



NATO Science for Peace and Security Series - C:
Environmental Security

Polarimetric Detection, Characterization, and Remote Sensing

Edited by
Michael I. Mishchenko
Yaroslav S. Yatskiv
Vera K. Rosenbush
Gorden Videen

 Springer



*This publication
is supported by:*

The NATO Science for Peace
and Security Programme

Polarimetric Detection, Characterization, and Remote Sensing

NATO Science for Peace and Security Series

This Series presents the results of scientific meetings supported under the NATO Programme: Science for Peace and Security (SPS).

The NATO SPS Programme supports meetings in the following Key Priority areas: (1) Defence Against Terrorism; (2) Countering other Threats to Security and (3) NATO, Partner and Mediterranean Dialogue Country Priorities. The types of meeting supported are generally "Advanced Study Institutes" and "Advanced Research Workshops". The NATO SPS Series collects together the results of these meetings. The meetings are co-organized by scientists from NATO countries and scientists from NATO's "Partner" or "Mediterranean Dialogue" countries. The observations and recommendations made at the meetings, as well as the contents of the volumes in the Series, reflect those of participants and contributors only; they should not necessarily be regarded as reflecting NATO views or policy.

Advanced Study Institutes (ASI) are high-level tutorial courses to convey the latest developments in a subject to an advanced-level audience

Advanced Research Workshops (ARW) are expert meetings where an intense but informal exchange of views at the frontiers of a subject aims at identifying directions for future action

Following a transformation of the programme in 2006 the Series has been re-named and re-organised. Recent volumes on topics not related to security, which result from meetings supported under the programme earlier, may be found in the NATO Science Series.

The Series is published by IOS Press, Amsterdam, and Springer, Dordrecht, in conjunction with the NATO Emerging Security Challenges Division.

Sub-Series

A. Chemistry and Biology	Springer
B. Physics and Biophysics	Springer
C. Environmental Security	Springer
D. Information and Communication Security	IOS Press
E. Human and Societal Dynamics	IOS Press

<http://www.nato.int/science>

<http://www.springer.com>

<http://www.iospress.nl>



Series C: Environmental Security

Polarimetric Detection, Characterization, and Remote Sensing

edited by

Michael I. Mishchenko

NASA Goddard Institute for Space Studies
New York, NY
USA

Yaroslav S. Yatskiv

Main Astronomical Observatory
National Academy of Sciences
Kyiv, Ukraine

Vera K. Rosenbush

Main Astronomical Observatory
National Academy of Sciences
Kyiv, Ukraine

and

Gorden Videen

US Army Research Laboratory
Adelphi, MD
USA



Springer

Published in cooperation with NATO Emerging Security Challenges Division

Proceedings of the NATO Advanced Study Institute on Special Detection
Technique (Polarimetry) and Remote Sensing
Kyiv, Ukraine
12–25 September, 2010

Library of Congress Control Number: 2011930098

ISBN 978-94-007-1638-4 (PB)
ISBN 978-94-007-1635-3 (HB)
ISBN 978-94-007-1636-0 (e-book)
DOI 10.1007/978-94-007-1636-0

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springer.com

Printed on acid-free paper

All Rights Reserved

© Springer Science + Business Media B.V. 2011

No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

Contents

Preface: polarimetric detection, characterization, and remote sensing	vii
<i>M. I. Mishchenko, Ya. S. Yatskiv, V. K. Rosenbush, and G. Videen</i>	
Contributors	xix
Stellar spectropolarimetry: basic principles, observing strategies, and diagnostics of magnetic fields	1
<i>S. Bagnulo</i>	
Reflection symmetry of a sphere's internal field and its consequences on scattering: behavior of the Stokes parameters	31
<i>M. J. Berg</i>	
Light scattering by large faceted particles	49
<i>A. G. Borovoi and N. V. Kustova</i>	
Regularization of inverse problems in atmospheric remote sensing	79
<i>A. Doicu, Th. Trautmann, and F. Schreier</i>	
Light scattering resonances in small particles with electric and magnetic optical properties	117
<i>B. García-Cámara, F. González, F. Moreno, and G. Videen</i>	
Laboratory measurements of light scattered by clouds and layers of solid particles using an imaging technique	137
<i>E. Hadamcik, J.-B. Renard, A. C. Levasseur-Regourd, and J. Lasue</i>	
High sensitivity polarimetry: techniques and applications	177
<i>J. Hough</i>	
Aerosol retrievals under partly cloudy conditions: challenges and perspectives	205
<i>E. Kassianov, M. Ovchinnikov, L. K. Berg, and C. Flynn</i>	
Astrophysical polarimetry in Ukraine	233
<i>N. N. Kiselev and M. I. Mishchenko</i>	
Photopolarimetric remote sensing of aggregates in cosmic dust	261
<i>L. Kolokolova</i>	

Astrobiological remote sensing with circular polarization	277
<i>L. Kolokolova, W. Sparks, and D. Mackowski</i>	
Inferring properties of dust within small bodies of the solar system through observations and simulations of the linear polarization of scattered solar light	295
<i>A.-C. Levasseur-Regourd</i>	
Semi-empirical BRDF and BPDF models applied to the problem of aerosol retrievals over land: testing on airborne data and implications for modeling of top-of-atmosphere measurements	313
<i>P. Litvinov, O. Hasekamp, B. Cairns, and M. Mishchenko</i>	
An estimation of surface albedo from the SEVIRI/MSG observing system by using POLDER BRDF measurements	341
<i>I. Pokrovsky, O. Pokrovsky, and J.-L. Roujean</i>	
Plasmonic spectroscopy of 2D densely packed and layered metallic nanostructures	383
<i>A. N. Ponyavina and S. M. Kachan</i>	
Opposition optical phenomena in planetary astrophysics: observational results	409
<i>V. K. Rosenbush and M. I. Mishchenko</i>	
Mueller-matrix characterization of biological tissues	437
<i>S. N. Savenkov</i>	
Plasmonic optical properties and the polarization modulation technique	473
<i>B. K. Serdega, S. P. Rudenko, L. S. Maksimenko, and I. E. Matyash</i>	
Inferring microstructure and turbulence properties in rain through observations and simulations of signal spectra measured with Doppler–polarimetric radars	501
<i>F. Yanovsky</i>	
Index	543

Preface

Polarimetric detection, characterization, and remote sensing

Michael Mishchenko^{1*}, Yaroslav S. Yatskiv², Vera K. Rosenbush²,
and Gorden Videen³

¹ NASA Goddard Institute for Space Studies, 2880 Broadway, New York, NY 10025, USA

² Main Astronomical Observatory of the National Academy of Sciences of Ukraine, 27
Akademika Zabolotnoho St., 03680 Kyiv, Ukraine

³ Army Research Laboratory, 2800 Powder Mill Road, Adelphi, Maryland, 20783, USA

The NATO Advanced Study Institute (ASI) on “Special Detection Technique (Polarimetry) and Remote Sensing” took place in Kyiv, Ukraine, 12–25 September 2010. The main focus of the meeting was photopolarimetry, a rapidly developing, multidisciplinary topic with numerous military, ecological remote-sensing, astrophysical, biomedical, and technological applications [1–9]. Typical remote-sensing instruments measure the total intensity of the light scattered by a system of interest. Although the results of such measurements can be extremely valuable, they carry only a fraction of potentially useful information contained in the scattered light. The remaining information is coded in the polarization state of the light. As the need for accurate optical characterization and diagnostic techniques is increasing, it is important to find improved ways of extracting the additional information contained within the polarization state of the measured light. Advanced polarimetric methodologies are currently being used to

- detect the presence of biological-warfare-agent aerosols that may threaten military and civilian populations,
- monitor environmental effects and climate of our own planet Earth,
- detect the extent of cancerous regions within the human body,
- characterize nano-structures on substrates, and
- characterize structural elements of astrophysical bodies like comets and satellites.

Various polarization techniques have largely been developed independently by small research groups within specific scientific disciplines. Because the field is extremely interdisciplinary and the number of research groups actively utilizing polarization information often represents only a small fraction of the scientists within a remote-sensing, *in situ*, or laboratory optical-characterization discipline,

* Corresponding author. E-mail: mmishchenko@giss.nasa.gov



The conference venue: resort *Koncha-Zaspa* located in the picturesque and quiet “green zone” of Kyiv.

no single conference or workshop had been organized to address this field in general. With any interdisciplinary field, it is important for the players to meet and interact, so that they can discuss and build on the methods that were successful, and especially to teach students these techniques. It was, therefore, anticipated that bringing this diverse group of scientists together to teach and discuss different aspects of this one specific topic would encourage future collaborative efforts among scientists working on similar problems in different fields of research who otherwise would not have this opportunity.

The ASI was organized in such a way that much of the key research was presented by experts most familiar with the respective major topics. Additional contributed presentations dealing with aspects related to the feature talks were made by scientists and students having deep working knowledge of the particular nuances. As with any new interdisciplinary line of research, it was anticipated—



ASI Director Yaroslav Yatskiv welcomes the participants at the Main Astronomical Observatory.



The conference facility of the resort *Koncha-Zaspa*.

correctly—that some of the key barriers in one discipline had been worked out in other disciplines. This made the specific ASI format especially appropriate and extremely beneficial to all participants. We then identified and discussed the key common problems that we are trying to solve in this field along with what research needs to be performed to acquire the knowledge and research techniques to solve these problems.

While the formal lectures, shorter oral talks, and three poster sessions were key components of the ASI, another critical component was providing opportunities for direct personal interactions. Not only was this important for lecturers to further clarify key points of their tutorials and provide detailed explanations on a

NASA **NATO OTAN** *Advanced Study Institute*

NATO ADVANCED STUDY INSTITUTE
"SPECIAL DETECTION TECHNIQUE (POLARIMETRY)
AND
REMOTE SENSING"
 Kyiv, Ukraine, 12–25 September 2010

CHAIRS
Michael Mishchenko
 (NASA Goddard Institute for Space Studies, New York, USA)
Yaroslav Yatskiv
 (Main Astronomical Observatory, NAS of Ukraine, Kyiv, Ukraine)

SCIENTIFIC ORGANIZING COMMITTEE
James Hough (University of Hertfordshire, Hatfield, UK.)
Hal Maring (NASA Headquarters, Washington, DC, USA)
M. Pinar Mengüç (University of Kentucky, Lexington, KY, USA;
 Ozyegin University, Istanbul, Turkey)
Vera Rosenbush (Main Astronomical Observatory, NAS of Ukraine, Kyiv, Ukraine)
Gorden Videen (US Army Research Laboratory, Adelphi, MD, USA)

TOPICS
 Basics of polarization and measurement techniques
 Astronomical and terrestrial applications
 Military and industrial applications
 Medical diagnostic techniques
 Novel methodologies

INVITED LECTURERS
Stefano Bagnulo (Armagh Observatory, UK)
Adrian Doicu (Remote Sensing Technology Institute, Germany)
Oleg Dubovik (University of Lille, France)
James Hough (University of Hertfordshire, UK)
Nikolai Khlebtsov (IEPPM, Russia)
Theodor Kostluk (NASA GSFC, USA)
Anny-Chantal Levasseur-Regourd (University of Paris, France)
Hal Maring (NASA HQ, Washington, DC, USA)
M. Pinar Mengüç (University of Kentucky, USA; Ozyegin University, Turkey)
Michael Mishchenko (NASA GISS, USA)
Fernando Moreno (University of Cantabria, Spain)
Karri Muinonen (University of Helsinki, Finland)
Yuriy Shkuratov (Kharkiv National University, Ukraine)
Gorden Videen (Army Research Laboratory, USA)
Nikolai Voshchinnikov (St. Petersburg University, Russia)
Yaroslav Yatskiv (Main Astronomical Observatory, Ukraine)

LOCAL CONTACT ADDRESS
 NATO ASI-2010
 Main Astronomical Observatory
 National Academy of Sciences of Ukraine
 27 Akademika Zabolotnoho St.
 03680 Kyiv
 UKRAINE
 Fax: +380 (44) 526 2147
 Tel: +380 (44) 526 0869
 E-mail: rosevera@mao.kiev.ua

VENUE
 Resort "Koncha-Zaspa"
 03084 Kyiv
 Stolychne Shose 215
 www.koncha-zaspa.kiev.ua

LOCAL ORGANIZING COMMITTEE
 Chair: Vera Rosenbush
 Secretary: Irina Kulyk
 Tamara Bulba
 Alexandra Ivanova
 Nikolai Kiselev
 Tanya Korsun
 Nadya Kostogryz
 Tanya Nikityuk

WEB-SITE
<http://www.giss.nasa.gov/staff/mishchenko/Ukraine-2010/nato-2010.htm>

SPONSORS
 NATO OTAN, The NATO Alliance for Peace and Prosperity, NASA, ESA, IAGLR, National Science Foundation

NATO ASI poster.

more personal level, but it was also imperative to provide ample time for communications and exchanges of ideas that could ultimately lead to long-term collaborations helping advance the field into the future. The Local Organizing Committee (LOC) of the ASI worked hard to arrange such activities, which included an opening reception, two formal conference dinners, a field trip to the Main Astronomical Observatory (MAO), a *JQSRT* Session and Reception, and two tours of Kyiv.

It was especially appropriate that this meeting took place in Ukraine. First of all, the 2010 ASI built on the great success of the 2003 ASI on “Photopolarimetry



The official ASI logo and book of abstracts.

in Remote Sensing” convened in Yalta [3,4]. Second of all, Ukrainian astrophysicists have performed pioneering research in both measuring and interpreting photopolarimetric signals from remote astronomical objects [9]. These techniques have subsequently been extended to numerous other fields, including environmental remote-sensing of aerosol loading and terrestrial changes, such as the effects of pollution, erosion, and desertification. In this regard, the field trip to MAO, concluded by a lively picnic, was particularly instructive and productive.

In summary, the goal of the ASI was to present high-level tutorial courses on the most recent advances in polarimetric detection, characterization, and remote sensing, including military and environmental monitoring as well as terrestrial, atmospheric, and biomedical characterization. We discussed and taught techniques developed in various disciplines to acquire information from the polarization signal of scattered electromagnetic waves. We identified techniques that have been



Gorden Videen, Hal Maring, and Pinar Mengüç (from left to right) enjoy their time in Kyiv.



Michael Mishchenko, Nikolai Khlebtsov, and Vera Rosenbush (from left to right) discuss planetary opposition phenomena during the picnic at MAO.

especially successful for various applications and the future needs of the research communities. It is hoped that the inclusion of researchers from various disciplines will lead to cross-pollination of ideas and foster collaborations that will improve research efficiency. We did our best to provide the necessary elements to commence fruitful collaborations, i.e., food, drink, and discussion, and are confident that the participants were able to turn this great opportunity to their advantage.

An integral part of the ASI was the presentation of two memorable peer awards. The first one, the Special Prize of the LOC, was given to Professor Jim Hough for his exceptional lecture course on high-sensitivity polarimetry and its applications. The second one, the Young Scientist Award of the *Journal of Quan-*

titative Spectroscopy and Radiative Transfer (JQSRT) was presented to Dr. Pavel Litvinov for his outstanding contributions to the field of electromagnetic scattering by particles. We wholeheartedly congratulate both awardees on the well-deserved distinctions. The award ceremony was part of the second conference dinner, which also featured an utterly brilliant and rousing concert given by the enthusiastic members of the Ukrainian Folk Band of the Kyiv Taras Shevchenko National University.

Given the growing importance of polarimetry and the large amount of useful material presented at the ASI, it was essential to archive the well-established and new knowledge in the form of appropriate publications available to the entire scientific community. By its very design, this volume contains only tutorial reviews of specific fields of research, with a minimum of original material. Recent original results as well as a few additional reviews will be published in a special issue of *JQSRT* in the form of full-size peer-reviewed papers.

We thank all lecturers for the willingness to contribute their time and extensive knowledge and to provide instructive and illuminating tutorials that formed the backbone of this ASI. We also thank all the student participants (ranging from actual PhD students to senior researchers) for their energy and enthusiasm without which this ASI would not have happened. Special recognition is owed to the MAO research staff who were responsible for all aspects of local organization, overcame numerous obstacles, and made possible all the good things that happened during the two memorable weeks in Kyiv. We also thank the management and staff of the resort *Koncha-Zaspa* for fine cooperation and warm hospitality at their beautiful facility. Last but not least, we thank all contributors to this volume for providing outstanding chapters and responding to our numerous editorial requests in a timely manner.

Acknowledgments

The primary support for the NATO ASI was provided by the NATO Science for Peace and Security Programme. Major funding was also contributed by the NASA Radiation Sciences Program managed by Dr. Hal Maring (NASA Headquarters). Supplementary funding was provided by the US Office of Naval Research Global, the US Army International Technology Center – Atlantic, and Elsevier. Several travel fellowships were awarded by the National Science Foundation of the USA.

The NATO ASI was also sponsored by the National Academy of Sciences of Ukraine (NASU) and its Main Astronomical Observatory. We thank executive members of the NASU Headquarters Academicians Anton G. Naumovets and Vadym M. Loktev for their kind support of the ASI.

Major logistical assistance was provided by Carl Codan, Patricia Formosa, and Nadia Zakharova of Sigma Space Partners, LLC (New York, NY). Nadia Zakharova provided comprehensive editorial assistance in the preparation of this book.



The second conference dinner featured Ukrainian folk dances and songs.



The lively *JQSR*T reception sponsored by Elsevier.

Any opinions, findings and conclusions or recommendations expressed in this volume are those of the authors and do not necessarily reflect the views of the NATO Science for Peace and Security Programme, NASA, the US Army International Technology Center – Atlantic, the US Office of Naval Research Global, Elsevier, or the US National Science Foundation.



Organizational structure of the NATO ASI

Directors

Michael Mishchenko *NASA Goddard Institute for Space Studies, New York, USA*
Yaroslav Yatskiv *Main Astronomical Observatory, Kyiv, Ukraine*

Scientific organizing committee

James Hough *University of Hertfordshire, Hatfield, UK*
Hal Maring *NASA Headquarters, Washington, DC, USA*
M. Pinar Mengüç *University of Kentucky, Lexington, KY, USA; and Ozyegin University, Istanbul, Turkey*
Vera Rosenbush *Main Astronomical Observatory, Kyiv, Ukraine*
Gorden Videen *US Army Research Laboratory, Adelphi, MD, USA*

Local organizing committee

Tamara Bul'ba
Alexandra Ivanova
Nikolai Kiselev
Tanya Korsun
Nadya Kostogryz
Irina Kulyk *Secretary*
Tanya Nikityuk
Vera Rosenbush *Chair*

Main lecturers

Stefano Bagnulo *Armagh Observatory, UK*
Adrian Doicu *Remote Sensing Technology Institute, Germany*
Oleg Dubovik *University of Lille, France*
James Hough *University of Hertfordshire, UK*
Vadym Kaydash *Kharkiv National University, Ukraine*
Nikolai Khlebtsov *Institute of Biochemistry and Physiology of Plants and Microorganisms, Russia*
Theodor Kostiuik *NASA Goddard Space Flight Center, USA*
Anny-Chantal Levasseur-Regourd *University of Paris, France*
Hal Maring *NASA Headquarters, USA*
M. Pinar Mengüç *University of Kentucky, USA; and Ozyegin University, Turkey*
Michael Mishchenko *NASA Goddard Institute for Space Studies, USA*
Karri Muinonen *University of Helsinki, Finland*
Gorden Videen *Army Research Laboratory, USA*
Nikolai Voshchinnikov *St. Petersburg University, Russia*
Yaroslav Yatskiv *Main Astronomical Observatory, Ukraine*

References

1. Videen, G., and M. Kocifaj, Eds., 2002: *Optics of Cosmic Dust* (Kluwer, Dordrecht, The Netherlands).
2. Hovenier, J. W., C. van der Mee, and H. Domke, 2004: *Transfer of Polarized Light in Planetary Atmospheres* (Springer, Berlin).



Concert and picnic at the Main Astronomical Observatory.



Working session of the ASI.

3. Videen, G., Ya. Yatskiv, and M. Mishchenko, Eds., 2004: *Photopolarimetry in Remote Sensing* (Kluwer, Dordrecht, The Netherlands).
4. Videen, G., Ya. S. Yatskiv, and M. I. Mishchenko, Eds., 2004: Special issue on photopolarimetry in remote sensing. *J. Quant. Spectrosc. Radiat. Transfer* **88**, 1–406.
5. Tuchin, V. V., L. V. Wang, and D. A. Zimnyakov, 2006: *Optical Polarization in Biomedical Applications* (Springer, Berlin).
6. Mishchenko, M. I., L. D. Travis, and A. A. Lacis, 2006: *Multiple Scattering of Light by Particles* (Cambridge University Press, Cambridge, UK).
7. Hoekstra, A., V. Maltsev, and G. Videen, Eds., 2007: *Optics of Biological Particles* (Springer, Dordrecht, The Netherlands).
8. Clarke, D., 2010: *Stellar Polarimetry* (Wiley-VCH, Weinheim, Germany).
9. Mishchenko, M. I., V. K. Rosenbush, N. N. Kiselev, et al., 2010: *Polarimetric Remote Sensing of Solar System Objects* (Akademperiodyka, Kyiv) (arXiv:1010.1171).

Contributors

Stefano Bagnulo, Armagh Observatory, College Hill, Armagh BT61 9DG, Northern Ireland, UK

Larry K. Berg, Pacific Northwest National Laboratory, Richland, WA 99352, USA

Matthew J. Berg, Mississippi State University, Department of Physics & Astronomy, Mississippi State, MS 39762, USA

Anatoli G. Borovoi, V. E. Zuev Institute of Atmospheric Optics, Academician Zuev Sq. 1, Tomsk 634021, Russia

Brian Cairns, NASA Goddard Institute for Space Studies, 2880 Broadway, New York, NY 10025, USA

Adrian Doicu, German Aerospace Center, Remote Sensing Technology Institute, Oberpfaffenhofen, Germany

Connor Flynn, Pacific Northwest National Laboratory, Richland, WA 99352, USA

Braulio García-Cámara, Grupo de Óptica, Departamento de Física Aplicada, Universidad de Cantabria, Avda. de los Castros s/n 39005, Santander, Spain

Francisco González, Grupo de Óptica, Departamento de Física Aplicada, Universidad de Cantabria, Avda. de los Castros s/n 39005, Santander, Spain

Edith Hadamcik, UPMC University Paris 06/CNRS, LATMOS-IPSL, 11 Bld D'Alembert, 78280 Guyancourt, France

Otto Hasekamp, SRON Netherlands Institute for Space Research, Sorbonnelaan 2, 3584 CA Utrecht, The Netherlands

James Hough, Centre for Astrophysics Research, University of Hertfordshire, College Lane, Hatfield AL10 9AB, UK

Svetlana M. Kachan, Belarusian National Technical University, Khmel'nitskogo Str. 9, Minsk, 220013, Belarus

Evgueni Kassianov, Pacific Northwest National Laboratory, Richland, WA 99352, USA

Nikolai N. Kiselev, Main Astronomical Observatory of the National Academy of Sciences of Ukraine, 27 Akademika Zabolotnoho St., 03680 Kyiv, Ukraine

Ludmilla Kolokolova, University of Maryland, College Park, MD 20742, USA

- Natalia V. Kustova**, V. E. Zuev Institute of Atmospheric Optics, Academician Zuev Sq. 1, Tomsk 634021, Russia
- Jeremie Lasue**, LANL, Space Science and Applications, ISR-1, Mail Stop D-466, Los Alamos, NM 87545, USA; and Lunar and Planetary Institute, 3600 Bay Area Blvd., Houston, TX 77058, USA
- Anny-Chantal Levasseur-Regourd**, UPMC Univ. Paris 06, UMR 8190 (LATMOS–CNRS), BC 102, 4 place Jussieu, 75005 Paris, France
- Pavel Litvinov**, SRON Netherlands Institute for Space Research, Sorbonnelaan 2, 3584 CA Utrecht, The Netherlands
- Mikhail Ovchinnikov**, Pacific Northwest National Laboratory, Richland, WA 99352, USA
- Daniel Mackowski**, Department of Mechanical Engineering, Auburn University, AL 36849, USA
- Lidiya S. Maksimenko**, V. Lashkaryov Institute of Semiconductor Physics, National Academy of Sciences of Ukraine, 45 Prospekt Nauky, 03028 Kyiv, Ukraine
- Igor E. Matyash**, V. Lashkaryov Institute of Semiconductor Physics, National Academy of Sciences of Ukraine, 45 Prospekt Nauky, 03028 Kyiv, Ukraine
- Michael I. Mishchenko**, NASA Goddard Institute for Space Studies, 2880 Broadway, New York, NY 10025, USA
- Fernando Moreno**, Grupo de Óptica, Departamento de Física Aplicada, Universidad de Cantabria, Avda. de los Castros s/n 39005, Santander, Spain
- Igor Pokrovsky**, Voeikov Main Geophysical Observatory, 7 Karbyshev Str., St.-Petersburg, 194021, Russia
- Oleg Pokrovsky**, Voeikov Main Geophysical Observatory, 7 Karbyshev Str., St.-Petersburg, 194021, Russia
- Alina N. Ponyavina**, Institute of Physics, National Academy of Sciences of Belarus, Nezavisimosti Ave. 68, Minsk, 220072, Belarus
- Jean-Baptiste Renard**, LPC2E/CNRS, 3A Avenue de la Recherche Scientifique, 45071 Orleans Cedex 2, France
- Vera K. Rosenbush**, Main Astronomical Observatory of the National Academy of Sciences of Ukraine, 27 Akademika Zabolotnoho St., 03680 Kyiv, Ukraine
- Jean-Louis Roujean**, CNRM/GAME (Météo-France/CNRS), 42, avenue Gaspard Coriolis, 31057 Toulouse Cedex, France
- Svetlana P. Rudenko**, V. Lashkaryov Institute of Semiconductor Physics, National Academy of Sciences of Ukraine, 45 Prospekt Nauky, 03028 Kyiv, Ukraine

Sergey N. Savenkov, Department of Radiophysics, Kyiv Taras Shevchenko University, Kyiv, Ukraine

Franz Schreier, German Aerospace Center, Remote Sensing Technology Institute, Oberpfaffenhofen, Germany

Boris K. Serdega, V. Lashkaryov Institute of Semiconductor Physics, National Academy of Sciences of Ukraine, 45 Prospekt Nauky, 03028 Kyiv, Ukraine

William Sparks, Space Telescope Science Institute, Baltimore, MD 21218, USA

Thomas Trautmann, German Aerospace Center, Remote Sensing Technology Institute, Oberpfaffenhofen, Germany

Gorden Videen, Army Research Laboratory, 2800 Powder Mill Road, Adelphi, Maryland, 20783, USA

Felix Yanovsky, National Aviation University, Prospect Komarova 1, 03680 Kyiv, Ukraine

Yaroslav S. Yatskiy, Main Astronomical Observatory of the National Academy of Sciences of Ukraine, 27 Akademika Zabolotnoho St., 03680 Kyiv, Ukraine



Tour of the Museum of Ukrainian Folk Architecture.

Stellar spectropolarimetry: basic principles, observing strategies, and diagnostics of magnetic fields

Stefano Bagnulo*

Armagh Observatory, College Hill, Armagh BT61 9DG, Northern Ireland, UK

Abstract. A large fraction of night-time astronomical polarimetric measurements is aimed at the detection of stellar magnetic fields. These kinds of measurements are often performed at the limit of instrumental capabilities, and it is of paramount importance to employ observational and data-reduction techniques that allow one to reach the highest possible precision. This chapter is a review of these techniques ranging from the basic principles to the presentation of recent results. I will start with the definition of the Stokes parameters and with an analytical characterization of the optical devices that are most commonly used to measure the polarization of light for objects of astrophysical interest. I will then discuss the observational and data-reduction techniques that help minimize the instrumental effects, allowing one at the same time to perform basic quality checks of the measured signals. Next, I will recapitulate the Zeeman effect and show how analyses of Stokes-parameter profiles can be exploited for the detection and modeling of stellar magnetic fields. Lastly, I will describe a method for field detection based on high-resolution spectropolarimetry and a method based on low-resolution spectropolarimetry.

Keywords: polarization, spectropolarimetry, stellar magnetic fields

1. Introduction

The study of stellar magnetism through circular spectropolarimetry is by far the most common application of polarimetric techniques to astronomical objects. The high-resolution spectropolarimeters ESPaDOnS of the 4-m Canada–France–Hawaii Telescope (CFHT), its twin Narval[†] of the 2-m Bernard Lyot Telescope of the Pic-du-Midi Observatory (France), and the low-resolution FORS spectrograph and imager (Appenzeller et al. 1998) of the 8-m ESO Very Large Telescope (VLT) are currently the workhorse instruments for this kind of studies. The high-resolution HARPS instrument (Mayor et al. 2003) of the ESO’s 3.6-m telescope

* Corresponding author. E-mail: sba@arm.ac.uk

[†] Both the ESPaDOnS and Narval instruments are described at <http://www.ast.obs-mip.fr/projets/espadons/espadons.html>

has been recently equipped with polarimetric optics, and its commissioning data have already shown that the instrument will be able to deliver high-quality results (Snik et al. 2010). During the last 10 years, about 1000 h of telescope time have been allocated for the investigation of stellar magnetic fields with the FORS instrument, while the ESPaDOnS and Narval instruments are currently employed for similar purposes at the rate of a few hundred hours per year. While the time dedicated to polarimetry represents a tiny fraction of all astronomical observations, whenever a photon travels across polarimetric optics of an astronomical spectrograph, the chances are that it will eventually be used to study the magnetic field of a star.

