region. Their neural network scheme appears able to capture main trends in the seasonal variabilities of the aerosol volume concentration, complex refractive index and SSA, although not their daily variability, possibly because of the limited amount of information available in MODIS reflectances for the simultaneous retrieval of a large number of aerosol parameters.

4.5.7 Volcanic ash Detection and Retrieval

A particular case of aerosol retrieval application is represented by the study of volcanic ash resulting from major eruptions. Near-real-time, operational monitoring of volcanic ash is necessary, given the impact ash has on civil aviation (Alexander 2013), and several algorithms, mostly based on look-up-tables (LUTs), have been developed in order to detect volcanic ash and retrieve its microphysical parameters in satellite imagery (Prata 1989a, b; Prata and Grant 2001). They generally use the brightness temperature difference between two thermal infrared bands (10–11 μm and 11.5–12.5 μm) as a signature to distinguish volcanic ash plumes from ordinary water clouds (the difference is negative for volcanic ash and positive for water clouds). Picchiani et al. (2011) proposed a neural network approach to ash detection and retrieval from MODIS data. The authors trained two separate neural networks, one for ash detection and one for the retrieval of the ash mass on the pixels identified as containing ash by the first network. The ash detection neural network used as inputs the brightness temperatures in MODIS channels 28, 31 and 32. The brightness temperature in channel 28 (centered on 7.3 μm) was used as a proxy for the water vapour column, thereby eliminating the need for an assumption on this parameter, whereas the channels 31 and 32 correspond to the spectral bands involved in the brightness temperature difference that identifies the presence of ash. The sign of the brightness temperature difference provided the target value for the neural network. The ash retrieval neural network uses the same channels as the ash detection network as inputs, and provides an estimate of the ash mass in t/km^2. Again, the LUT algorithm was used to generate the target data for the training. The authors tested their neural network scheme on MODIS images acquired during a number of eruptions of Mt. Etna, in Italy. The percentage of correct detections varied between 75 and 98%, the percentage of missed detections between 6 and 12%, and the percentage of false

positives was varied between 12 and 64% depending on the considered image. The RMS errors of ash the ash mass retrievals was reportedly around 0.3 t/km^2.

An approach similar to that described above is discussed in Piscini et al. (2014). More precisely, this latter study aims at using MODIS data to estimate, in addition to the ash mass, the ash optical thickness at 11 μm, the ash effective radius and the SO_2 concentration in the ash plume. To this end, the authors extended the number of MODIS channels used by the neural network. They performed a sensitivity study to assess which channels – among the complete set of 36 MODIS channels – are the most important for the retrieval of each of the considered ash parameters, which led to a selection of 28 channels. As case study, the authors chose the famous eruption of the Icelandic volcano Eyjafjallajökull in 2010. Comparing the neural network retrievals to those performed using standard algorithms they found a reasonable consistency in AOT, ash mass and effective radius (RMS errors around 0.3, 1.3–1.8 t/km^2 and 0.8–0.9 μm with correlation coefficients around 0.90-0.95), whereas the retrieval capability for SO_2 was more limited.

A different neural-network-based scheme for the detection of volcanic ash in MODIS images is described in Gray and Bennartz (2015). The input vector of this neural network scheme includes the brightness temperature at 11 μm and the brightness temperature differences 12–11 μm, 11–8.6 μm, and 11–7.3 μm. The output quantity is a binary flag (ash/no ash). The training data are generated by matching a number of MODIS granules with ash concentrations determined using the Hybrid Single Particle Lagrangian Integrated Trajectory (HYSPLIT) model (Stein et al. 2015). A second neural network is trained to detect SO_2-rich ash. For this latter neural network the authors report a correct identification of SO_2-rich ash in 99.7% or cases, with a false detection rate of 1.6% and a missed detection rate of 0.3%. For the general ash detection network, a correct detection rate of 93.2% is reported, with false detection rate of 12.3% and a missed detection rate of 6.8%.

4.5.8 Neuro-Variational Retrievals

As a last example of application of neural networks to aerosol retrievals we will discuss the so-called "neuro-variational" retrieval schemes (Jamet et al. 2005). In these retrieval schemes, neural networks are not used as stand-alone algorithms, but they are used as part of a more general retrieval architecture. In some of the cases we will review, the neural network is used to build an approximate forward model, that is later minimized through conventional, iterative techniques. In other cases the neural network provides a first guess for an iterative algorithm.

The idea of inverting a neural-network-based radiative transfer model through iterative techniques has been first proposed by Jamet et al. (2005) for the joint retrieval of atmospheric and ocean parameters from SeaWiFS measurements. The authors trained three multilayer perceptron neural networks:

1. a neural network to estimate the aerosol contribution to the top-of-atmosphere reflectance;
2. a network to estimate the diffuse transmittance of the atmosphere;
3. a network to estimate the water-leaving reflectance.

The aerosol properties were described by means of the AOT at 865 nm, the exponent of a power-law size distribution and the complex refractive index, whereas the ocean was described by means of the pigment concentration and a scattering parameter. Solar and viewing zenith angles and relative azimuth angles were used as additional inputs for the neural network. The results of the neural network simulations were deemed by the authors compatible with their purpose. Anyhow, the task of estimating the aerosol contribution to the TOA reflectance proved more difficult then the task of estimating the transmittance and the water-leaving reflectance. The results of the three neural networks were combined to produce an estimate of the total TOA reflectance minus the Rayleigh contribution. This was then used as forward model in the inversion of real SeaWiFS data. The retrievals were compared to operational SeaWiFS retrievals, and spatial patterns in the retrieved parameters as those produced by the operational algorithm were observed, but with a negative bias in the chlorophyll-a concentration. However, validation against in-situ chlorophyll-a measurements revealed that the chlorophyll concentrations retrieved by the neuro-variational scheme were more realistic than those of the SeaWiFS algorithm. A more systematic validation of the method against in-situ measurements is presented in Brajard et al. (2006b), where it is shown that the neuro-variational scheme gave smaller retrieval errors than the operational method for water-leaving reflectances at 443 and 490 nm and chlorophyll-a concentrations, and larger error for AOT at 865 nm. An update to the approach presented above is described in Brajard et al. (2006a), where broader variability ranges in some of the aerosol parameters are used to create the training set, with the main goal of improving the retrieval capabilities in cases of absorbing aerosols. The application of the updated method to Medium Resolution Imaging Spectroradiometer (MERIS) data is presented in Brajard et al. (2012). Diouf et al. (2013) further refined the method by performing the iterative retrieval including a first-guess provided by a modification of the SOM method by Niang et al. (2006) discussed in Sect. 4.5.6. The method – that the authors call SOM-NV (SOM-Neuro-variational) – was applied to SeaWiFS images and the AOT retrievals were validated against AERONET measurements at Dakar and Cape Verde, and a good agreement was found over a range of AOTs between 0 and 0.8, that goes beyond the maximum AOT at which the operational SeaWiFS product is defined (0.35). Similar patterns between SOM-NV and the operational SeaWiFS algorithm are reported in the retrieved chlorophyll-a.

An alternative type of neuro-variational algorithm is proposed in Di Noia et al. (2015). In this method, a multilayer perceptron neural network is trained to retrieve AOT at 550 nm, aerosol effective radius and complex refractive index (assumed wavelength-independent) for a fine and a coarse mode of an assumed bimodal log-normal size distribution, from ground-based measurements of skylight radiance and degree of linear polarization at multiple wavelengths and viewing angles. The result

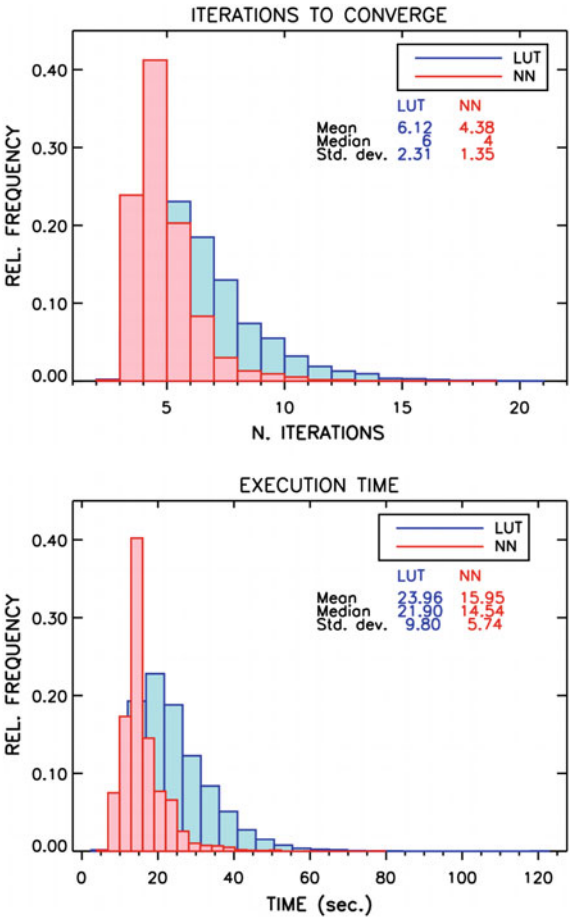

Fig. 4.4 Histograms of the number of iterations needed for convergence (above) and of the total retrieval time (below) for a Phillips–Tikhonov retrieval algorithm initialized with a LUT-based first guess (blue) and the same algorithm initialized using a neural network retrieval as first guess (red), obtained by a test over 5000 synthetic data. After Di Noia et al. (2015)

of the neural network retrieval is then used as first guess in an iterative algorithm based on Phillips–Tikhonov regularization as described in Hasekamp et al. (2011). The effect of using neural network retrievals as first guess in the iterative algorithm was studied by comparing the retrievals obtained using the modified method to those obtained using the original version of the iterative algorithm, in which the first guess is provided by a LUT. The comparison was performed on synthetic data and on real measurements produced by the groundSPEX spectropolarimeter (van Harten et al. 2014). The main observed results were a remarkable increase in the number of retrievals that passed the convergence criterion based on the goodness-of-fit χ^2 of the final retrieval (a retrieval was considered successful if the final χ^2 was below 2), as well as a decrease in the total retrieval time and in the number of iterations required to converge (Fig. 4.4).

The retrievals performed by applying the algorithm to groundSPEX measurements were compared to co-located AERONET measurements at Cabauw performed with

high temporal resolution over three days in July 2013 and one day in September 2013. The neuro-variational retrievals proved capable of following the daily cycles of most of the aerosol parameters with satisfactory accuracy, and an improved agreement with AERONET, compared to the original version of the algorithm, was observed in particular for the imaginary part of the refractive index. An inspection of the retrievals of this parameter produced by the neural network alone (that is, prior to the application of the iterative algorithm) revealed that the accuracy of the final retrievals was not substantially different from that of the neural network retrievals. An improvement in the accuracy of the AOT was observed, instead, as a result of applying the iterative retrieval after the neural network first guess. The extension of this approach to satellite spectropolarimetric measurements is currently under development.

4.6 Conclusions

In this paper we have discussed the application of machine learning algorithms, such as neural networks and support vector machines, to aerosol and cloud remote sensing problems. In the first part of the paper we have tried to summarize the theoretical foundations of neural networks and support vector machines, and to explain the conditions under which these can be applied to remote sensing problems. In the second part of the paper we have reviewed a number of works published in literature, that describe the application of machine learning methods to several remote sensing problems.

The first class of applications we have considered is the use of supervised neural networks to approximate radiative transfer calculations. Such an application is justified by the universal approximation properties of neural network, that is, their ability to approximate any continuous function on a compact set to an arbitrary accuracy. We have seen examples regarding the approximation of vector radiative transfer models, radiative transfer calculations in disperse media and 3D radiative transfer calculations with inhomogeneous clouds. We have discussed two problems regarding the approximation of forward models using neural networks: the difficulties with the approximation of forward models requiring a very large number of input parameters and the problem of approximating the derivative of a forward model with respect to input parameters, which is very important if the forward model has to be used in a retrieval algorithm.

After discussing the application of neural networks to forward modeling, we have examined their application to classification and retrieval problems in the realm of cloud and aerosol remote sensing. The applications we discussed include cloud detection in multispectral images, cloud type classification, retrieval of cloud properties, as well as aerosol typing and retrieval of aerosol microphysical properties. Most of the applications we have reviewed are based on multilayer perceptron networks, but we have also discussed a number of methods based on radial basis function networks, self-organizing maps and support vector machines. We have tried to emphasize that

a large variety of ways exist to apply machine learning algorithms to cloud and aerosol remote sensing. Some of the schemes we have reviewed derive information on aerosols and clouds from the spectral (and/or angular) properties of remote sensing signals, while other schemes make use of textural information, i.e. information regarding the spatial variability of these signals within an image. Spectral and textural information may be also combined in an integrated approach, and we have reviewed an example in which this is done (Miller and Emery 1997). Several methods are as well possible for generating training sets for machine learning algorithms. We have seen schemes entirely trained using synthetic data and schemes where coincident measurements of the input and target quantities of interest are used. In the final part of the paper we have emphasized that machine learning methods can also be combined with "classical" retrieval schemes, in so-called "neuro-variational" approaches, and we have reviewed a number of such approaches. In some cases, supervised neural networks are used as approximate forward models and inverted through iterative retrieval schemes. In other cases, supervised or unsupervised neural networks are used to generate a first guess for an iterative scheme employing a full radiative transfer model.

References

Aires F, Schmitt M, Chédin A, Scott N (1999) The "weight smoothing" regularization of MLP for Jacobian stabilization. IEEE Trans Neural Netw 10(6):1502–1510. https://doi.org/10.1109/72.809096

Alexander D (2013) Volcanic ash in the atmosphere and risks for civil aviation: A study in European crisis management. Int J Disaster Risk Sci 4(1):9–19. https://doi.org/10.1007/s13753-013-0003-0

Bartlett PL (1997) For valid generalization, the size of the weights is more important than the size of the network. In: Jordan MI, Petsche T (eds) Advances in neural information processing systems 9. Proceedings of neural information processing systems (NIPS) conference 1996, MIT Press, Cambridge, MA, United States, pp 134–140

Bellman R (1961) Adaptive control processes. A guided tour. Princeton University Press, Princeton

Berdnik VV, Gallyamova GI (2012) Calculation of the reflectance and transmittance of a disperse layer by the neural network method. Opt Spectrosc 112(4):618–623. https://doi.org/10.1134/S0030400X12040078

Berdnik VV, Loiko VA (2009) Retrieval of size and refractive index of spherical particles by multiangle light scattering: neural network method application. Appl Opt 48(32):6178–6187. https://doi.org/10.1364/AO.48.006178

Berdnik VV, Loiko VA (2016) Neural networks for particle parameter retrieval by multi-angle light scattering. In: Kokhanovsky A (ed) Light scattering reviews 10: light scattering and radiative transfer. Springer, Berlin, pp 291–340. https://doi.org/10.1007/978-3-662-46762-6_7

Berdnik VV, Mukhamedyarov RD, Loiko VA (2004) Sizing of soft spheroidal particles by multiangle scattered light intensity data: application of neural networks. J Quant Spectrosc Radiat Transf 89:279–289. https://doi.org/10.1016/j.jqsrt.2004.05.028

Bishop CM (1995a) Neural networks for pattern recognition. Oxford University Press, New York

Bishop CM (1995b) Training with noise is equivalent to Tikhonov regularization. Neural Comput 7(1):108–116. https://doi.org/10.1162/neco.1995.7.1.108

Bös S, Amari S (1999) Annealed online learning in multilayer neural networks. In: Saad D (ed) On-line learning in neural networks. Cambridge University Press, New York, pp 209–229. https://doi.org/10.1017/CBO9780511569920.011

Boucher O, Randall D, Artaxo P, Bretherton C, Feingold G, Forster P, Kerminen V-M, Kondo Y, Liao H, Lohmann U, Rasch P, Satheesh SK, Sherwood S, Stevens B, Zhang XY (2013) Clouds and aerosols. In: Stocker TF, Qin D, Plattner K-F, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds) Climate change 2013: the physical science basis. contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, pp 571–657

Brajard J, Jamet C, Moulin C, Thiria S (2006a) Use of a neuro-variational inversion for retrieving oceanic and atmospheric constituents from satellite ocean colour sensor: application to absorbing aerosols. Neural Netw. 19:178–185. https://doi.org/10.1016/j.neunet.2006.01.015

Brajard J, Jamet C, Moulin C, Thiria S (2006b) Validation of a neuro-variational inversion of ocean colour images. Adv Space Res 38:2169–2175. https://doi.org/10.1016/j.asr.2006.03.039

Brajard J, Santer R, Crépon M, Thiria S (2012) Atmospheric correction of MERIS data for case-2 waters using a neuro-variational inversion. Remote Sens Environ 126:51–61. https://doi.org/10.1016/j.rse.2012.07.004

Chevallier F, Chéruy F, Scott NA, Chédin A (1998) A neural network approach for a fast and accurate computation of a longwave radiative budget. J Appl Meteorol 37:1385–1397. https://doi.org/10.1175/1520-0450(1998)037<1385:ANNAFA>2.0.CO;2

Cornet C, Isaka H, Guillemet B, Szczap F (2004) Neural network retrieval of cloud parameters of inhomogeneous clouds from multispectral and multiscale radiance data: feasibility study. J Geophys Res 109(D12):203. https://doi.org/10.1029/2003JD004186

Cornet C, Buriez J-C, Riédi J, Isaka H, Guillemet B (2005) Case study of inhomogeneous cloud parameter retrieval from MODIS data. Geophys Res Lett 32(D13):807. https://doi.org/10.1029/2005GL022791

Cover TM (1965) Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. IEEE Trans Electron Comput EC-14:326–334. https://doi.org/10.1109/pgec.1965.264137

Cybenko G (1989) Approximation by superpositions of a sigmoidal function. Math Control Signal Syst 2(4):303–314. https://doi.org/10.1007/BF02551274

de Haan JF, Bosma PB, Hovenier JW (1987) The adding method for multiple scattering calculations of polarized light. Astron Astrophys 183:371–391

Di Noia A, Hasekamp OP, van Harten G, Rietjens JHH, Smit JM, Snik F, Henzing JS, de Boer J, Keller CU, Volten H (2015) Use of neural networks in ground-based aerosol retrievals from multi-angle spectropolarimetric observations. Atmos Meas Tech 8:281–299. https://doi.org/10.5194/amt-8-281-2015

Diouf D, Niang A, Brajard J, Crépon M, Thiria S (2013) Retrieving aerosol characteristics and sea-surface chlorophyll from satellite ocean color multi-spectral sensors using a neural-variational method. Remote Sens Environ 130:74–86. https://doi.org/10.1016/j.rse.2012.11.002

Dubovik O, King MD (2000) A flexible inversion algorithm for retrieval of aerosol optical properties from Sun and sky radiance measurements. J Geophys Res 105(D16):20,673–20,796. https://doi.org/10.1029/2000JD900282

Dubovik O, Herman M, Holdak A, Lapyonok T, Tanré D, Deuzé JL, Ducos F, Sinyuk A, Lopatin A (2011) Statistically optimized inversion algorithm for enhanced retrieval of aerosol properties from spectral multi-angle polarimetric satellite observations. Atmos Meas Tech 4:975–1018. https://doi.org/10.5194/amt-4-975-2011

Faure T, Isaka H, Guillemet B (2001a) Mapping neural network computation of high-resolution radiant fluxes of inhomogeneous clouds. J Geophys Res 106(D14):14,961–14,973. https://doi.org/10.1029/2001JD900058

Faure T, Isaka H, Guillemet B (2001b) Neural network analysis of the radiative interaction between neighboring pixels in inhomogeneous clouds. J Geophys Res 106(D13):14,465–14,484. https://doi.org/10.1029/2000JD900686

Faure T, Isaka H, Guillemet B (2001c) Neural network retrieval of cloud parameters of inhomogeneous and fractional clouds: feasibility study. Remote Sens Environ 77:123–128. https://doi.org/10.1016/S0034-4257(01)00199-7

Faure T, Isaka H, Guillemet B (2002) Neural network retrieval of cloud parameters from high-resolution multispectral radiometric data: a feasibility study. Remote Sens Environ 80:285–296. https://doi.org/10.1016/S0034-4257(01)00310-8

Fisher RA (1922) On the mathematical foundations of theoretical statistics. Philos Trans R Soc Lond A 222:309–368. https://doi.org/10.1098/rsta.1922.0009

Funahashi K-I (1989) On the approximate realization of continuous mappings by neural networks. Neural Netw 2(3):183–192. https://doi.org/10.1016/0893-6080(89)90003-8

Gray TM, Bennartz R (2015) Automatic volcanic ash detection from MODIS observations using a back-propagation neural network. Atmos Meas Tech 8:5089–5097. https://doi.org/10.5194/amt-8-5089-2015

Hagan MT, Menhaj MB (1994) Training feedforward networks with the Marquardt algorithm. IEEE Trans Neural Netw 5(6):989–993. https://doi.org/10.1109/72.329697

Han B, Vucetic S, Braverman A, Obradovic Z (2006) A statistical complement to deterministic algorithms for the retrieval of aerosol optical thickness from radiance data. Eng Appl Artif Intel 19:787–795. https://doi.org/10.1016/j.engappai.2006.05.009

Haralick RM, Shanmugam K, Dinstein I (1973) Textural features for image classification. IEEE Trans Syst Man Cybern SMC-3(6):610–621. https://doi.org/10.1109/TSMC.1973.4309314

Hasekamp OP, Litvinov P, Butz A (2011) Aerosol properties over the ocean from PARASOL multiangle photopolarimetric measurements. J Geophys Res 116(D14):204. https://doi.org/10.1029/2010JD015469

Haykin S (1999) Neural networks: a comprehensive foundation. Prentice Hall, Upper Saddle River

Hertz J, Krogh A, Palmer RG (1991) Introduction to the theory of neural computation. Westview Press, Boulder

Holzer-Popp T, Schroedter-Homscheidt M, Breitkreutz H, Martynenko D, Klüser L (2008) Improvements of synergetic aerosol retrieval for ENVISAT. Atmos Chem Phys 8:7651–7672. https://doi.org/10.5194/acp-8-7651-2008

Hornik K (1991) Approximation capabilities of multilayer feedforward networks. Neural Netw 4(2):251–257. https://doi.org/10.1016/0893-6080(91)90009-T

Hornik K, Stinchcombe M, White H (1989) Multilayer feedforward networks are universal approximators. Neural Netw 2(5):359–366. https://doi.org/10.1016/0893-6080(89)90020-8

Hornik K, Stinchcombe M, White H (1990) Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. Neural Netw 3(5):551–560. https://doi.org/10.1016/0893-6080(90)90005-6

Huang G-B, Zhu Q-Y, Siew C-K (2005) Extreme learning machine: theory and applications. Neurocomputing 70(1):489–501. https://doi.org/10.1016/j.neucom.2005.12.126

Ishimaru A, Marks RJ II, Tsang L, Lam CM, Park DC, Kitamura S (1990) Particle-size distribution determination using optical sensing and neural networks. Opt Lett 15(21):1221–1223. https://doi.org/10.1364/OL.15.001221

Jamet C, Thiria S, Moulin C, Crépon M (2005) Use of a neurovariational inversion for retrieving oceanic and atmospheric constituents from ocean color imagery: a feasibility study. J Atmos Ocean Technol 22:460–475. https://doi.org/10.1175/JTECH1688.1

Jang J-D, Viau AA, Anctil F, Bartholomé, (2006) Neural network application for cloud detection in SPOT VEGETATION images. Int J Remote Sens 27(4):719–736. https://doi.org/10.1080/01431160500106892

Kahn R, Banerjee D, McDonald D (2001) Sensitivity of multiangle imaging to natural mixtures of aerosols over ocean. J Geophys Res 106(D16):18,219–18,238. https://doi.org/10.1029/2000JD900497

Kecman V (2001) Learning and soft computing. Support vector machines, neural networks and fuzzy logic model. MIT Press, Cambridge

Key J, Maslanik JA, Schweiger AJ (1989) Classification of merged AVHRR and SMMR Arctic data with neural networks. Photogramm Eng Remote Sens 55(9):1331–1338

Kohonen T (1982) Self-organized formation of topologically correct feature maps. Biol Cybern 43:59–69. https://doi.org/10.1007/bf00337288

Kox S, Bugliaro L, Ostler A (2014) Retrieval of cirrus cloud optical thickness and top altitude from geostationary remote sensing. Atmos Meas Tech 7:3233–3246. https://doi.org/10.5194/amt-7-3233-2014

Krasnopolsky VM (2007) Reducing uncertainties in neural network Jacobians and improving accuracy of neural network emulations with NN ensemble approaches. Neural Netw. 20(4):454–461. https://doi.org/10.1016/j.neunet.2007.04.008

Lee J, Weger RC, Sengupta SK, Welch RM (1990) A neural network approach to cloud classification. IEEE Trans Geosci Remote Sens 28(5):846–855. https://doi.org/10.1109/36.58972

Lee Y, Lin Y, Wahba G (2004a) Multicategory support vector machines. Theory and application to the classification of microarray data and satellite radiance data. J Am Stat Assoc 99:67–81. https://doi.org/10.1198/016214504000000098

Lee Y, Wahba G, Ackerman SA (2004b) Cloud classification of satellite radiance data by multicategory support vector machines. J Atmos Ocean Tech 21:159–169. https://doi.org/10.1175/1520-0426(2004)021<0159:CCOSRD>2.0.CO;2

Leshno M, Lin VY, Pinkus A, Schocken S (1993) Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. Neural Netw 6(6):861–867. https://doi.org/10.1016/S0893-6080(05)80131-5

Levy RC, Remer LA, Mattoo S, Vermote EF, Kaufman YJ (2007) Second-generation operational algorithm: Retrieval of aerosol properties over land from inversion of Moderate Resolution Imaging Spectroradiometer spectral reflectance. J Geophys Res 112(D13):211. https://doi.org/10.1029/2006JD007811

Loyola DG, Thomas W, Livschitz Y, Ruppert T, Albert P, Hollmann R (2007) Cloud properties derived from GOME/ERS-2 backscatter data for trace gas retrieval. IEEE Trans Geosci Remote Sens 45(9):2747–2758. https://doi.org/10.1109/TGRS.2007.901043

Loyola DG, Thomas W, Spurr RJD, Mayer B (2010) Global patterns in daytime cloud properties derived from GOME backscatter UV-VIS measurements. Int J Remote Sens 31(16):4295–4318. https://doi.org/10.1080/01431160903246741

Marchuk GI, Mikhailov GA, Nazaraliev MA, Darbinjan RA, Kargin BA, Elepov BS (1980) The Monte Carlo methods in atmospheric optics. Springer, Berlin

Marshak A, Davis A, Cahalan RF, Wiscombe W (1998) Nonlocal independent pixel approximation: direct and inverse problems. IEEE Trans Geosci Remote Sens 36(1):192–205. https://doi.org/10.1109/36.655329

Mazzoni D, Garay MJ, Davies R, Nelson D (2007) An operational MISR pixel classifier using support vector machines. Remote Sens Environ 107:149–158. https://doi.org/10.1016/j.rse.2006.06.021

McCulloch W, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys 5(4):115–133. https://doi.org/10.1007/BF02478259

Mercer J (1909) Functions of positive and negative type, and their connection with the theory of integral equations. Phil. Trans. Royal Soc. A 209:415–446. https://doi.org/10.1098/rsta.1909.0016

Miller SW, Emery WJ (1997) An automatic neural network cloud classifier for use over land and ocean surfaces. J Appl Meteorol 36:1346–1362. https://doi.org/10.1175/1520-0450(1997)036<1346:AANNCC>2.0.CO;2

Møller MF (1993) A scaled conjugate gradient algorithm for fast supervised learning. Neural Netw 6(4):525–533. https://doi.org/10.1016/S0893-6080(05)80056-5

Niang A, Gross L, Thiria S, Badran F, Moulin C (2003) Automatic neural classification of ocean colour reflectance spectra at the top of the atmosphere with introduction of expert knowledge. Remote Sens Environ 86:257–271. https://doi.org/10.1016/S0034-4257(03)00113-5

Niang A, Badran F, Moulin C, Crépon M, Thiria S (2006) Retrieval of aerosol type and optical thickness over the Mediterranean from SeaWiFS images using an automatic neural classification method. Remote Sens Environ 100:82–94. https://doi.org/10.1016/j.rse.2005.10.005

Nocedal J, Wright SJ (1999) Numerical optimization. Springer series in operations research. Springer, New York

Park J, Sandberg IW (1991) Universal approximation using radial-basis-function networks. Neural Comput 3(2):246–257. https://doi.org/10.1162/neco.1991.3.2.246

Park J, Sandberg IW (1993) Approximation and radial-basis-function networks. Neural Comput 5(2):305–316. https://doi.org/10.1162/neco.1993.5.2.305

Phillips DL (1962) A technique for the numerical solution of certain integral equations of the first kind. J Assoc Comput Mach 9(1):84–97. https://doi.org/10.1145/321105.321114

Picchiani M, Chini M, Corradini S, Merucci L, Sellitto P, Del Frate F, Stramondo S (2011) Volcanic ash detection and retrievals using MODIS data by means of neural networks. Atmos Meas Tech 4:2619–2631. https://doi.org/10.5194/amt-4-2619-2011

Piscini A, Picchiani M, Chini M, Corradini S, Merucci L, Del Frate F, Stramondo S (2014) A neural network approach for the simultaneous retrieval of volcanic ash parameters and SO_2 using MODIS data. Atmos Meas Tech 7:4023–4027. https://doi.org/10.5194/amt-7-4023-2014