Magnetic fields are present in a large variety of stars across the entire Hertzsprung–Russell (HR) diagram and play a fundamental role during the various evolutionary stages, from the collapse of the proto-stellar material through the end of the star’s life either as a white dwarf or a neutron star.

In low-mass main-sequence stars (up to ~ 1.5 solar masses), the presence of a magnetic field is often inferred from observations of solar-like-activity phenomena, such as dark spots and prominences. Magnetic fields are thought to be nearly ubiquitous in low-mass stars and originate within the star’s convective envelopes through a dynamo process. Among higher-mass stars, magnetic fields are more rarely observed, but when present, they often have a higher strength and a simpler morphology than in lower-mass stars. The dynamo hypothesis is hardly consistent with the observed high strengths, with the fact that only few early-type stars are magnetic, and with the lack of any obvious correlation between field strength and rotation rate. Therefore, the presence of a magnetic field in intermediate- to high-mass stars is commonly interpreted in terms of the fossil theory, according to which the observed fields are the remnants of a field existing in the earlier stages of stellar evolution: either a weak interstellar field present where the cloud was collapsing into a proto-star, or a field generated by a dynamo when the star was in the pre-main-sequence phase. The fact that no correlation exists between the rotation period and the magnetic field and that only a small percentage of upper-main-sequence stars is magnetic is naturally explained in terms of the amount of flux trapped during the star formation or early evolution. It is clear that understanding the role of magnetic fields in stellar evolution requires observational constraints everywhere in the HR diagram, a goal toward which much important progress has been achieved during the last decade.

The current state of our knowledge of stellar magnetism is thoroughly reviewed by Donati and Landstreet (2009), while a deep theoretical background is presented by Mestel (1999). This chapter is an introduction to the theoretical concepts, the astronomical instrumentation, and the observing and modeling techniques for the studies of stellar magnetic fields intended for a reader familiar with the use of *linear* polarimetry as a remote-sensing tool for the study of planetary surfaces and particles, but less familiar with the astrophysical application of *circular* spectropolarimetry. This review is limited to the remote-sensing techniques developed for the study of magnetic fields of non-degenerate stars. The Sun is a

well studied magnetic star, but the respective observing and modeling techniques differ from those commonly applied to stars; therefore, solar spectropolarimetry will be left outside the scope of this review.

I will start with the definition of the Stokes parameters (Section 2), summarize the basic principle underlying the measurement of polarized radiation (Section 3), and show how the analysis of the Zeeman effect on the Stokes-parameter profiles of the spectral lines allows one to detect and model magnetic fields of non-degenerate stars (Section 4). Finally, I will review the diagnostic techniques introduced over the last few decades (Section 5). This chapter cannot deal with all theoretical, instrumental, and observational details, but is intended to give a fairly complete overview of the relevant key points. Selected references to the literature will help to refine the general picture.

2. Definition of the Stokes parameters

The measurement of polarimetric quantities lacks a well-established standardization. While there is little risk of ambiguity in the definition of the fraction of linear polarization, ambiguities in the definition of its position angle and of all the Stokes parameters except the intensity are more the rule than the exception. These ambiguities are discussed in some detail, e.g., by Clarke (1974), Landi Degl’Innocenti et al. (2007), and Bagnulo et al. (2009). In the following, I will recall the definitions that are most commonly adopted in night-time optical astronomy.

Defining the Stokes parameters requires a preliminary choice of the reference direction x pertaining to the plane perpendicular to the direction of the propagation of the radiation. This choice is somewhat arbitrary, yet its definition is crucial for linear-polarization measurements. In stellar astronomy, it is common to choose the reference direction x as the celestial meridian passing through the observed object. In solar observations one can adopt the direction of the tangent to the solar limb (for prominence observations) or the direction passing through the observed point and the center of the solar disk (for sunspots observations). For observations of asteroids and other objects of the solar system, it is common to adopt the normal to the scattering plane (identified by the great circle passing through the object itself and the Sun).

We then consider a *right-handed* reference system (x, y, z) with the z -axis directed along the direction of propagation of the electromagnetic wave. It is possible to define the Stokes parameters following two different approaches. In the first one, the Stokes parameters are expressed as statistical averages of bilinear products of the components of the electric field vector along two perpendicular axes x and y . Rigorous definitions are given, e.g., in Landi Degl’Innocenti and Landolfi (2004). The second approach is based on the concept of ideal filters for linear and circular polarization (e.g., Shurcliff 1962). Both definitions are consistent with the schematic visualization in Fig. 1. The Stokes parameter Q is the difference between the intensity of the radiation with the electric field vector oscillating along

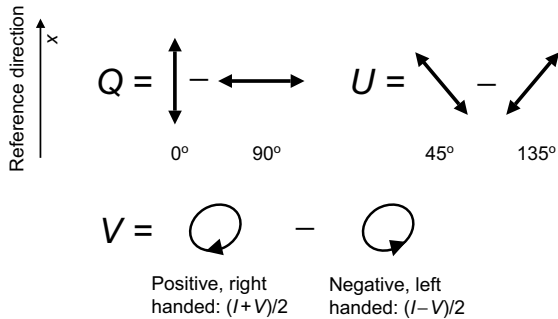


Fig. 1. Geometrical visualization of the definition of the Stokes parameters according to Shurcliff (1962) and Landi Degl’Innocenti and Landolfi (2004). The observer is looking toward the source.

the x -axis and that with the electric field oscillating along the y -axis. The Stokes parameter U is the difference between the intensity of the radiation with the electric field oscillating at 45° and that with the electric field oscillating at 135° with respect to the reference direction x (angles are reckoned counterclockwise from the reference direction when looking at the source). The fraction of linear polarization is given by

$$P_L = \frac{\sqrt{Q^2 + U^2}}{I} \quad (1)$$

and the position angle Θ is obtained by inverting the relationships

$$\begin{aligned} \frac{Q}{I} &= P_L \cos 2\Theta, \\ \frac{U}{I} &= P_L \sin 2\Theta. \end{aligned} \quad (2)$$

Inverting Eq. (2) requires some attention, and the position angle should be calculated as

$$\Theta = \frac{1}{2} \arctan \left(\frac{U}{Q} \right) + \Theta_0, \quad (3)$$

where $\Theta_0 = 0$ if $Q > 0$ and $U \geq 0$; $\Theta_0 = \pi$ if $Q > 0$ and $U < 0$; $\Theta_0 = \pi/2$ if $Q < 0$ †.

The Stokes parameter V is given by the difference between the right-handed and left-handed circular polarizations, defined so that at a fixed point in space, the tip of the electric field vector carried by a beam having positive circular polariza-

† Some programming language routines include the function `atan2`, which accepts two arguments, and should correctly include the constant Θ_0 .

tion rotates clockwise, as seen by an observer looking at the source of radiation. Conversely, the tip of the electric field vector carried by a beam having negative circular polarization rotates counterclockwise, as seen by an observer looking at the source.

In the majority of applications, it is customary to normalize the Stokes parameters Q , U , and V to I , and adopt the reduced parameters $P_Q = Q/I$, $P_U = U/I$, and $P_V = Q/V$.

The Stokes parameters provide a full description of the ellipse drawn by the tip of the electric field vector in the plane perpendicular to the direction of wave propagation. In particular, the position angle Θ of Eq. (3) is the angle that the major semi-axis of the ellipse forms with the reference direction x , reckoned counterclockwise; the ellipse's major and minor semi-axes a and b , expressed in dimensionless units, are given by (e.g., Landi Degl'Innocenti et al. 2007)

$$\begin{aligned} a &= \frac{1}{2}[(1 - P_V)^{1/2} + (1 + P_V)^{1/2}], \\ b &= \frac{1}{2}|(1 - P_V)^{1/2} - (1 + P_V)^{1/2}|. \end{aligned} \quad (4)$$

3. Measurement of the Stokes parameters in night-time optical astronomy

Polarimeters include one or more modulators (retarder wave plates or Fresnel rhombs) that can be either rapidly rotated or put in a series of fixed positions. Modulators are followed by a linear polarizer, typically a Wollaston prism.

In this section, I will first introduce the properties of the individual optical elements in the ideal case, following the approach of Landi Degl'Innocenti and Landolfi (2004). Then I will present the actual scheme that is most commonly adopted in imaging polarimeters and spectropolarimeters currently used in night-time astronomy. For a wider overview of the different techniques and instrument setups employed in astronomy see, e.g., Tinbergen (1996), Hough (2005, 2011), and references therein.

Waveplates are optical elements with two principal axes, one called the fast axis and the other one called the slow axis, characterized by two different refractive indices. The refractive index of the fast axis is smaller than the refractive index of the slow axis, so that linearly polarized light with its electric field vector parallel to the fast axis travels faster than linearly polarized light with its electric field vector parallel to the slow axis. A quarter-wave plate and a half-wave plate produce a phase retardation between the components of the electric fields along the fast and slow axes equal to $\pi/2$ and π , respectively. Wave plates are made of birefringent crystals (e.g., calcite); the birefringence properties depend on the wavelength and on the thickness of the crystal. Fresnel rhombs produce similar results as quarter-wave plates. The phase retardation is not due to the birefringence properties of the material, but rather is introduced by a double internal reflection. Fresnel rhombs can operate over a wider wavelength range than retarder wave plates.

A filter for linear polarization (called *polarizer*) is a device that can be interposed along a radiation beam and which, by definition, is totally transparent to the component of the electric field along a given direction perpendicular to the direction of propagation (the transmission axis of the polarizer), and totally opaque to the component of the electric field in the orthogonal direction.

Ideally, linear polarization can be measured using a polarizer set at different position angles on the sky (0° and 90° for the Stokes parameter Q ; 45° and 135° for the Stokes parameter U). A filter transmitting positive circular polarization can be realized with the combination of a quarter-wave plate followed by a polarizer whose transmission axis is rotated counterclockwise (looking at the source of radiation) by an angle of 45° with respect to the direction of the fast axis of the plate. A transmitting negative circular polarization has a quarter-wave plate followed by a polarizer whose transmission axis is rotated by an angle of -45° . Obviously, when the linear analyzer is aligned in the reference direction, a filter transmitting positive circular polarization is obtained by setting the fast axis of the retarder wave plate at an angle of -45° with respect to the reference direction, looking at the source of radiation, and a filter transmitting negative circular polarization is obtained by setting the fast axis of the retarder wave plate at an angle of $+45^\circ$ with respect to the reference direction.

Instead of a linear polarizer, it is often advantageous to use a Wollaston prism, which can be thought as a combination of two linear polarizers: the incoming radiation is split into two beams which define an *exit* plane, or *principal* plane. The two beams are polarized, one in the direction parallel to that plane (which we will call the *parallel* beam) and one in the direction perpendicular to it (which we will call the *perpendicular* beam). In principle, linear polarization can be measured using only a Wollaston prism and rotating the Wollaston or the entire instrument, but this will not be convenient for observations of extended sources. A linear polarizer filter is more conveniently realized by using a rotatable half-wave plate followed by a Wollaston prism.

For an arbitrary configuration of the polarimetric optics, in the geometrical scenario of Fig. 2, a detector that follows a retarder wave plate and a linear polarizer will measure the following signal (Landi Degl'Innocenti and Landolfi 2004):

$$S(\alpha, \beta, \gamma) \propto \frac{1}{2} [I + (Q \cos 2\alpha + U \sin 2\alpha) \cos(2\beta - 2\alpha) - (Q \sin 2\alpha - U \cos 2\alpha) \sin(2\beta - 2\alpha) \cos \gamma + V \sin(2\beta - 2\alpha) \sin \gamma], \quad (5)$$

where α as the angle between the reference direction and the fast axis of the retarder wave plate, measured counterclockwise from the reference direction; β is the position angle of the transmission axis of the linear polarizer measured counterclockwise from the reference direction; and γ is the phase retardance introduced by the retarder wave plate. If the analyzer is a Wollaston prism then this will be inserted into the beam in such a way that its parallel beam is parallel to the reference axis. In this way, the principal beam and the perpendicular beam will be as-

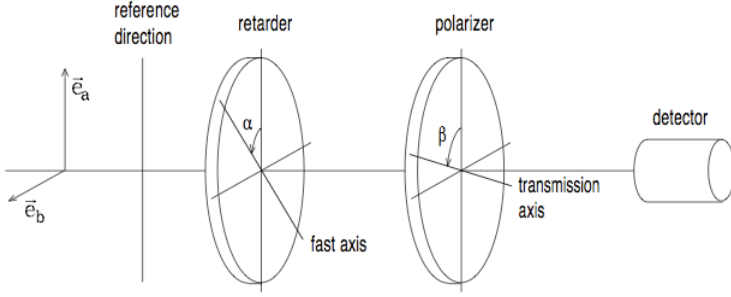


Fig. 2. Schematic representation of an ideal polarimeter (from Landi Degl’Innocenti and Landolfi 2004).

sociated with the β values 0° and 90° , respectively.

We define $G(\alpha, \gamma)$ as the ratio of the difference and the sum of the signals corresponding to the parallel and perpendicular beams measured with the retarder wave plate with a retardance phase γ at a position angle α :

$$G(\alpha, \gamma) = \frac{S(\alpha, \beta = 0^\circ, \gamma) - S(\alpha, \beta = 90^\circ, \gamma)}{S(\alpha, \beta = 0^\circ, \gamma) + S(\alpha, \beta = 90^\circ, \gamma)}. \quad (6)$$

The measurement of the reduced Stokes parameter $P_Q = Q/I$ can ideally be performed by setting the position angle α of the retarder wave plate to 0° or to any multiple of 45° , whereas the measurement of the reduced Stokes parameter $P_U = U/I$ can be performed by setting the angle α to 22.5° or to any angle differing from 22.5° by a multiple of 45° . The measurement of the reduced Stokes parameter $P_V = V/I$ can be performed by setting the angle α to -45° or to any angle differing from -45° by a multiple of 90° :

$$\begin{aligned} G(\alpha = 0^\circ, \gamma = \pi) &= -G(\alpha = 45^\circ, \gamma = \pi) = G(\alpha = 90^\circ, \gamma = \pi) = \dots \\ &= -G(\alpha = 315^\circ, \gamma = \pi) = P_Q, \\ G(\alpha = 22.5^\circ, \gamma = \pi) &= -G(\alpha = 67.5^\circ, \gamma = \pi) = G(\alpha = 112.5^\circ, \gamma = \pi) = \dots \\ &= -G(\alpha = 337.5^\circ, \gamma = \pi) = P_U, \\ G(\alpha = -45^\circ, \gamma = \pi/2) &= -G(\alpha = 45^\circ, \gamma = \pi/2) = G(\alpha = 135^\circ, \gamma = \pi/2) \\ &= -G(\alpha = 225^\circ, \gamma = \pi/2) = P_V. \end{aligned} \quad (7)$$

Equation (5) may be regarded as the analytical description of an “ideal polarimeter” realized with a retarder wave plate (or a Fresnel rhomb) and a Wollaston prism, while Eqs. (7) give the simplest observing recipes. In all practical applications, the behavior of a “real polarimeter” will deviate from the ideal case for a number of reasons. The three most important effects to be considered are the following:

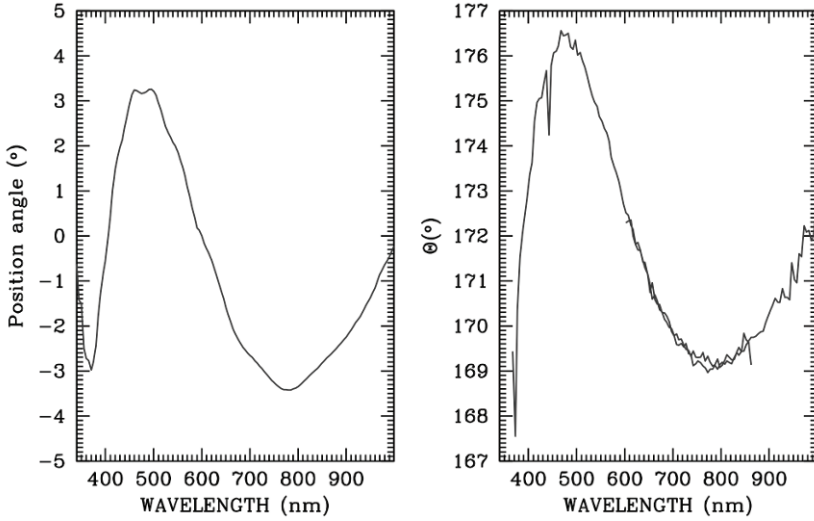


Fig. 3. Left-hand panel: position angle of the fast axis of the half-wave plate of the ESO VLT FORS instrument (set arbitrarily to zero for $\lambda = 6000 \text{ \AA}$) as measured in the laboratory. Right-hand panel: the position angle of the standard star for linear polarization Ve 6-23 observed with the FORS, obtained using Eq. (8) and without correcting for the chromatism of the retarder wave plate.

1. The transmission function in the principal beam will differ from the transmission function of the perpendicular beam, even for an unpolarized source. Accurate flat-fielding can mitigate this problem, but experience shows that it is hard to achieve accuracy better than 2%–3%.
2. Both the retardance angle and the position angle of the retarder wave plate depend on the wavelength and may deviate from the nominal values up to several degrees. As an example, the left-hand panel of Fig. 3 shows the position angle of the fast axis of the half-wave plate of the FORS instrument as a function of wavelength, arbitrarily setting to zero the value assumed at $\lambda = 6000 \text{ \AA}$.
3. In spectropolarimetric measurements, a less-than-perfect wavelength calibration of the two beams split by the Wollaston prism will introduce a spurious signal of polarization in spectral lines. As an example, let us consider an unpolarized spectral line approximated by a Gaussian curve. Suppose that the wavelength calibration introduces a spurious offset equal to $1/10$ of its FWHM between the parallel and perpendicular beams, which is probably not common but still represents a realistic scenario. Figure 4 shows that while the two Gaussian curves are nearly overlapping, the reduced Stokes-parameter profiles obtained with Eqs. (7) would show a spurious polarization signal of a $\sim 10\%$ amplitude.

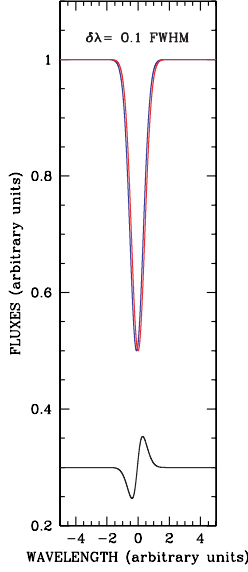


Fig. 4. Two Gaussian curves offset by 0.1 FWHM (top curves) and their difference divided by their sum (bottom curve, offset by +0.3 for display purposes).

These three problems are largely overcome by combining observations obtained with the retarder wave plate set to at least two different position angles. There exist two methods, “the difference method” and the “ratio method”, which are both based on the same observing recipes: circular polarization measurements should be obtained with the quarter-wave plate at position angles turned in steps of 90° starting from $\alpha = -45^\circ$; linear polarization measurements should be obtained with the half-wave plate turned in steps of 22.5° starting from $\alpha = 0^\circ$. Denoting by f^\parallel the flux measured in the parallel beam ($\beta = 0^\circ$) and by f^\perp the flux measured in the perpendicular beam ($\beta = 90^\circ$), following the difference method, the reduced Stokes parameters are then obtained as

$$P_X = \frac{1}{2N} \sum_{j=1}^N \left[\left(\frac{f^\parallel - f^\perp}{f^\parallel + f^\perp} \right)_{\alpha_j} - \left(\frac{f^\parallel - f^\perp}{f^\parallel + f^\perp} \right)_{\alpha_j + \Delta\alpha} \right]. \quad (8)$$

Following the ratio method, the reduced Stokes parameters are calculated as

$$P_X = \frac{\left[\prod_{j=1}^N \left(\frac{f^\parallel}{f^\perp} \right)_{\alpha_j} \left(\frac{f^\perp}{f^\parallel} \right)_{\alpha_j + \Delta\alpha} \right]^{1/(2N)} - 1}{\left[\prod_{j=1}^N \left(\frac{f^\parallel}{f^\perp} \right)_{\alpha_j} \left(\frac{f^\perp}{f^\parallel} \right)_{\alpha_j + \Delta\alpha} \right]^{1/(2N)} + 1}. \quad (9)$$

In both Eqs. (8) and (9), for $X=Q$, α_j belongs to the set $\{0^\circ, 90^\circ, 180^\circ, 270^\circ\}$ and $\Delta\alpha = 45^\circ$; for $X=U$, α_j belongs to the set $\{22.5^\circ, 112.5^\circ, 202.5^\circ, 292.5^\circ\}$ and $\Delta\alpha = 45^\circ$; for $X=V$, α_j belongs to the set $\{-45^\circ, 135^\circ\}$ and $\Delta\alpha = 90^\circ$. N represents the number of pairs of observations obtained at two different position angles.

Bagnulo et al. (2009) have analytically shown that the two methods give identical results to the first order. For circular polarization measurements, both methods allow one to cancel out the spurious effects introduced by deviations of the angles α , β , and γ from their nominal values, as well as by imperfections of the flat-fielding corrections and changes of the sky transparency. For linear polarization measurements, both methods compensate for deviations of the angles β and γ from their nominal values, as well as for differences in the transmission functions of the two beams and changes in sky transparency. However, it is still necessary to correct for the deviations of the position angle of the retarder wave plate, α , from its nominal value. Assuming, for instance, that the position angle of a linearly polarized source is constant with wavelength, the position angle obtained from observations reduced either with Eq. (8) or (9) will be wavelength-modulated in the same way as the curve of the half-wave plate position angle, and centered about the real value of the position angle of the star's polarization. The right-hand panel of Fig. 3 shows the position angle of the standard star for linear polarization Ve 6–23, observed with the FORS instrument, and obtained using the “difference method”. More details are given in Bagnulo et al. (2009) and Fossati et al. (2007).

Potentially, either the difference or the ratio method permits one to measure the Stokes parameters up to 10^{-4} accuracy even when the instrument is not fully characterized (which is common for the currently available facilities). However, there are still a number of additional instrumental effects that cannot be easily corrected. For instance, the ratio between the transmission functions of the two beams may vary when the modulator is rotated at different angles (especially in the case of fiber-fed instruments). This will introduce a spurious, although nearly featureless, polarization signal. If this is the case, polarimetric measurements of stellar spectral lines can still be achieved by assuming that the continuum is not polarized (which is generally true for the spectra of non-degenerate stars) and performing a kind of “rectification” of the Stokes-parameter profiles. For measurements of polarization in the continuum, a fiber-fed instrument is less suitable than a slit spectropolarimeter. Other unwanted effects include fringing introduced by the retarder wave plate (Aitkin and Hough 2001), which is largely mitigated by the use of Fresnel rhombs; the cross-talk from linear to circular polarization (or vice versa), which is typically introduced by the optics preceding the polarimeter, such as the instrument collimator, or the atmospheric dispersion corrector (e.g., Bagnulo et al. 2009); the fact that the slit jaws may absorb light polarized parallel to the jaws more efficiently than light polarized perpendicularly to the jaws, or that the grism transmission function may depend on the polarization of the incoming light. For a more thorough discussion of these and other effects, see Keller (2002).

In the majority of the astronomical observatories, the full characterization and correction of all these “minor” effects is generally neglected, since it would re-

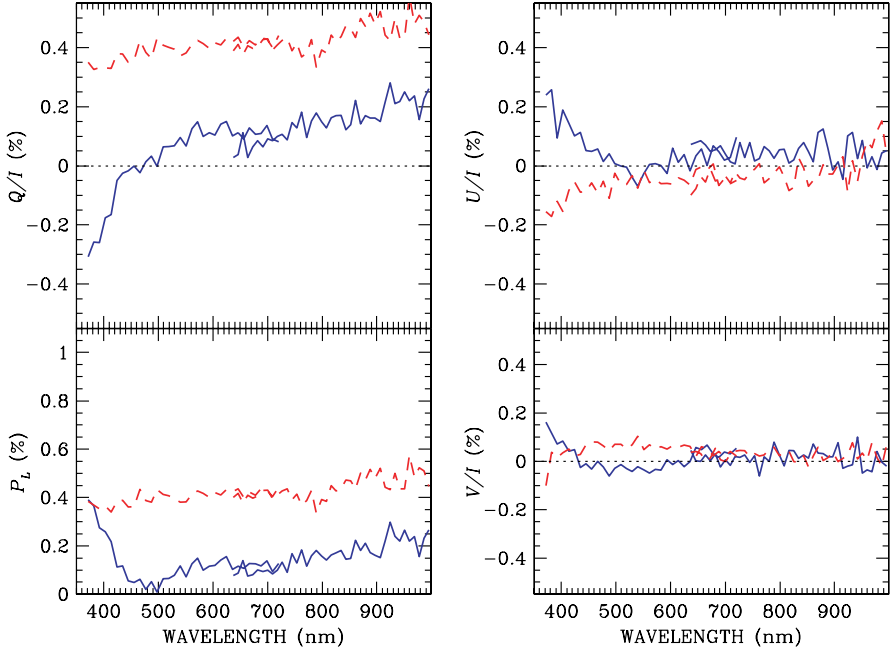


Fig. 5. Observed reduced Stokes-parameter profiles of HD 215373. Solid and dashed curves refer to the measurements obtained with the 0° and 90° instrument position angles on the sky, respectively. The observations were performed with gratings 300V and 300I.

quire a huge investment in terms of technical resources, compared to a relatively small pressure from the user side. Astronomers should check the instrument performance through observations of well known polarized and unpolarized standard stars. For point sources that do not vary on a short-time scale, a simple observing strategy that may help to assess the reliability of a polarimetric measurement is to repeat the same observations setting the instrument at different position angles on the sky. The reduced Stokes parameters P'_Q , P'_U , and P'_V measured at a position angle on the sky χ should be obtained from the reduced Stokes parameters P_Q , P_U , and P_V measured with the instrument position angle aligned to the North celestial pole using the following relationships:

$$\begin{aligned}
 P'_Q &= P_Q \cos 2\chi + P_U \sin 2\chi, \\
 P'_U &= -P_Q \sin 2\chi + P_U \cos 2\chi, \\
 P'_V &= P_V.
 \end{aligned}
 \tag{10}$$

Any deviation from this prediction will point to instrument problems. [Figure 5](#) shows the Stokes-parameter profiles of the star HD 215377, from which we expect a small signal of linear polarization and a zero signal of circular polarization. This star was observed for a technical test with the FORS2 instrument of the ESO VLT

while setting the instrument at position angles 0° and 90° on the sky. In the ideal case, we would expect:

$$\begin{aligned}
 P_Q(\text{PA} = 90^\circ) &= -P_Q(\text{PA} = 0^\circ), \\
 P_U(\text{PA} = 90^\circ) &= -P_U(\text{PA} = 0^\circ), \\
 P_L(\text{PA} = 90^\circ) &= P_L(\text{PA} = 0^\circ), \\
 P_V(\text{PA} = 90^\circ) &= P_V(\text{PA} = 0^\circ).
 \end{aligned}
 \tag{11}$$

Figure 5 shows that the P_Q profile obtained with the instrument set at 90° on the sky is clearly not symmetric about zero with respect to the profile obtained with the instrument set at 0° . This phenomenon can be explained by a (small) instrumental polarization in Q that depends on the instrument position angle, at the level of $\sim 0.2\%$. As a consequence, the measured fraction of linear polarization depends on the instrument position angle. Figure 5 shows also that the P_V profiles observed at the two different position angles on the sky are roughly symmetric about zero. This is qualitatively consistent with the hypothesis that the observed small signal of circular polarization is due to a cross-talk from linear to circular polarization. The source has an intrinsic signal in Q a small fraction of which is transformed into V by the optics preceding the polarimeter. When the instrument is rotated by 90° , the incoming linear polarization reverses its sign and is partially transformed into a V signal with the opposite sign to that in the previous case. This cross-talk problem is discussed in more detail by Bagnulo et al. (2009).

In conclusion, extreme care has to be taken in the interpretation of unexpected small circular polarization signals in the source that are intrinsically linearly polarized. For the circular polarization of spectral lines due to the presence of a magnetic field, the cross-talk problem works in the other direction. The intrinsic circular polarization signal in a spectral line due to a magnetic field is usually much larger than then the intrinsic linear polarization; hence the observed Stokes parameters Q and U may be potentially polluted by instrumental cross-talk from the Stokes parameter V .