Prata AJ (1989a) Infrared radiative transfer calculations for volcanic ash clouds. Geophys Res Lett 16(11):1293–1296. https://doi.org/10.1029/GL016i011p01293

Prata AJ (1989b) Observations of volcanic ash clouds in the 10–12 μm window using AVHRR/2 data. Int J Remote Sens 10(4):751–761. https://doi.org/10.1080/01431168908903916

Prata AJ, Grant IF (2001) Retrieval of microphysical and morphological properties of volcanic ash plumes from satellite data: application to Mt Ruapehu, New Zealand. Q J R Meteorol Soc 127:2153–2179. https://doi.org/10.1002/qj.49712757615

Radosavljevic V, Vucetic S, Obradovic Z (2010) A data-mining technique for aerosol retrieval across multiple accuracy measures. IEEE Geosci Remote Sens Lett 7(2):411–415. https://doi.org/10.1109/LGRS.2009.2037720

Riedmiller M (1994) Advanced supervised learning in multi-layer perceptrons - from backpropagation to adaptive learning algorithms. Comput Stand Interfaces 16:265–278. https://doi.org/10.1016/0920-5489(94)90017-5

Ristovski K, Vucetic S, Obradovic Z (2012) Uncertainty analysis of neural-network-based aerosol retrieval. IEEE Trans Geosci Remote Sens 50(2):409–414. https://doi.org/10.1109/TGRS.2011.2166120

Robert CP (2007) The Bayesian choice. From decision-theoretic foundations to computational implementation. Springer, New York. https://doi.org/10.1007/0-387-71599-1

Rodgers CD (2000) Inverse methods for atmospheric sounding: theory and practice. World Scientific, London

Rosenblatt F (1958) The perceptron: a probabilistic model for information storage and organization in the brain. Psychol Rev 65:386–408. https://doi.org/10.1037/h0042519

Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. Nature 323:533–536. https://doi.org/10.1038/323533a0

Russell PB, Bergstrom RW, Shinozuka Y, Clarke AD, De Carlo PF, Jimenez JL, Livingston JM, Redemann J, Dubovik O, Strawa A (2010) Absorption Angstrom exponent in AERONET and related data as an indicator of aerosol composition. Atmos Chem Phys 10:1155–1169. https://doi.org/10.5194/acp-10-1155-2010

Russell PB, Kacenelenbogen M, Livingston JM, Hasekamp OP, Burton SP, Schuster GL, Johnson MS, Knobelspiesse KD, Redemann J, Ramachandran S, Holben B (2013) A multiparameter aerosol classification method and its application to retrievals from spaceborne polarimetry. J Geophys Res 119(16):9838–9863. https://doi.org/10.1002/2013JD021411

Saponaro G, Kolmonen P, Karhunen J, Tamminen J, de Leeuw G (2013) A neural network algorithm for cloud fraction estimation using NASA-Aura OMI VIS radiance measurements. Atmos Meas Tech 6:2301–2309. https://doi.org/10.5194/amt-6-2301-2013

Schölkopf B, Smola AJ (2002) Learning with Kernels. Support vector machines, regularization, optimization and beyond. MIT Press, Cambridge

Shettle EP, Fenn RW (1979) Models for the aerosols of the lower atmosphere and the effects of humidity variations on their optical properties. Environmental research paper, vol 676. Air Force Laboratory, Hanscom AFB, MA, United States

Spurr RJD, Kurosu TP, Chance KV (2001) A linearized discrete ordinate radiative transfer model for atmospheric remote-sensing retrieval. J Quant Spectrosc Rad Transf. 68:689–735. https://doi.org/10.1016/S0022-4073(00)00055-8

Stein AF, Draxler RR, Rolph GD, Stunder BJB, Cohen MD, Ngan F (2015) NOAA's HYSPLIT atmospheric transport and dispersion modeling system. Bull Am Meteorol Soc 96:2059–2077. https://doi.org/10.1175/BAMS-D-14-00110.1

Taravat A, Del Frate F, Cornaro C, Vergari S (2015) Neural networks and support vector machine algorithms for automatic cloud classification of whole-sky ground-based images. IEEE Geosci Remote Sens Lett 12(3):666–670. https://doi.org/10.1109/LGRS.2014.2356616

Taylor M, Kazadzis S, Tsekeri A, Gkikas A, Amiridis V (2014) Satellite retrieval of aerosol microphysical and optical parameters using neural networks: a new methodology applied to the Sahara desert dust peak. Atmos Meas Tech 7:3151–3175. https://doi.org/10.5194/amt-7-3151-2014

Tikhonov AN, Arsenin VY (1977) Solutions of Ill-posed problems. Winston & Sons, Washington

Tuia D, Verrelst J, Pérez-Cruz F, Camps-Valls G (2011) Multioutput support vector regression for remote sensing biophysical parameter estimation. IEEE Geosci Remote Sens Lett 8(4):804–808. https://doi.org/10.1109/LGRS.2011.2109934

Ulanowski Z, Wang Z, Kaye P, Ludlow IK (1998) Application of neural networks to the inverse light scattering problem for spheres. Appl Opt 37(18):4027–4033. https://doi.org/10.1364/AO.37.004027

van Harten G, de Boer J, Rietjens JHH, Di Noia A, Snik F, Volten H, Smit JM, Hasekamp OP, Henzing JS, Keller CU (2014) Atmospheric aerosol characterization with a ground-based SPEX spectropolarimetric instrument. Atmos Meas Tech 7:4341–4351. https://doi.org/10.5194/amt-7-4341-2014

Vapnik VN (1995) The nature of statistical learning theory. Springer, New York

Vapnik VN (1998) Statistical learning theory. Wiley, New York

Vucetic S, Han B, Mi W, Obradovic Z (2008) A data-mining approach for the validation of aerosol retrievals. IEEE Geosci Remote Sens Lett 5(1):113–117. https://doi.org/10.1109/LGRS.2007.912725

Wang Z, Ulanowski Z, Kaye PH (1999) On solving the inverse scattering problem with RBF neural networks: Noise-free case. Neural Comput Appl 8:177–186. https://doi.org/10.1007/s005210050019

Werbos PJ (1974) Beyond regression: new tools for prediction and analysis in the behavioral sciences. PhD thesis, Harvard University, Cambridge, MA, United States

Yhann SR, Simpson JJ (1995) Application of neural networks to AVHRR cloud segmentation. IEEE Trans Geosci Remote Sens 33(3):590–604. https://doi.org/10.1109/36.387575

Chapter 5
Stereogrammetric Shapes of Mineral Dust Particles

Olli Jokinen, Hannakaisa Lindqvist, Konrad Kandler, Osku Kemppinen and Timo Nousiainen

5.1 Introduction

This chapter considers shape models which are retrieved from scanning electron microscope (SEM) images of real dust particles. These models describe the real shape of the particle as measured from the images with automated image processing techniques. A general description of methods for retrieving the shape from SEM stereo images is provided together with our own developments. The current implementation of our shape retrieval process starts with finding a sparse set of corresponding points between one image and another image which has been tilted with respect to the first one. The sparse correspondences are used to refine the image orientations and to estimate the 3-D object points in a bundle adjustment. The correspondences are densified using least squares image matching techniques resulting in a dense surface model. A volumetric model is obtained by adding the backside of the particle by mirroring.

O. Jokinen (✉)
Antreantie 19 C, 02140, Espoo, Finland
e-mail: olli.t.jokinen@gmail.com

H. Lindqvist · O. Kemppinen · T. Nousiainen
Finnish Meteorological Institute, P.O. Box 503, 00101 Helsinki, Finland
e-mail: hannakaisa.lindqvist@fmi.fi

O. Kemppinen
e-mail: okemppin@phys.ksu.edu

T. Nousiainen
e-mail: timo.nousiainen@fmi.fi

K. Kandler
Institut fuer Angewandte Geowissenschaften, Technische Universität Darmstadt, Schnittspahnstr 9, 64287 Darmstadt, Germany
e-mail: kandler@geo.tu-darmstadt.de

A. Kokhanovsky (ed.), *Springer Series in Light Scattering*, Springer Series in Light Scattering, https://doi.org/10.1007/978-3-319-70796-9_5

From light scattering perspective, the use of model shapes derived directly from those of real particles presents a paradigm change. For example, in Earth sciences, where target particles would be readily accessible for detailed analysis, all climate models and remote sensing applications employ simplistic model shapes such as spheres, spheroids, ellipsoids or blocks of cubes for atmospheric dust particles in their radiative treatments. More sophisticated shape models, such as deformed spheroids (Gasteiger et al. 2011), agglomerates (Zubko et al. 2013), concave fractal polyhedra (Liu et al. 2012), spatial Poisson-Voronoi tessellation (Ishimoto et al. 2010) or random flakes (Nousiainen et al. 2009) have been considered in specific modelling studies, but these shape models are still descriptive: they aim at mimicking some morphological features of real dust particles qualitatively, rather than being derived directly from the real, observed shapes. The Gaussian random sphere (GRS) model by Muinonen et al. (1996) is a notable exception, as its shape statistics can, and indeed have been derived directly from observed silhouettes of real dust particles (e.g., Nousiainen et al. 2003; Veihelmann et al. 2006; Muñoz et al. 2007). Even with GRS, however, the link to real shapes is only statistical and does not necessarily result in realistically shaped model particles. The GRS model is also limited to shapes that can be expressed using a spherical harmonics expansion. All these modeling approaches have been recently reviewed by Nousiainen and Kandler (2015).

The use of stereogrammetry to derive the real shapes of mineral dust particles offers one route for assessing the performance of the presently used modeling approaches. Indeed, several such comparisons are presented here. Further investigations include effects of particle surface roughness on the scattering properties. Obtaining the real shapes of dust particles is a key ingredient in computing the single-scattering properties of real dust particles accurately and consistently. However, real dust particles also are inhomogeneous and may possess morphological features that cannot be derived through stereogrammetry. Such features include, for example, porous cavities and other internal structures that have been shown to be present in real dust particles (Jeong and Nousiainen 2014). Derivation of the model shapes directly from observed individual dust particles opens a new frontier with great potential in the optical modeling, but its full exploitation requires much additional research. The chapter thus concludes with some ideas for further improvement of the shape retrieval algorithms in view of the light scattering computations.

5.2 SEM Image Capturing

Stereogrammetric shape retrieval is based on imaging the particle from at least two different viewpoints with an appropriate imaging system. Since atmospheric dust particles are of the size of a few microns and we want to observe surface details of the nanometer scale, the resolution of an optical microscope is not enough but a scanning electron microscope needs to be applied. A suitable resolution, reaching 2 nm at high magnifications, and a large depth of field make the SEM advantageous for imaging dust particles. An atomic force microscope could be a viable alternative and for that,

the reader is referred to, e.g., Vijendran et al. (2007) and Woodward (2014). The imaging principle of a SEM is based on focusing a beam of electrons to image the particle and to obtain information about its structure and composition. To improve the image quality, the particle samples can be sputter-coated with thin conductive layers (e.g. gold, platinum, or carbon). For example, gold coating produces fine structures, which may help the matching algorithms on homogeneous areas. Secondary (SE) and backscatter (BSE) electron images are typically collected. The SE signal amplitude is a function of local surface orientation with respect to the incident beam while the BSE signal is mainly correlated with the local average atomic weight (Goldstein et al. 2003). The acceleration voltage affects how deep the primary beam electrons penetrate into the particle and it has also an influence on the visibility of surface details in the images (Richards et al. 2000). Other factors affecting the image quality include working distance, scan rate, spot size, magnification, and aperture (Goodhew et al. 2001). The influence of these on the 3-D reconstruction is analyzed by Carli et al. (2011).

For depth recovery, at least two SEM images must be captured. The stage is tilted between the image acquisitions in order to obtain a disparity between corresponding points which is different at different heights of the object surface. Moving the object only on a plane would lead to a uniform disparity in a parallel projection imaging geometry (Zolotukhin et al. 2013). The horizontal disparity is defined as a difference between the horizontal image coordinate of a point in one image and the horizontal image coordinate of its corresponding point in another tilted image. The vertical disparity is defined respectively. In eucentric tilting, the intersection of the primary electron beam with the specimen defines the center of tilting (Electron Microscopy Sciences 2016). The tilting can be, however, performed around an arbitrary line which is taken into account when the 3-D object is reconstructed from the corresponding image observations. The tilt angles can be computed from the images in an auto-calibration process (Danzl et al. 2007).

For magnifications larger than 1000×, the imaging geometry can be well approximated by a parallel projection, where the projection center is at infinity and the image rays are parallel (Cornille et al. 2003). Accurate calibration of the SEM includes correcting for image distortions with a non-parametric model (Cornille et al. 2003). This model can be estimated from a pure translational image sequence followed by a sequence of arbitrary motions of a planar calibration target textured with a random speckle pattern. The depth coordinate can be calibrated with a multiple-step heights artefact (Carli 2010). An external calibration with an atomic force microscope is also possible.

5.3 Shape Retrieval

The geometrical representation of a stereogrammetric shape model is given by a set of voxels filling the volume of the particle. The voxel representation is derived from surface measurements consisting of a set of 3-D points on the object surface and

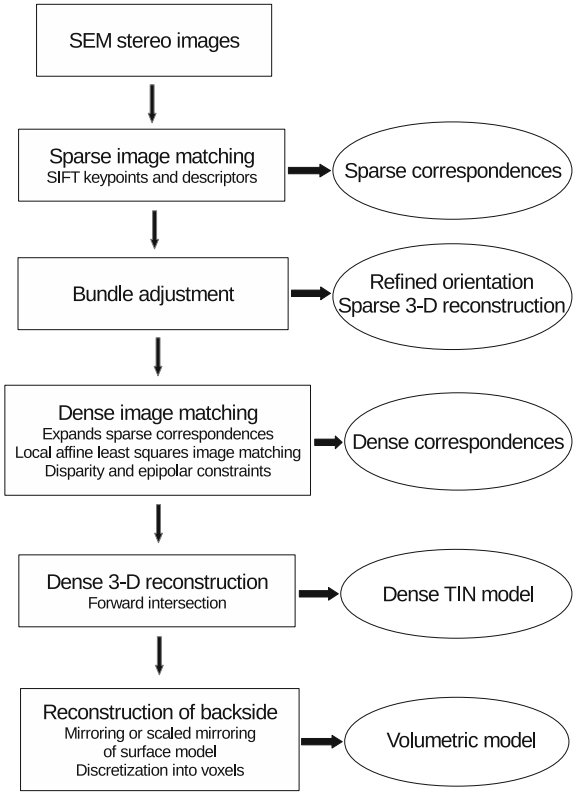

Fig. 5.1 Diagram of our shape retrieval process

triangulated into a mesh. For the determination of the 3-D points, the main effort is to determine corresponding points between the two SEM images forming a stereo pair. The denser the point set is, the more detailed is the retrieved shape model. Dense image matching aims at establishing correspondences down to pixel level resolution. Initial estimates for the dense correspondences are obtained from sparse image matching, where the matching is performed only in sparsely located interest points extracted from the images. These sparse correspondences are also used to refine the relative orientation of the images, which is essential for subsequent dense image matching and accurate 3-D point computation. The different steps of our shape retrieval algorithm are outlined in Fig. 5.1 and described in detail in the following with a review of other alternative approaches.

5.3.1 Sparse Image Matching

The task in sparse image matching is to find for a set of interest points in the first image the corresponding points in the second image. The interest points may be divided into two classes. The first one includes corner points such as detected by Harris detector (Harris and Stephens 1988) or Förstner operator (Förstner and Gülch 1987). For a corner point in the first image, the search for the corresponding corner point in the second image is typically based on finding the point which maximizes the normalized cross-correlation (NCC) computed between image windows centered at the corner point in the first image and each nearby corner point in the second image (Kammerud 2005). The NCC can be improved with stretch correlation (Lacey et al. 1998), which stretches and shears the second image window and thus accounts for surface rotations caused by changing views. The location of the corresponding point is refined to the sub-pixel accuracy by least squares matching within the image windows.

The second class of interest points consists of detectors and descriptors, which are invariant to various image transformations such as scale and rotation, and changes in illumination or 3-D viewpoint (Lowe 2004). The detectors aim to locate points with a local coordinate system which have high repeatability under the transformations while the descriptors should also be highly distinctive to describe the image content in the local neighborhood. The descriptors may be computed, e.g., from local image gradient orientations, magnitudes, and their histograms (Lowe 2004). The performance of various detectors and descriptors is evaluated by Mikolajczyk and Schmid (2005) and Mikolajczyk et al. (2005) including SIFT (Scale Invariant Feature Transform) (Lowe 2004), Harris-Affine and Hessian-Affine (Mikolajczyk and Schmid 2004), MSER (Maximally Stable Extremal Regions) (Matas et al. 2002), EBR (Edge-Based Regions) and IBR (Intensity Extrema-Based Regions) (Tuytelaars and Van Gool 2004), and salient regions (Kadir et al. 2004). There exists also a fast version of SIFT called SURF (Speeded-Up Robust Features) (Bay et al. 2008) utilizing integral images. The matching of interest points between two images is based on the descriptor vectors. The point with the most similar descriptor is the correct match. In order to improve the matching, the images may be divided into sub-images and the matching performed within corresponding sub-images. The positions of sub-images in the second image should be shifted according to rough horizontal and vertical disparities, which can be single values for the whole image visually estimated from the images (Lindqvist et al. 2014). SIFT keypoints have been applied to SEM images by Lindqvist et al. (2014) and Zolotukhin et al. (2013) and SURF keypoints by Roy et al. (2012).

5.3.2 Image Orientation and Sparse Reconstruction

In parallel projection geometry, the measured image coordinates (x_{nk}, y_{nk}) of point k in image n in metric units are related to the 3-D object coordinates (X_k, Y_k, Z_k) by (Hemmleb et al. 1996)

$$\begin{pmatrix} x_{nk} \\ y_{nk} \end{pmatrix} + \mathbf{v}_{nk} = R_n^{12} \begin{pmatrix} X_k - X_{0n} \\ Y_k - Y_{0n} \\ Z_k \end{pmatrix} \tag{5.1}$$

where R_n^{12} denotes the first two rows of a rotation matrix R_n, $\mathbf{t}_n = [X_{0n}\, Y_{0n}]^T$ is a translation in the XY-plane, and $\mathbf{v}_{nk}$ is a residual vector related to the image observations. The metric image coordinates are obtained from the measured pixel coordinates (I_{nk}, J_{nk}) according to $x_{nk} = s(J_{nk} - j_0)$, $y_{nk} = s(i_0 - I_{nk})$, where s is a known scale provided with the SEM image and (i_0, j_0) is the center point of the SEM image.

The exterior orientations of the SEM images given by R_n and $\mathbf{t}_n$ and the 3-D coordinates of the object points corresponding to the matched interest points are solved in a bundle adjustment. The bundle adjustment requires initial estimates for the unknown parameters. The rotation matrices are parametrized by three angles describing rotations around the coordinate axes. Assuming that the tilting of the specimen stage occurs around a line approximately parallel to the Y-axis, the tilt angle of the SEM image provides an initial estimate for the rotation angle around this axis and the other two angles are initialized with zeros. Since the image coordinates have been centered to the middle of the image, where the rotation center is often located nearby, the translation parameters are also given zero initial values. There exist several methods to compute the 3-D coordinates from matched image points, which have different assumptions on the imaging geometry such as eucentric tilting (Piazzesi 1973; Burkhardt 1981) or that the tilting occurs around a line parallel to one of the image coordinate axes (Themelis et al. 1990). The latter one is adopted here to give initial estimates for the object points according to

$$X_k^0 = (-x_{1k} \sin\phi_2 + x_{2k} \sin\phi_1)/\sin(\phi_1 - \phi_2) \tag{5.2}$$

$$Y_k^0 = (y_{1k} + y_{2k})/2 \tag{5.3}$$

$$Z_k^0 = (-x_{1k} \cos\phi_2 + x_{2k} \cos\phi_1)/\sin(\phi_1 - \phi_2) \tag{5.4}$$

where ϕ_1 and ϕ_2 are the tilt angles around the Y-axis of the specimen stage of the first and second SEM image, respectively.

Let the components of the right hand side of (5.1) be denoted by x'_{nk} and y'_{nk}. In bundle adjustment, the merit function to be minimized with respect to the orientation parameters and object coordinates is given by

$$f_1 = \sum_{k=1}^{K} \sum_{n=1}^{2} w_k [(x_{nk} - x'_{nk})^2 + (y_{nk} - y'_{nk})^2] \tag{5.5}$$

where w_k are weights used to discard incompatible corresponding points. The weight w_k equals one if all the four residuals related to point k in x and y directions in both images are below an adaptive threshold getting tighter as the iteration proceeds according to a scheme based on the statistical distribution of all the residuals below the current threshold similarly as in (Zhang 1994). Otherwise w_k equals zero. The nonlinear least squares minimization problem is solved using the Levenberg-Marquardt algorithm. In order to cope with the datum deficiency, the orientation of the first image is kept fixed so that the algorithm results in refined exterior orientation of the second image and sparse 3-D reconstruction of compatible corresponding points.

5.3.3 Dense Image Matching

Early methods (Kammerud 2005; Lindqvist et al. 2014) for densifying the correspondences between the SEM images relied on edge features typically extracted with the Canny detector (Canny 1986). The edge areas usually contain enough texture to enable NCC or least squares based image matching. An initial estimate for the corresponding point can be interpolated from the sparse correspondences at the interest points (Lindqvist et al. 2014). A dense disparity map for the whole image may be produced by fitting piecewise polynomials to the disparity values established at the edge pixels (Lacey et al. 1998). The accuracy of such an interpolated disparity map is however limited. Better results are obtained when the horizontal and vertical disparities are determined at every pixel using image matching techniques. This is called dense image matching, which is indeed possible in textured areas of the images.

Methods for dense image matching include global and local ones. The global methods typically minimize an energy function, which sums the pixel-wise matching costs and contains a global smoothness term supporting piecewise smooth disparity selection (Hirschmüller 2008). A multi-resolution approach with a pixel-wise disparity search space facilitates and speeds up the matching (Rothermel et al. 2012). The local methods are based on comparing the intensity values within local image windows with possible constraints on the change of disparity from pixel to pixel. The epipolar constraint reduces the search for the corresponding point to a 1-D problem. The images may be also rectified to the normal case (Morgan et al. 2004) so that the corresponding point in the second image is located on the same row or column as the point in the first image. For recent state-of-the-art and comparison of several dense image matching methods, the reader is referred to Cavegn et al. (2014), Haala (2013), Remondino et al. (2014), and Remondino and Zhang (2006). Also, commercial software exists for dense image matching of SEM images (Electron Microscopy

Sciences 2016; Lucideon 2016; Zeiss 2016). In the following, a local method based on expanding the correspondences is proposed as inspired by the work of Furukawa and Ponce (2010), Kannala and Brandt (2007), and Ylimäki et al. (2012).

The main idea here is thus to gradually expand the correspondences from the interest points, which were compatible in the bundle adjustment, to their neighborhoods until the whole image has been filled up. Since a semi-dense disparity map may be enough for the light scattering computations, the disparities are determined in a grid with a spacing of m pixels, where $m \geq 1$. The locations of the interest points in the first image are rounded to the nearest grid points, which yield the first seed points. The unique grid points in the 8-connected neighborhoods of all the seed points where no disparity has been yet estimated are selected. For each grid point (I_{1k}, J_{1k}) in the first image, an initial estimate for the corresponding point is given by $(I_{2k}^0, J_{2k}^0) = (I_{1k} - d_I^0, J_{1k} - d_J^0)$, where d_I^0, d_J^0 are initial estimates for the vertical and horizontal disparities at (I_{1k}, J_{1k}), interpolated from the disparities at the interest points. The interpolation is realized by triangulating the horizontal and vertical disparities at the interest points into TIN (triangulated irregular network) models and looking the values of the TIN surfaces at the grid point in question. The position of the corresponding point is refined by affine least squares image matching to be described below. The refined correspondence is accepted if both horizontal and vertical disparities in the new point differ from their values in the seed point less than a threshold T_d. This controls that the disparities change smoothly. Moreover, the distance of the corresponding point from the epipolar line is required to be lower than another threshold T_e. The epipolar line can be defined as a projection of the image ray of the point in the first image onto the second image. It is computed by selecting two 3-D points from the image ray of the point in the first image according to the inverse of (5.1), projecting these points onto the second image using (5.1), and computing a line through the two image points. Successfully matched correspondences which satisfy the disparity and epipolar constraints define new seed points in the first image. The process is then started again from expanding the seed points to their neighborhoods. Finally, after all the grid points have been tested, the grid points where the correspondence was rejected are reconsidered. These points are often located in areas where the disparities do not change smoothly like near edges and where the points that were started to be expanded from neighboring interest points meet. For each point to be reconsidered, its neighboring points, which have been originally expanded from other interest points than the point which is reconsidered, are selected as potential new seed points. The reconsidered correspondence is accepted if it satisfies the disparity constraint with the disparity at one or more of the new seed points and if it also satisfies the epipolar constraint. The remaining points are reconsidered again with a new initial estimate for the corresponding point given by the corresponding point of the neighboring seed point plus the difference between the point in the first image and its seed point. The corresponding point is thus initially estimated to have moved equally as the point in the first image. The procedure thereafter is as before, including affine least squares image matching and disparity and epipolar constraints. All the new seed points in the neighborhood are tested until a valid solution is found, if any.

The details of affine least squares image matching are as follows. Let (I_{1k}, J_{1k}) be a grid point in the first image, the approximate corresponding point of which is (I^0_{2k}, J^0_{2k}) in the second image. Let g_1 be a template image centered at (I_{1k}, J_{1k}) and denote the pixels of the template by lower case letters $(i_1, j_1) \in \Omega$, where Ω is the domain of the template of size p pixels squared in the first image. Note that the template image is a patch of the original image despite of the grid spacing m. The template image is matched against the second image g_2 within a search window centered at (I^0_{2k}, J^0_{2k}). An affine transformation which maps the coordinates of the template image to the coordinates of the search image is defined as

$$\begin{pmatrix} i_{12} \\ j_{12} \end{pmatrix} = \begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix} \begin{pmatrix} i_1 - I_{1k} + I^0_{2k} \\ j_1 - J_{1k} + J^0_{2k} \end{pmatrix} + \begin{pmatrix} a_5 \\ a_6 \end{pmatrix} \tag{5.6}$$

$$g_{12} = g_1 + a_7 \tag{5.7}$$

where $\mathbf{a} = [a_1 \ldots a_7]^T$ are parameters to be estimated. The gray levels g_2 are bilinearly interpolated to the intermediate pixels (i_{12}, j_{12}) and denoted by $\tilde{g}_2$. The merit function to be minimized is given by

$$f(\mathbf{a}) = \sum_{\Omega} [w(i_1, j_1)(g_{12}(i_1, j_1) - \tilde{g}_2(i_{12}, j_{12}))]^2 / \sum_{\Omega} (w(i_1, j_1))^2 \tag{5.8}$$

where $w(i_1, j_1)$ are weights used to discard incompatible pixels, where the gray levels g_{12} and $\tilde{g}_2$ differ more than an adaptive threshold getting tighter as the iteration proceeds according to a scheme proposed by Zhang (1994). The nonlinear minimization problem is solved using the Levenberg-Marquardt algorithm. Once the iteration has converged, the refined corresponding point is obtained from

$$\begin{pmatrix} I_{2k} \\ J_{2k} \end{pmatrix} = \begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix} \begin{pmatrix} I^0_{2k} \\ J^0_{2k} \end{pmatrix} + \begin{pmatrix} a_5 \\ a_6 \end{pmatrix} \tag{5.9}$$

5.3.4 Dense Reconstruction

A dense 3-D reconstruction is computed using forward intersection. For each pair of corresponding points (I_{1k}, J_{1k}) and (I_{2k}, J_{2k}) resulting from dense image matching, the image coordinates are first scaled to metric units (x_{1k}, y_{1k}) and (x_{2k}, y_{2k}). Equation (5.1) is then written for both points with previously estimated exterior orientations. The 3-D coordinates (X_k, Y_k, Z_k) are solved from these equations in the least squares sense. The 3-D coordinates of the interest points, which were estimated in sparse reconstruction, are added to the point set and the whole set is triangulated into a dense TIN model, which describes the object surface visible in both images.

5.3.5 *Volumetric Model*

The challenges in converting a triangulated surface into a closed volume are best depicted by the example in Fig. 5.2, top row, where the retrieved TIN surface of a dolomite particle is viewed from four sides. The network of points is dense at the top of the particle but some points are also retrieved from the edges and even from the plane where the particle is placed upon. The latter points are efficiently excluded by setting a minimum height h_0 for the approved points. A suitable value for h_0 is unique for each shape and is recommended to be evaluated visually. The visible boundary in BSE images can be used for additional reference. In Fig. 5.2, $h_0 = 1.0\,\mu\text{m}$, and the resulting shape after excluding the points below h_0 is presented in the middle row. The number of points below the stage level was only 0.02% of the total number but, if not excluded, these points can have a large impact on the volumetric shape.