4. The Zeeman effect on Stokes-parameter profiles of spectral lines

The main diagnostic tool for the detection of magnetic fields in non-degenerate stars is the analysis of the Zeeman effect on the Stokes-parameter profiles of spectral lines, which I will quickly and qualitatively illustrate in this section. The Zeeman effect is properly explained in terms of quantum mechanics, and its description can be found in numerous textbooks and reviews (see, e.g., Eisberg and Resnick 1974; Donati and Landstreet 2009). However, the classical approach adopted in this section has the advantage of offering a simpler and more intuitive visualization and is probably more suitable in the context of this review.

Note that above the critical regime of ~ 1 MG, the quadratic Zeeman effect becomes important, and line formation can be explained only in quantum-mechanical terms. For $B > 50$ MG, the magnetic field and Coulomb forces are of comparable strengths, and only a few numerical calculations have been performed to ex-

plore the physics of line formation in this regime. These complications should be taken into account for objects such as (some) magnetic white dwarfs or cataclysmic variables, but can be safely neglected in non-degenerate stars.

The process of line formation in a magnetic atmosphere can be classically interpreted in terms of an atomic oscillator in the presence of a magnetic field. Observed from the direction perpendicular to the magnetic field vector, the motion can be decomposed into three oscillations, one equal to the natural frequency ν_0 of the oscillator (the π component), and two at the frequencies $\nu_0 \pm \nu_L$ (the σ components), where ν_L is the Larmor frequency. The oscillations are linear, the π component is parallel to the magnetic field vector, and the σ components are perpendicular to the magnetic field vector. When observing in the direction parallel to the magnetic field vector, the frequency of the oscillator appears split into two σ components, in circular and opposite motions, and at frequencies $\nu_0 \pm \nu_L$, respectively. Figure 6 shows how this manifests in terms of Stokes-parameter profiles of a spectral line formed in presence of a magnetic field. Let us recall that the radiation field, proportional to the vector products $\mathbf{r} \times (\mathbf{r} \times \mathbf{a})$ (where \mathbf{r} is the unit vector that points from the observer to the oscillator and \mathbf{a} is the acceleration vector of the charged particle), has the same direction as the particle acceleration, and that an absorption line is polarized in the opposite direction to that of an emission line. If the magnetic field is perpendicular to the line of sight then the spectral line appears split into a π component polarized in the direction perpendicular to the magnetic field, and into two σ components at frequencies $\nu_0 \pm \nu_L$ polarized in the direction parallel to the magnetic field. The circular polarization is zero. If the x -axis of the reference system is taken along the direction of the magnetic field, all linear polarization is represented by the Stokes parameter Q [§], while its profile is positive in the σ components and negative in the π component. A spectral line formed in the presence of a magnetic field parallel to the line of sight will be split into two σ components at frequencies $\nu_0 \pm \nu_L$, circularly polarized in opposite directions, and the linear polarization is zero. For an arbitrarily oriented magnetic field vector, the line components are elliptically polarized.

This quick summary suggests that, roughly speaking, the Stokes parameter I is mainly sensitive to the magnetic field strength, the Stokes parameters Q and U are sensitive to the transverse components of the magnetic field, and the Stokes parameter V is sensitive to the longitudinal component of the magnetic field.

Figure 6 refers to the contribution to the Stokes-parameter profiles of a spectral line given by a single element of the stellar surface. The stellar case is more complicated, as one has to integrate the contributions from all stellar surface elements, taking into account a magnetic field that changes across the visible stellar disk, and (usually) a Doppler shift between surface elements. The polarization signal is noticeably diluted, since the contributions coming from different stellar

[§] We neglect here the so called magneto-optic effects (e.g., Landolfi and Landi Degl'Innocenti 1982) which are responsible for a signal in the Stokes parameter U also for a magnetic field parallel to the reference direction x .

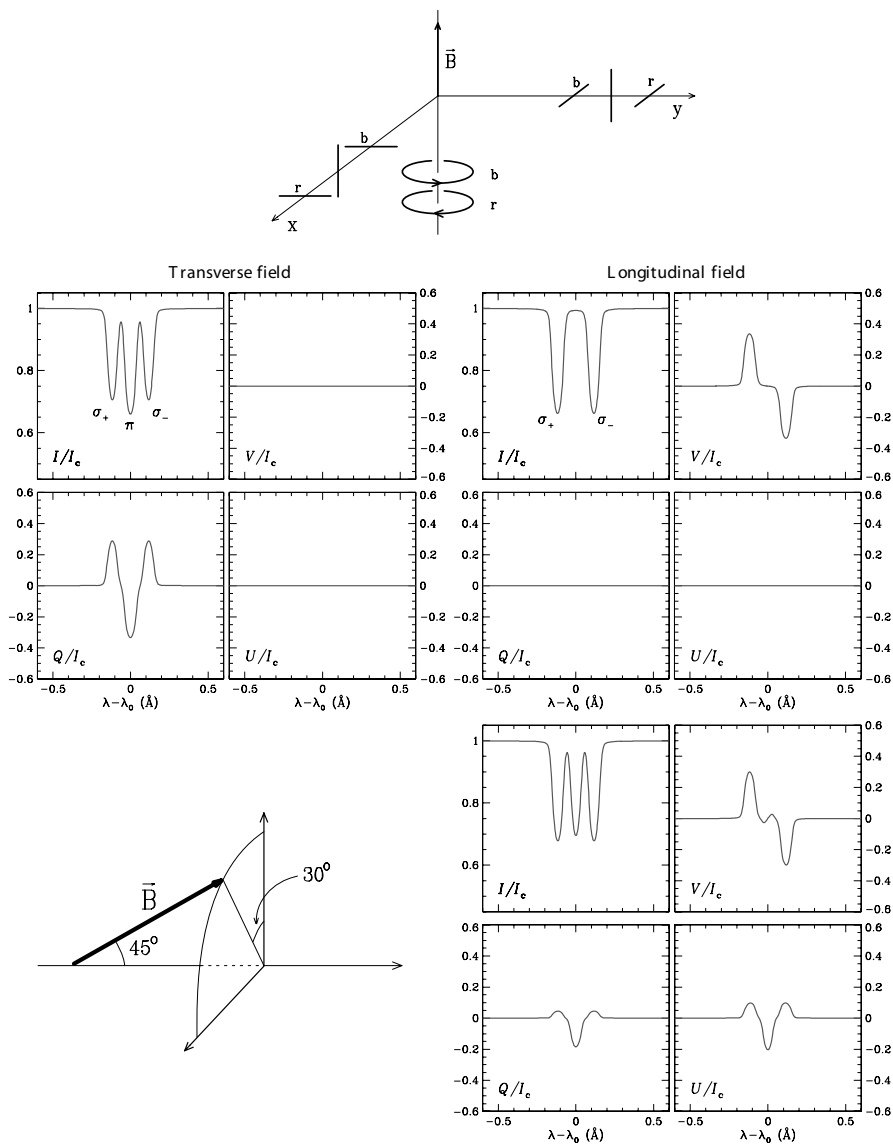


Fig. 6. The classical interpretation of the Zeeman effect and the process of line formation in a magnetic atmosphere. The left-hand middle panel shows the Stokes-parameter profiles of the transverse field, the right-hand middle panel refers to the longitudinal field, and the bottom right-hand panel refers to a magnetic field oriented as shown in the bottom left-hand panel.

regions tend to cancel each other. Furthermore, stars rotate, and rotational broadening tends to wash out the Zeeman effect. In unpolarized light, line splitting is

observed only in stars with low $v \sin i$ (the projected equatorial velocity) and a strong magnetic field. The rule of thumb is that Zeeman splitting in the Stokes parameter I can be detected when the average field modulus, expressed in kG, is greater than $v \sin i$, expressed in kms^{-1} . Rotational broadening is less critical for the Stokes parameters Q , U , and V , and a magnetic field with a strength of a few hundred G can still be detected in circular polarization even in stars with $v \sin i$ up to several tens of kms^{-1} . So far, the fastest rotating non-degenerate star in which a field has been detected is the B2 variable-type HD 142184, with a mean longitudinal field varying between -2kG and -1kG and $v \sin i = 270 \text{ kms}^{-1}$ (Grunhut et al. 2010b), yet detections in stars with $v \sin i > 100 \text{ kms}^{-1}$ are much less common than in more slowly rotating stars. Finally, another effect that dilutes the polarization signal of a stellar line is the finite spectral resolution of the instrument employed for the observations. Metal lines have generally widths of the order of 10^{-1} \AA , and for their spectropolarimetric observations an instrument with a resolving power of several tens of thousands is required. Instead, a spectral resolution of about 1000 will be sufficient to detect circular polarization in the hydrogen Balmer lines which extend for several tens of \AA .

Figure 7 shows the Stokes-parameter profiles predicted for a synthetic strong metal line in a star with $v \sin i = 10 \text{ kms}^{-1}$, with a 10 kG dipolar field tilted at 45° with respect to the rotation axis, and displayed without taking into account the blurring introduced by the instrument profile. At zero phase, when the dipole axis is directed toward the observer, both the field modulus and the longitudinal field strength, averaged over the visible stellar disk, are highest. The Stokes parameter I shows a broadening due to both the Doppler and the Zeeman effect, the latter being responsible for a hint of line splitting. At a 0.25 phase, when the magnetic axis is perpendicular to the line of sight, the mean field modulus is minimal and the mean longitudinal field is zero. Rotational broadening of the Stokes parameter I dominates over Zeeman splitting. Note that even if the mean longitudinal field is zero, the Stokes parameter V still shows a non-zero profile. This is explained by the fact that the contributions to the line profile originating in the regions around the north magnetic pole are red-shifted, while the contributions from the regions formed close to the negative magnetic pole (opposite in sign) are blue-shifted by the Doppler effect, hence their contributions to the Stokes parameter V do not cancel each other. This phenomenon is called the “cross-over” effect, and was first observed in a magnetic star by Babcock (1951). Finally, note that in the example of Fig. 6, the amplitude of the Stokes-parameter Q and U profiles are less than half the amplitude of the Stokes-parameter V profiles. For a more complex magnetic configuration, the linear polarization signal, which depends on the components of the magnetic field perpendicular to the line of sight, becomes substantially smaller than the circular polarization signal and, as a matter of fact, is only rarely detected in real stars.

The Stokes-parameter profiles of Figs. 6 and 7 were synthesized using the analytical solution given by Unno (1956) for a Milne–Eddington atmosphere. This is a simplified approach to a very complex problem. Even when the presence of a

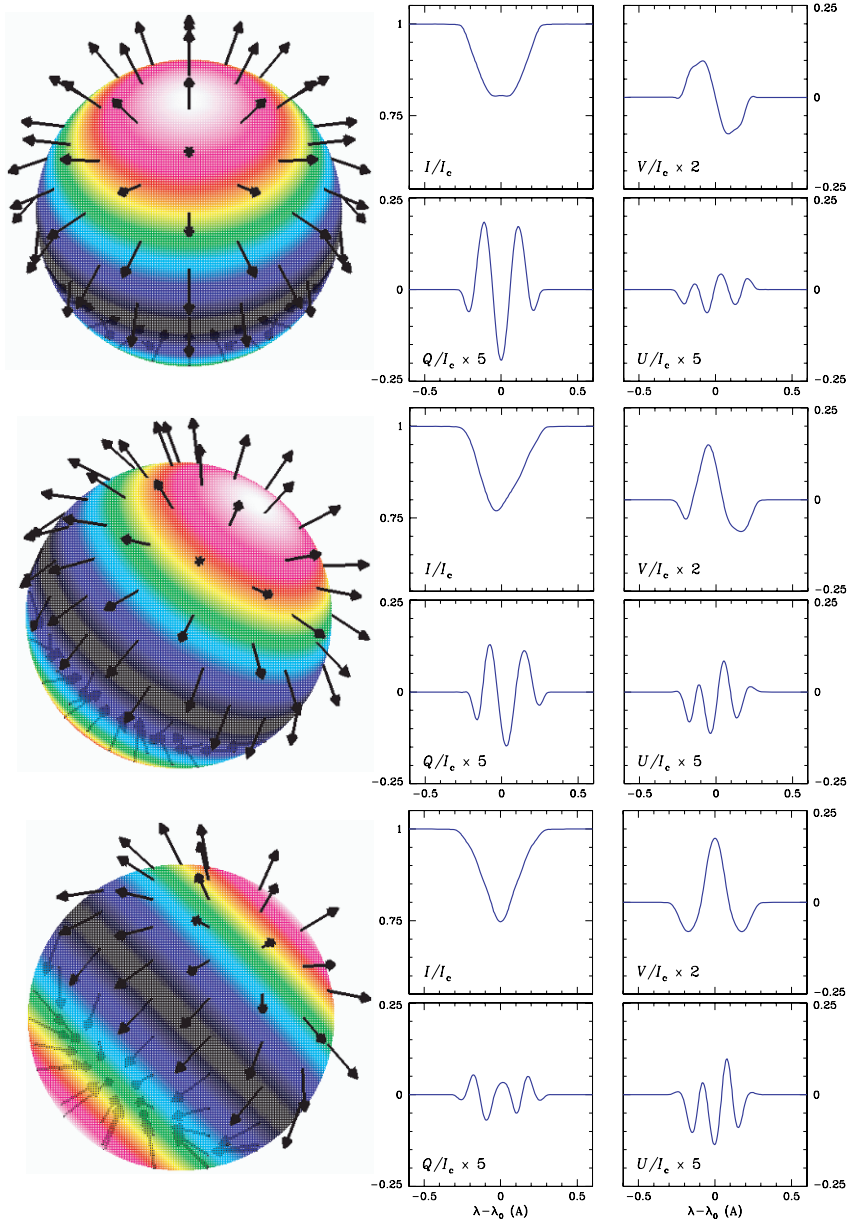


Fig. 7. Predicted Stokes profiles of a synthetic metal line (normal Zeeman triplet) for a star rotating with $v \sin i = 10 \text{ km s}^{-1}$ and with a dipolar magnetic field. The star's rotation axis is perpendicular to the line of sight, and the dipole axis is tilted at 45° with respect to the rotation axis. The field strength at the magnetic poles (red-whitish area) is 10 kG, and it is half this value at the magnetic equator (dark blue regions). From top to bottom, panels refer to rotation phases 0.000, 0.125, and 0.250, respectively.

magnetic field represents the only departure from the simple plane-parallel atmosphere model, a proper synthesis of Stokes-parameter profiles requires the use of sophisticated model atmospheres, using realistic values of stellar parameters (such as effective temperature and gravity), and possibly taking into account inhomogeneous distributions of the chemical elements. This “direct” problem, i.e., the calculation of Stokes-parameter profiles given a certain model of a magnetic star, is relatively well understood and may be solved under a certain number of realistic assumptions. Atmosphere modeling and radiative transfer in a magnetic stellar atmosphere are topics that are covered in a vast literature ranging from textbooks with the basic physical principles to advanced scientific papers presenting sophisticated mathematic algorithms. Having in mind practical applications, a possible starting point to get a deeper insight of these themes is the review and comparison of three radiative codes presented by Wade et al. (2001).

Finally, in the presence of a relatively strong magnetic field (typically at least 10 kG), the Paschen–Back effect may affect the Stokes-parameter profiles of certain spectral lines, increasing the equivalent width of the Stokes parameter I and weakening the Q , U , and V profiles. For more details see Landolfi et al. (2001), Kochukhov (2008), and Stift et al. (2008).

In the next section I will discuss the challenging inverse problem, i.e., the recovery of information about a stellar magnetic morphology from a series of observations of Stokes-parameter profiles.

5. Detection and modeling of stellar magnetic fields

Before describing the techniques for the detection and modeling of stellar magnetic fields, it is useful to make a preliminary note. Observations of stellar magnetic fields are strongly biased toward a relatively small fraction (5% to 10%) of A- and B-type stars that are chemically peculiar. Because of diffusion processes (Michaud 1970), the photospheric abundances of some elements (mainly iron peak elements and rare-earths) differ, in some case dramatically, from the values measured for the Sun (e.g., Preston 1974). These stars are commonly referred to as Ap and Bp stars. They tend to rotate much more slowly than the “chemically normal” A- and B-type stars, some of them having a rotation period of weeks, months, years, or even decades (e.g., Stępień 2000). Most Ap and Bp stars exhibit the presence of a magnetic field organized at a large scale, stable at least over a time-scale of several decades, and with a typical strength between a few tens and a few tens of thousand of Gauss. Their magnetic field is not symmetric about the rotation axis, so that the observer sees a magnetic configuration that changes with time according to the period of the star’s rotation (measured, e.g., via photometric techniques).

The physics of the atmospheres of Ap and Bp stars is probably just an important niche of stellar astrophysics, but their magnetic nature makes them a laboratory of great importance for the study of stellar magnetism. By comparison, late M-dwarf stars are faint objects, and their spectra are dominated by molecular lines

which are still difficult to model theoretically in the presence of a magnetic field; solar-type stars have a magnetic field concentrated in relatively small spots of opposite polarity: integrated over the visible stellar disk, the polarization signal is very weak and difficult to measure; massive stars (early B- and O-type stars) possess much weaker magnetic fields than Ap stars, rotate very fast, and the shape of their spectral lines is often distorted by the presence of stellar winds, which further hampers the detection and modeling of their magnetic fields. It does not come as a surprise that the first star other than the Sun for which a magnetic field was detected for the first time was an Ap star (Babcock 1947), and that Ap and Bp stars are still the primary objects outside the solar system where the magnetic field is best observed and studied. Mainly in the last decade, our knowledge of the stellar magnetic fields has tremendously expanded to include virtually every other object of the HR diagram, mostly thanks to the introduction of highly efficient instruments and new sophisticated methods for data analysis. Of particularly great scientific impact are the current studies of magnetic fields in massive stars (O- and B-type, see Wade et al. 2010) and M-type stars (Morin et al. 2010; Donati et al. 2006).

Stellar magnetic fields are usually detected via observations of either high- or low-resolution polarized spectra. High-resolution spectropolarimeters (e.g., ES-PaDOnS, Narval, and HARPS) have typically a spectral resolution $R = \lambda/\Delta\lambda$ between 60 000 and 100 000 and allow one to detect spectral signatures of individual metal lines. Low-resolution spectropolarimeters (such as FORS) have a spectral resolution of the order of few hundreds to few thousands and are more suitable for the detection of the broad H and He lines, as shown in Fig. 8. Alternative detection methods are based either on the measurement of Zeeman splitting in the Stokes parameter I observed in slowly rotating magnetic stars (e.g., Babcock 1960; Mathys et al. 1997); on narrow-band spectropolarimetry of H Balmer lines (Angel and Landstreet 1970; Borra and Landstreet 1980); on broadband linear polarization measurements (Leroy 1995); and on measurements of the Zeeman broadening of spectral lines observed in unpolarized light (Preston 1971; Robinson 1980; Mathys 1995).

It is natural to classify the observing programs aimed at the study of stellar magnetism into two broad categories: surveys and single-target oriented. Surveys are aimed at generally assessing the occurrence of magnetic fields in a certain class of stars. The target list of a survey consists of a fairly large number of stars that will be observed two or three times using circular spectropolarimetry. Ideally, data analyses will reveal whether magnetism plays an important role in the physics of a certain class of stars. Two or three observations will not tell much about the magnetic morphology of individual stars. A detailed modeling requires at least a dozen of full sets of high-resolution Stokes-parameter I , Q , U , and V profiles sampling the entire star's rotation cycle. Low-resolution spectropolarimetric data do not provide sufficient details to allow proper modeling.

Regardless of the adopted technique, stellar magnetic field measurements tend to be time consuming. For a quantitative estimate, it is necessary to specify what

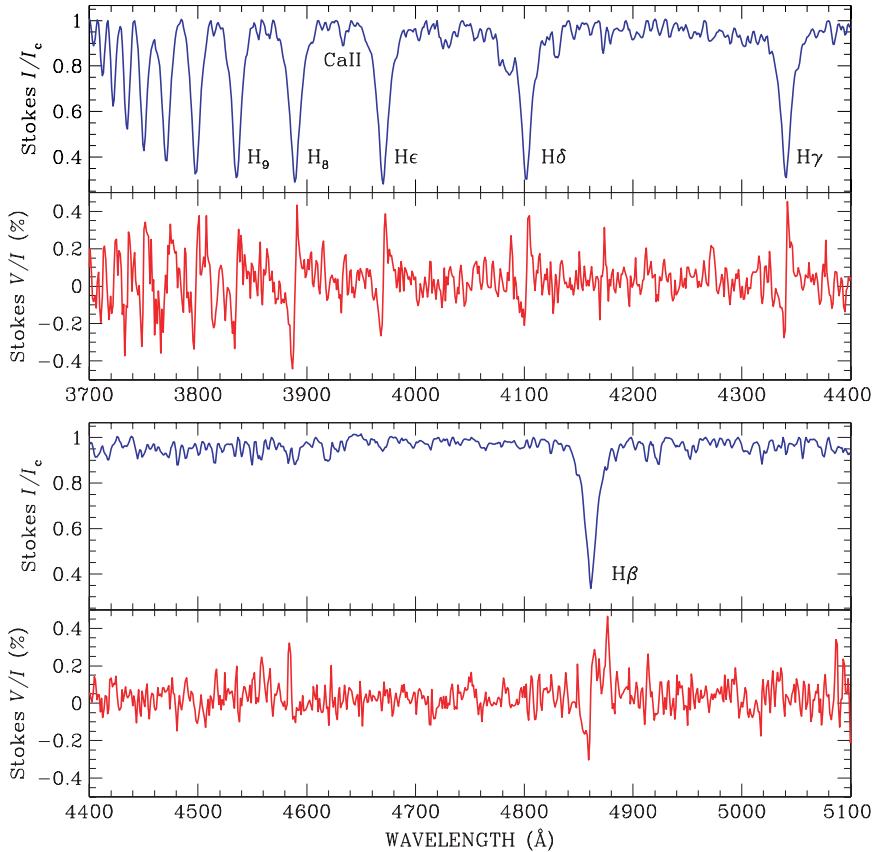


Fig. 8. The low-resolution polarized spectrum of the magnetic Ap star HD 94660 obtained with the ESO VLT FORS1 instrument (after Bagnulo et al. 2002). The Stokes parameter I is normalized to the pseudo-continuum for display purposes.

accuracy is required, the target magnitude, the telescope mirror size and telescope–instrument efficiency, and the type of stellar spectrum, i.e., whether it is rich in spectral lines, and how rotational broadening compares to Zeeman broadening. The key point is that the error bars of Stokes-parameter profiles are given by the inverse of the signal-to-noise ratio (SNR), and that we typically aim at measuring features of Stokes-parameter Q , U , and V profiles with an amplitude of the order of 10^{-3} (or smaller). Practically speaking, one needs to collect spectra with a SNR of at least a few hundreds per spectral bin, and to use special techniques to combine the information from several spectral lines and further increase the “effective” SNR up to values of one thousand or more. In terms of telescope time, we are speaking about 0.5 h (including overheads) to measure, with a few G error bar, the longitudinal field of a bright star, a couple of hours for a complete Stokes-parameter I, Q, U, V set for a not-so-bright star, and up to several hours for

extreme cases of a very weak field or very faint stars. Performing an accurate modeling of a couple of stars is likely to require roughly the same amount of telescope time as a reasonably broad survey.

5.1. Field detection

For survey purposes, one is generally interested in translating the observed Stokes-parameter profiles into a simple number that gives an indication about the overall field strength.

Linear polarization is sensitive to the averages of the two components of the magnetic field perpendicular to the line of sight. The contributions to the Stokes-parameter Q and U profiles from the various regions of the stellar disk tend to cancel each other more readily than the contributions to the Stokes parameter V , which is sensitive to the longitudinal component only. Therefore, the linear polarization signal of a magnetic star is substantially weaker than the circular polarization signal. Even if linear polarization is detected, it is impossible to infer information about the global structure of the magnetic field without a model of the field morphology. Therefore, magnetic field surveys of non-degenerate stars are generally based on observations of Stokes-parameter I and V profiles only.

The theory of radiative transfer in a magnetic atmosphere provides us with a couple set of equations that relate some features of the Stokes-parameter profiles I and V to the so called *mean longitudinal magnetic field* $\langle B_z \rangle$, i.e., the component of the magnetic field along the line of sight, averaged over the stellar disk. One equation tells us that the mean longitudinal field is proportional to the first-order moment of the Stokes parameter V about the line center (or, in a mathematically equivalent way, to the splitting of the same spectral line observed in left- and right-hand circular polarization):

$$\langle B_z \rangle = - \frac{1}{C_z g_{\text{eff}} \lambda_0^2} \frac{\int (\lambda - \lambda_0) \frac{V(\lambda)}{I_c} d\lambda}{\int \left(1 - \frac{I(\lambda)}{I_c} \right) d\lambda}, \quad (12)$$

where g_{eff} is the effective Lande factor, $C_z = 4.67 \times 10^{-13} \text{ A}^{-1} \text{ G}^{-1}$, I_c is the intensity in the continuum, and λ_0 is the central wavelength of the spectral line. The second equation tells us that the longitudinal field is proportional to the ratio of the Stokes parameter V to the derivative of the Stokes parameter I with respect to wavelength:

$$\frac{V}{I} = - g_{\text{eff}} C_z \lambda^2 \frac{1}{I} \frac{dI}{d\lambda} \langle B_z \rangle. \quad (13)$$

For more details on how Eqs. (12) and (13) are derived, see, e.g., Mathys (1989).

Equation (12) is generally applied to metal lines observed with high resolution, while Eq. (13) is generally applied to hydrogen (or helium) Balmer lines observed with low-resolution spectropolarimetry. [Figure 9](#) shows how Eq. (13) can

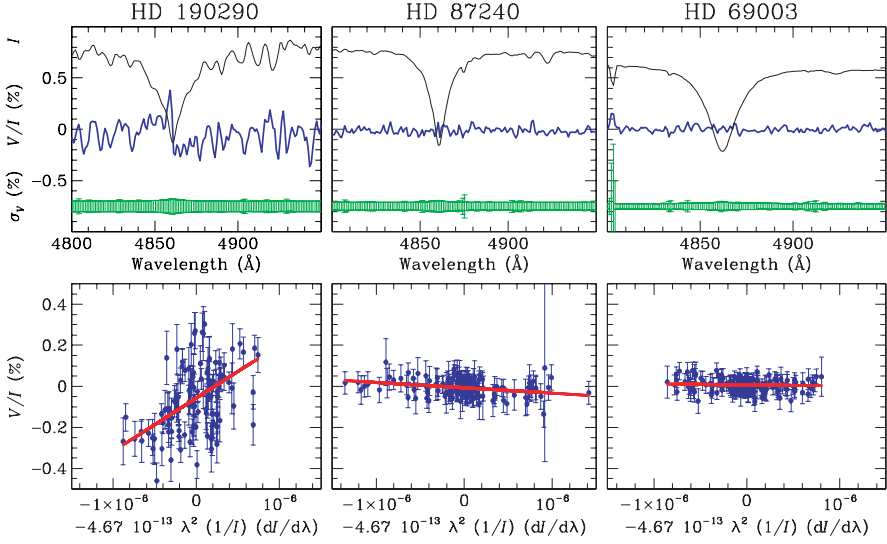


Fig. 9. Examples of application of Eq. (13) to the H Balmer lines for three early-type stars. The top panels show the spectral region around H β . The mean longitudinal magnetic field is obtained by applying the least-square technique to the various spectral bins around the H Balmer lines (bottom panels). From the left-hand to the right-hand panel, the measured longitudinal field is 2429 ± 110 G, -257 ± 58 G, and -48 ± 53 G, respectively. After Baguolo et al. (2006).

be used to determine the longitudinal field via a least-square technique. Both Eqs. (12) and (13) have the advantage of being very simple and the disadvantage of being strictly valid only under a certain number of assumptions that in fact are generally not satisfied. The main assumptions are that the observed spectral line is very weak, or that the Zeeman splitting is much smaller than the intrinsic line broadening; for a typical metal line, this means that the field strength at the stellar surface is $\ll 1$ kG. However, numerical simulations have shown that both formulas retain a statistical value, i.e., they still provide a reasonable estimate of the longitudinal field when it is inferred from a fairly large amount of spectral lines. For hydrogen lines, the weak-field-regime approximation holds for field strength up to ~ 10 kG; therefore, Eq. (13) could in principle provide a reasonable estimate of the longitudinal field even when applied to one Balmer line only. However, the validity of Eq. (13) rests on a description of the line formation accounting only for the Zeeman splitting of the atomic levels, whereas in the formation of the H Balmer lines there are also significant contributions from the linear Stark effect due to the surrounding charged particles and from the Lorentz effect due to the thermal motion of the hydrogen atoms in the magnetic fields (Mathys et al. 2000). The validity of Eq. (13) applied to H Balmer is supported mainly by the rough consistency with the field estimates obtained by applying Eq. (12) to metal lines.