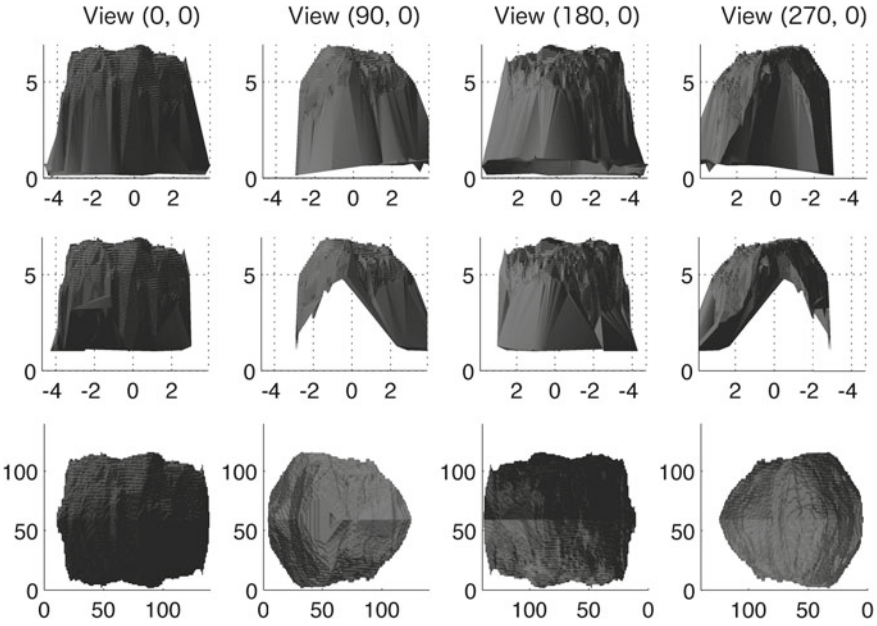

Fig. 5.2 An example of the construction of a volumetric model based on the stereogrammetrically retrieved surface of the dolomite particle. The top row shows the TIN surface from the side (declination 0°), from four azimuthal angles. The middle row shows the same TIN surfaces after the removal of the points that were likely to be from the stage instead of the particle surface. In this case, we excluded all points below the height of 1.0 μm. Perspectives of the volumetric model are collected in the bottom row. For dolomite, the volumetric model was made using a 1:1 mirroring of the upper surface with respect to a horizontal plane. This 3-D shape consists of about 800 000 volume elements. The units in the top and middle rows are microns and in the bottom row indices of the volume elements

Since the stereogrammetric method was, thus far, only used to retrieve the upper hemisphere of the particle surface, assumptions about the other hemisphere cannot be avoided when constructing the volumetric model. Lindqvist et al. (2014) applied two methods: mirroring and scaled mirroring. Both methods construct a lower hemisphere that is mirror symmetric to the upper; the only difference is the height of the mirroring plane which determines the scaled height of the lower hemisphere, so that the aspect ratio of the particle is conserved. After the removal of the stage level points, the mirroring plane can be set to any desired level, and then mirror and scale the points that are above the level. An example of 1:1 mirroring is shown in Fig. 5.2, bottom row. It is important to set the mirroring plane in such a level that enough of the retrieved surface points lie above that level; in the example of Fig. 5.2, only 4.8% of the points are lost due to the choice of the mirroring level.

We note that the particle surface and its mirror give only an approximation of the full 3-D particle as obviously only those parts of the upper surface are covered which are visible to both SEM images. Also, the particle boundary areas are usually not well covered and would require more images from oblique viewpoints to be fully modeled.

However, with a well-defined surface on both hemispheres, it is next straightforward to discretize the space between the surfaces into small volume elements. Volume discretization is obviously necessary only if the light scattering method employed is a volume integration method such as the discrete-dipole approximation (DDA) (e.g., Yurkin and Hoekstra 2011), or if the particle is internally inhomogeneous and the volume fractions and locations of the inhomogeneities need to be transferred to the light scattering method. The size of the individual volume elements defines the resolution for surface details of the volumetric model, because the dense TIN surface model has higher resolution for surface details than the maximum reasonable resolution for the volumetric model, which is, in turn, largely dictated by the light scattering method and the available computing resources.

In the generation of the volumetric model from the stereogrammetrically retrieved surface, there are several phases which would benefit from a detailed study dedicated to analyzing how the choices (e.g., stage level, mirroring) made in the shape generation propagate to the light scattering results. We performed a limited number of such studies in the early stages of this work, using a calcite particle as the test subject. First, we experimented with the selection height of the stage level, and found that the degree of linear polarization $-S_{21}/S_{11}$ was the most sensitive to it. The angular dependence of the element changed especially in the intermediate scattering angles, when the stage level was set too high, so that points from the particle surface were deleted in addition to the stage. For lower stage level choices where most or all of the excluded points were from the stage, the difference in scattering was minor. Second, we tested different levels of scaled mirroring so that the height of the particle remained constant, and we set the mirroring plane at $h_{\max}/2$, $2h_{\max}/5$, and $h_{\max}/4$, and also constructed an extreme case where the lower side was a flat, horizontal

plane. In other words, we obtained the lower part using 1:1, 3:2, 3:1, and 1:0 mirroring, respectively. We found, not surprisingly, that the particle with the flat lower hemisphere differed from the others both in the scattered intensity and polarization properties. The mirrorings showed most differences at scattering angles $\theta > 90°$, and more for the polarization elements than for intensity. Further studies have to be performed on the artifacts introduced by scaled mirroring, because this study was done using only one shape, and isolating the impact of the artifacts from that of a changing particle volume is challenging.

5.3.6 Examples of Retrieved Shapes

In this section, the performance of the shape retrieval algorithm outlined in Fig. 5.1 is illustrated with four mineral dust particles, namely a calcite, a dolomite, a silicate, and an aggregate particle. The particles were selected from a Saharan mineral dust sample collected during the SAMUM campaign (Heintzenberg 2009; Scheuvens et al. 2011) over Morocco in 2006. For electron microscopy, the particle samples were sputter-coated with a thin gold layer. The single particles were imaged with an FEI environmental scanning electron microscope (ESEM) Quanta 200, equipped with a field emission gun, at two different angles by tilting the specimen stage at a working distance of 10 mm. Secondary and backscatter electron images were collected. An acceleration voltage of 15 kV with a "spot size 3" was found optimal for image quality, resulting in a nominal lateral resolution better than 3 nm. One of the SEM images for each particle is shown in Fig. 5.3.

As a preprocessing step for shape retrieval, the area covered by the particle was manually segmented from one of the SEM images of each particle. SIFT keypoints extracted within the whole image were applied in sparse matching and for solving the image orientations in a bundle adjustment with sparse 3-D reconstruction. Dense image matching was then performed based on expanding the sparse correspondences to the whole segmented area with a grid spacing of $m = 1$ pixel, i.e., the dense correspondences were established at every pixel. The thresholds in dense image

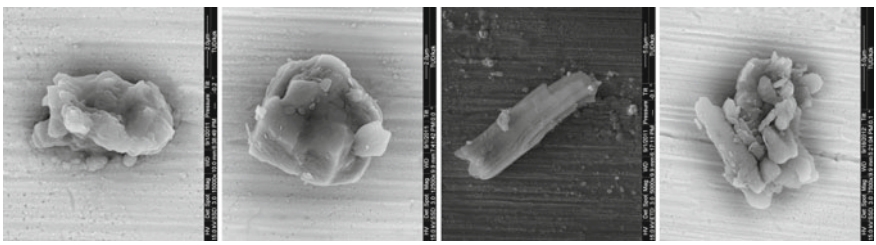

Fig. 5.3 SEM images (from left to right) of a calcite, a dolomite, a silicate, and an aggregate particle. The particles are of the size of 5–20 microns

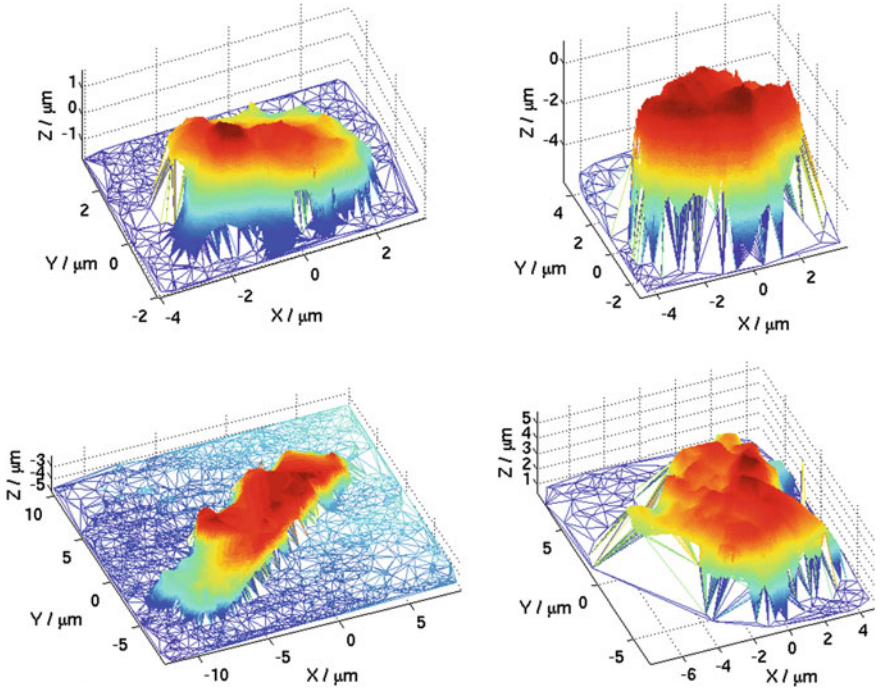

Fig. 5.4 Dense TIN models (row-wise from left to right) of the calcite, dolomite, silicate, and aggregate particle. The colour is proportional to the Z-coordinate value

matching were $T_d = 5$ pixels and $T_e = 5$ pixels and the size of the image template was $p = 61$ pixels squared for all the particles (the whole image is of the size of 2048 × 1887 pixels). The reconstructed dense surface models are illustrated in Fig. 5.4 and the volumetric ones in Fig. 5.5.

It may be realized that the surface models are very detailed without any outliers visible. No manual editing has been performed but the process is fully automated excluding the segmentation step. The algorithms were implemented using matlab software and the computations were performed in a PC with 3.30 GHz processor. The total computing times varied from 90 min to five hours for the different particles with most time spent in dense image matching. The number of matched points varies from approximately half a million (silicate particle) and 800 000 (calcite particle) to one million (dolomite and aggregate particles). The sparsely located points on the stage result from matched SIFT keypoints. The volumetric models clearly show the overall geometry of the particles.

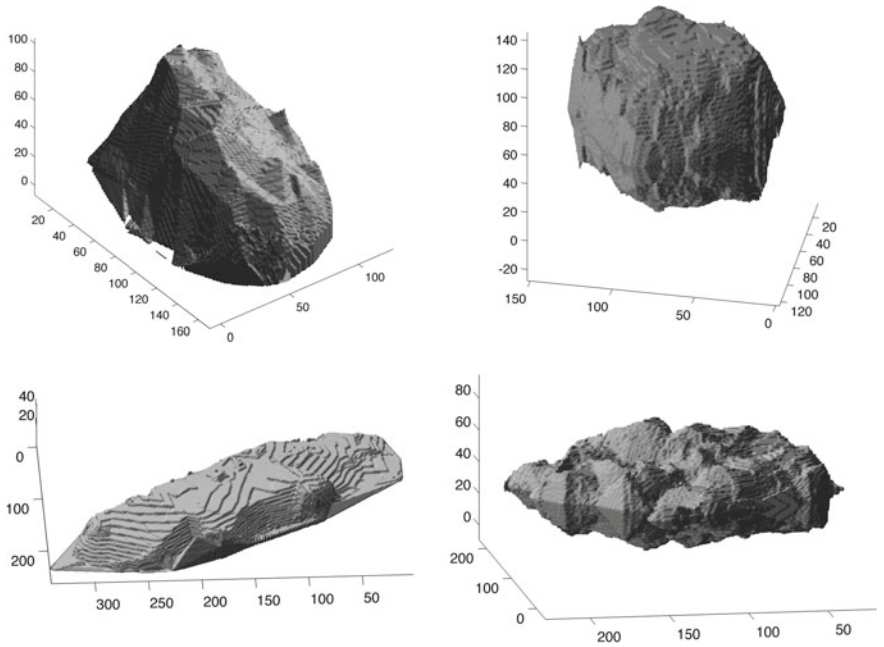

Fig. 5.5 Volumetric models (row-wise from left to right) of the calcite, dolomite, silicate, and aggregate particle. The units of the coordinate axes are indices of the volume elements

5.4 Application to Light Scattering

For light scattering purposes, particle shape parameters have been analyzed directly from images already in 1984, when Hill et al. (1984) measured the aspect ratios of mineral dust particles. Although this was still far from stereogrammetry, the approach was fundamentally similar to what we discuss here, as the starting point was to take measurements of the real dust particles and use those as the basis for the dust shape model. The fundamental difference, however, is that they derived simple parameters to describe the shapes, while we derive the actual shapes. As the computational methods for light scattering have become more powerful and versatile, the selection of possible shape models, for example for mineral dust, has expanded tremendously. Still, a mathematical shape model, such as a spheroid that is described with its aspect ratio, is but an idealization of a real particle, and the light scattering results obtained using such models are thus questionable. The stereogrammetric shape retrieval method offers a novel possibility to acquire the shape model directly from the imaged particle, with considerably more detail than the strongly simplified ones used before. Stereogrammetric technique was pioneered in light scattering applications by Lindqvist et al. (2014). They analyzed pairs of scanning electron microscope images of four micrometer-sized mineral dust particles using stereogrammetry, built the three-dimensional models directly based on the retrieved

shapes, and computed their light scattering using the discrete-dipole approximation (DDA). A central question of the entire application is: how different are these stereogrammetric models compared to simplified mathematical shape models, from light scattering perspective? We will address this question by comparing the stereogrammetric models to spheres, spheroids, and Gaussian random spheres, which all are widely used shape models for mineral dust particles.

5.4.1 Scattering by Stereogrammetrically Modeled Mineral Dust

We present light scattering results for four stereogrammetrically retrieved particle models that represent a calcite, a dolomite, a silicate, and an aggregate particle, all a few micrometers in diameter. The particles are the same as in Fig. 5.3 but the shape models are adopted from Lindqvist et al. (2014) where the dense mapping procedure was not yet implemented. The shapes thus have the same overall geometries, but lack in small-scale details. The particles have inhomogeneous compositions that have been investigated using energy-dispersive spectroscopy and taken into account in the volumetric models by manually setting the refractive indices for each volume element (see Lindqvist et al. (2014), for details). As concluded by Lindqvist (2013) and Lindqvist et al. (2014), the inhomogeneous compositions did not have as large an impact to scattering than the different shapes in this case (likely because the refractive indices of the different minerals were typically similar, except for hematite, but its volume fraction was in all cases less than 2%), so we can attribute most of the particle-to-particle differences in scattering to arise from differences in shapes.

Light scattering computations were made using the DDA code ADDA, version 1.1 (Yurkin and Hoekstra 2011). ADDA is a volume integral method for computing the electromagnetic interaction of the incident radiation with a wavelength-scale particle, and solves for the total scattered field. This interaction can be described mathematically with a 4×4 scattering matrix S, where most of the information is stored in the following six matrix elements: S_{11}, which denotes the total scattered intensity for incident unpolarized light; and different polarization elements S_{21}, S_{22}, S_{33}, S_{34}, and S_{44}. These elements are functions of the size parameter $x = 2\pi a_{\mathrm{eq}}/\lambda$, where a_{eq} is the radius of the particle (for nonspherical particles, the radius of an equivalent-volume sphere is here used), and λ is the wavelength of light. The results shown here are computed for a visible wavelength of $\lambda = 550$ nm and using particle radii from 0.04 μm to 1.4 μm, corresponding to size parameters from $x = 0.5$ to $x = 16$. The true sizes of the four particles were around the highest x considered, but we expect the retrieved shape to be a realistic dust model even if scaled to smaller or larger particle sizes.

Size parameter dependence of the scattering matrix elements is presented in Figs. 5.6, 5.7, 5.8 and 5.9 for the calcite, dolomite, silicate, and aggregate particle, respectively. Overall, the overarching scientific finding that these results yield

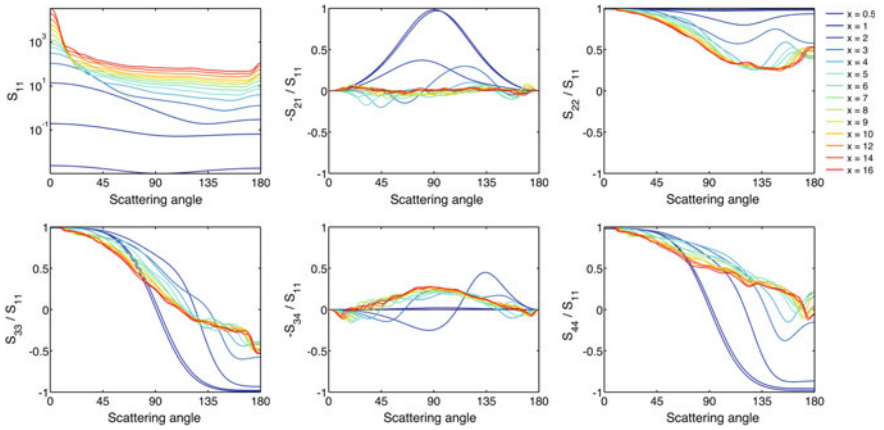

Fig. 5.6 Size parameter dependence of the six scattering matrix elements for the stereogrammetric calcite particle

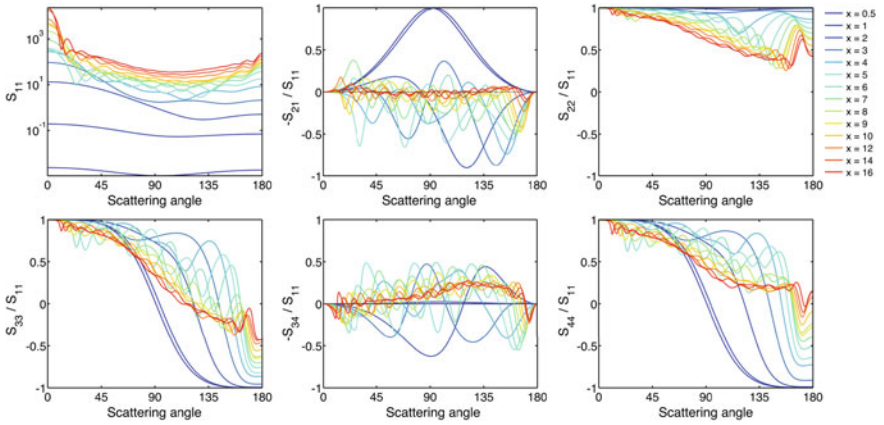

Fig. 5.7 As Fig. 5.6, but for the dolomite particle

is that the differences in scattering by micron-sized real dust particles are surprisingly large in all of the scattering matrix elements. The size dependence of scattering varies to some extent between particle shapes although has similarities as well: for the polarization elements, the results show shape dependence for size parameters $x \geq 3$ in general, but the convergence at large x depends on the shape. While the very thin and elongated silicate particle shows convergence in its polarization properties already at $x \geq 5$, the more equidimensional dolomite particle does not converge even at the largest x considered. Scattered intensity S_{11} develops similarly as a function of size in all cases: intensity increases with x throughout all scattering angles, and shows a nonlinear increase near backscattering at larger size parameters regardless of the shape. Notable in the S_{11} results is that only the calcite and aggregate particles

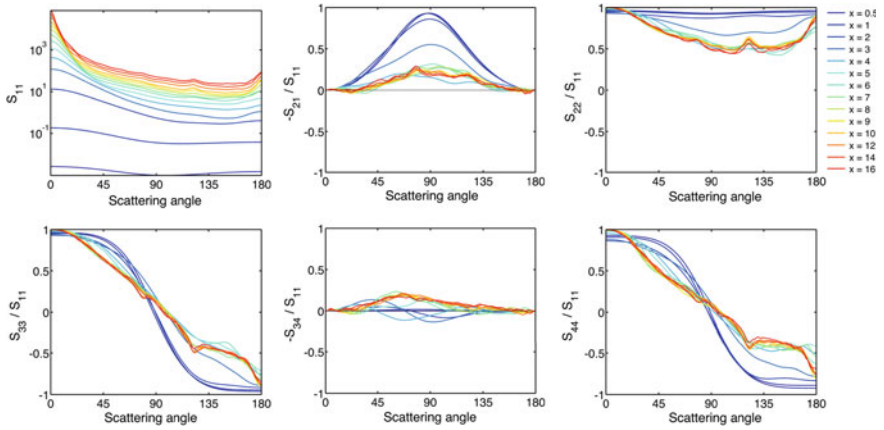

Fig. 5.8 As Fig. 5.6, but for the silicate particle

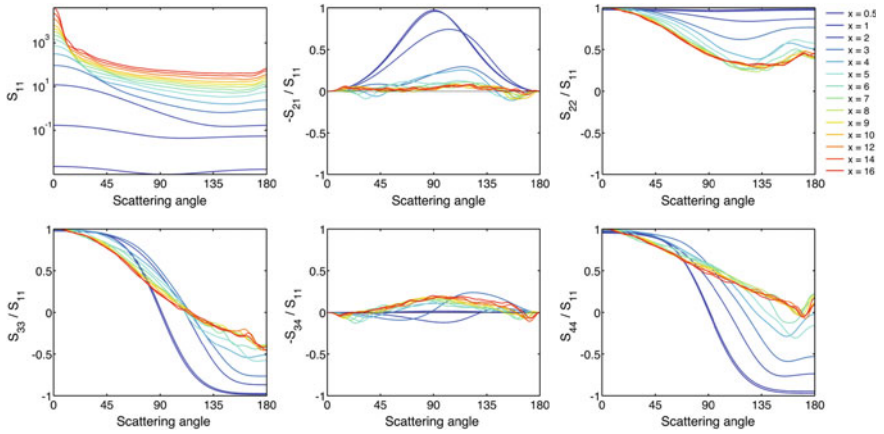

Fig. 5.9 As Fig. 5.6, but for the aggregate particle

produce a typical mineral dust phase function with a featureless and flat side scattering and a minor increase towards backscattering.

5.4.2 Comparison to Simplified Shape Models

Stereogrammetric shape retrieval is, despite its beauty, a highly elaborate method for obtaining a model for a single particle. Therefore, we see its benefits first and foremost in validation of simplified shape models that are more effective in both modeling the shape and light scattering, but may have their inaccuracies and inconsistencies that are not necessarily well known (Nousiainen and Kandler 2015). Here,

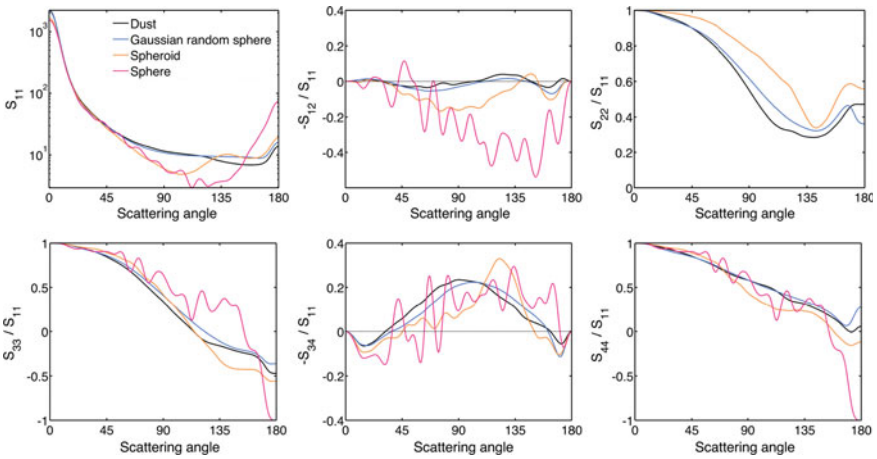

Fig. 5.10 Angular dependence of the scattering matrix elements for a lognormal size distribution of stereogrammetric calcite particles (black line). For comparison, integration over a similar size distribution of simplified shape models of a sphere (pink line), a spheroid (orange line), and a Gaussian random sphere (blue line) are also shown

we demonstrate comparisons of stereogrammetric models to simplified models in terms of light scattering. As stereogrammetric models we consider our four example dust particles, while as simplified models we consider a sphere, a spheroid, and a Gaussian random sphere, the last of which is a stochastic shape model for irregular particles (Muinonen et al. 1996). To eliminate one dimension of variables and thus facilitate comparisons, the scattering results are here integrated over a lognormal size distribution with a cross-section-weighted average radius $r_{\text{eff}} = 0.82\,\mu\text{m}$ (Lindqvist et al. 2014; Mishchenko and Travis 1998). All the simplified model particles have volumes equal to the stereogrammetric shapes (mass conservation). An aspect ratio for each spheroidal model was derived from the stereogrammetric images: the silicate particle turned out to be the most elongated with an aspect ratio of 0.29, while the dolomite was closest to an equidimensional shape with an aspect ratio of 0.91. The shape parameters for the Gaussian random sphere model (σ and the correlation function, see Muinonen et al. (1996) for details) were retrieved from three orthogonal, randomly chosen intersections of each stereogrammetric particle.

For the calcite particle, the comparisons are shown in Fig. 5.10. The size-integrated scattering properties of the stereogrammetric shape are smooth and quite typical for mineral dust. The Gaussian random sphere geometry models scattering remarkably successfully for all the elements, only biased somewhat high for S_{22}/S_{11} and low for $-S_{12}/S_{11}$. Spheroids of the same aspect ratio as the particle cannot mimic the scattering properties, and fail in particular at intermediate to large scattering angles for all elements considered. The distribution of spheres is not even comparable because of the strong resonances characteristic to monodisperse spheres have not balanced out after the size integration.

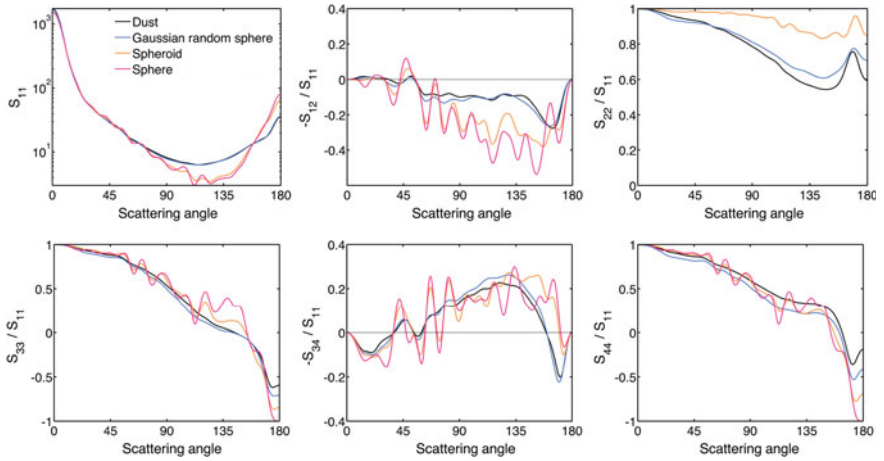

Fig. 5.11 As Fig. 5.10, but for the dolomite particle and the simplified shape models based on that particle

The dolomite particle scattering comparisons (Fig. 5.11) reveal that this particle is much closer to a spherical particle, probably because of its aspect ratio close to unity. Although the examples in Fig. 5.2 do not show a visual resemblance to a sphere, the interferences in the monodisperse scattering in Fig. 5.7 are a clear indication that the dolomite particle shape supports similar interference mechanisms that take place inside a sphere or other regularly shaped particles (Muinonen et al. 2011). The size-integrated results confirm this, in particular with the asymmetric phase function with a distinct minimum, and an almost entirely negative $-S_{12}/S_{11}$. Then again, S_{22}/S_{11} that deviates from unity is an unmistakable indication of nonsphericity. The Gaussian random spheres are also almost spherical ($\sigma = 0.084$) for this particle but deviate from a sphere enough so that the S_{22}/S_{11} of the Gaussian random sphere geometry closely matches the angular dependence of S_{22}/S_{11} by the stereogrammetic shape.

The silicate particle is different from the other three dust particles in that it is very elongated and thin. To begin with, the simplified shape models employed in this study are not ideal for modeling the shapes of such particles. This is demonstrated from light scattering perspective in Fig. 5.12, where the simplified models are clearly unable to model the scattering matrix of the silicate particle. Due to the elongation and the fact that we use volume equivalence for determining the particle size, the silicate stereogrammetric particle has two of the three dimensions significantly larger than the other model shapes, which is evident from the $S_{11}(\theta = 0°)$ that is almost an order of magnitude higher for the silicate than for the spherical model.

The size-integrated scattering by the aggregate particle (Fig. 5.13) is, like calcite,"traditional" mineral dust scattering. For the aggregate, the spherical and spheroidal particles cannot be sed as proxies in scattering computations, and even model particles based on the Gaussian random sphere geometry scatter here notably differently to the original shape, potentially due to the lack of an inhomogeneous composition or a non-optimal shape parameter retrieval.