Donati et al. (1997) introduced a kind of cross-correlation technique dubbed Least Square Deconvolution (LSD). This technique, normally applied to circular-polarization spectra, is based on the approximation of Eq. (13) and on the assumption that the Stokes parameter I has the same shape for all lines (this approximation is strictly true in the optically thin case, but is reasonable for many photospheric lines). By combining several spectral lines (typically >100), one obtains the mean profile of the Stokes parameters I and V , with a SNR approximately scaling with the square root of the number of spectral lines used for the deconvolution. The longitudinal field is then measured from the LSD profiles of a single “deconvolved” line via Eq. (12). This technique represents a break-through diagnostic tool for weak stellar magnetic fields, since it allows one to firmly detect circular polarization signals even when the Stokes-parameter V profiles of individual lines are indistinguishable from noise. LSD is nowadays routinely applied to all data acquired with high resolution. Further theoretical considerations and developments of this technique have recently been presented by Semel et al. (2009) and Kochukhov et al. (2010).

Surveys based on Eq. (12) applied to a number of spectral lines observed with high resolution include, e.g., a general catalogue of magnetic stars by Babcock (1958) and a survey of Ap stars by Mathys (1991). Surveys based on Eq. (13) applied to low-resolution data include, for instance, those of open-cluster Ap stars (Bagnulo et al. 2006), Herbig Ae/Be stars by Wade et al. (2007), central stars of planetary nebulae (Jordan et al. 2005), hot sub-dwarfs (O’Toole et al. 2005), and RR Lyrae stars (Kolenberg and Bagnulo 2009). The surveys based on LSD techniques include, e.g., searches for magnetic fields in A-, B-, and F-type stars (Shorlin et al. 2002), M-type stars (Donati et al. 2008; Morin et al. 2008), slowly pulsating B, β Cephei, and B-type emission-line stars (Silvester et al. 2009), and late-type supergiant stars (Grunhut et al. 2010a).

5.2. Field modeling

Until a few years ago, the great majority of model simulations of stellar magnetic fields were mostly based on the interpretation of the longitudinal field measurements as a function of the star’s rotation phase, with no attempt to reproduce the Stokes-parameter profiles which were generally observed with a relatively low SNR. Stibbs (1950) first explained the observed variability of the longitudinal field in terms of the so-called oblique rotator model, i.e., a star with a dipole field tilted at an angle with respect to the star’s rotation axis, like the one in Fig. 7. The observed magnetic field curve is interpreted using the following formula:

$$\langle B_z \rangle = \frac{15+u}{20(3-u)} B^d [\cos i \cos \beta + \sin i \sin \beta \cos(f - f_0)], \quad (14)$$

where B^d is the field strength at the positive pole of the dipole, i is the tilt angle of the rotation axis with respect to the line of sight, β is the inclination of the dipole with respect to the rotation axis, u is the limb-darkening coefficient (typically

~ 0.5), and f is the rotation phase, defined such as $f = f_0$ when the positive pole of the dipolar field points toward the observer (for a full characterization of the oblique rotator model including a dipole and an arbitrarily oriented planar quadrupole, see Landolfi et al. 1998). The inversion problem, if based on longitudinal field measurements only, is ill posed, but the inclination angle i can be obtained independently of the magnetic field curve if $v \sin i$, the star's rotation period, and the star's radius are known. In this case, the best model fit to the longitudinal field curve based on the first-order Fourier expansion allows one to obtain B^d and β . Examples of systematic applications of this method are, for instance, the studies by Borra and Landstreet (1980) and Hubrig et al. (2011). A catalogue of magnetic field curves gathered from the literature is given in Bychkov et al. (2005).

The longitudinal field curve is mainly sensitive to the dipolar component of the magnetic field. A quadrupole, if present, would give a contribution as small as 1/10 of the dipolar contribution (Schwarzschild 1950; Landolfi et al. 1998), and the contributions from higher order terms would be even smaller. Modeling based on longitudinal field measurements, even when formally accounting for contributions of a higher order than the dipole, allows one to characterize only the dipolar component of the star's magnetic field. From the analysis of longitudinal field measurements only, one is strongly tempted to conclude that the great majority of magnetic fields of early-type stars are well represented by a dipole, which is not confirmed by more refined analyses, as I will discuss later. The results of the modeling of the longitudinal field may be useful at best if one is interested in the stellar magnetosphere, where the dipole contribution dominates over those of higher-order terms. The modeling results may be used for instance to study the interaction between the magnetic field and stellar wind in O-type stars, but they are inadequate to characterize the field in the stellar photosphere, e.g., to study the role of the magnetic field in the diffusion of the chemical elements in the photosphere of Ap and Bp stars.

More sophisticated modeling can be performed by including additional kinds of magnetic observables when available, such as, e.g., the mean field modulus measured from the Zeeman splitting of spectral lines. The earliest attempts in that direction made, e.g., by Wolff and Wolff (1970), Preston (1970), and Huchra (1972) pointed already at the conclusion that the simple dipolar model is not sufficient to represent the magnetic morphology of Ap stars. A more recent example is given in Fig. 10, which shows the results of modeling of the longitudinal field and field modulus of the magnetic Ap star 53 Camelopardalis (Bagnulo et al. 2001) obtained under the assumption that the magnetic field is represented by a second-order multipolar expansion (i.e., the superposition of an arbitrarily oriented dipole and planar quadrupole). Although in this case the fit to the magnetic observables is reasonably good, the best-fit model predictions for the detailed Stokes-parameter profiles still do *not* agree satisfactorily with the observed Stokes-parameter profiles (Bagnulo et al. 2001).

During the last decade, the quality of spectropolarimetric observations has improved substantially, and the modeling of the observed Stokes-parameter profiles

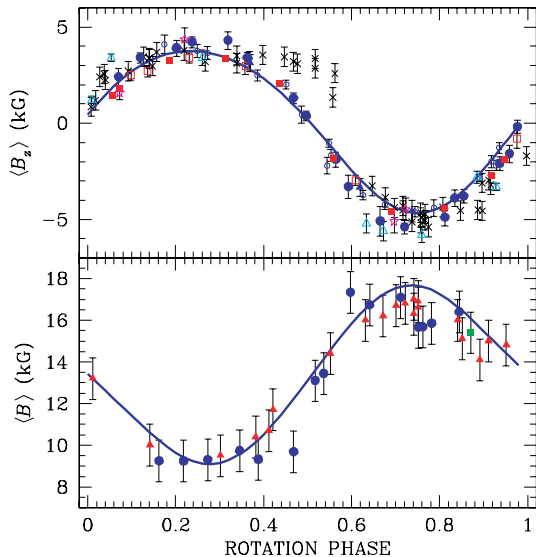


Fig. 10. Observations of the mean longitudinal field modulus (top panel) and mean field modulus (bottom panel) of the magnetic Ap star 53 Cam. Different symbols correspond to different observing sets gathered from the literature, as summarized in Bagnulo et al. (2001). The solid curves represent the best fit to the observations obtained assuming that the field is represented by the superposition of a dipole and a quadrupole.

(or their LSD profiles) has become routine. The simplest modeling approach is similar to that used for the modeling of the magnetic quantities, that is, to assume a parametrized model for the magnetic field and find the best-model parameters by applying the least-square technique to the observed Stokes-parameter profiles (instead of the magnetic observables). The inversion technique may be pretty robust, but the parameterization may be inadequate to describe the complexity of a real stellar magnetic configuration. An alternative, more successful approach consists of interpreting the observations without making any *a priori* assumption on the magnetic configuration and using a regularization function to converge toward a stable solution. An example of this “Magnetic Doppler Imaging” (MDI) has been presented by Piskunov and Kochukhov (2002) and Kochukhov and Piskunov (2002) and applied to several magnetic Ap stars, e.g., by Kochukov et al. (2004) and Kochukov and Wade (2010). A similar modeling technique, generally called “Zeeman Doppler Imaging”, implemented by Donati (2001) and based on earlier studies by Brown et al. (1991) and Donati and Brown (1997), involves the decomposition of the field into poloidal and toroidal components which are both expressed as spherical harmonic expansions. Numerous applications include, e.g., studies of the magnetic topologies of mid and late M-type stars (Morin et al. 2008, 2010) as well as G-type stars (e.g., Petit et al. 2005), and the modeling of the recently discovered magnetic field in the A-type star Vega (Petit et al. 2010).

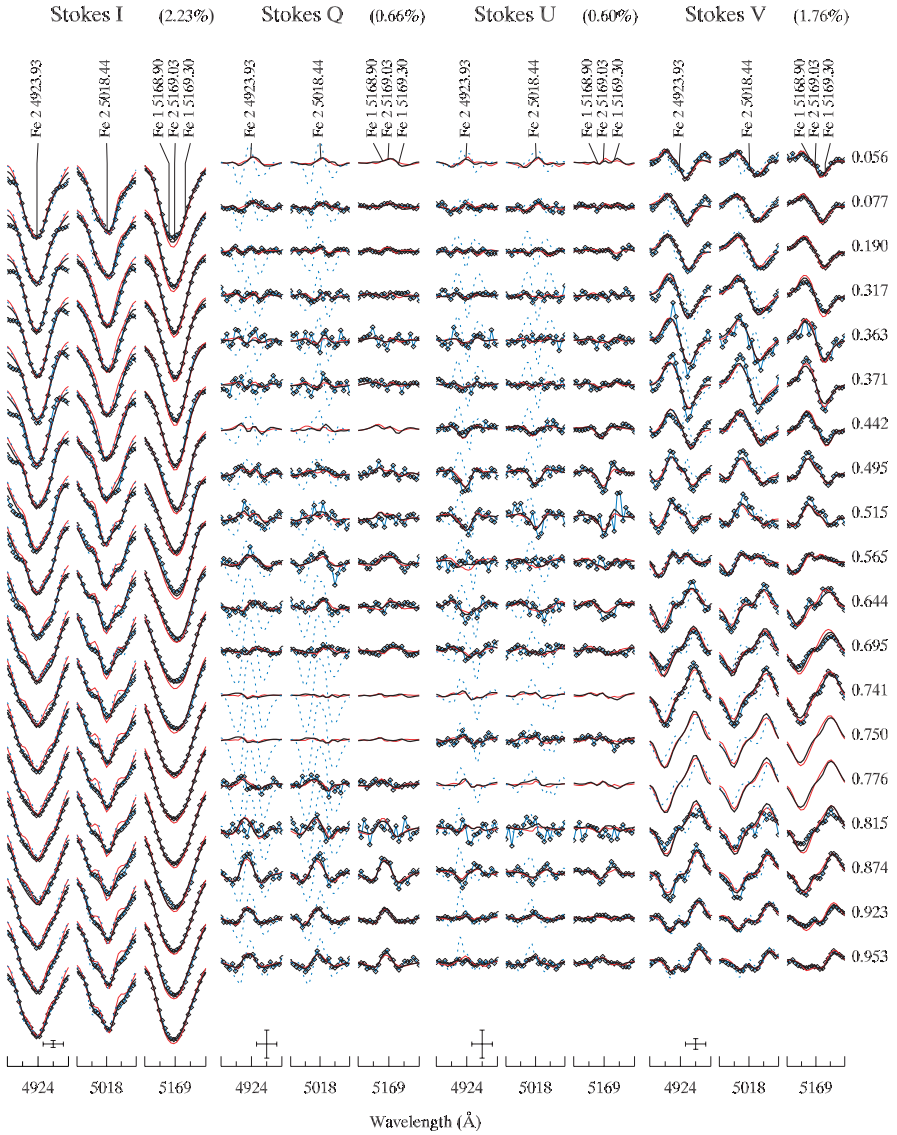


Fig. 11. Observed (diamonds) and modeled (curves) Stokes-parameter I, Q, U, V profiles of three spectral lines of the magnetic Ap star 53 Cam, normalized to the continuum intensity. The blue curves represent the best-fit model obtained assuming that the magnetic configuration can be described by the superposition of a dipole and a quadrupole. The thin red curves represent the best-fit model obtained with the MDI assuming a homogeneous Fe distribution. The thick black curves represent the best-fit model obtained with MDI by letting the Fe distribution vary over the stellar surface. The star's rotation phase is given on the right-hand side of the panels. The bars at the lower left of each panel show the horizontal and vertical scale (1 Å and 5% of the Stokes continuum intensity I , respectively). After Kochukhov et al. (2004).

Figure 11 shows the Stokes-parameter profiles of three spectral lines observed in the Ap star 53 Cam, the best-fit model profiles obtained under the assumption that the magnetic field be represented by a second-order multipolar expansion, and the best-fit model profiles obtained with the MDI. Note, however, that, while it is clear that the MDI delivers much better results than the technique based on a parametrized model, it is more prone to interpret systematic data errors in terms of physical features rather than returning a poor fit to the observations.

Simultaneously with the magnetic configuration, the MDI can reconstruct the surface distribution of the chemical elements, which is often found to be inhomogeneous in early-type magnetic stars. In some respect, spectropolarimetry and the MDI allow one to observe stellar disks with a spatial resolution enormously higher than what could be obtained with any modern interferometric technique.

6. Conclusions

Just over a decade ago, spectropolarimetric studies of stellar magnetic fields were still a somewhat limited niche of stellar astrophysics. For a long time, magnetism had been generically considered an important physical agent in several phenomena of astrophysical interest, but the characterization of its detailed role was lacking a comparison with the observations. This is no longer true. New efficient instruments and new detection and modeling techniques make it now possible to perform “laboratory experiments”. Stars provide an enormous variety of different physical conditions, and polarimetric observations of these objects provide much stronger constraints on models than simple flux measurements.

The current studies of magnetic fields will play a fundamental role as benchmarks for further applications of spectropolarimetric techniques. Spectropolarimetry is a promising technique for the characterization of exoplanets (e.g., Stam 2008) and may even reveal itself as an important tool for the search of extraterrestrial life (Sparks et al. 2009; Sterzik et al. 2010). These new and exciting branches of astrophysical research will strongly benefit from the experience acquired and from the techniques developed for the detection and modeling of stellar magnetic fields.

References

- Aitken, D. K., and J. H. Hough, 2001: Spectral modulation, or ripple, in retardation plates for linear and circular polarization. *Publ. Astron. Soc. Pacific* **113**, 1300–1305.
- Angel, J. R. P., and J. D. Landstreet, 1970: Magnetic observations of white dwarfs. *Astrophys. J.* **160**, L147–L152.
- Appenzeller, I., K. Fricke, W. Fürtig, et al., 1998: Successful commissioning of FORS1 – the first optical instrument on the VLT. *Messenger* **94**, 1–6.
- Babcock, H. W., 1947: Zeeman effect in stellar spectra. *Astrophys. J.* **105**, 105–119.
- Babcock, H. W., 1951: The magnetically variable star HD 125248. *Astrophys. J.* **114**, 1–36.
- Babcock, H. W., 1958: A catalogue of magnetic stars. *Astrophys. J. Suppl. Ser.* **3**, 141–210.
- Babcock, H. W., 1960: The 30 kiloGauss magnetic field of HD 215441. *Astrophys. J.* **132**, 521–532.

- Bagnulo, S., T. Szeifert, G. A. Wade, et al., 2002: Measuring magnetic fields of early-type stars with FORS1 at the VLT. *Astron. Astrophys.* **389**, 191–201.
- Bagnulo, S., G. A. Wade, J.-F. Donati, et al., 2001: A study of polarized spectra of magnetic CP stars: predicted vs. observed Stokes *IQUV* profiles for β CrB and 53 Cam. *Astron. Astrophys.* **369**, 889–907.
- Bagnulo, S., J. D. Landstreet, E. Mason, et al., 2006: Searching for links between magnetic fields and stellar evolution. I. A survey of magnetic fields in open cluster A- and B-type stars with FORS1. *Astron. Astrophys.* **450**, 777–791.
- Bagnulo, S., M. Landolfi, J. D. Landstreet, et al., 2009: Stellar spectropolarimetry with retarder waveplate and beam splitter devices. *Publ. Astron. Soc. Pacific* **121**, 993–1015.
- Borra, E., and J. D. Landstreet, 1980: The magnetic fields of the AP stars. *Astrophys. J. Suppl. Ser.* **42**, 421–445.
- Brown, S., J.-F. Donati, D. E. Rees, and M. Semel, 1991: Zeeman–Doppler imaging of solar-type and AP stars. IV – Maximum entropy reconstruction of 2D magnetic topologies. *Astron. Astrophys.* **250**, 463–474.
- Bychkov, V. D., L. V. Bychkova, and J. Madej, 2005: A catalog of stellar magnetic rotational phase curves. *Astron. Astrophys.* **430**, 1143–1154.
- Clarke, D., 1974: Polarimetric definitions. In T. Gehrels, Ed., *Planets, Stars, and Nebulae: Studied with Photopolarimetry* (University of Arizona Press, Tucson, AZ), pp. 45–53.
- Donati, J.-F., 2001: Imaging the magnetic topologies of cool active stars. *Lect. Not. Phys.* **573**, 207–231.
- Donati, J.-F., and S. F. Brown, 1997: Zeeman–Doppler imaging of active stars. V. Sensitivity of maximum entropy magnetic maps to field orientation. *Astron. Astrophys.* **326**, 1135–1142.
- Donati, J.-F., and J. D. Landstreet, 2009: Magnetic fields of nondegenerate stars. *Annu. Rev. Astron. Astrophys.* **47**, 333–370.
- Donati, J.-F., M. Semel, B. D. Carter, et al., 1997: Spectropolarimetric observations of active stars. *Mon. Not. R. Astron. Soc.* **291**, 658–682.
- Donati, J.-F., T. Forveille, A. Collier-Cameron, et al., 2006: The large-scale axisymmetric magnetic topology of a very-low-mass fully convective star. *Science* **311**, 633–635.
- Donati, J.-F., J. Morin, P. Petit, et al., 2008: Large-scale magnetic topologies of early M dwarfs. *Mon. Not. R. Astron. Soc.* **390**, 545–560.
- Eisberg, R., and R. Resnick, 1974: *Quantum Physics of Atoms, Molecules, Solids, Nuclei and Particles* (Wiley, New York).
- Fossati, L., S. Bagnulo, E. Mason, and E. Landi Degl’Innocenti, 2007: Standard stars for linear polarization observed with FORS1. In C. Sterken, Ed., *The Future of Photometric, Spectrophotometric and Polarimetric Standardization* (Astronomical Society of Pacific, San Francisco), pp. 503–507.
- Grunhut, J. H., G. A. Wade, D. A. Hanes, and E. Alecian, 2010a: Systematic detection of magnetic fields in massive, late-type supergiants. *Mon. Not. R. Astron.* **408**, 2290–2297.
- Grunhut, J. H., G. A. Wade, T. Rivinius, et al., 2010b: Discovery of the most rapidly-rotating, non-degenerate, magnetic massive star by the MiMeS collaboration. arXiv:1009.3246.
- Hough, J. H., 2005: Polarimetry techniques at optical and infrared wavelengths. In A. Adamson, C. Aspin, C. J. Davis, and T. Fujiyoshi, Eds., *Astronomical Polarimetry: Current Status and Future Directions* (Astronomical Society of Pacific, San Francisco), pp. 3–14.

- Hough, J., 2011: High sensitivity polarimetry: techniques and applications. In M. I. Mishchenko, Ya. S. Yatskiv, V. K. Rosenbush, and G. Videen, Eds., *Polarimetric Detection, Characterization, and Remote Sensing* (Springer, Berlin), pp. 177–204.
- Hubrig, S., I. Ilyin, M. Schoeller, et al., 2011: First magnetic field models for recently discovered magnetic beta Cephei and slowly pulsating B stars. *Astrophys. J.* **726**, L5–L9.
- Huchra, J., 1972: An analysis of the magnetic field of 53 Camelopardalis and its implications for the decentered-dipole rotator model. *Astrophys. J.* **174**, 435–438.
- Jordan, S., K. Werner, and S. J. O’Toole, 2005: Discovery of magnetic fields in central stars of planetary nebulae. *Astron. Astrophys.* **432**, 273–279.
- Keller, C. U., 2002: Instrumentation for astrophysical spectropolarimetry. In J. Trujillo-Bueno, F. Moreno-Insertis, and F. Sánchez, Eds., *Astrophysical Spectropolarimetry* (Cambridge University Press, Cambridge, UK), pp. 303–354.
- Kochukhov, O., 2008: The Paschen–Back effect in the Li I 6708 Å line and the presence of lithium in cool magnetic Ap stars. *Astron. Astrophys.* **483**, 557–566.
- Kochukhov, O., and N. Piskunov, 2002: Doppler Imaging of stellar magnetic fields. II. Numerical experiments. *Astron. Astrophys.* **388**, 868–888.
- Kochukhov, O., and G. A. Wade, 2010: Magnetic Doppler imaging of α 2 Canum Venaticorum in all four Stokes parameters. Unveiling the hidden complexity of stellar magnetic fields. *Astron. Astrophys.* **513**, A13.
- Kochukhov, O., S. Bagnulo, G. A. Wade, et al., 2004: Magnetic Doppler imaging of 53 Camelopardalis in all four Stokes parameters. *Astron. Astrophys.* **414**, 613–632.
- Kochukhov, O., V. Makaganiuk, and N. Piskunov, 2010: Least-squares deconvolution of the stellar intensity and polarization spectra. *Astron. Astrophys.* **524**, A5.
- Kolenberg, K., and S. Bagnulo, 2009: Observational constraints on the magnetic field of RR Lyrae stars. *Astron. Astrophys.* **498**, 543–550.
- Landi Degl’Innocenti, E., and M. Landolfi, 2004: *Polarization in Spectral Lines* (Kluwer Academic Publishers, Dordrecht).
- Landi Degl’Innocenti, E., S. Bagnulo, and L. Fossati, 2007: Polarimetric standardization. In *The Future of Photometric, Spectrophotometric and Polarimetric Standardization* (Astronomical Society of Pacific, San Francisco), pp. 495–502.
- Landolfi, M., and E. Landi Degl’Innocenti, 1982: Magneto-optical effects and the determination of vector magnetic fields from Stokes profiles. *Solar Phys.* **78**, 355–364.
- Landolfi, M., S. Bagnulo, and M. Landi Degl’Innocenti, 1998: Modelling of magnetic fields of CP stars. I. A diagnostic method for dipole and quadrupole fields from Stokes I and V observations. *Astron. Astrophys.* **338**, 111–121.
- Landolfi, M., S. Bagnulo, M. Landi Degl’Innocenti, and E. Landi Degl’Innocenti, 2001: The Paschen–Back effect on fine and hyperfine structure: iImpact on polarized spectra of Ap and Bp stars. In G. Mathys, S. K. Solanki, and D. T. Wickramasinghe, Eds., *Magnetic Fields Across the Hertzsprung–Russell Diagram* (Astronomical Society of Pacific, San Francisco), pp. 349–352.
- Leroy, J.-L., 1995: Linear polarimetry of AP stars. V. A general catalogue of measurements. *Astron. Astrophys. Suppl. Ser.* **114**, 79–104.
- Mathys, G., 1989: The observation of magnetic fields in nondegenerate stars. *Fund. Cosmic Phys.* **13**, 143–308.
- Mathys, G., 1991: Spectropolarimetry of magnetic stars. II – The mean longitudinal magnetic field. *Astron. Astrophys. Suppl. Ser.* **89**, 121–157.
- Mathys, G., 1995: Spectropolarimetry of magnetic stars. V. The mean quadratic magnetic field. *Astron. Astrophys.* **293**, 746–763.

- Mathys, G., S. Hubrig, J. D. Landstreet, et al., 1997: The mean magnetic field modulus of AP stars. *Astron. Astrophys. Suppl. Ser.* **123**, 353–402.
- Mathys, G., C. Stehlé, S. Brillant, and T. Lanz, 2000: The physical foundations of stellar magnetic field diagnosis from polarimetric observations of hydrogen lines. *Astron. Astrophys.* **358**, 1151–1156.
- Mayor, M., F. Pepe, D. Queloz, et al., 2003: Setting new standards with HARPS. *The Messenger* **114**, 20–24.
- Mestel, L., 1999: *Stellar Magnetism* (Clarendon, Oxford).
- Michaud, G., 1970: Diffusion processes in peculiar stars. *Astrophys. J.* **170**, 641–658.
- Morin, J., J.-F. Donati, P. Petit, et al., 2008: Large-scale magnetic topologies of mid M dwarfs. *Mon. Not. R. Astron. Soc.* **390**, 567–581.
- Morin, J., J.-F. Donati, P. Petit, et al., 2010: Large-scale magnetic topologies of late M dwarfs. *Mon. Not. R. Astron. Soc.* **407**, 2269–2286.
- O’Toole, S. J., S. Jordan, S. Friedrich, and U. Heber, 2005: Discovery of magnetic fields in hot subdwarfs. *Astron. Astrophys.* **437**, 227–234.
- Petit, P., J.-F. Donati, and M. Aurière, 2005: Large-scale magnetic field of the G8 dwarf ξ Bootis A. *Mon. Not. R. Astron. Soc.* **361**, 837–849.
- Petit, P., F. Lignières, G. A. Wade, et al., 2010: The rapid rotation and complex magnetic field geometry of Vega. *Astron. Astrophys.* **523**, A41.
- Piskunov, N., and O. Kochukhov, 2002: Doppler imaging of stellar magnetic fields. I. Techniques. *Astron. Astrophys.* **381**, 736–756.
- Preston, G. W., 1970: The large magnetic field of HD 126515 and its implication for the rigid-rotator model of magnetic stars. *Astrophys. J.* **160**, 1059–1070.
- Preston, G. W., 1971: The mean surface fields of magnetic stars. *Astrophys. J.* **164**, 309–315.
- Preston, G. W., 1974: The chemically peculiar stars of the upper main sequence. *Annu. Rev. Astron. Astrophys.* **12**, 257–277.
- Robinson, R. D., Jr., 1980: Magnetic field measurements on stellar sources – a new method. *Astrophys. J.* **239**, 961–967.
- Schwarzschild, M., 1950: Zeeman shifts for stellar dipoles and quadrupoles with inclined axes. *Astrophys. J.* **112**, 222–223.
- Semel, M., J. C. Ramírez Vélez, M. J. Martínez González, et al., 2009: Multiline Zeeman signatures through line addition. *Astron. Astrophys.* **504**, 1003–1009.
- Shorlin, S. L. S., G. A. Wade, J.-F. Donati, et al., 2002: A highly sensitive search for magnetic fields in B, A and F stars. *Astron. Astrophys.* **392**, 637–652.
- Shurcliff, W. A., 1963: *Polarized Light: Selected Reprints* (American Institute of Physics, New York).
- Silvester, J., C. Neiner, H. F. Henrichs, et al., 2009: On the incidence of magnetic fields in slowly pulsating B, β Cephei and B-type emission-line stars. *Mon. Not. R. Astron. Soc.* **398**, 1505–1511.
- Snik, F., O. Kochukhov, N. Piskunov, et al., 2010: The HARPS polarimeter. arXiv1010.0397.
- Spark, W. B., J. H. Hough, L. Kolokolova, et al., 2009: Circular polarization in scattered light as a possible biomarker. *J. Quant. Spectrosc. Radiat. Transfer* **110**, 1771–1779.
- Stam, D. M., 2008: Spectropolarimetric signatures of Earth-like extrasolar planets. *Astron. Astrophys.* **482**, 989–1007.
- Stępień, K., 2000: Loss of angular momentum of magnetic Ap stars in the pre-main sequence phase. *Astron. Astrophys.* **353**, 227–238.

- Sterzik, M., S. Bagnulo, A. Azua, et al., 2010: Astronomy meets biology: EFOC2 and the chirality of life. *The Messenger* **142**, 25–27.
- Stibbs, D. W. N., 1950: A study of the spectrum and magnetic variable star HD 125248. *Mon. Not. R. Astron. Soc.* **110**, 395–404.
- Stift, M. J., F. Leone, and E. Landi Degl’Innocenti, 2008: Modelling the incomplete Paschen–Back effect in the spectra of magnetic Ap stars. *Mon. Not. R. Astron. Soc.* **385**, 1813–1819.
- Tinbergen, J., 1996: *Astronomical Polarimetry* (Cambridge University Press, Cambridge, UK).
- Unno, W., 1956: Line formation of a normal Zeeman triplet. *Publ. Astron. Soc. Japan* **8**, 108–125.
- Wade, G. A., S. Bagnulo, O. Kochukhov, et al., 2001: LTE spectrum synthesis in magnetic stellar atmospheres. The intergreement of three independent polarised radiative transfer codes. *Astron. Astrophys.* **374**, 265–279.
- Wade, G. A., S. Bagnulo, D. Drouin, et al., 2007: A search for strong, ordered magnetic fields in Herbig Ae/Be stars. *Mon. Not. R. Astron. Soc.* **376**, 1145–1161.
- Wade, G. A., E. Alecian, D. A. Bohlender, et al., 2010: The MiMeS project: overview and current status. arXiv1009.3563.
- Wolff S. C., and R. J. Wolff, 1970: The magnetic field of β Coronae Borealis. *Astrophys. J.* **160**, 1049–1058.