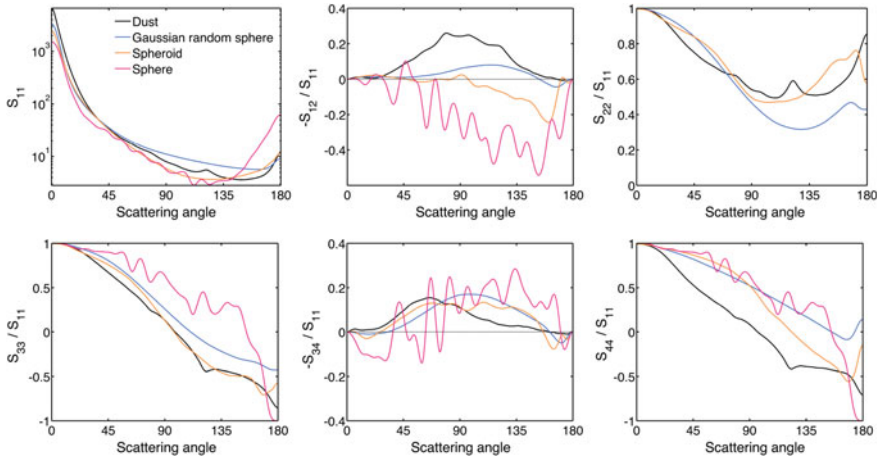

Fig. 5.12 As Fig. 5.10, but for the silicate particle and the simplified shape models based on that particle

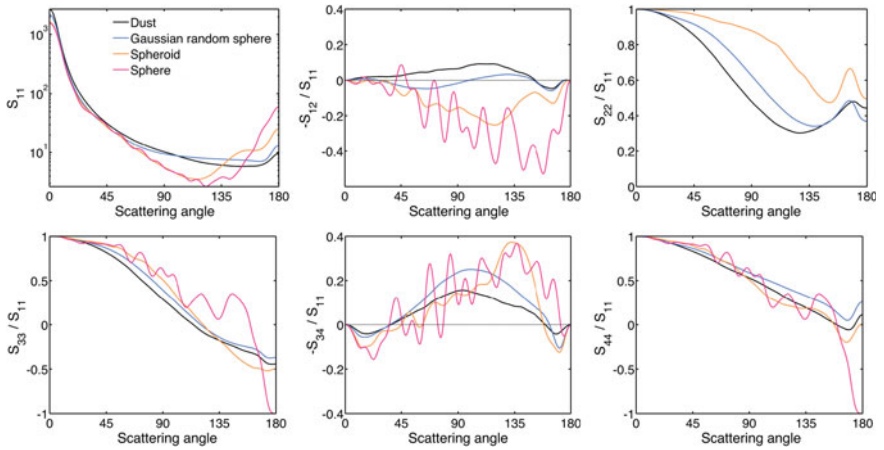

Fig. 5.13 As Fig. 5.10, but for the aggregate particle and the simplified shape models based on that particle

5.4.3 Impact of Surface Roughness on Scattering

The stereogrammetrically retrieved particle models are applied to investigate the impact of surface roughness on the optical properties of real dust particles. The results of introducing additional surface roughness, while conserving mass and overall shape, has been studied by Kemppinen et al. (2015). Their roughening algorithm was based on Monte Carlo collision of rays with the surface elements of the targets that created small mounds or craters at the collision points. Examples of various

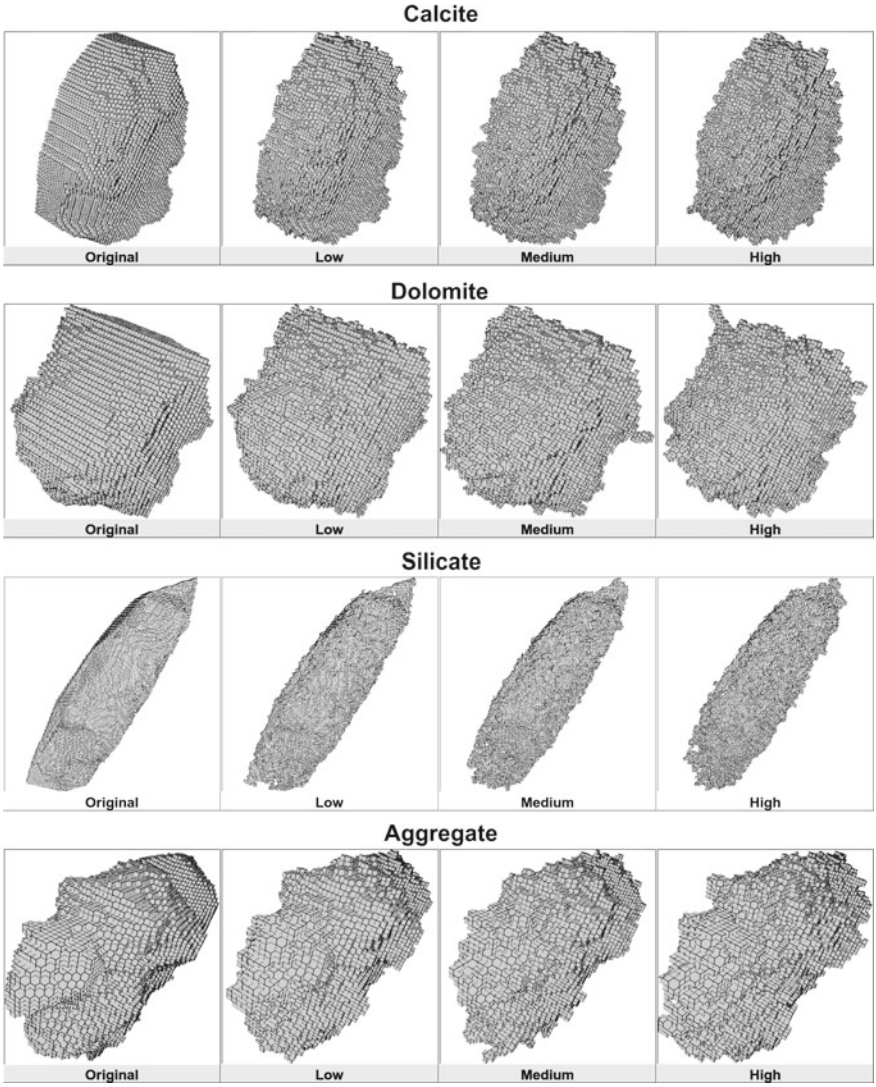

Fig. 5.14 Volume element representations of the various roughening stages of the stereogrammetric particle models. For each particle, the unroughened stage is shown along with three stages of increasing roughening: low, medium and high

stages of the roughened stereogrammetric particles are shown in Fig. 5.14. Each row depicts the advancing stages of surface roughening for one of the particles. The overall shape is not affected significantly, but surface details are clearly distorted at the later stages of roughening.

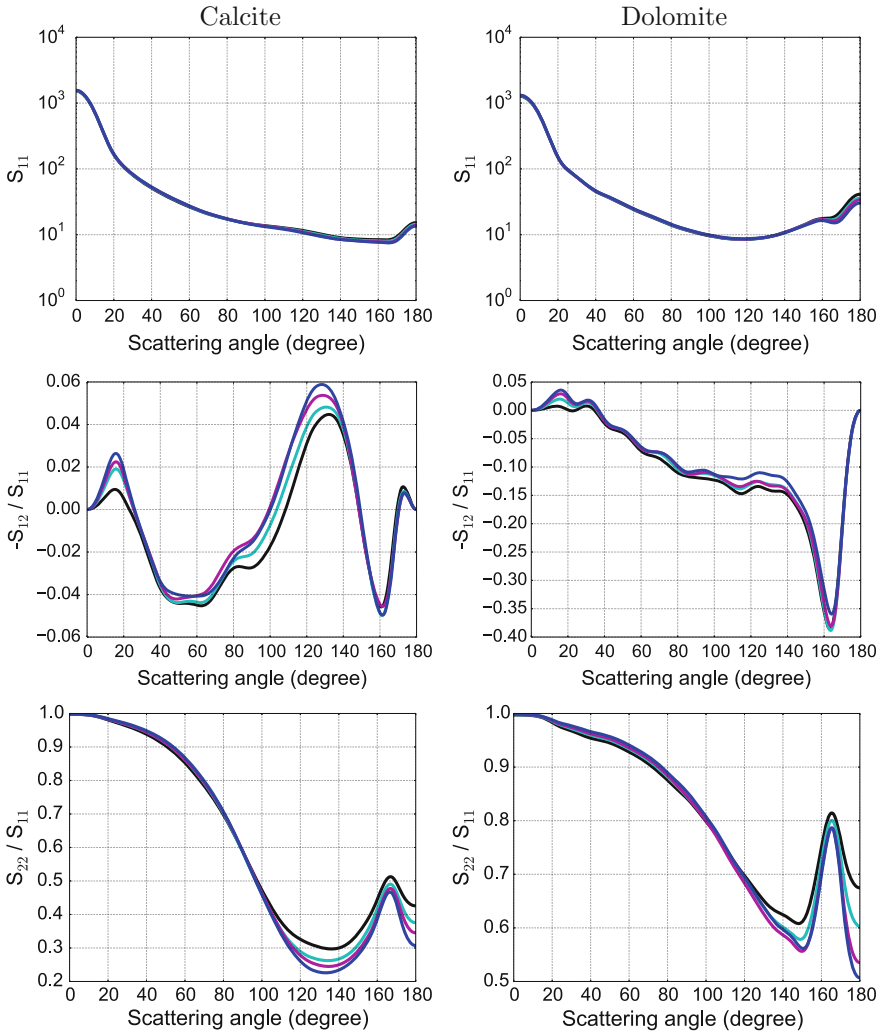

Fig. 5.15 Scattering matrix elements S_{11}, $-S_{12}/S_{11}$, and S_{22}/S_{11} of different roughening stages for the stereogrammetric models of the calcite and dolomite particles. Black is the scattering function of the original particle, and cyan, magenta, and blue correspond to the low, medium, and high roughening cases, respectively

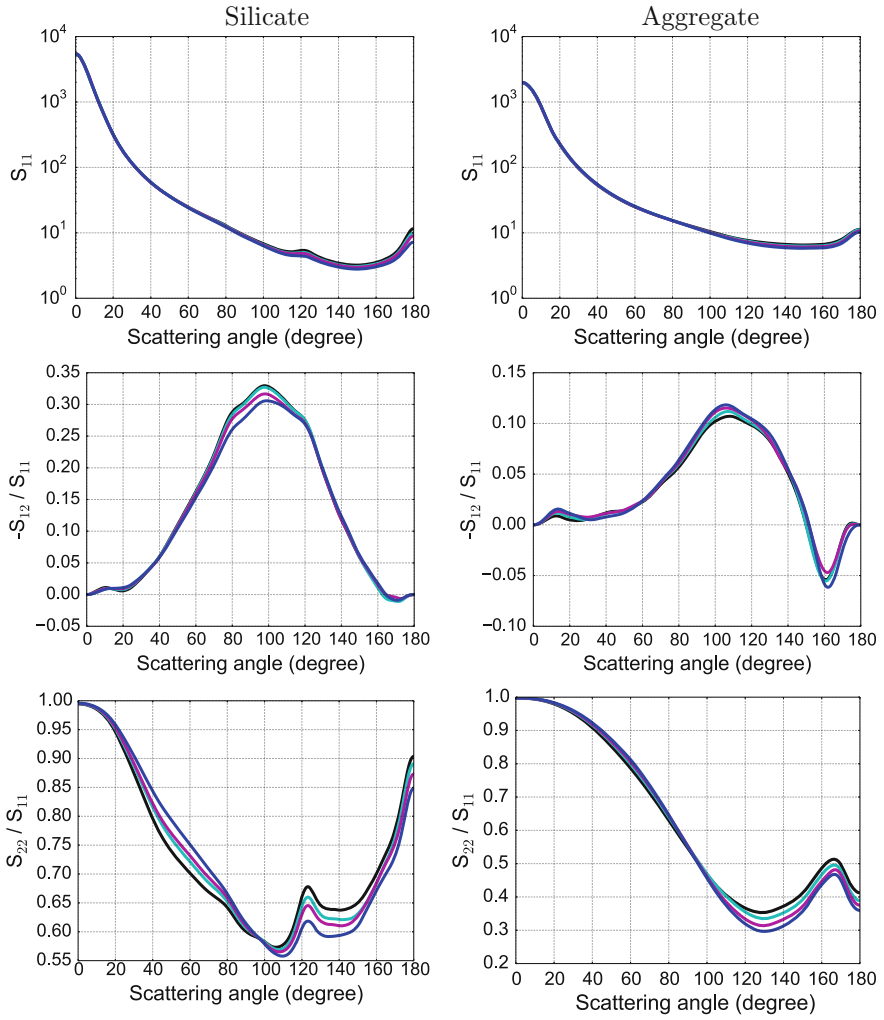

Fig. 5.16 As Fig. 5.15, but for the stereogrammetric models of the silicate and aggregate particles

The level of surface roughness seems to have a systematic impact for all the scattering matrix elements, and the effect is seen in all the particles studied. The results are summarized as follows: S_{11}, S_{22}/S_{11}, and $-S_{34}/S_{11}$ are decreased, while $-S_{12}/S_{11}$, S_{33}/S_{11}, and S_{44}/S_{11} are increased. For the diagonal elements, the effects are generally the most pronounced at the backscattering angles; whereas, for the off-diagonal elements, the effects are greatest at the side-scattering angles. The impact of roughening on scattering is illustrated in Figs. 5.15 and 5.16 for S_{11}, $-S_{12}/S_{11}$, and S_{22}/S_{11}. It is clearly seen that the level of roughening has a quantitative effect on scattering. It is also seen that the effects are relatively modest. However, when

considering lidar quantities, such as lidar ratio and linear depolarization ratio, the effects can be significant (not shown). It is also worth noting that it seems that particles of larger sizes are affected more strongly by the additional surface roughness.

5.5 Conclusions and Future Work

This chapter has presented methods for retrieving the real 3-D shapes of atmospheric dust particles from SEM images. The methods were fully automated, excluding a manual segmentation of the particle from its background and selection of the mirroring plane in volumetric modeling. The experimental results showed that highly detailed models can be recovered with the proposed dense image matching techniques. The resulting models are fundamentally different from the simplistic or descriptive shape models that have been exclusively employed in dust single-scattering modeling until now. Most importantly, the stereogrammetric models are directly based on real particles and can capture both their large-scale and smaller-scale shape features that are unique to each particle.

Given that the stereogrammetrically modeled shapes are based on real particles, light scattering studies made using these model particles are arguably among the best available references for single scattering by individual, wavelength-scale particles. The first such study concentrated on mineral dust particles, where light scattering by the stereogrammetric model particles was simulated using the discrete-dipole approximation. Simulations for four mineral dust particles of different types suggest that the type can largely affect the optical properties. The most significant differences in the scattering properties between the dust particle types are seen in the side and backscattering angular region — the scattering angles that are highly relevant for, e.g., lidar and satellite-based remote sensing. All scattering matrix elements have shown differences to some degree: from the polarization elements, the most sensitive to particle type are $-S_{12}/S_{11}$ and S_{22}/S_{11}, while the smallest differences are seen in $-S_{34}/S_{11}$. Reference results like these are crucial when selecting a suitable single-scattering model for an application. Because the retrieval of a stereogrammetric model geometry and the light scattering computations are elaborate and resource-consuming processes, a simplified (mathematical) model may be more convenient for practical applications. Particles such as spheres or spheroids would be particularly desirable, but have been shown to bear little resemblance in their size-integrated single-scattering properties with respect to the stereogrammetric dust particles.

In addition to providing very detailed reference material for light scattering research, the stereogrammetric models can also be used as a realistic baseline for sensitivity studies. One such study has been made concerning surface roughness. Added small-scale surface roughness seems to have a systematic and consistent impact on light scattering by irregular, realistically shaped particles. Largest effects are seen at backscattering and side-scattering angles. Particularly large impacts are seen with regards to lidar quantities and for large particles.

Despite its obvious benefits in light scattering research, the stereogrammetric method still has many limitations and future challenges to overcome before it can retrieve truly real-like particle models. Currently, a major limitation is the lack of a proper technique to image the backside of the particle. Thus, the surface of only one hemisphere of the particle is retrieved. Even though it is possible to construct the other hemisphere by assuming symmetries in the particle shape (e.g., mirror symmetry), it is not an ideal approach and will not provide, e.g., the true particle volume (or mass). To get the full 3-D shape imaged and retrieved, the particle should be turned upside down and imaged also from multiple oblique viewpoints. Research is then needed to register the models from all the viewpoints to the same coordinate system and to merge them into a consistent 3-D representation of the particle.

Another drawback of the stereogrammetric method is the fundamental disability to look into internal cavities and deep crevices. Only such features that are fully exposed to both imaging directions can be mapped. This limits the selection of particles to preferably compact particles with identifiable surface features. Aggregate particles and particles with internal structure (e.g., porosity or other material inhomogeneity) are particularly challenging. Internal structures can be revealed, e.g., by focused ion beam milling, but including that information explicitly in the model is yet another challenge.

In all its beauty, stereogrammetry is not a sufficient method to capture all the physical characteristics of small particles. Light scattering can be very dependent on the optical properties of the material, i.e., its complex refractive index. Even though certain mineralogical compositions may be deduced based on their characteristic shapes, identification of particle composition and structure requires supplemental methods. For example, energy-dispersive X-ray spectroscopy can derive a localized chemical composition (atomic composition map) of a particle. Electron diffraction analysis can be used to determine the mineralogical composition in small selected areas. Also, electron energy loss spectroscopy may yield information on bonding state at spots of a particle. If these techniques are combined with a slicing of the particle (e.g., by focused ion beam milling) or if abrasive techniques like (nano-scale) secondary ion mass spectrometry are used, three-dimensional compositional and structural information can be retrieved. In principle, all localized analytical methods may add up to the compositional and structural picture of a single particle, but probably no technique is currently able to describe all of the particle in detail. As a result, interpretative steps are still necessary.

Overall, incorporating stereogrammetry into light scattering modeling has renewed the field and certainly brought it closer to capturing the actual variety of the shapes of particles in nature. Despite its current limitations, the method is already applicable to selected light scattering studies, where it helps us learn more about the complex interaction of light and a particle. After future improvements to the method, the stereogrammetric approach will ultimately allow us to connect the morphological properties to optical ones more accurately and fundamentally than has previously been possible.

References

Bay H, Ess A, Tuytelaars T, Van Gool L (2008) Speeded-Up Robust Features (SURF). Comput Vis Image Und 110(3):346–59

Burkhardt R (1981) Die stereoskopische Ausmessung elektronenmikroskopischer Bildpaare und ihre Genauigkeit. Methodensammlung der Elektronenmikroskopie 10:1–59

Canny J (1986) A computational approach to edge detection. IEEE Trans Pattern Anal 8(6):679–98

Carli L (2010) 3D-SEM Metrology for Coordinate Measurements at the Nanometer Scale. PhD Thesis, Technical University of Denmark, 266 p

Carli L, Genta G, Cantatore A, Barbato G, De Chiffre L, Levi R (2011) Uncertainty evaluation for three-dimensional scanning electron microscope reconstructions based on the stereo-pair technique. Meas Sci Technol 22(035103):1–11

Cavegn S, Haala N, Nebiker S, Rothermel M, Tutzauer P (2014) Benchmarking high density image matching for oblique airborne imagery. Int Arch Photogramm Remote Sens Spatial Inform Sci XL-3: 45–52

Cornille N, Garcia D, Sutton MA, McNeill SR, Orteu J-J (2003) Automated 3-D reconstruction using a scanning electron microscope. In: SEM Annual Conference Proceedings - Exposition on Experiment Applied Mechanics, 8 p

Danzl R, Mershon WJ, Helmli F, Scherer S (2007) 3D Reconstruction and matching of fractured surfaces using images from scanning electron microscopes. Microsc Microanal 13(S02):1666–7

Electron Microscopy Sciences. http://www.emsdiasum.com/microscopy/products/sem/3dSEM.aspx. Accessed 17 March 2016

Förstner W, Gülch E (1987) Fast operator for detection and precise location of distinct points, corners and centres of circular features. In: Proceedings of the ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data, pp 281–305

Furukawa Y, Ponce J (2010) Accurate, dense, and robust multiview stereopsis. IEEE Trans Pattern Anal 32(8):1362–76

Gasteiger J, Wiegner M, Groß S, Freudenthaler V, Toledano C, Tesche M, Kandler K (2011) Modelling lidar-relevant optical properties of complex mineral dust aerosols. Tellus B 63(4):725–41

Goldstein JI, Newbury DE, Joy D, Lyman C, Echlin P, Lifshin E, Sawyer L, Michael J (2003) Scanning Electron Microscopy and X-Ray Microanalysis, 3rd edn. Kluwer Academic / Plenum Publishers, New York

Goodhew P, Humphreys J, Beanland R (2001) Electron Microscopy and Analysis, 3rd edn. Taylor & Francis group

Haala N (2013). The landscape of dense image matching algorithms. In: Fritsch D (ed) Photogrammetric Week '13, Wichmann, Berlin/Offenbach, pp 271–84

Harris C, Stephens M (1988) A combined corner and edge detector. Alvey Vis Conf 147–51

Heintzenberg J (2009) The SAMUM-1 experiment over Southern Morocco: overview and introduction. Tellus B 61(1):2–11

Hemmleb M, Albertz J, Schubert M, Gleichmann A, Köhler JM (1996) Digital microphotogrammetry with the scanning electron microscope. Int Arch Photogramm Remote Sensing XXXI, Part B5:225–30

Hill SC, Hill AC, Barber PW (1984) Light scattering by size/shape distributions of soil particles and spheroids. Appl Opt 23(7):1025–31

Hirschmüller H (2008) Stereo processing by semiglobal matching and mutual information. IEEE Trans Pattern Anal 30:328–41

Ishimoto H, Zaizen Y, Uchiyama A, Masuda K, Mano Y (2010) Shape modeling of mineral dust particles for light-scattering calculations using the spatial Poisson-Voronoi tessellation. J Quant Spectrosc Radiat Transfer 111:2434–43

Jeong GY, Nousiainen T (2014) TEM analysis of the internal structures and mineralogy of Asian dust particles and the implications for optical modeling. Atmos Chem Phys 14:7233–54

Kadir T, Zisserman A, Brady M (2004) An affine invariant salient region detector. ECCV 404–16

Kammerud CR (2005) Comparison of Microscopic Modalities for the 3D Model Reconstruction of Nanodevices. MSc Thesis, The University of Tennessee, Knoxville, 106 p

Kandler K, Schütz L, Deutscher C, Hofmann H, Jäckel S, Knippertz P, Lieke K, Massling A, Schladitz A, Weinzierl B, Zorn S, Ebert M, Jaenicke R, Petzold A, Weinbruch S (2009) Size distribution, mass concentration, chemical and mineralogical composition, and derived optical parameters of the boundary layer aerosol at Tinfou, Morocco, during SAMUM 2006. Tellus B 61(1):32–50

Kannala J, Brandt SS (2007) Quasi-dense wide baseline matching using match propagation. CVPR, 1–8

Kemppinen O, Nousiainen T, Lindqvist H (2015) The impact of surface roughness on scattering by realistically shaped wavelength-scale dust particles. J Quant Spectrosc Radiat Transfer 150C:55–67

Lacey AJ, Thacker NA, Crossley S, Yates RB (1998) A multi-stage approach to the dense estimation of disparity from stereo SEM images. Image Vision Comput 16(5):373–83

Lindqvist H (2013) Atmospheric Ice and Dust: From Morphological Modeling to Light Scattering. PhD Thesis, Finnish Meteorological Institute Contributions 102, 154 p

Lindqvist H, Jokinen O, Kandler K, Scheuvens D, Nousiainen T (2014) Single scattering by realistic, inhomogeneous mineral dust particles with stereogrammetric shapes. Atmos Chem Phys 14:143–57

Liu C, Panetta RL, Yang P, Macke A, Baran AJ (2012) Modeling the scattering properties of mineral aerosols using concave fractal polyhedra. Appl Opt 52(4):640–52

Lowe D (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vision 60(2):91–110

Lucideon. http://www.ceram.com/testing-analysis/techniques/three-dimensional-scanning-electron-microscopy-3dsem. Accessed 17 March 2016

Matas J, Chum O, Urban M, Pajdla T (2002). Robust wide baseline stereo from maximally stable extremal regions. BMVC, 384–93

Mikolajczyk K, Schmid C (2004) Scale and affine invariant interest point detectors. Int J Comput Vision 60(1):63–86

Mikolajczyk K, Schmid C (2005) A performance evaluation of local descriptors. IEEE Trans Pattern Anal 27(10):1615–30

Mikolajczyk K, Tuytelaars T, Schmid C, Zisserman A, Matas J, Schaffalitzky F, Kadir T, Van Gool L (2005) A comparison of affine region detectors. Int J Comput Vision 65(1/2):43–72

Mishchenko MI, Travis LD (1998) Capabilities and limitations of a current FORTRAN implementation of the T-matrix method for randomly oriented, rotationally symmetric scatterer. J Quant Spectrosc Radiat Transfer 60(3):309–24

Morgan M, Kim K, Jeong S, Habib A (2004) Indirect epipolar resampling of scenes using parallel projection modeling of linear array scanners. Int Arch Photogramm Remote Sensing XXXV, Part B3:58–63

Muinonen K, Nousiainen T, Fast P, Lumme K, Peltoniemi JI (1996) Light scattering by Gaussian random particles: Ray optics approximation. J Quant Spectrosc Radiat Transfer 55(5):577–601

Muinonen K, Tyynelä J, Zubko E, Lindqvist H, Penttilä A, Videen G (2011) Polarization of light backscattered by small particles. J Quant Spectrosc Radiat Transfer 112(13):2193–212

Muñoz O, Volten H, Hovenier JW, Nousiainen T, Muinonen K, Guirado D, Moreno F, Waters LBFM (2007) Scattering matrix of large Saharan dust particles: experiments and computations. J Geophys Res 112:D13215

Nousiainen T, Kandler K (2015) Light scattering by atmospheric mineral dust particles. In: Kokhanovsky A(ed) Light Scattering Reviews, vol 9. Springer, Berlin, pp 3–52

Nousiainen T, Zubko E, Niemi JV, Kupiainen K, Lehtinen M, Muinonen K, Videen G (2009) Single-scattering modeling of thin, birefringent mineral dust flakes using the discrete-dipole approximation. J Geophys Res 114:D07207

Nousiainen T, Muinonen K, Räisänen P (2003) Scattering of light by large Saharan dust particles in a modified ray optics approximation. J Geophys Res 108:4025

Piazzesi G (1973) Photogrammetry with the scanning electron microscope. J Phys E Sci Instrum 6(4):392–6
Remondino F, Spera MG, Nocerino E, Menna F, Nex F (2014) State of the art in high density image matching. Photogramm Rec 29(146):144–66
Remondino F, Zhang L (2006) Surface reconstruction algorithms for detailed close-range object modeling. Int Arch Photogramm Remote Sensing Spatial Inf Sci XXXVI, Part 3:117–23
Richards RG, Wieland M, Textor M (2000) Advantages of stereo imaging of metallic surfaces with low voltage backscattered electrons in a field emission scanning electron microscope. J Microsc 199(2):115–23
Rothermel M, Wenzel K, Fritsch D, Haala N (2012). SURE: Photogrammetric surface reconstruction from imagery. In: Proceedings of the LC3D Workshop, Berlin
Roy S, Meunier J, Marian AM, Vidal F, Brunette I, Costantino S (2012). Automatic 3D reconstruction of quasi-planar stereo Scanning Electron Microscopy (SEM) images. In: 34th Annual International Conference IEEE EMBS, pp 4361–4364
Scheuvens D, Kandler K, Küpper M, Lieke K, Zorn SR, Ebert M, Schütz L, Weinbruch S (2011) Individual-particle analysis of airborne dust samples collected over Morocco in 2006 during SAMUM 1. Tellus B 63(4):512–30
Schultz MA (2004) Multiple View 3D Reconstruction of Micro- to Nano-Scopic Specimens. MSc Thesis, The University of Tennessee, Knoxville, 122 p
Themelis G, Chikwembani S, Weertman J (1990) Determination of the orientation of Cu-Bi grain boundary facets using a photogrammetric technique. Mater Charact 24:27–40
Tuytelaars T, Van Gool L (2004) Matching widely separated views based on affine invariant regions. Int J Comput Vision 59(1):61–85
Veihelmann B, Nousiainen T, Kahnert M, van der Zande WJ (2006) Light scattering by small feldspar particles simulated using the Gaussian random sphere geometry. J Quant Spectrosc Radiat Transfer 100:393–405
Vijendran S, Sykulska H, Pike WT (2007) AFM investigation of Martian soil simulants on micromachined Si substrates. J Microsci 227(3):236–45
Woodward X (2014) The Three Dimensional Shape and Roughness of Mineral Dust. Master's report, Michigan Technological University, 20 p
Ylimäki M, Kannala J, Holappa J, Heikkilä J, Brandt SS (2012) Robust and accurate multi-view reconstruction by prioritized matching. In: Proceedings of the 21st International Conference Pattern Recognition, pp 2673–6
Yurkin MA, Hoekstra AG (2011) The discrete-dipole approximation code ADDA: capabilities and known limitations. J Quant Spectrosc Radiat Transfer 112(13):2234–47
Zeiss. http://www.zeiss.com/microscopy/en_de/products/scanning-electron-microscopes/upgrades/3d-surface-modelling.html. Accessed 17 March 2016
Zhang Z (1994) Iterative point matching for registration of free-form curves and surfaces. Int J Comput Vision 13(2):119–52
Zolotukhin AA, Safonov IV, Kryzhanovskii KA (2013) 3D reconstruction for a scanning electron microscope. Pattern Recogn Image Anal 23(1):168–74
Zubko E, Muinonen K, Muñoz O, Nousiainen T, Shkuratov Y, Sun W, Videen G (2013) Light scattering by feldspar particles: comparison of model agglomerate debris particles with laboratory samples. J Quant Spectrosc Radiat Transfer 131:175–87

Index

A. Kokhanovsky (ed.), *Springer Series in Light Scattering*, Springer Series in Light Scattering, https://doi.org/10.1007/978-3-319-70796-9