From left to right: Matthew Berg, Stefano Bagnulo, and Pavel Litvinov



From left to right: Anatoli Borovoi, Nikolai Rogovtsov, Victor Tishkovets, and Oleg Dubovik.

Reflection symmetry of a sphere's internal field and its consequences on scattering: behavior of the Stokes parameters

Matthew J. Berg*

Mississippi State University, Department of Physics & Astronomy, Mississippi State,
MS 39762, USA

Abstract. In this chapter we examine the reflection symmetries of the field inside a spherical particle. Consequences of this symmetry are identified that relate to the angular behavior of the far-field scattered wave's polarization state. To do this, a connection between the internal and scattered waves is developed in terms of a microphysical model in which each volume element in a particle is associated with a wavelet contributing to the scattered wave. The influence of the symmetry manifests through cancellation between vector-components of the wavelets at points in the planes defining the symmetry. Thus, a conceptual explanation emerges accounting for aspects of the angular structure of the polarization state. The utility of this explanation is demonstrated by predicting features in the polarimetric behavior of nonspherical particles, even though such particles are beyond the mathematical scope of the initial analysis. Specifically, the angular degree of linear polarization is shown to differ from that of the incident wave for a Rayleigh-sized, absorbent, spheroidal particle provided that its orientation breaks the reflection symmetry present for a spherical particle. If however, this same particle is symmetrically oriented, the degree of linear polarization does not change.

Keywords: symmetry, Stokes parameters, Rayleigh scattering, Mie theory, internal field, volume integral equation

1. Introduction

The concept of symmetry is fundamental to many areas of science, and can be useful to understand the nature of some physical phenomena. For example, so-called dipole transitions between quantum states in single-electron atoms can be allowed or prohibited depending on the spatial symmetry, or parity, of the associated wavefunctions; this, in-part, leads to an explanation of the emission spectra of elemental gasses like hydrogen (Bransden and Joachain 2003). In electromagnetic scattering, the symmetry of a particle's shape influences the spatial structure of the fields internal and external to the particle. Describing this connection – between

* Corresponding author. E-mail: matt.berg@msstate.edu

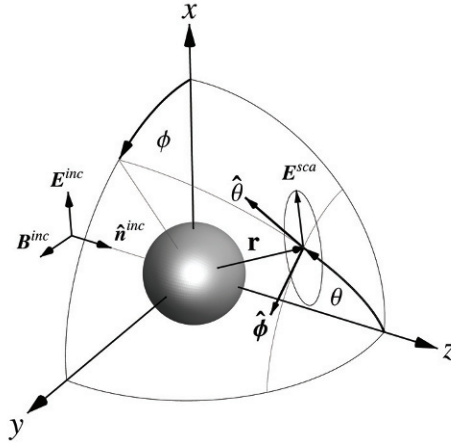


Fig. 1. Scattering arrangement. Shown at the tip of \mathbf{r} is the scattered electric field \mathbf{E}^{sca} along with its polarization ellipse, which is discussed in Section 4.

particle symmetry and scattering – is hardly a new enterprise and many excellent references exist showing various derivations and applications of this connection, see, e.g., van de Hulst (1981), Baum and Kritikos (1995), Hovenier (1969, 1970), Hovenier and Mackowski (1998), Hu et al. (1987), Nghiem et al. (1992), Ren et al. (1994), Schultz et al. (1999), Waterman (1971), and Yueh et al. (1994).

The purpose of this chapter is to revisit the basic problem of scattering from a spherical particle and examine the polarization state of the scattered wave. Unlike previous work on this topic, this will be done from an uncommon microphysical perspective where the particle's internal wave is regarded as the source of the scattered wave (Mishchenko et al. 2002; Tyynela et al. 2007; Berg 2008; Berg et al. 2010). As a consequence, the particle symmetry is connected to features in the angular structure of the polarization state through a concept-based model involving wave interference. This model leads to predictions regarding similar behavior that should be observed in nonspherical particles. Using the Discrete Dipole Approximation (DDA), these nonspherical particles are investigated and found to display the suspected polarimetric behavior.

2. Scattering arrangement

To begin, consider a uniform sphere of radius R and refractive index m located at the origin and surrounded by vacuum. A linearly polarized plane wave of wavelength λ traveling along the z -axis in the $\hat{\mathbf{n}}^{inc}$ direction is incident on this particle. The wave number is $k = 2\pi/\lambda$ and the polarization direction is along the x -axis, see Fig. 1. The electric and magnetic fields of the incident wave at the observation point \mathbf{r} are

$$\mathbf{E}^{inc}(\mathbf{r}) = E_0^{inc} \exp(ikr\hat{\mathbf{r}} \cdot \hat{\mathbf{n}}^{inc}) \hat{\mathbf{x}}, \quad (1)$$

$$\mathbf{B}^{\text{inc}}(\mathbf{r}) = \frac{k}{\omega} \hat{\mathbf{n}}^{\text{inc}} \times \mathbf{E}^{\text{inc}}(\mathbf{r}), \quad (2)$$

respectively, where E_0^{inc} is the real-valued wave amplitude and $i = (-1)^{1/2}$. All field quantities in this work share the same harmonic time dependence $\exp(-i\omega t)$, where $\omega = kc$, and $c = (\varepsilon_0 \mu_0)^{-1/2}$ is the speed of light with ε_0 and μ_0 being the permittivity and permeability of free space, respectively. For conciseness, this time dependence is suppressed.

The so-called ‘‘scattering arrangement’’ referred to throughout this work is defined as the geometric configuration formed by the particle, the polarization, and the propagation directions of the incident wave. The particle's surface S separates space into two volumes, those internal and external to the surface, denoted V^{int} and V^{ext} , respectively. The particle is surrounded by a large imaginary spherical surface S_l , also centered on the origin. The observation point \mathbf{r} is restricted to points on S_l , and it is assumed that the size of S_l , given by its radius R_l , is large enough that \mathbf{r} satisfies the far-field conditions of Mishchenko (2006). The scattered wave will then be an outward-traveling, transverse, spherical wave, the polarization of which can be described by the conventional Stokes parameters.

3. Reflection symmetry of the internal field

The Mie internal fields can be cast into a form that readily displays their reflection symmetry. First, the electric field inside the particle is expressed as

$$\mathbf{E}^{\text{int}}(\mathbf{r}) = \sum_{n=1}^{\infty} E_n [c_n \mathbf{M}_n(\mathbf{r}) - i d_n \mathbf{N}_n(\mathbf{r})], \quad \mathbf{r} \in V^{\text{int}} \quad (3)$$

and the magnetic field follows as

$$\mathbf{B}^{\text{int}}(\mathbf{r}) = -\frac{1}{\omega} \nabla \times \mathbf{E}^{\text{int}}(\mathbf{r}), \quad \mathbf{r} \in V^{\text{int}}. \quad (4)$$

In Eq. (3), $E_n = i^n E_0^{\text{inc}} (2n+1)/[n(n+1)]$ and the functions \mathbf{M}_n and \mathbf{N}_n are vector spherical wave functions. These functions are usually formulated in spherical coordinates and explicit expressions for them are given in Bohren and Huffman (1983). The expansion coefficients c_n and d_n are determined by the boundary conditions at S and depend only on the sphere's size parameter kR and refractive index m . Because of their independence on \mathbf{r} , c_n and d_n are not explicitly given here. The connection between \mathbf{E}^{int} and \mathbf{B}^{int} allows one to consider only the electric field when analyzing the waves, and thus no further mention of the magnetic field will be made.

The internal electric field can be expressed in Cartesian coordinates by first defining the radial, polar and azimuthal functions

$$R_n(\mathbf{r}) = E_n [c_n \mathbf{M}_n(\mathbf{r}) - i d_n \mathbf{N}_n(\mathbf{r})] \cdot \hat{\mathbf{r}}, \quad (5)$$

$$\Theta_n(\mathbf{r}) = E_n [c_n \mathbf{M}_n(\mathbf{r}) - i d_n \mathbf{N}_n(\mathbf{r})] \cdot \hat{\boldsymbol{\theta}}, \quad (6)$$

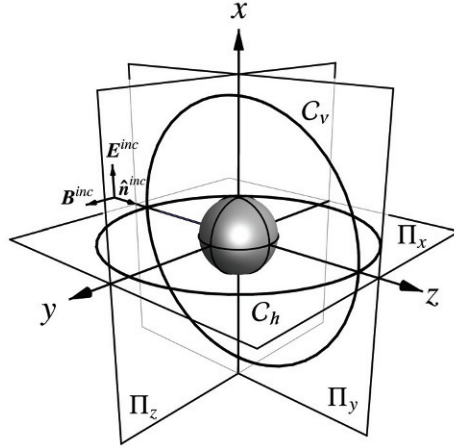


Fig. 2. Diagram showing the Π_x , Π_y , and Π_z planes and the C_h and C_v contours.

$$\Phi_n(\mathbf{r}) = E_n [c_n \mathbf{M}_n(\mathbf{r}) - i d_n \mathbf{N}_n(\mathbf{r})] \cdot \hat{\boldsymbol{\phi}}, \quad (7)$$

respectively. Then, the rectangular functions X_n , Y_n , and Z_n are defined as

$$X_n(\mathbf{r}) = R_n(\mathbf{r}) \sin \theta \cos \phi + \Theta_n(\mathbf{r}) \cos \theta \cos \phi - \Phi_n(\mathbf{r}) \sin \phi, \quad (8)$$

$$Y_n(\mathbf{r}) = R_n(\mathbf{r}) \sin \theta \sin \phi + \Theta_n(\mathbf{r}) \cos \theta \sin \phi + \Phi_n(\mathbf{r}) \cos \phi, \quad (9)$$

$$Z_n(\mathbf{r}) = R_n(\mathbf{r}) \cos \theta - \Theta_n(\mathbf{r}) \sin \theta, \quad (10)$$

Equations (8)–(10) now give the internal field in Cartesian coordinates,

$$\mathbf{E}^{\text{int}}(\mathbf{r}) = \sum_{n=1}^{\infty} [X_n(\mathbf{r}) \hat{\mathbf{x}} + Y_n(\mathbf{r}) \hat{\mathbf{y}} + Z_n(\mathbf{r}) \hat{\mathbf{z}}]. \quad (11)$$

To describe the reflection symmetry of \mathbf{E}^{int} and of the scattering arrangement, let Π_x , Π_y , and Π_z denote the y - z , x - z , and x - y planes through the origin, respectively. The intersection of S_l with the Π_x and Π_y planes defines the horizontal C_h and vertical C_v contours. These contours often have practical significance: in simple scattering measurements, the detector is confined to one of these contours, which then reside in the so-called horizontal and vertical scattering planes (Chu 1991; Hovenier 2000). **Figure 2** shows this configuration of planes and contours, and inspection of this figure demonstrates that the scattering arrangement is invariant under reflection about Π_x and Π_y . It is not, however, invariant under reflection about Π_z because of the propagation direction of the incident wave.

By expressing \mathbf{r} in Cartesian coordinates, the reflection symmetry of \mathbf{E}^{int} is revealed through the behavior of the rectangular functions X_n , Y_n , and Z_n under the transformations $x \rightarrow -x$ and $y \rightarrow -y$. Upon making these transformations in Eqs. (8)–(10), one can show that

$$X_n(x, y, z) = X_n(-x, y, z) = X_n(x, -y, z), \quad (12)$$

$$Y_n(x, y, z) = -Y_n(-x, y, z) = -Y_n(x, -y, z), \quad (13)$$

$$Z_n(x, y, z) = -Z_n(-x, y, z) = Z_n(x, -y, z). \quad (14)$$

Because the components of the field must be continuous inside of the particle, Eqs. (13) and (14) provide the additional conditions that

$$Y_n(\mathbf{r}) = 0, \quad \text{for } x = 0 \text{ or } y = 0, \quad (15)$$

$$Z_n(\mathbf{r}) = 0, \quad \text{for } x = 0. \quad (16)$$

Equation (12) shows that the x -component of the internal field is invariant under reflection about Π_x and Π_y , whereas Eq. (13) shows that the y -component of the field switches sign upon reflection about these planes. Equation (14) demonstrates that the z -component of the field is invariant under reflection about Π_y , but switches sign about Π_x . Equations (15) and (16) show that the y -component vanishes at points on Π_x and Π_y , while the z -component vanishes in Π_x ; this means that the field can have only an x -component in the Π_x plane. Moreover, owing to Eq. (12), the field magnitude must have reflection symmetry about the z -axis in this plane. To demonstrate all of these symmetries, a numerical implementation of the Mie solution is used to generate the plots in Fig. 3, which show the normalized field magnitude and direction in the Π_x , Π_y , and Π_z planes for a spherical particle with $kR = 12$ and $m = 1.33 + i0$.

4. A microphysical model

The scattered electric field \mathbf{E}^{sca} at observation points in the far-field zone is formally related to \mathbf{E}^{int} through the Volume Integral Equation (VIE),

$$\mathbf{E}^{\text{sca}}(\mathbf{r}) = \frac{\exp(ikr)}{r} \frac{k^2}{4\pi} (m^2 - 1) (\mathbf{\bar{I}} - \hat{\mathbf{r}} \otimes \hat{\mathbf{r}}) \cdot \int_{V^{\text{int}}} \mathbf{E}^{\text{int}}(\mathbf{r}') \exp(-ik\mathbf{r}' \cdot \hat{\mathbf{r}}) d\mathbf{r}', \quad (17)$$

where $\mathbf{\bar{I}}$ is the Cartesian identity dyad and $\hat{\mathbf{r}} \otimes \hat{\mathbf{r}}$ is the dyad formed by the dyadic product of $\hat{\mathbf{r}}$ with itself (Mishchenko et al. 2006).

Because the far-field scattered wave is transverse, the magnetic field is then given by

$$\mathbf{B}^{\text{sca}}(\mathbf{r}) = \frac{k}{\omega} \hat{\mathbf{r}} \times \mathbf{E}^{\text{sca}}(\mathbf{r}), \quad (18)$$

and thus both \mathbf{E}^{sca} and \mathbf{B}^{sca} are tangent to the imaginary surface S_l . The electric field is then fully described by its projections along the $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\phi}}$ unit vectors at \mathbf{r} . In general, the polarization state is elliptical, meaning that the tip of the \mathbf{E}^{sca} vector will trace out an ellipse, the polarization ellipse, in time at \mathbf{r} , recall Fig. 1 (Jackson 1999; Born and Wolf 1999). This ellipse graphically describes the polarization state and is characterized by three quantities: its ellipticity, defined as

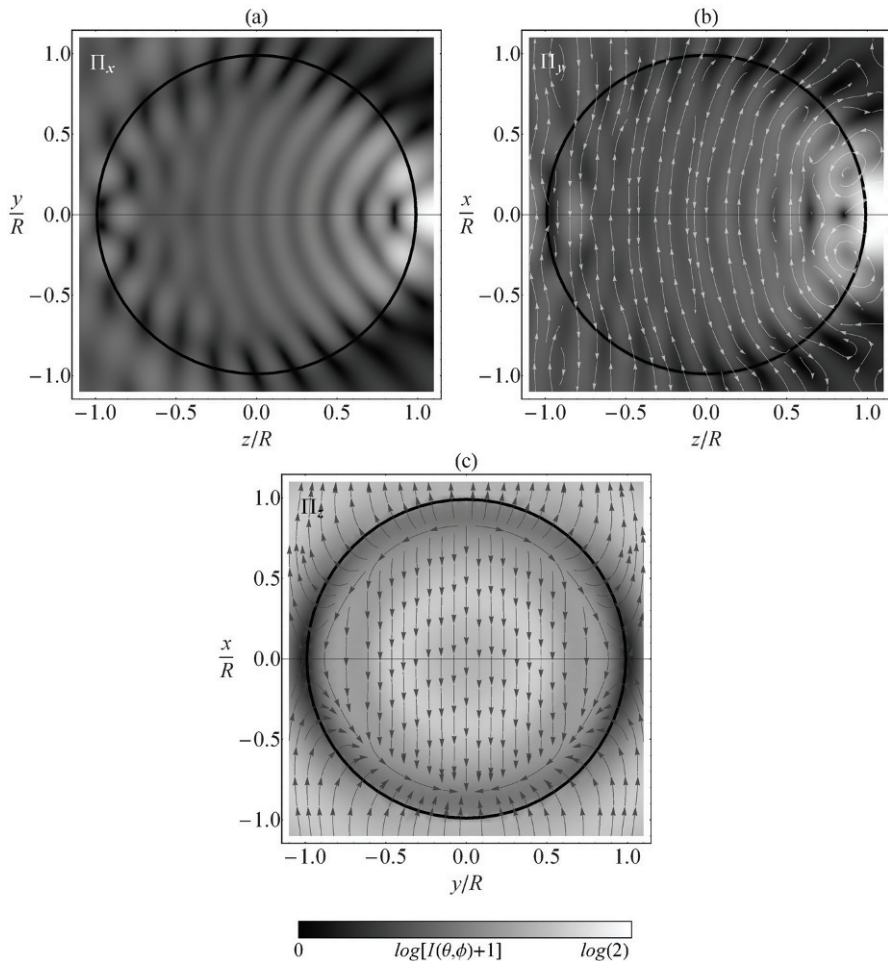


Fig. 3. Electric field inside and surrounding a spherical particle. The size parameter and refractive index are $kR = 12$ and $m = 1.33 + i0$, respectively. Panel (a) shows the normalized field-magnitude in the Π_x plane. No field vectors are shown here since the field is normal to this plane, see Eqs. (15) and (16). Panels (b) and (c) show the field in the Π_y and Π_z planes, respectively.

the ratio of its minor and major axes, its orientation in the $\hat{\theta} - \hat{\phi}$ plane, and the sense of rotation of \mathbf{E}^{sca} in time, i.e., either right or left-handed with respect to the direction of propagation (Mishchenko et al. 2002).

Now, consider dividing the sphere into many infinitesimal volume elements ΔV , located at \mathbf{r}_i . The size of an element is taken to be sufficiently small that the electric field throughout it is uniform. Then from Eq. (17)

$$\mathbf{E}^{\text{sca}}(\mathbf{r}) = \frac{\exp(ikr)}{r} \frac{k^2}{4\pi} (m^2 - 1) \lim_{\Delta V \rightarrow 0} \sum_i (\mathbf{I} - \hat{\mathbf{r}} \otimes \hat{\mathbf{r}}) \cdot \mathbf{E}^{\text{int}}(\mathbf{r}_i) \exp(-ik\mathbf{r}_i \cdot \hat{\mathbf{r}}) \Delta V, \quad (19)$$

where the sum runs over the locations \mathbf{r}_i of every volume element in the sphere. Next, let $c_o = k^2(m^2 - 1)\Delta V / 4\pi$ and define the following quantities:

$$z_{i\theta}(\hat{\mathbf{r}}) = c_o \{ [E_x^{\text{int}}(\mathbf{r}_i) \cos \phi + E_y^{\text{int}}(\mathbf{r}_i) \sin \phi] \cos \theta - E_z^{\text{int}}(\mathbf{r}_i) \sin \theta \} \exp(-ik\mathbf{r}_i \cdot \hat{\mathbf{r}}), \quad (20)$$

and

$$z_{i\phi}(\hat{\mathbf{r}}) = c_o [E_y^{\text{int}}(\mathbf{r}_i) \cos \phi - E_x^{\text{int}}(\mathbf{r}_i) \sin \phi] \exp(-ik\mathbf{r}_i \cdot \hat{\mathbf{r}}), \quad (21)$$

where E_x^{int} , E_y^{int} , and E_z^{int} are the Cartesian components of the internal field. These quantities, $z_{i\theta}$ and $z_{i\phi}$, in Eqs. (20) and (21) have physical significance (discussed below) and yield the following form for the scattered field:

$$\mathbf{E}^{\text{sca}}(\mathbf{r}) = \frac{\exp(ikr)}{r} \sum_i [z_{i\theta}(\hat{\mathbf{r}}) \hat{\boldsymbol{\theta}} + z_{i\phi}(\hat{\mathbf{r}}) \hat{\boldsymbol{\phi}}]. \quad (22)$$

Equation (22) reveals the meaning of the far-field form of the VIE (Mishchenko 2006): the internal electric field within a volume element acts as the source of an outward-traveling transverse spherical wave, or wavelet for short. The amplitudes of the vector components of the field of the i^{th} wavelet are given by the complex-valued quantities $z_{i\theta}$ and $z_{i\phi}$, which depend on the location of the volume element \mathbf{r}_i and the direction to the observation point $\hat{\mathbf{r}}$. The scattered field is then the superposition of all the wavelet fields radiated out to \mathbf{r} , i.e. Eq. (22). *This wavelet-based description of the field forms the basis for a microphysical perspective of the scattering process.* One can think of this model essentially as Huygens' principle, except formulated in exact full-vector form.

5. Consequences of the symmetry

Consider placing the observation point somewhere on the C_h contour. With \mathbf{r} on this contour, the azimuthal angle can have two values, $\phi = \pi/2$ and $\phi = 3\pi/2$, and one finds from Eq. (20) that

$$z_{i\theta}(\hat{\mathbf{r}}) = -c_o [E_z^{\text{int}}(\mathbf{r}_i) \sin \theta \mp E_y^{\text{int}}(\mathbf{r}_i) \cos \theta] \exp(-ik\mathbf{r}_i \cdot \hat{\mathbf{r}}), \quad (23)$$

and from Eq. (21),

$$z_{i\phi}(\hat{\mathbf{r}}) = \mp c_o E_x^{\text{int}}(\mathbf{r}_i) \exp(-ik\mathbf{r}_i \cdot \hat{\mathbf{r}}), \quad (24)$$

where the top and bottom signs in Eq. (23) correspond to $\phi = \pi/2$ and $\phi = 3\pi/2$, respectively. Next, the sum over the sphere's volume in Eq. (22) for the θ -component of the field is split in two. One sum covers the hemisphere for internal points with $x \geq 0$ and the other sum covers the hemisphere with $x < 0$, i.e.,

$$\sum_i z_{i\theta}(\hat{\mathbf{r}}) = \sum_j^{x \geq} z_{j\theta}(\hat{\mathbf{r}}) + \sum_k^{x <} z_{k\theta}(\hat{\mathbf{r}}), \quad (25)$$

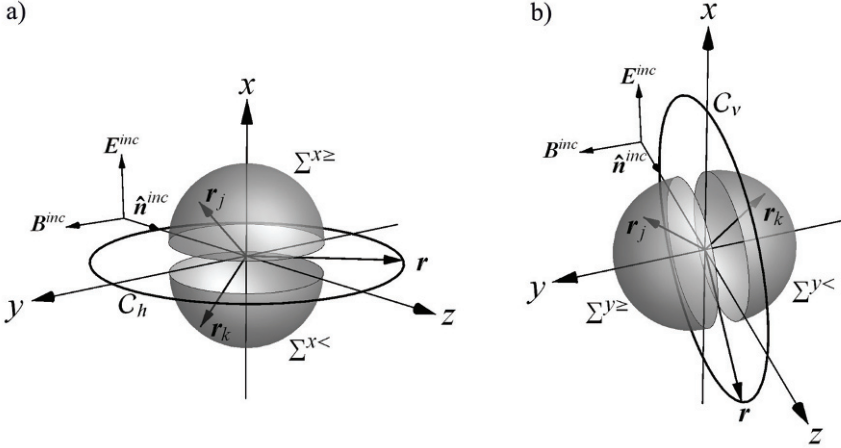


Fig. 4. Hemispheres used to derive Eqs. (27) and (32). The sphere is separated to delineate the hemispheres. The internal locations \mathbf{r}_j and \mathbf{r}_k are mirror points about (a) the Π_x plane and (b) the Π_y plane.

where the points \mathbf{r}_j and \mathbf{r}_k are mirror points about Π_x , see Fig. 4. From Eqs. (13) and (14) one sees that the y and z -components of the internal field appearing in Eq. (23) change sign upon reflection about Π_x . This means that the sums over the hemispheres in Eq. (25) cancel each other *term-by-term*, i.e.,

$$z_{j\theta}^{x\geq}(\hat{\mathbf{r}}) = -z_{j\theta}^{x<}(\hat{\mathbf{r}}), \quad (26)$$

where $z_{j\theta}^{x\geq}$ and $z_{j\theta}^{x<}$ represent the phasors associated with the hemisphere sums in Eq. (25). Then from Eqs. (22) and (24),

$$\mathbf{E}^{\text{scat}}(\mathbf{r}) = \mp c_0 \frac{\exp(ikR_l)}{R_l} \sum_i E_x^{\text{int}}(\mathbf{r}_i) \exp(-ik\mathbf{r}_i \cdot \hat{\mathbf{r}}) \hat{\boldsymbol{\phi}}, \quad \mathbf{r} \in C_h, \quad (27)$$

where the top and bottom signs correspond to $\phi = \pi/2$ and $\phi = 3\pi/2$, respectively.

Equation (27) shows that *for observation points on the C_h contour, the scattered wave is determined solely by the x -component of the sphere's internal electric field*. The symmetry of the field leads to the cancellation of the θ -components of the wavelet amplitudes, which carry the dependence on the other Cartesian components of the field. Moreover, the direction of the scattered field for observation points on this contour is only along $\hat{\boldsymbol{\phi}}$, which is parallel or antiparallel to the $\hat{\mathbf{x}}$ direction on C_h . This demonstrates that *the scattered wave for points on C_h has the same (linear) polarization state as the incident wave*.

Next consider the observation point on the C_v contour. Here $\phi = 0$ and $\phi = \pi$. Now Eqs. (20) and (21) become

$$z_{i\theta}(\hat{\mathbf{r}}) = -c_o [E_z^{\text{int}}(\mathbf{r}_i) \sin\theta \mp E_x^{\text{int}}(\mathbf{r}_i) \cos\theta] \exp(-ik\mathbf{r}_i \cdot \hat{\mathbf{r}}) \quad (28)$$

and

$$z_{i\phi}(\hat{\mathbf{r}}) = \pm c_o E_y^{\text{int}}(\mathbf{r}_i) \exp(-ik\mathbf{r}_i \cdot \hat{\mathbf{r}}), \quad (29)$$

where the top and bottom signs in Eqs. (28) and (29) correspond to $\phi = 0$ and $\phi = \pi$, respectively.

The sum in Eq. (22) is again split in two as above, except here the hemispheres correspond to $y \geq 0$ and $y < 0$, see Fig. 4. Then, for the ϕ -component,

$$\sum_i z_{i\phi}(\hat{\mathbf{r}}) = \sum_j^{y \geq} z_{j\phi}(\hat{\mathbf{r}}) + \sum_k^{y <} z_{k\phi}(\hat{\mathbf{r}}), \quad (30)$$

where \mathbf{r}_j and \mathbf{r}_k are now mirror points about Π_y . From the behavior of the y -component of \mathbf{E}^{int} in Eq. (13) and with Eq. (29), one finds that

$$z_{j\phi}^{x \geq}(\hat{\mathbf{r}}) = -z_{j\phi}^{x <}(\hat{\mathbf{r}}), \quad (31)$$

which, from Eqs. (22) and (29), demonstrates that *there is no contribution to the scattered field on C_v from the y -component of the internal field*. Then,

$$\mathbf{E}^{\text{sca}}(\mathbf{r}) = -c_o \frac{\exp(ikR_l)}{R_l} \sum_i [E_z^{\text{int}}(\mathbf{r}_i) \sin\theta \mp E_x^{\text{int}}(\mathbf{r}_i) \cos\theta] \exp(-ik\mathbf{r}_i \cdot \hat{\mathbf{r}}) \hat{\boldsymbol{\theta}}, \quad \mathbf{r} \in C_v, \quad (32)$$

where again the top and bottom signs correspond to $\phi = 0$ and $\phi = \pi$, respectively. This result demonstrates that *the scattered field on C_v has only a θ -component and hence is linearly polarized*.

Using the Stokes-parameter formalism, the polarization state of the scattered wave is expressed below in terms of the wavelet amplitudes, Eqs. (20) and (21). These parameters consist of four quantities I , Q , U , and V having units of power and can thus be directly measured. Explicit expressions for the parameters and a description of their measurement are given in Mishchenko et al. (2006). The quantity I describes the intensity of the scattered wave and the overall size of the polarization ellipse; whereas, Q and U collectively describe its orientation and ellipticity, and V describes the sense of rotation of the field vector.

In casting the Stokes parameters in terms of Eqs. (20) and (21), one obtains a double sum over the sphere's volume. To explicitly build the symmetries of the internal field into these parameters, the double sums are separated into component sums covering volume elements contained in the four regions of the sphere's interior bounded by the intersection of the Π_x and Π_y planes with the sphere's surface. These wedge-shaped regions, or quadrants for short, are labeled by the Roman numerals I–IV and the volume of each is denoted $V^I - V^{IV}$.

Using Eqs. (12)–(14), the double sums over the four quadrants are reduced to a single double-sum over *only the first* quadrant V^I . To explain how this is done, consider two arbitrary volume-element locations $\mathbf{r}_j^{(I)}$ and $\mathbf{r}_k^{(I)}$ in V^I . Let $\mathbf{r}_j^{(II)}$,

$\mathbf{r}_j^{(\text{III})}, \mathbf{r}_j^{(\text{IV})}$, and $\mathbf{r}_k^{(\text{II})}, \mathbf{r}_k^{(\text{III})}, \mathbf{r}_k^{(\text{IV})}$ be the locations obtained in the remaining three quadrants by performing successive reflections of $\mathbf{r}_j^{(\text{I})}$ and $\mathbf{r}_k^{(\text{I})}$ about the Π_x and Π_y planes. Next, let $z_{j\theta}^{(\text{I})}, z_{j\phi}^{(\text{I})}, z_{k\theta}^{(\text{I})}, z_{k\phi}^{(\text{I})}$ be the wavelet amplitudes of Eqs. (20) and (21) evaluated at the points $\mathbf{r}_j^{(\text{I})}$ and $\mathbf{r}_k^{(\text{I})}$, respectively. Lastly, let $z_{j\theta}^{(\text{II})}, z_{j\phi}^{(\text{II})}, \dots, z_{k\theta}^{(\text{IV})}, z_{k\phi}^{(\text{IV})}$ be the wavelet amplitudes $z_{j\theta}$ and $z_{j\phi}$ evaluated at the reflected points $\mathbf{r}_j^{(\text{II})}, \mathbf{r}_j^{(\text{III})}, \mathbf{r}_j^{(\text{IV})}$, and similarly for the k subscript terms. Because the points in the $\text{V}^{\text{II}} - \text{V}^{\text{IV}}$ quadrants are related to the points in the V^{I} quadrant through the same reflection operations that define the symmetry of the internal field, the wavelet amplitudes in the $\text{V}^{\text{II}} - \text{V}^{\text{IV}}$ quadrants are directly related to the amplitudes in only the V^{I} quadrant. With these considerations, the Stokes parameters can be expressed as Eqs. (33)–(36) below as double sums over only the V^{I} quadrant,

$$I(\hat{\mathbf{r}}) = \frac{1}{2} \sqrt{\frac{\epsilon_0}{\mu_0}} \sum_{j,k}^{(\text{I})} \sum_{\alpha,\beta} \{ z_{j\theta}^{(\alpha)}(\hat{\mathbf{r}}) [z_{k\theta}^{(\beta)}(\hat{\mathbf{r}})]^* + z_{j\phi}^{(\alpha)}(\hat{\mathbf{r}}) [z_{k\phi}^{(\beta)}(\hat{\mathbf{r}})]^* \}, \quad (33)$$

$$Q(\hat{\mathbf{r}}) = \frac{1}{2} \sqrt{\frac{\epsilon_0}{\mu_0}} \sum_{j,k}^{(\text{I})} \sum_{\alpha,\beta} \{ z_{j\theta}^{(\alpha)}(\hat{\mathbf{r}}) [z_{k\theta}^{(\beta)}(\hat{\mathbf{r}})]^* - z_{j\phi}^{(\alpha)}(\hat{\mathbf{r}}) [z_{k\phi}^{(\beta)}(\hat{\mathbf{r}})]^* \}, \quad (34)$$

$$U(\hat{\mathbf{r}}) = \frac{1}{2} \sqrt{\frac{\epsilon_0}{\mu_0}} \sum_{j,k}^{(\text{I})} \sum_{\alpha,\beta} \{ z_{j\theta}^{(\alpha)}(\hat{\mathbf{r}}) [z_{k\phi}^{(\beta)}(\hat{\mathbf{r}})]^* + z_{j\phi}^{(\alpha)}(\hat{\mathbf{r}}) [z_{k\theta}^{(\beta)}(\hat{\mathbf{r}})]^* \}, \quad (35)$$

$$V(\hat{\mathbf{r}}) = \frac{i}{2} \sqrt{\frac{\epsilon_0}{\mu_0}} \sum_{j,k}^{(\text{I})} \sum_{\alpha,\beta} \{ z_{j\phi}^{(\alpha)}(\hat{\mathbf{r}}) [z_{k\theta}^{(\beta)}(\hat{\mathbf{r}})]^* + z_{j\theta}^{(\alpha)}(\hat{\mathbf{r}}) [z_{k\phi}^{(\beta)}(\hat{\mathbf{r}})]^* \}, \quad (36)$$

where $\alpha, \beta \in \{\text{I}, \text{II}, \text{III}, \text{IV}\}$, and the notation $\Sigma^{(\text{I})}$ means that the sum covers the first quadrant V^{I} only, and $[\dots]^*$ denotes complex conjugation.

By making the transformations $\phi \rightarrow \pi - \phi$ and $\phi \rightarrow 2\pi - \phi$ in Eqs. (33)–(36), which describe the reflection of the *observation point* about the Π_x and Π_y planes, respectively, one can show that

$$I(\theta, \phi) = I(\theta, \pi - \phi) = I(\theta, 2\pi - \phi), \quad (37)$$

$$Q(\theta, \phi) = Q(\theta, \pi - \phi) = Q(\theta, 2\pi - \phi), \quad (38)$$

$$U(\theta, \phi) = -U(\theta, \pi - \phi) = -U(\theta, 2\pi - \phi), \quad (39)$$

$$V(\theta, \phi) = -V(\theta, \pi - \phi) = -V(\theta, 2\pi - \phi). \quad (40)$$

The symmetry conditions Eqs. (37)–(40) result from the *term-by-term* equality in Eqs. (33)–(36) under the reflection transformations. That this is so is a direct consequence of the internal field's reflection symmetry. Equations (39) and (40) and the requisite continuity of the Stokes parameters provides the extra conditions

$$U(\theta, 0) = U(\theta, \pi/2) = U(\theta, \pi) = U(\theta, 3\pi/2) = 0, \quad (41)$$

$$V(\theta, 0) = V(\theta, \pi/2) = V(\theta, \pi) = V(\theta, 3\pi/2) = 0. \quad (42)$$

Equations (37)–(42) are also obtained in Hovenier (1970), Yueh et al. (1994), and van de Hulst (1981) except using different methods. There the analysis concerns

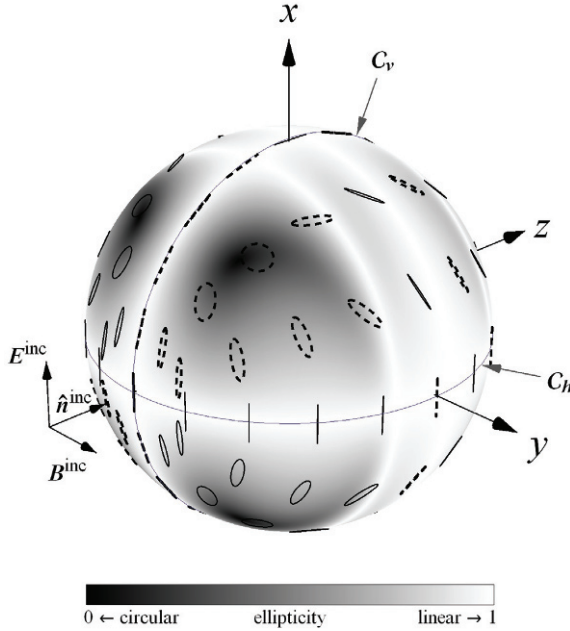


Fig. 5. Polarization state of the scattered wave for a spherical particle with $kR = 4$ and $m = 1.25 + i0$. The plot shows the polarization ellipses at various points on S_i . The thick dashed ellipses correspond to left-handed rotation of the field; whereas, thin solid ellipses correspond to right-handed rotation. The ellipticity is indicated by gray shades on S_i . Darker (lighter) shades indicate more circular (linear) polarization.

either the scattered field's structure only or the elements of a matrix transformation that maps the incident onto the scattered wave. Notice that Eqs. (41) and (42) reveal that the scattered wave is linearly polarized in the planes of reflection symmetry. Moreover, these properties of the Stokes parameters require *all* of the reflection symmetries of the internal field derived in Section 3. Therefore, at least from a mathematical viewpoint, this polarimetric behavior relates only to spherical particles and is not necessarily expected to apply to more arbitrary situations, e.g., nonspherical particles. Section 8 however will demonstrate otherwise.

6. Illustration of the polarimetric effect of reflection symmetry

Here the polarization state for a spherical particle with $kR = 4$ and $m = 1.25 + i0$ is examined to demonstrate graphically the effect of the symmetries discussed above. The size and refractive index of the sphere is chosen such that the polarimetric structures are not too complicated to examine. Other kR and m values have been investigated and give the same qualitative results.

Figure 5 shows the polarization state of the scattered wave on the imaginary surface S_i over mostly the backscattering directions. Lighter shades correspond to

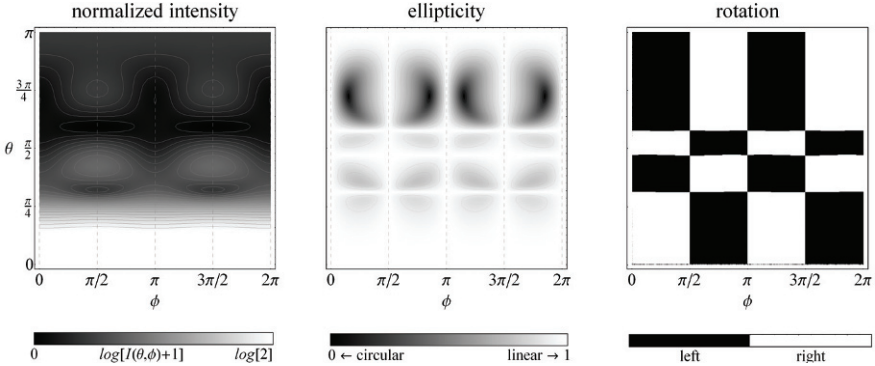


Fig. 6. Intensity, ellipticity, and rotation of the scattered wave. The intensity is normalized to the forward direction $\theta=0$ and gray shades are assigned in log scale as indicated. The sphere is the same as in Fig. 5.

more linear polarization; whereas, darker shades correspond to more circular. Polarization ellipses are shown for various points on S_l to aid interpretation. The left-handed rotation of an ellipse is indicated by a thick dashed line and the right-handed rotation is indicated by a thin solid line. Inspection of this figure shows that the wave is linearly polarized along the C_h and C_v contours as indicated by Eqs. (27) and (32).

Figure 6 displays the intensity, ellipticity and rotation of the scattered wave on S_l for the same particle as in Fig. 5, except all directions are shown. The intensity displays reflection symmetry about both Π_x and Π_y , recall Fig. 2. Notice from Fig. 5 that the scattered wave is, in general, elliptically polarized in all directions except those contained in the Π_x and Π_y planes. The ellipticity displays reflection symmetry about these planes and shows that the wave is more linearly polarized in the forward directions as compared to the backward directions. The plot of the rotation reveals that it changes sense upon reflection about either plane. Moreover, the angular distribution of the rotation-sense appears to be conserved in that for each region of right-handed rotation, there is another region that is the mirror image to the former about Π_x or Π_y that shows left-handed rotation.

Every reflection symmetry in Figs. 5 and 6 is explained by the analysis in Section 5. Among other things, these figures show that the scattered wave is linearly polarized along the C_h and C_v contours. This can be understood from Eq. (42), where $V=0$ for directions lying in the Π_x and Π_y planes. Moreover, the orientation of the (linear) polarization ellipse is along either the $\hat{\theta}$ - or $\hat{\phi}$ -direction, since $U=0$ from Eq. (41). Figure 6 shows that the intensity of the wave is invariant upon reflection about Π_x and Π_y , which is due to the invariance of Eq. (37) upon reflection. The change in the wave's rotation and the orientation of its polarization ellipse between directions reflected about Π_x and Π_y is accounted for by the negative sign appearing in Eqs. (39) and (40).

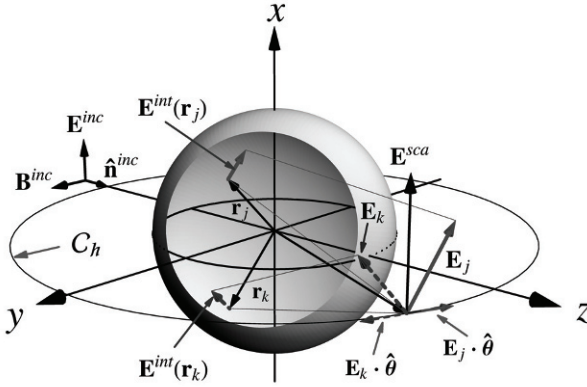


Fig. 7. Illustration of the wavelet-component cancellation concept. Internal mirror-points \mathbf{r}_j and \mathbf{r}_k about the Π_x plane are shown along with an example of the internal field vectors $\mathbf{E}^{\text{int}}(\mathbf{r}_j)$ and $\mathbf{E}^{\text{int}}(\mathbf{r}_k)$ that satisfy the symmetry conditions of Eqs. (12)–(16). Each of these internal wavelets contribute \mathbf{E}_j and \mathbf{E}_k to the scattered field where the cancellation of their θ -components yields the observed linear polarization.

Comparing the rotation and ellipticity in Fig. 6 reveals that the wave becomes linearly polarized where angular regions of opposing rotation meet. There are also distinct points where the ellipticity becomes circular. These points of circular polarization and the lines of linear polarization appear to be examples of the so-called “C - points” and “L - lines,” respectively, of Nye (1983, 1997).

7. Interpretation

As is seen above, the microphysical perspective afforded by the VIE is able to connect the reflection symmetries of the internal field to the structure of the polarization state. This connection can be given a clearer physical meaning by remembering that the quantities $z_{i\theta}$ and $z_{i\phi}$ represent the amplitudes of the θ and ϕ - components of each wavelet's contribution to the scattered wave. For example, suppose that the observation point resides on the C_h contour as shown in Fig. 7. Here, two wavelet locations at points \mathbf{r}_j and \mathbf{r}_k are shown where the points are mirror-image twins about Π_x . The electric field vectors shown at the points, $\mathbf{E}^{\text{int}}(\mathbf{r}_j)$ and $\mathbf{E}^{\text{int}}(\mathbf{r}_k)$, obey the symmetry properties of Eqs. (12)–(16). The contributions to the scattered field from these wavelets are shown at \mathbf{r} and labeled \mathbf{E}_j and \mathbf{E}_k , respectively, along with their projection onto $\hat{\theta}$. These projections are given by $z_{j\theta}$ and $z_{k\theta}$, except for a factor of $\exp(ikr)/r$ which is a constant on C_h . The figure demonstrates that these components of the wavelet's contribution to the scattered field are opposed and thus cancel each other; it is only the ϕ -components then that contribute to the scattered wave, and these in turn depend only on the x -component of the internal field. A similar cancellation concept can be used to explain the linear polarization state observed on the C_v contour.

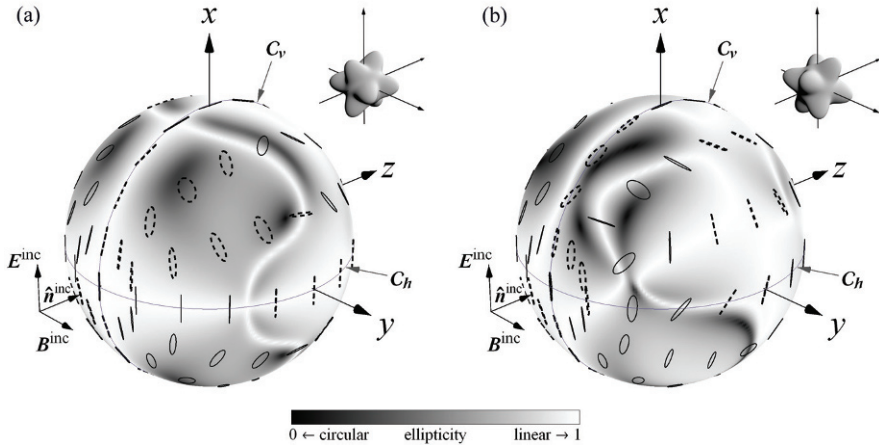


Fig. 8. Symmetry breaking and its effect on the polarization state. In (a) the particle's orientation forms a scattering arrangement with the same planes of reflection symmetry as the spherical particles above; recall Fig. 2. The insets show the particle in these orientations.

From this wavelet-cancellation viewpoint, one can now understand that *the ellipticity of the polarization state is due to the break in symmetry of the scattering arrangement that results from taking the observation point off the C_h or C_v contours*. This break in symmetry prevents the delicate cancellation of wavelet-components that otherwise gives rise to linear polarization and introduces a relative phase shift between the components.

8. Implications

Although the analysis above relates specifically to spherical particles, the conclusions derived from the wavelet-cancellation concept have broader implications. First, consider a nonspherical particle such as the star-shaped one shown in Fig. 8. This particle shape has multiple planes of reflection symmetry, and depending on its orientation with respect to the incident wave, these planes can be degenerate with those associated with the spherical particles above. If this is the case then the wavelet-cancellation concept must apply even though the particle is nonspherical, and hence, is beyond the mathematical scope of the preceding analysis. Given that the scattering arrangement's reflection symmetry is the same as the spherical particles, the scattered wave will be linearly polarized along C_h and C_v . The reason for this is the same as earlier: each differential volume element in the particle can be paired with its mirror-image twin across a reflection plane and the vector components of the associated wavelets exactly cancel via destructive interference, except along the polarization direction of the incident wave. If however, the orientation of the particle is changed such that its planes of reflection symmetry no

longer coincide with Π_x and Π_y , this cancellation will not occur. Consequently, the scattered wave can become elliptically polarized.

Figure 8 shows an example of the effect that this orientation-induced break in symmetry can have on the polarization state. The star-shaped particle is shown in two orientations, (a) and (b). Here the scattered wave is calculated using the DDA, see, e.g. Yurkin and Hoekstra (2007) and Berg (2008). In (a), the particle's planes of reflection symmetry are degenerate with those in the spherical-particle arrangement. Consequently, the polarization is linear along C_h and C_v , as predicted by the wavelet-cancellation model. By rotating the particle in (b), the reflection symmetry of the arrangement is broken and one can see that the polarization state becomes elliptical along these contours.

Another implication of the wavelet-cancellation concept regards the effect of absorption in small particles. Provided that $kR \ll 1$, the field inside a spherical or spheroidal nonabsorbing particle will be uniform and directed along the polarization direction of the incident wave.[†] This is an example of the so-called Rayleigh limit and is independent of the orientation of the particle (Bohren and Huffman 1983). Equations (20), (21), and (36) show that for such a case, i.e., $\mathbf{E}^{\text{int}} = E_x^{\text{int}} \hat{\mathbf{x}}$, where E_x^{int} is a constant, the handedness of the wave vanishes for all directions. This means that the scattered wave is linearly polarized in *all* directions; a result familiar from dipole radiation, which is essentially what the Rayleigh limit represents. Figure 9a demonstrates this behavior for a small prolate-spheroidal particle in an orientation without reflection symmetry. The particle's aspect ratio – the ratio of its major to minor axes – is 0.5 and its size parameter and refractive index is $kR = 0.1$ and $m = 1.25 + i0$, respectively. As with the star-shaped particle, the scattered wave is calculated using the DDA.

If the particle is absorbing, the internal field is nonuniform. Compared to its nonabsorbing counterpart, absorption causes the field to become more tangential to the particle's inner surface and to decay in magnitude with distance from the geometrically illuminated side (Jackson 1999). This effect is schematically depicted in Fig. 9b along with the polarization state for the same spheroidal particle as in Fig. 9a, except here the refractive index is $m = 1.25 + i0.75$ and the particle is oriented with its axis of rotational symmetry along $\hat{\mathbf{n}}^{\text{inc}}$. This orientation yields a scattering arrangement with the same reflection symmetry as the star-shaped particle in Fig. 8a. As before, a consequence of this symmetry should be linear polarization of the scattered wave along the C_h and C_v contours. The polarization plot in Fig. 9b shows that this is indeed case.

Now suppose that this Rayleigh-sized absorbing particle is rotated to break the reflection symmetry of the scattering arrangement. This means that the wavelet-component cancellation no longer occurs for points on the C_h and C_v contours and the scattered wave will become elliptically polarized. This situation is shown in Fig. 9c where the same absorbing particle in Fig. 9b is used but rotated to break

[†] Note that the internal field is necessarily uniform only for spherical and spheroidal particles, and in general depends on particle shape even in the Rayleigh limit, see Stratton (1941).

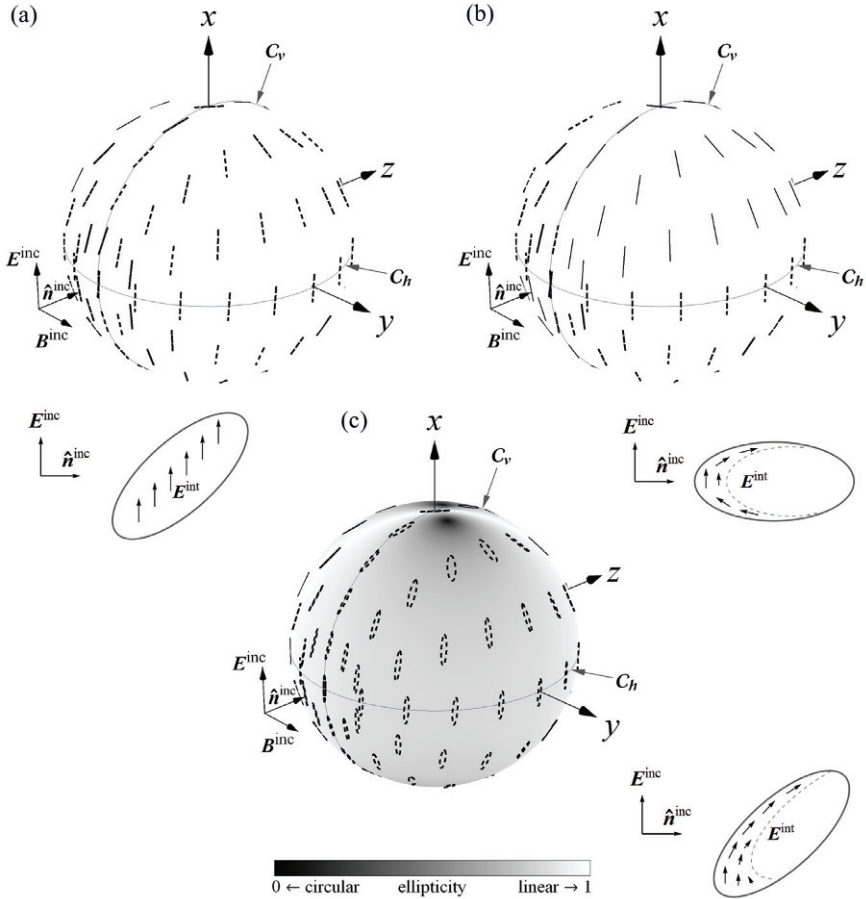


Fig. 9. Polarimetric influence of absorption in the Rayleigh limit. The particle here is a prolate spheroid with an aspect ratio of 0.5 and size parameter of $kR = 0.1$. In (a), the particle is nonabsorbing with a refractive index of $m = 1.25 + i0$ and is in an orientation that breaks the reflection symmetry of the scattering arrangement. The reflection symmetry is restored in (b) and the particle is now absorbing with $m = 1.25 + i0.75$. This absorbing particle is then rotated in (c) to again break the reflection symmetry of the scattering arrangement.

the symmetry. Qualitatively, this reorientation causes the attenuated internal field to skew asymmetrically as depicted. This asymmetry causes an imbalance in the wavelet-component cancellation that would otherwise be responsible for linear polarization along C_h and C_v .

The polarimetric behavior seen in Fig. 9 is surprising because one would typically expect particle-shape related effects to disappear in the far-field as the size parameter shrinks to the Rayleigh limit: the scattered and incident waves should

share the same polarization state due to the dipole-like character of scattering by particles in the Rayleigh regime. Although it is perhaps obvious that the internal field becomes nonuniform in an absorbing spheroidal particle, it is less clear how this nonuniformity affects the polarization state, if it does so at all. Figure 9 reveals that it is not the nonuniformity of the internal field alone that accounts for the appearance of elliptical polarization, but that the reflection symmetry of the scattering arrangement must also be broken.

9. Conclusion

By regarding the wave inside of a spherical particle as the source of the scattered wave, this work connects the particle's reflection symmetry to general features in the angular structure of the far-field polarization state. This is done by expressing the scattered wave as the superposition of wavelets originating from the differential volume elements in the particle, in close analogy to Huygens' principle. The result is a so-called microphysical model that can relate particle shape to features in the polarization state by revealing special cancellation effects between the wavelets. For example, provided that the incident wave is linearly polarized, one finds that the scattered wave is also linearly polarized in directions contained in the planes of reflection symmetry. This is independent of the particle size or refractive index. In other directions however, the wave is elliptically polarized due to relative phase-shifts between the wavelets introduced by the break in symmetry caused by taking the observation point off of the symmetry planes. Moreover, some components of the internal field are shown to make no *direct* contribution to the scattered wave for certain high-symmetry directions.

All of the polarimetric symmetry-related properties found in this work have been previously described by others, although without any mention of the role of the internal wave. The true utility of this work, however, derives from an intuition engendered by the microphysical model allowing one to understand the origin of features in the polarimetric profile for spherical particles. As a consequence, one is able to predict the appearance of similar features for nonspherical particles *without recourse to any calculation or simulation*.

Acknowledgments

The author is thankful for the helpful discussions with Michael Mishchenko, O. Larry Weaver, Chris Sorensen, Amit Chakrabarti, and Thomas Wriedt.

References

- Baum, C. E., and N. H. Kritikos, 1995: *Electromagnetic Symmetry* (Taylor and Francis, Bristol).
- Berg, M. J., 2008: A microphysical model of scattering, absorption, and extinction in electromagnetic theory. Ph.D. dissertation (Kansas State University, Manhattan, KS).
- Berg, M. J., C. M. Sorensen, and A. Chakrabarti, 2010: Explanation of the patterns in Mie scattering. *J. Quant. Spectrosc. Radiat. Transfer* **111**, 782–794.

- Bohren, C. F., and D. R. Huffman, 1983: *Absorption and Scattering of Light by Small Particles* (Wiley, New York).
- Born, M., and E. Wolf, 1999: *Principles of Optics* (Cambridge University Press, Cambridge, UK).
- Bransden, B. H., and C. J. Joachain, 2003: *Physics of Atoms and Molecules* (Pearson, Essex).
- Chu, B., 1991: *Laser Light Scattering* (Academic Press, New York).
- Hovenier, J. W., 1969: Symmetry relationships for scattering of polarized light in a slab of randomly oriented particles. *J. Atmos. Sci.* **26**, 488–499.
- Hovenier, J. W., 1970: Principles of symmetry for polarization studies of planets. *Astron. Astrophys.* **7**, 86–90.
- Hovenier, J. W., 2000: Measuring scattering matrices of small particles at optical wavelengths. In M. I. Mishchenko, J. W. Hovenier, and L. D. Travis, Eds., *Light Scattering by Nonspherical Particles: Theory, Measurements, and Applications* (Academic Press, San Diego), pp. 355–365.
- Hovenier, J. W., and D. W. Mackowski, 1998: Symmetry relations for forward and backward scattering by randomly oriented particles. *J. Quant. Spectrosc. Radiat. Transfer* **60**, 483–492.
- Hu, C.-H., G. W. Kattawar, M. E. Parkin, and P. Herb, 1987: Symmetry theorems on the forward and backward scattering Muller matrices for light scattering from nonspherical dielectric scatterer. *Appl. Opt.* **26**, 4159–4173.
- Jackson, J. D., 1999: *Classical Electrodynamics* (Wiley, New York).
- Mishchenko, M. I., 2006. Far-field approximation in electromagnetic scattering. *J. Quant. Spectrosc. Radiat. Transfer* **100**, 268–276.
- Mishchenko, M. I., L. D. Travis, and A. A. Lacis, 2002: *Scattering, Absorption, and Emission of Light by Small Particles* (Cambridge University Press, Cambridge, UK).
- Mishchenko, M. I., L. D. Travis, and A. A. Lacis, 2006: *Multiple Scattering of Light by Particles: Radiative Transfer and Coherent Backscattering* (Cambridge University Press, Cambridge, UK).
- Nghiem, S. V., S. H. Yueh, R. Kwok, and F. K. Li, 1992: Symmetry properties in polarimetric remote sensing. *Radio Sci.* **27**, 693–711.
- Nye, J. F., 1983: Lines of circular polarization in electromagnetic fields. *Proc. R. Soc. London A* **389**, 279–290.
- Nye, J. F., 1997: Line singularities in wave fields. *Phil. Trans. R. Soc. London A* **355**, 2065–2069.
- Ren, K. F., G. Grehan, and G. Gouesbet, 1994: Symmetry relations in generalized Lorenz–Mie theory. *J. Opt. Soc. Am. A* **11**, 1812–1817.
- Schultz, F. M., K. Stamnes, and J. J. Stamnes, 1999: Point-group symmetries in electromagnetic scattering. *J. Opt. Soc. Am. A* **16**, 853–865.
- Stratton, J. A., 1941: *Electromagnetic Theory* (McGraw Hill, New York).
- Tyynela, J., E. Zubko, G. Videen, and K. Muinon, 2007: Interrelating angular scattering characteristics to internal electric fields for wavelength-sized spherical particles. *J. Quant. Spectros. Radiat. Transfer* **106**, 520–534.
- van de Hulst, H. C., 1981: *Light Scattering by Small Particles* (Dover, New York).
- Waterman, P. C., 1971: Symmetry, unitarity, and geometry in electromagnetic symmetry. *Phys. Rev. D* **3**, 825–839.
- Yueh, S. H., R. Kwok, and S. V. Nghiem, 1994: Polarimetric scattering and emission properties of targets with reflection symmetry. *Radio Sci.* **29**, 1409–1420.
- Yurkin, M. A., and A. G. Hoekstra, 2007: The discrete dipole approximation: an overview and recent developments. *J. Quant. Spectrosc. Radiat. Transfer* **106**, 558–589.

Light scattering by large faceted particles

A. G. Borovoi* and N. V. Kustova

V. E. Zuev Institute of Atmospheric Optics, Academician Zuev Sq. 1, Tomsk 634021, Russia

Abstract. The scattered field for a large faceted particle, that is, a particle bounded by a set of plane surfaces, is expanded in a series over multiplicity of reflection/refraction events inside the particle. As a result, the scattered field is represented as a set of plane-parallel beams in the near zone of the particle. Then every beam is easy to describe analytically at an arbitrary distance from the particle. In particular, the beams undergo the Fraunhofer diffraction in the wave zone. In the far zone, a simple relationship between the geometric-optics and physical-optics Mueller matrices is considered that allows one to use both of them. In the majority of problems of practical interest, only one or a small number of the beams makes the main contribution to the measured quantities. In these cases, the geometric-optics and physical-optics Mueller matrices can be calculated either analytically or numerically by means of appropriate fast codes. In particular, it is shown that for particles with a refractive index of about 1.3 such as ice crystals, the main contribution to the phase function for randomly oriented particles is caused by the beams with one internal reflection. This fact allows us to parameterize the phase function for ensembles of such particles. Furthermore, the nature of the backscattering peak well-known for ice crystal clouds is discussed in the framework of the geometric-optics Mueller matrix. Finally, an efficient algorithm for the calculation of diffraction near the forward-scattering direction for ensembles of faceted particles is developed which corresponds to the physical-optics approximation.

Keywords: light scattering, Fraunhofer diffraction, backscatter, ice crystals

1. Introduction

The challenging problem of light scattering by nonspherical and large, relative to the incident wavelength, particles has been attacked over the previous decades from two opposite perspectives. The first approach is to solve the Maxwell equations directly by means of a numerical computer code. Here, the T -matrix, finite-difference time-domain (FDTD), discrete dipole approximation (DDA), and other methods have been widely explored (e.g., [15,16,21]). For all these methods, there are applicability limits on maximal particle sizes depending on the computer

* Corresponding author. E-mail: borovoi@iao.ru

power. At present, this limit is reached at, say, $x \sim 20$, where $x = (\text{particle size}) / (\text{incident wavelength})$.

However, not all particles of practical interest obey this condition. In the case of $x \gg 1$, another approach largely insensitive to particle size is well known, viz., the method of geometric optics. However, in spite of its intuitive obviousness and simplicity, this approximation proves to be computationally demanding as well (see, e.g., [11]).

Among the particle types to which the geometric-optics approximation has been applied, ice crystals seem to be of particular practical importance because cirrus clouds consisting of such particles affect significantly the radiation balance of the Earth and, consequently, its climate. In general, a theory of light scattering by a large faceted particle, i.e., a particle bounded by a set of plane faces such as a crystal, is independent of the specific particle shape. Therefore, theoretical equations considered in this chapter are applicable to any faceted particle, but actual numerical results are presented only for ice crystals.

Our approach to light scattering by a large faceted particle is based on a series expansion of the scattered field in a set of beams formed in the near zone of the particle. Then the beams propagate away from the particle, forming the Fraunhofer diffraction patterns. Each beam is described analytically at an arbitrary distance from the particle. In numerous problems of practical interest, only one type of beam (or a small number of types) provides the dominant contribution. In these cases, we do not need to take into account the rest of the beams, and the problem is solved by deriving simple analytical equations that can be handled with efficient numerical algorithms.

In general, geometric optics represents the electromagnetic field as a set of ray trajectories allowing one to treat the electromagnetic field from the corpuscular point of view. In this treatment, photons propagate along a ray trajectory, and their polarization can be characterized by the Stokes parameters. We should emphasize that the term “photon” is used here only for brevity and intuitive consistency similarly, for example, to quantum mechanics wherein high-energy particles are often treated as corpuscles. The concept of a photon in the physical sense of quantum electrodynamics is not implied in any way. The corpuscular treatment is also convenient for the development of fast and efficient computer codes simulating various light propagation and scattering processes. This technique is referred to as ray tracing.

Numerous papers have been devoted to calculations of the scattering (Mueller) matrix for ice crystal particles by use of the ray tracing technique [5,7,9,12–14,17–20,22,23]. Here, every ray undergoing several refractions and/or reflections is characterized by the Stokes parameters or, more generally, by the 4×4 Mueller matrix. Then the Mueller matrices for all the rays leaving a crystal in the same direction are added. Such a matrix obtained for the whole particle is hereinafter referred to as the geometric-optics matrix.

A shortcoming of the geometric-optics matrix is that it ignores completely the wave nature of light. In other words, the geometric-optics matrix neglects such ba-

sis phenomena of physical optics as diffraction and interference. Therefore, there is a need to develop a physical-optics approximation defined as an approach wherein the diffraction and interference concepts are accounted for in a rigorous and consistent way. A purpose of this paper is to consider certain steps in this direction. In particular, we demonstrate that the physical-optics approximation is easily derived for faceted particles.

It is worthwhile to note that the majority of authors calculating the Mueller matrices for ice crystals with ray tracing algorithms have included the wave phenomena into consideration by using various, generally rough, approximations. Since these authors were interested in the case of randomly oriented particles, these approximations were justified by the assumption that averaging over particle orientations would essentially extinguish the interference phenomenon. In particular, the simplest approximation for the inclusion of the wave phenomena has been to replace the delta-function peak of the geometric-optics approach appearing in the forward-scattering direction. This delta-function peak is caused by two physical phenomena. The first one is diffraction by the particle projection taking place for particles of arbitrary shapes. The second one is the diffraction caused by parallel facets of pristine crystals. The parallel facets are equivalent to a plane-parallel plate resulting in the propagation of certain outgoing rays in exactly the forward direction. The simplest way to include this diffraction is to replace the delta function by the well-known Fraunhofer diffraction pattern from a certain circle [7,9,14,19,22]. The area of the diffraction circle is usually assumed to be the area of the particle projection on the plane normal to the forward scattering direction.

A more rigorous treatment of the wave phenomena in the scattering problem demands more sophisticated approaches. For this purpose, one needs to abandon the corpuscular treatment of radiation and associate the geometric-optics rays directly with the electromagnetic field. Undoubtedly, it is not difficult to characterize theoretically a ray trajectory by the electric and magnetic field vectors rather than the Stokes parameters, or by the complex-valued 2×2 amplitude (Jones) matrix rather than the 4×4 Mueller matrix. However, such simulations of the electromagnetic field require much increased computer resources and have not been possible so far. Still, some earlier publications considered the implementation of the Jones-matrix formalism for a ray trajectory. In particular, Cai and Liou [7] used this approach to simplify their theoretical considerations, while still obtaining numerical results via the use of conventional algorithms based on the corpuscular treatment of radiation. Similarly, Takano and Jayaweera [22] applied the Jones matrices in order to prove symmetry relations for the geometric-optics Mueller matrix.

An obvious approach is to divide the light scattering problem into two steps. The first one is to find the scattered field either inside the particle or at a surface surrounding it. This surface can coincide with the particle boundary or can be an arbitrary surrounding surface located in the near zone of the particle, i.e., at a distance $R \ll a^2/\lambda$. Here, a is the typical particle size and λ is the wavelength. Since

the scattered field in the wave zone $R \gg a^2/\lambda$ is of conventional interest, the second step is the calculation of the wave-zone field. It is well known that for a given near-zone field, the wave-zone field can be rigorously calculated as a convolution of the near-zone field with the Green function of the Maxwell equations. If one uses the near-zone field determined inside the particle, this convolution is an integral over the particle volume with the volume Green function; otherwise one deals with an integral over a surface with the surface Green function.

Now, turning back to the first step, one can state that, for large particles, the near-zone field can be calculated approximately, but to a good accuracy, within the framework of geometric optics only. Further, in this approach, the near-zone field splits into two parts. One of them is formed by the refracted and reflected rays, and the other part is the shadow-forming field created by a jump of the incident radiation along the particle contour. This approach is quite general and obvious, and its implementations can be found in various fields of optics, radiophysics, etc. For example, one of the present authors (A.B.) has successfully used these ideas to treat the problems of both single and multiple scattering of light by atmospheric precipitation, by an ensemble of closely packed large particles, etc. [1].

As to the problem of light scattering by large crystal particles, Muinonen [17] was the first to propose the above-mentioned ideas. He called this approach the Kirchhoff approximation because he considered the convolution of the near-zone field with the surface Green function in the form of the Kirchhoff equation. At the same time, his computer algorithm did not allow him to calculate the near zone field, and he eventually reduced his numerical calculations to a convolution of the geometric optics Mueller matrix with the Fraunhofer diffraction pattern for a circle with an area equal to the area of the particle projection. A similar procedure was applied by him with coauthors [18] to the special cases of near-forward and near-backward scattering.

The paper by Yang and Liou [24] was the first wherein the near-zone and wave-zone fields were calculated numerically for large ice crystal particles in a physical-optics approximation. The authors considered theoretically the electromagnetic field at the particle surface by the use of ray trajectories and the respective Jones matrices. Then the field was transformed into surface currents by means of rather cumbersome formulas and, finally, these currents were transformed into the wave-zone field using an exact electrodynamics equation known as the electromagnetic equivalence theorem. They validated their technique against the conventional finite-difference-time-domain method for two-dimensional (2D) and three-dimensional (3D) randomly oriented particles. More recently, these authors reported another method [25], the ray-by-ray integration algorithm, wherein they used a 3D convolution of the near-zone field with the volume Green function to obtain the wave-zone field; the near-zone field inside a crystal was derived, as before, by means of the Jones matrices for ray trajectories. It seems clear that both methods calculating the scattered electromagnetic field or, equivalently, the Jones scattering matrix for the whole particle, require much more computer resources than the conventional ray-tracing method. Moreover, in [24,25] the authors repre-

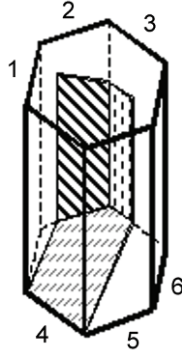


Fig. 1. Facet tracing by a hexagonal column.

sented numerically the near-zone field as a set of ray tubes called wavelets or localized waves. In this way, they had to choose the transverse sizes of the tubes as small as $\lambda/2\pi$. This algorithm was found to be computationally costly, not to mention uncertain shapes of the ray tubes discussed in [25].

In our treatment of the problem of light scattering by large faceted particles, we represent the near-zone field at or near the surface of a faceted particle as a set of plane-parallel beams leaving the particle [2]. Here, unlike [24] and [25], every beam is a clearly defined physical object with a finite transverse size and known shape and spatial location. Specifically, any beam is produced, at the first step, by reflection or refraction of the incident light by a whole illuminated particle facet. In this case, we apply the conventional geometric optics concepts to the beam. In particular, the beam is refracted and propagated inside the particle according to Snell's law and Fresnel's coefficients, its transverse shape being determined by the projection of the whole crystal facet on the plane normal to the propagation direction. At the second step, this beam is split by one or several other crystal facets into more reflected and refracted beams, etc. (Fig. 1).

Our facet-tracing algorithm allows us to obtain directly all the beams exiting the crystal. For the beams thus obtained, the problem of taking into account both the diffraction and interference for the wave-zone field becomes obvious. In fact, we arrive at a well studied problem of the Fraunhofer diffraction for a plane-parallel beam of a polygonal shape that has been solved analytically [8].

2. Series expansion of scattered fields into geometrical trajectories

2.1. Near zone ($R \ll ka^2$)

In the general problem of light scattering by large particles, a distinctive feature of faceted particles as compared to nonspherical particles of arbitrary shape is that in the obvious expansion of the scattered electromagnetic field over multiplicity of reflection/refraction events, every term of the series becomes a plane-

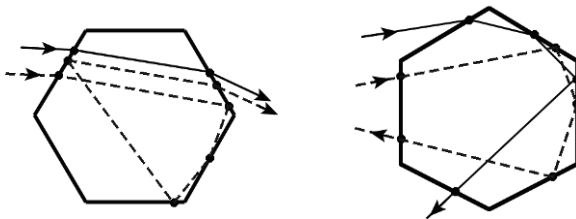


Fig. 2. Examples of ray trajectories.

parallel beam that can be readily described analytically. These analytical expressions are summarized below.

The total electric field $\mathbf{E}(\mathbf{r})$ in the near zone of a crystal becomes a superposition of the incident wave and the outgoing plane-parallel beams:

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}_0(\mathbf{r}) + \sum_j \mathbf{E}_j(\mathbf{r}), \quad (1)$$

where $\mathbf{r} = (x, y, z)$ is the position vector of the observation point, and the incident plane wave with an amplitude \mathbf{E}^0 and wavelength λ is assumed to propagate along, say, the positive z -axis:

$$\mathbf{E}_0(\mathbf{r}) = \mathbf{E}^0 \exp(ikz), \quad (2)$$

with $i = (-1)^{1/2}$ and $k = 2\pi/\lambda$ being the wave number.

Each plane-parallel beam $\mathbf{E}_j(\mathbf{r})$ in Eq. (1) is characterized by the following parameters:

1. the *ray trajectory*, i.e., if all facets are numbered as 1, 2, 3, ... then the subscript j means a sequence of arbitrary length of the numbers like 2132... describing the trajectory. An example of such trajectories is shown in Fig. 2;
2. the *propagation direction* \mathbf{n}_j . This direction is the same for all rays forming the beam;
3. the *beam shape* (including its location on the particle surface). The beam shape can be determined by the shape function $\eta_j(\boldsymbol{\rho})$ defined in any plane perpendicular to \mathbf{n}_j as

$$\eta_j(\boldsymbol{\rho}) = \begin{cases} 1 & \text{inside the beam,} \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

These shapes are always polygons. Examples of shapes of outgoing beams are presented in Fig. 3. It is worthwhile to note that Eq. (3) is valid only for non-absorbing particles. Otherwise the function $\eta_j(\boldsymbol{\rho})$ is variable across the beam instead of being given by Eq. (3);

4. the *complex-valued 2D amplitude* \mathbf{E}_j describing polarization. \mathbf{E}_j is defined in a plane perpendicular to the direction \mathbf{n}_j . Note that the vector \mathbf{E}_j

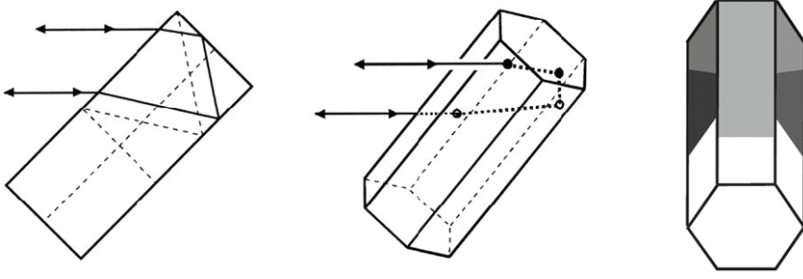


Fig. 3. Ray trajectories and associated plane-parallel beams.

is exactly the same for all the rays constituting the beam because of the beam plane-parallel nature. Examples of ray trajectories and their plane-parallel beams are shown in Fig. 3.

The polarization of each beam \mathbf{E}_j is described analytically as follows. Let an incident ray with the amplitude \mathbf{E}^0 propagate along the direction \mathbf{n}_0 . Choose a 3D right-handed orthogonal coordinate system with axes 1, 2, and 3, in which axis 3 is directed along the vector \mathbf{n}_0 and the corresponding orthogonal components of the amplitude \mathbf{E}^0 are denoted by subscripts 1 and 2. If the ray impinges on a plane with a normal \mathbf{n} then it produces, in the reference plane $(\mathbf{n}_0, \mathbf{n})$ through the vectors \mathbf{n}_0 and \mathbf{n} , two rays called reflected and refracted rays. The propagation directions of these new rays are defined by the well-known Snell's law. Introduce similar 3D right-handed coordinate systems for each of these rays, wherein axes 3 correspond to their propagation directions, a common axis, say axis 2, is perpendicular to the reference plane, and the parallel axes 1 are found according to the right-hand rule for each coordinate system. In these reference-plane coordinate systems, a reflection/refraction event is described by a diagonal matrix

$$\mathbf{F} = \begin{bmatrix} F_1(\beta) & 0 \\ 0 & F_2(\beta) \end{bmatrix}, \quad (4)$$

where F_1 and F_2 are the Fresnel reflection/refraction coefficients for the parallel and perpendicular components, respectively, depending on the incidence angle β and the refractive index. Finally, the amplitude of a ray after a reflection/refraction event is described by the product $\mathbf{E}' = \mathbf{F}\mathbf{L}\mathbf{E}^0$, where

$$\mathbf{L} = \begin{bmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{bmatrix} \quad (5)$$

is the rotation matrix transforming the vector \mathbf{E}^0 from an initial coordinate system to the appropriate reference-plane coordinate system; α is the angle between, say, planes (1,3) for the incident and refracted rays.

If an exiting ray undergoes p reflection/refraction events, its polarization is described by a sequential product

$$\mathbf{P}_j = \mathbf{L}_{p+1} \mathbf{F}_p \mathbf{L}_p \dots \mathbf{F}_2 \mathbf{L}_2 \mathbf{F}_1 \mathbf{L}_1 \quad (6)$$

called the polarization matrix. Here, the last rotation matrix \mathbf{L}_{p+1} transforms a 2D vector from the last reference-plane coordinate system to the scattering-plane coordinate system. By analogy with the former, the latter corresponds to the scattering plane $(\mathbf{n}_0, \mathbf{n}_j)$ containing the initial \mathbf{n}_0 and outgoing \mathbf{n}_j directions. We also denote the parallel and perpendicular components relative to the scattering plane by subscripts 1 and 2, respectively.

In addition to the matrices transforming the polarization of the electromagnetic wave along a trajectory, we need to take into account the optical path along the trajectory. We determine a phase shift $\xi_j = kl_j$ relative to a reference trajectory (say, a trajectory which connects a point light source in the direction $-\mathbf{n}_0$ with the center of the crystal and then propagates in the direction \mathbf{n}_j), and l_j is the optical path for this reference trajectory, i.e., l_j is the geometrical length of the internal part of the trajectory multiplied by the refractive index. Finally, the outgoing beam is described by

$$\mathbf{E}_j(\mathbf{r}) = \exp(i\xi_j) \eta_j(\boldsymbol{\rho}) \mathbf{P}_j \mathbf{E}^0. \quad (7)$$

Equation (7) depends explicitly on the amplitude of the incident wave \mathbf{E}^0 . It is more convenient to omit the vector \mathbf{E}^0 and to characterize the beam in terms of the 2×2 amplitude (Jones) matrix \mathbf{j}_j defined by the following equations:

$$\mathbf{E}_j = \eta_j \mathbf{j}_j \mathbf{E}^0, \quad (8)$$

$$\mathbf{j}_j = \exp(i\xi_j) \mathbf{P}_j. \quad (9)$$

Equation (8) describes both the rays and plane-parallel beams that are formed by reflection/refraction events on the particle surface. In addition, there is one more plane-parallel beam that must be included in the superposition of Eq. (1). Specifically, the total field $\mathbf{E}(\mathbf{r})$ should include the so-called shadow-forming plane-parallel beam [1,2,17] given by

$$\mathbf{E}_{sh}(\mathbf{r}) = \begin{cases} -\mathbf{E}_0(\mathbf{r}) & \text{inside the shadow,} \\ 0 & \text{otherwise.} \end{cases} \quad (10)$$

The direction of propagation of the shadow-forming beam is given by the propagation direction \mathbf{n}_0 of the incident plane wave, and its shape function corresponds (with a multiplicative factor of -1) to the crystal projection on the $z = \text{constant}$ plane.

Note that refraction events can also create plane-parallel beams $\mathbf{E}_j(\mathbf{r})$ that propagate in the direction \mathbf{n}_0 . These beams appear when two opposite facets correspond to a part of a plane-parallel plate relative to the incident wave. However, the shadow-forming beam of Eq. (10) appears independently of such beams. It is caused only by the particle contour, i.e., by the break of the amplitude of the incident plane wave along the particle contour. As an additional illustration of the existence of the shadow forming field, let us mention a well-known fact that the ex-

tion cross section for a large particle is equal to twice its geometric cross-sectional area. Indeed, according to the energy conservation law, the normalized energy flux of all the reflected/refracted beams, i.e., the geometric-optics scattering cross section, is obviously equal to the area of the particle projection on the plane $z = \text{constant}$, provided that there is no absorption inside the particle. It is just the shadow-forming beam of Eq. (10) that adds the same quantity to the total extinction cross section. As a consequence, the extinction cross section turns out to be equal to twice the area of the particle projection.

Polarization can be equivalently described in the language of the Stokes vectors and Mueller matrices. Indeed, any 2×2 matrix \mathbf{A} transforming a transverse electromagnetic wave \mathbf{E} is converted into an equivalent 4×4 matrix $\tilde{\mathbf{A}}$ for the respective Stokes vectors via the equation

$$\tilde{\mathbf{A}} = \Gamma(\mathbf{A} \otimes \mathbf{A}^*)\Gamma^{-1}, \quad (11)$$

where \otimes denotes the Kronecker matrix product, the asterisk denotes a complex-conjugate value, and the matrix Γ is given by

$$\Gamma = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & -1 & -1 & 0 \\ 0 & -i & i & 0 \end{bmatrix}, \quad \Gamma^{-1} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & -1 & i \\ 0 & 0 & -1 & -i \\ 1 & -1 & 0 & 0 \end{bmatrix}. \quad (12)$$

Instead of Eqs. (6) and (9), the change of light polarization along a ray trajectory can be described by the following successive product:

$$\mathbf{m}_j = \tilde{\mathbf{L}}_{p+1} \tilde{\mathbf{F}}_p \tilde{\mathbf{L}}_p \dots \tilde{\mathbf{F}}_1 \tilde{\mathbf{L}}_1, \quad (13)$$

where $\tilde{\mathbf{F}}$ and $\tilde{\mathbf{L}}$ are the appropriate matrices of Eq. (6) transformed by means of Eq. (11):

$$\tilde{\mathbf{F}} = \frac{1}{2} \begin{bmatrix} |F_1|^2 + |F_2|^2 & |F_1|^2 - |F_2|^2 & 0 & 0 \\ |F_1|^2 - |F_2|^2 & |F_1|^2 + |F_2|^2 & 0 & 0 \\ 0 & 0 & F_1 F_2^* + F_1^* F_2 & F_1 F_2^* - F_1^* F_2 \\ 0 & 0 & -F_1 F_2^* + F_1^* F_2 & F_1 F_2^* + F_1^* F_2 \end{bmatrix}, \quad (14)$$

$$\tilde{\mathbf{L}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\alpha & \sin 2\alpha & 0 \\ 0 & -\sin 2\alpha & \cos 2\alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (15)$$

Denoting the Stokes vectors $\mathbf{I} = [I \ Q \ U \ V]^T$ of the incident and scattered fields as $\mathbf{I}^0(\mathbf{r})$ and $\mathbf{I}(\mathbf{r})$, respectively, with T standing for ‘‘transposed’’, we obtain an equation similar to Eq. (8):

$$\mathbf{I}_j(\mathbf{r}) = \eta_j(\boldsymbol{\rho}) \mathbf{m}_j \mathbf{I}^0. \quad (16)$$

It worthwhile to note that there is a difference between the conventional ray-tracing procedures and our Eq. (16). Indeed, the former usually include the energy conservation law in each refraction event [5,7,9,11–14,17–20,22,23]. This can be done readily by an additional multiplicative factor $\cos \beta' / \cos \beta$ for each refraction event, where β and β' are the incident and refraction angles, respectively. In fact, this factor takes into account the change of cross sections for a ray tube. Such an approach allows one to reduce the computer time significantly. On the other hand, one loses the information on beam shapes, and the problem of a rigorous consideration of diffraction, for example, proves to be overwhelming. On the contrary, our Eq. (16) is based on rigorous wave equations and does not necessitate any additional factors.

To conclude this subsection, let us briefly discuss the domain of applicability of the primordial plane-beam representation of the near-zone field. Are there additional physical-optics phenomena that could cause significant corrections to the plane-beam representation? Two phenomena could be of importance. First of all, one could consider the diffraction of the plane-parallel beams when they propagate from one facet to another inside the crystal. However, such diffraction phenomena are usually negligible for simple particle shapes, wherein longitudinal and transverse typical sizes do not differ much. Another typical physical-optics feature is an evanescent wave created by a sharp corner and or rib. It is unlikely that evanescent waves can cause any significant contribution on top of the plane-parallel-beam background. Moreover, in practice, there is a substantial uncertainty in the knowledge of sizes and shapes of actual faceted particles that smoothes out any weak effects. Therefore, the plane-parallel representation of the near-zone field appears to be a reasonably simple, reliable, and adequate approximation for the case of large faceted particles ($a \gg \lambda$).

2.2. Far zone ($R \gg a$)

In the preceding subsection, we have determined the scattered wave in the near zone $R \ll ka^2$. As the scattered wave propagates away from the particle, it undergoes specific transformations with distance R . In this subsection, we consider the wave in the far zone determined by the condition $R \gg a$.

In general, any scattered field at a large distance from the particle evolves into a diverging spherical wave, its Stokes parameters decreasing as $1/R^2$, i.e.,

$$\mathbf{I}(R, \mathbf{n}) = \mathbf{D}(\mathbf{n})/R^2, \quad (17)$$

where $R \rightarrow \infty$. Here, the Cartesian coordinates $\mathbf{r} = (x, y, z)$ are replaced by spherical coordinates (R, \mathbf{n}) originating inside the particle, R is the distance to the far-zone observation point, and \mathbf{n} is the unit vector in the direction of the observation point. In the far zone, the scattered wave at any spatial point propagates in just one scattering direction \mathbf{n} . In Eq. (17), the vector-parameter $\mathbf{D}(\mathbf{n})$ describes the distribution of the Stokes parameters over the scattering directions \mathbf{n} . Any solution of the wave scattering problem involves the calculation of the function

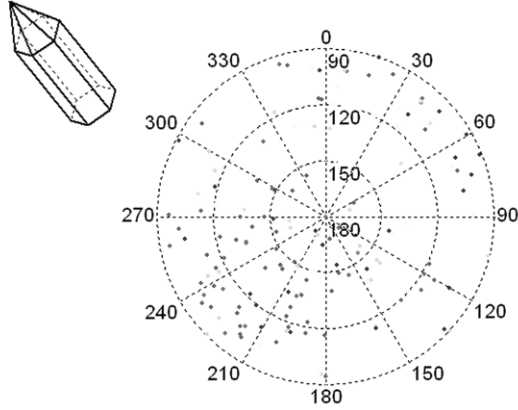


Fig. 4. Example of the geometric-optics Mueller matrix for a fixed particle orientation.

$\mathbf{D}(\mathbf{n})$. It is convenient to relate the Stokes parameters of the incident, \mathbf{I}_0 , and scattered, $\mathbf{D}(\mathbf{n})$, waves by a 4×4 matrix \mathbf{M} called the scattering or Mueller matrix:

$$\mathbf{D}(\mathbf{n}) = \mathbf{M}(\mathbf{n})\mathbf{I}^0. \quad (18)$$

In our case of plane-parallel beams, the propagation direction of a beam in the near zone \mathbf{n}_j proves to be the scattering direction in the far zone. Replacing the Cartesian coordinates \mathbf{r} of the near zone by the spherical coordinates (R, \mathbf{n}) of the far zone, Eq. (16) is transformed into the following equation:

$$\mathbf{I}_j(R, \mathbf{n}) = \delta(\mathbf{n} - \mathbf{n}_j) s_j \mathbf{m}_j \mathbf{I}^0 / R^2, \quad (19)$$

where δ is the angular delta function and s_j is the area of the transverse section of the beam:

$$s_j = \int \eta_j(\boldsymbol{\rho}) d\boldsymbol{\rho}. \quad (20)$$

As a consequence, the Mueller matrix for a faceted particle in the far zone is a superposition of delta-function contributions:

$$\mathbf{M}(\mathbf{n}) = \sum_j s_j \mathbf{m}_j \delta(\mathbf{n} - \mathbf{n}_j). \quad (21)$$

Equation (21) solves the problem of the analytical description of far-zone scattering in terms of the parameters of the near-zone scattered field. The matrix of Eq. (21) is called the geometric-optics Mueller matrix and is exemplified in Fig. 4.

2.3. Wave zone ($R \gg ka^2$)

In reality, any plane-parallel beam formed in the near zone is transformed during its propagation away from the scattering particle. As is well known, at a dis-

tance $R \approx ka^2$ it undergoes the Fresnel diffraction, wherein the beam edges become smoothed. Then, at a distance $R \gg ka^2$, the Fraunhofer diffraction takes place which is characterized by a steady angular pattern of the width of about λ/a . In the preceding subsection, we defined the far zone by the condition $R \gg a$, wherein the diffraction phenomena were neglected. Now we define the concept of the far zone more precisely by the following inequality: $ka^2 \gg R \gg a$.

In the wave zone $R \gg ka^2$, the scattered field is a diverging spherical wave given by second term on the right-hand side of the formula for the total field:

$$\mathbf{E}(R, \mathbf{n}) = \mathbf{E}^0 \exp(ikz) + \frac{\exp(ikR)}{R} \mathbf{E}^s(\mathbf{n}), \quad (22)$$

where the center of the particle is assumed to coincide with the origin $\mathbf{r} = \mathbf{0}$ of the Cartesian coordinate system $\mathbf{r} = (x, y, z)$, and \mathbf{n} is the scattering direction. In Eq. (22), the scattering or Jones matrix \mathbf{J} of the whole particle transforms the amplitude \mathbf{E}^0 of the incident field into the amplitude \mathbf{E}^s of the spherical scattered wave:

$$\mathbf{E}^s(\mathbf{n}) = \mathbf{J}(\mathbf{n}) \mathbf{E}^0. \quad (23)$$

The 2D amplitude \mathbf{E}^s is conventionally found in the 3D right-handed scattering-plane coordinate systems defined previously. Recall that the scattering plane $(\mathbf{n}_0, \mathbf{n})$ contains both the incidence, \mathbf{n}_0 , and the scattering, \mathbf{n} , direction, axis 3 corresponds to the scattering direction, and components 1 and 2 describe the parallel and perpendicular components of the vector \mathbf{E}^s relative to the scattering plane, respectively.

In the wave zone, unlike in the geometric-optics approach presented in the previous subsection, the propagation direction of a near-zone beam \mathbf{n}_j becomes the center of a diffraction pattern of the scattered field \mathbf{E}_j . The diffracted field is determined by the vector Fraunhofer diffraction equation (see Eq. (9.96) of [10]):

$$\mathbf{E}_j(\mathbf{n}) = \frac{ik e^{ikR}}{2\pi R} [\mathbf{n} \times (\mathbf{n}_j \times \mathbf{j}_j \mathbf{E}^0)] \int \eta_j(\boldsymbol{\rho}) e^{ik(\mathbf{n}_j - \mathbf{n}) \cdot \boldsymbol{\rho}} d\boldsymbol{\rho}, \quad (24)$$

where \mathbf{n} is the scattering direction. The scalar factor in Eq. (24) is the well-known scalar Fraunhofer diffraction pattern

$$f_j(\mathbf{n}) = \frac{ik}{2\pi} \iint \eta_j(\boldsymbol{\rho}) \exp[ik(\mathbf{n}_j - \mathbf{n}) \cdot \boldsymbol{\rho}] d\boldsymbol{\rho}. \quad (25)$$

Let us define a right-handed triplet of unit vectors $\mathbf{t} \times \mathbf{f} = \mathbf{n}$ on a unit sphere of scattering directions \mathbf{n} . These vectors are as follows:

$$\mathbf{n} = \begin{bmatrix} \sin \theta \cos \varphi \\ \sin \theta \sin \varphi \\ \cos \theta \end{bmatrix}, \quad \mathbf{t} = \begin{bmatrix} \cos \theta \cos \varphi \\ \cos \theta \sin \varphi \\ -\sin \theta \end{bmatrix}, \quad \mathbf{f} = \begin{bmatrix} -\sin \varphi \\ \cos \varphi \\ 0 \end{bmatrix}, \quad (26)$$

where θ and φ are the zenith and azimuth angles in a spherical coordinate system,

respectively, while the vectors \mathbf{t} and \mathbf{f} are directed along meridians and parallels on the sphere, respectively. Assume that the geometric-optics scattering direction \mathbf{n}_j is described by its zenith and azimuth angles Θ and Φ corresponding to the triplet of similar unit vectors:

$$\mathbf{N} = \begin{bmatrix} \sin \Theta \cos \Phi \\ \sin \Theta \sin \Phi \\ \cos \Theta \end{bmatrix}, \quad \mathbf{T} = \begin{bmatrix} \cos \Theta \cos \Phi \\ \cos \Theta \sin \Phi \\ -\sin \Theta \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} -\sin \Phi \\ \cos \Phi \\ 0 \end{bmatrix}. \quad (27)$$

Then the expression in the square brackets in Eq. (24) can be rewritten as follows:

$$[\mathbf{n} \times (\mathbf{n}_j \times \mathbf{j}_j \mathbf{E}^0)] = \begin{bmatrix} -\mathbf{F} \cdot \mathbf{f} & \mathbf{T} \cdot \mathbf{f} \\ \mathbf{F} \cdot \mathbf{t} & -\mathbf{T} \cdot \mathbf{t} \end{bmatrix} \mathbf{j}_j \mathbf{E}^0 \equiv \mathbf{Q}_j \mathbf{j}_j \mathbf{E}^0. \quad (28)$$

Finally, we derive the following equation for the Jones matrix:

$$\mathbf{J}(\mathbf{n}) = \sum_j \mathbf{J}_j(\mathbf{n}) = \sum_j \mathbf{Q}_j \mathbf{j}_j f_j(\mathbf{n}). \quad (29)$$

Equation (29) solves the problem of an analytical expression of the scattered field in the wave zone of a faceted particle in terms of the parameters of near-zone plane-parallel beams.

While the Jones matrices are more physically transparent and convenient for theoretical derivations, experimentally measured quantities are usually described by Stokes vectors and, consequently, by Mueller matrices. The Mueller matrix for a whole particle in the wave zone is readily obtained from of Eqs. (11) and (12):

$$\mathbf{M}(\mathbf{n}) = \Gamma [\mathbf{J}(\mathbf{n}) \otimes \mathbf{J}^*(\mathbf{n})] \Gamma^{-1}. \quad (30)$$

For a given Jones matrix, the matrix of Eq. (30) is referred to as a pure Mueller matrix. We split the pure Mueller matrix into two components:

$$\mathbf{M}(\mathbf{n}) = \mathbf{M}_{dif}(\mathbf{n}) + \mathbf{M}_{inter}(\mathbf{n}), \quad (31)$$

where

$$\mathbf{M}_{dif}(\mathbf{n}) = \sum_j \Gamma [\mathbf{J}_j(\mathbf{n}) \otimes \mathbf{J}_j^*(\mathbf{n})] \Gamma^{-1} = \sum_j \tilde{\mathbf{Q}}_j \mathbf{m}_j |f_j(\mathbf{n})|^2, \quad (32)$$

$$\mathbf{M}_{inter}(\mathbf{n}) = \sum_{j \neq l} \Gamma [f_j(\mathbf{n}) (\mathbf{Q}_j \mathbf{j}_j) \otimes (\mathbf{Q}_l \mathbf{j}_l^*) f_l^*(\mathbf{n})] \Gamma^{-1}. \quad (33)$$

It is obvious that the first component takes into account only the diffraction of the initial near-zone plane-parallel beams and will be called the diffraction Mueller matrix. In Eq. (32), \mathbf{m}_j turns out to be the Mueller matrix of a near-zone beam determined by Eq. (13). Consequently, the second component describes contributions created by the interference among the diffracted beams; it is called for brevity the interference Mueller matrix.

As is well known, diffraction is significant within a cone of angles about λ/a wide. If a detector does not resolve such angles, the diffraction can be ignored. In this case we can replace

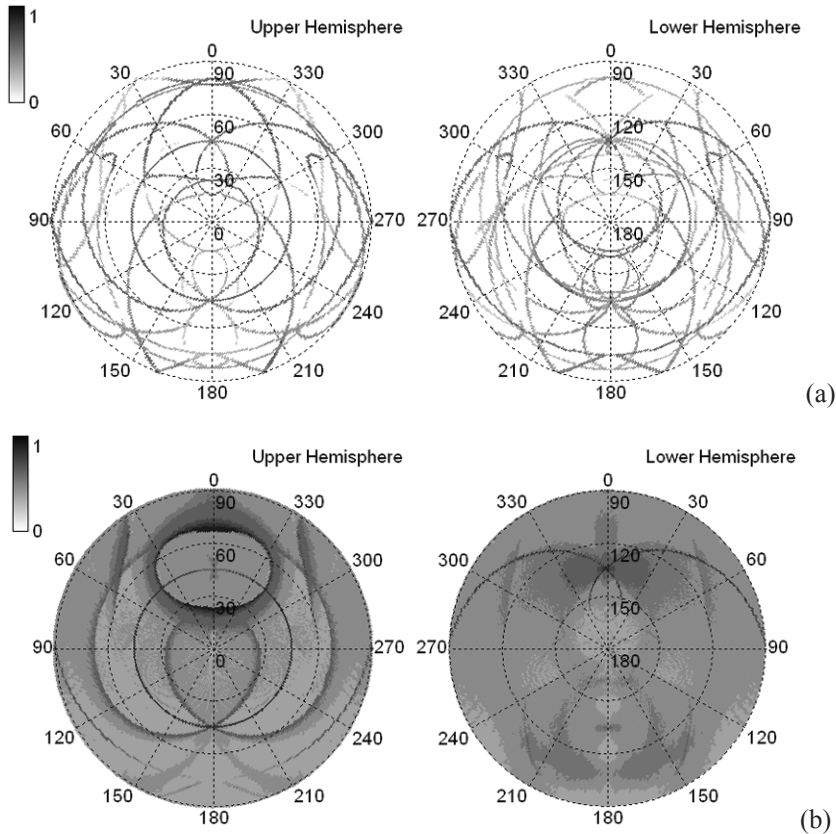


Fig. 5. Examples of geometric-optics Mueller matrices for hexagonal ice columns averaged over random horizontal orientations: (a) Parry or 1D orientation; (b) 2D orientation.

$$|f_j(\mathbf{n})|^2 \approx s_j \delta(\mathbf{n} - \mathbf{n}_j), \quad (34)$$

so the diffraction Mueller matrix of Eq. (32) becomes equal to the geometric-optics Mueller matrix determined by Eq. (21), while the interference matrix of Eq. (33) vanishes.

2.4. Averaging over particle orientations

In the preceding subsections, we considered light scattering by a particle having a fixed orientation. For the geometric-optics Mueller matrix, a change of the particle orientation results in a shift of the dots such as those shown in Fig. 4. During the process of averaging over particle orientations, these delta functions can be either smoothed into a regular function, or certain singularities can survive in discrete directions and/or along curves on the sphere of scattering directions. For ex-

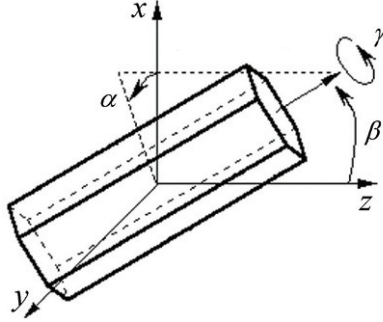


Fig. 6. Particle orientation.

ample, Fig. 5 shows the geometric-optics Mueller matrix for a hexagonal ice column with its main axis (i.e., the straight line connecting both centers of the hexagons) directed horizontally and the entire column rotated about the vertical direction. In Fig. 5a, the column is rotated only about the vertical direction (so called Parry orientations). As a result, every beam draws a curve on the scattering-direction sphere, and the Mueller matrix is singular along these curves. In Fig. 5b, the column is rotated additionally about its main axis. We see that some curves of Fig. 5a are smoothed into regular functions of the scattering direction, but other singularities survive.

Let us now consider the averaging of the Mueller matrix over an arbitrary distribution of particle orientations in space. In general, the spatial orientation of a particle is described by three independent variables, for example the Euler angles α (azimuth angle), β (zenith angle), and γ (rotation angle), as shown in Fig. 6. Note that although Fig. 6 depicts a hexagonal column, all formulas are valid for any particle shape.

If the incident wave propagates along the positive z -axis, the scattering direction \mathbf{n} is conveniently determined by the zenith, θ , and azimuth, φ , angles relative to the $+z$ and $+x$ axes, respectively. Formally, the Mueller matrix \mathbf{M} of Eq. (30) becomes a function of five variables: $\mathbf{M}(\theta, \varphi | \alpha, \beta, \gamma)$. However, the rotation of a particle around the z -axis described by the orientation azimuth angle α does not change the scattering process directly but leads to rather simple analytical transformations. Therefore, we can exclude the variable α by means of analytical equations, and the scattering process can be described numerically as a function of four variables.

To separate the analytical dependence on α , it is convenient to introduce the Mueller matrix $\boldsymbol{\mu}$ corresponding to the case of $\alpha = 0$:

$$\boldsymbol{\mu}(\theta, \varphi | \beta, \gamma) \equiv \mathbf{M}(\theta, \varphi | \alpha = 0, \beta, \gamma). \quad (35)$$

For an arbitrary azimuth angle α , the Mueller matrix is expressed through the reduced Mueller matrix via a shift by $\varphi - \alpha$, like a scalar function:

$$\mathbf{M}(\theta, \varphi | \alpha, \beta, \gamma) = \boldsymbol{\mu}(\theta, \varphi - \alpha | \beta, \gamma). \quad (36)$$

Thus, it is the reduced Mueller matrix of four variables $\boldsymbol{\mu}(\theta, \varphi | \beta, \gamma)$ that contains all essential information about the scattering action.

The probability distribution $P(\alpha, \beta, \gamma)$ over particle orientations should be conventionally normalized to unity:

$$\int \frac{d\alpha}{2\pi} \frac{\sin \beta d\beta}{2} \frac{d\gamma}{2\pi} P(\alpha, \beta, \gamma) = 1, \quad (37)$$

where $P(\alpha, \beta, \gamma) = 1$ for randomly oriented particles. The averaging of Eqs. (35) and (36) results in the following equations:

$$\langle \mathbf{M}(\theta, \varphi) \rangle = \int_0^{2\pi} \frac{d\alpha}{2\pi} p(\alpha) \langle \boldsymbol{\mu}_\alpha(\theta, \varphi - \alpha) \rangle, \quad (38)$$

$$\langle \boldsymbol{\mu}_\alpha(\theta, \varphi) \rangle = \int_0^\pi \frac{\sin \beta d\beta}{2} \int_0^{2\pi} \frac{d\gamma}{2\pi} p_\alpha(\beta, \gamma) \boldsymbol{\mu}(\theta, \varphi | \beta, \gamma), \quad (39)$$

where the conventional decomposition of the probability distribution $P(\alpha, \beta, \gamma)$ into the probability distribution over the azimuth angle $p(\alpha)$ and the conditional distribution $p_\alpha(\beta, \gamma)$ over the angles β and γ at a given α is used:

$$p(\alpha) = \int_0^\pi \frac{\sin \beta d\beta}{2} \int_0^{2\pi} \frac{d\gamma}{2\pi} P(\alpha, \beta, \gamma),$$

$$p_\alpha(\beta, \gamma) = P(\alpha, \beta, \gamma) / p(\alpha). \quad (40)$$

In the particular case of random 3D orientations, $P(\alpha, \beta, \gamma) = p(\alpha) = p_\alpha(\beta, \gamma) = 1$, and the average Mueller matrix becomes a function of only the zenith scattering angle θ :

$$\langle \mathbf{M}(\theta) \rangle = \int_0^\pi \frac{\sin \beta d\beta}{2} \int_0^{2\pi} \frac{d\gamma}{2\pi} \int_0^{2\pi} \frac{d\varphi'}{2\pi} \boldsymbol{\mu}(\theta, \varphi' | \beta, \gamma). \quad (41)$$

It is the function $\langle \mathbf{M}(\theta) \rangle$ that has been presented in numerous papers.

The statistical average of the geometric-optics Mueller matrix of Eq. (21) for any scattered beam can be found analytically, as we demonstrate below. Denote the Mueller matrix of a beam as follows:

$$\boldsymbol{\mu}_j(\theta, \varphi | \beta, \gamma) = s_j \mathbf{m}_j \delta(\mathbf{n} - \mathbf{n}_j) \equiv \boldsymbol{\mu}_j(\mathbf{n} | \mathbf{B}). \quad (42)$$

Formally, the set of orientation parameters $\mathbf{B} = (\beta, \gamma)$ corresponds to the surface of a unit sphere such as the scattering-direction sphere $\mathbf{n} = (\theta, \varphi)$. A change of the variables \mathbf{B} leads to a movement of the point \mathbf{n}_j on the sphere \mathbf{n} accompanied by a change of the values s_j and \mathbf{m}_j . Thus, during a change in the particle orientation, a

beam maps the surface \mathbf{B} onto the surface \mathbf{n} . The probability distribution over the surface \mathbf{B} corresponds to the conditional probability $P(\mathbf{B}) = p_\alpha(\beta, \gamma)$ of Eq. (40) with the following normalization:

$$\int P(\mathbf{B}) d\mathbf{B} = \int_0^\pi \frac{\sin \beta d\beta}{2} \int_0^{2\pi} \frac{d\gamma}{2\pi} p_\alpha(\beta, \gamma) = 1. \quad (43)$$

Omitting the subscript α for brevity, we obtain:

$$\begin{aligned} \langle \boldsymbol{\mu}_j(\mathbf{n}) \rangle &= \int \boldsymbol{\mu}_j(\mathbf{n} | \mathbf{B}) P(\mathbf{B}) d\mathbf{B} = \int s_j(\mathbf{n}_j) \mathbf{m}_j(\mathbf{n}_j) \delta(\mathbf{n} - \mathbf{n}_j) P(\mathbf{B}) d\mathbf{B} \\ &= \int s_j(\mathbf{n}_j) \mathbf{m}_j(\mathbf{n}_j) P[\mathbf{B}(\mathbf{n}_j)] \delta(\mathbf{n} - \mathbf{n}_j) \frac{d\mathbf{n}_j}{d\mathbf{n}} d\mathbf{B} = \frac{s_j(\mathbf{n}) \mathbf{m}_j(\mathbf{n}) P[\mathbf{B}(\mathbf{n})]}{d\mathbf{n}/d\mathbf{B}}. \end{aligned} \quad (44)$$

Thus, the average reduced Mueller matrix for a beam at an arbitrary point \mathbf{n} turns out to be a product of three factors. The first one is the matrix $s_j \mathbf{m}_j$ for the particle orientation corresponding to the scattered beam appearing at the point \mathbf{n} . The second factor is the probability density $P(\mathbf{B})$ for this orientation. And the third one is the scaling factor $d\mathbf{B}/d\mathbf{n}$ of the mapping in the vicinity of this orientation; it was moved into the denominator for convenience only.

While $s_j \mathbf{m}_j$ and $P(\mathbf{B})$ are usually smooth functions of the scattering direction \mathbf{n} , the factor $d\mathbf{n}/d\mathbf{B}$ can sometimes vanish, thereby resulting in singularities of a Mueller matrix along certain curves on the sphere \mathbf{n} . Well-known examples of such singularities are the halos observed in atmospheric optics.

Let us proceed to the physical-optics Mueller matrix of Eqs. (30)–(33). Here, upon the statistical averaging of Eq. (44), the delta function should be replaced by the Fraunhofer diffraction function $|f_j(\mathbf{n} - \mathbf{n}_j)|^2$. This implies that for a fixed j th beam, the diffraction Mueller matrix is a convolution of the geometric-optics Mueller matrix with a certain diffraction matrix:

$$\langle \mathbf{M}_{dif}(\mathbf{n}) \rangle = \int \mathbf{F}(\mathbf{n}, \mathbf{n}') \langle \mathbf{M}_G(\mathbf{n}') \rangle d\mathbf{n}'. \quad (45)$$

This equation is valid for scattering directions in which only one or a few beams provide the dominant contribution. As for the interference matrix of Eq. (33), it depends on pairs of beams, and so Eq. (45) becomes more complicated.

3. Randomly oriented particles

3.1. Statistical approach to light scattering by convex ice crystals

Direct calculations of scattering matrices for ensembles of faceted particles with various shapes, sizes, and orientations meet two difficulties. First, if the particle shape differs from the simplest one like, for example, a hexagonal column or plate, the calculation of the scattering matrix becomes costly and involved. Second, our knowledge of real shape distributions is, as a rule, poor. Therefore, in-

stead of direct calculations, a statistical approach to the problem of light scattering may be more promising. Such an approach is described in this section.

In the geometric optics approximation, if absorption is neglected, the scattering cross section is equal to the area s of the particle projection onto the plane normal to the direction of the incident light. After averaging over random particle orientations, s is replaced by the average area $\langle s \rangle$, and the distribution of the scattered energy over scattering directions is described by the phase function $p(\theta)$, where θ is the angle between the incidence and scattering directions (i.e., the scattering angle). We assume the following normalization:

$$2\pi \int_0^\pi p(\theta) \sin \theta d\theta = 1.$$

The basis for our statistical approach is a series expansion of $p(\theta)$ over the multiplicity of internal reflections inside the particle:

$$p(\theta) = \sum_{m=0}^{\infty} a_m p_m(\theta),$$

where

$$\sum_{m=0}^{\infty} a_m = 1 \quad \text{and} \quad 2\pi \int_0^\pi p_m(\theta) \sin \theta d\theta = 1. \quad (46)$$

By use of the ray-tracing or facet-tracing codes, the series (46) is easily evaluated numerically. The index m in Eq. (46) is interpreted as the number of internal collisions of a photon. Thus, a_m means the probability for every photon hitting a crystal to exit it after m internal collisions. The case $m=0$ corresponds to the immediate reflection from any facet without penetration inside the particle. Then, for example, the case $m=2$ means a refraction through the first facet followed by a reflection from the second facet followed by a refraction exit through the third facet. The partial phase function p_m describes the probability density for a photon to exit the particle at the scattering angle θ after m internal collisions.

Since the physical reason for creating photons with a given multiplicity m is the same for any crystal shape, the partial phase functions $p_m(\theta)$ should be similar for differently shaped randomly oriented crystals and should depend weakly on the crystal shape. Indeed, for the $m=0$ photons, it can be easily proven that the reflection from any convex randomly oriented particle does not depend on the particle shape. These photons correspond exactly to those reflected from a sphere, and the function $p_0(\theta)$ is determined by the Fresnel reflection coefficients for a plane surface. The respective probability a_0 is equal to 6.29%, and the value $a_0 p_0(\theta)$ is shown in Fig. 7. Here, the refractive index is assumed to be 1.31.

As is seen in Fig. 7, for crystals with the refractive index 1.31, $m=1$ photons dominate since they constitute up to 70% of the scattered light. These photons correspond to traversing a wedge determined by the dihedral angle α_i between the first and second facets encountered by a photon. For a given crystal shape, the

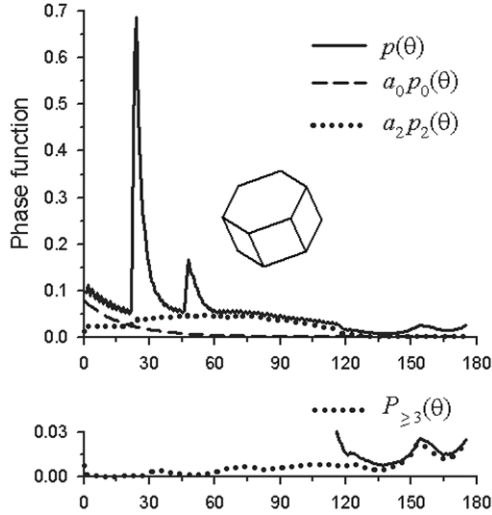


Fig. 7. The top panel depicts the phase function for a randomly oriented hexagonal ice plate with a height-to-diameter ratio of 1/2 (solid curve) and its components with zero (dashed curve) and two (dotted curve) internal collisions of photons. The lower panel uses finer vertical resolution scale to show the cumulative contribution from photons with three and more collisions.

set of dihedral angles α_i is known. It is convenient to expand the partial phase function $p_1(\theta)$ according to the set of dihedral angles α_i :

$$p_1(\theta) = \sum_i b_i p_{1i}(\theta),$$

where

$$\sum_i b_i = 1 \quad \text{and} \quad 2\pi \int_0^\pi p_{1i}(\theta) \sin \theta d\theta = 1. \quad (47)$$

Here, like in Eq. (46), b_i means the probability for a photon to traverse the i th dihedral angle. As is well known, the photons transmitted through a randomly oriented unbounded wedge are distributed within a narrow angular peak called the halo. For an ice crystal, the wedges have finite sizes, thereby causing a dependence of the halos $p_{1i}(\theta)$ on the particle shape. However, this dependence is rather weak, as illustrated in Fig. 8 for the traditional case of hexagonal columns and plate, where

$$Q = H/D$$

is the aspect ratio, H being the distance between the hexagonal bases and D being their diameter. Thus, we conclude that it is the coefficients a_i and b_i rather than the partial phase functions that depend significantly on the crystal shape. We note that

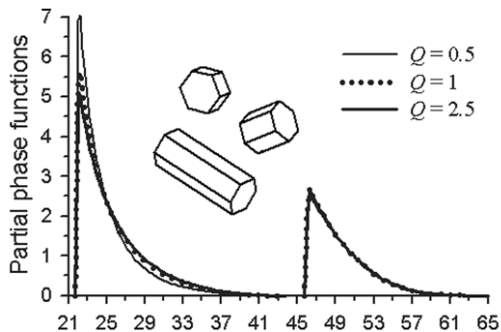


Fig. 8. Partial phase functions of the 22° and 46° halos ($p_{11}(\theta)$ and $p_{12}(\theta)$, respectively) for hexagonal ice columns and plates.

the case of $\alpha_0 = 0^\circ$ corresponding to parallel plane facets is also included in Eq. (47). In this case, $p_{10}(\theta)$ is not a halo but rather a delta function in the forward direction.

It is obvious that the coefficients b_i are related to the weights w_i of the dihedral angles determined, for a given crystal shape, by the formula

$$w_i = \frac{\sum_{k,l \neq i}^N S_k S_l}{\sum_{k,l=1}^N S_k S_l}, \quad (48)$$







where N is the number of crystal facets, S_k is the area of the k th facet, and the first sum gathers all the terms corresponding to the same dihedral angle. For ensembles of crystals with various shapes, the statistical weights $\langle w_i \rangle$ can also be calculated. The set of weights $\langle w_i \rangle$ is an effective means of characterizing such ensembles in lieu of shape descriptions. The details of the relationship between the coefficients b_i and weights $\langle w_i \rangle$ have not been calculated yet.

The photons with higher orders of collisions can be expected to “forget” the shape of the crystal. Indeed, the partial phase function $p_2(\theta)$ reveals a relatively smooth dependence on the scattering angle, as shown in Fig. 7. As to the cumulative contribution

$$P_{\geq 3}(\theta) = \sum_{m=3}^{\infty} b_m p_m(\theta),$$

it is small and approximately constant, as illustrated in the lower panel of Fig. 7. We note that Fig. 7 is a traditional histogram calculated with a bin width of $\Delta\theta = 1^\circ$. Here, the forward peak described by the delta function and coefficient b_0 is omitted. The use of the linear rather than the conventional logarithmic scale for the vertical axis is intended to emphasize negligible differences between the phase function and $P_{\geq 3}(\theta)$ as well as the relatively small magnitude of the function $P_{\geq 3}(\theta)$. We also mention that the function $P_{\geq 3}(\theta)$ has a singularity in the backward

Table 1. Cumulative probabilities (in %) for photons to exit ice crystals of different shapes after m internal collisions

Q	0.1	0.5	1	2.5	5	1.5
						
A_0	6.29	6.29	6.29	6.29	6.29	6.29
A_1	75.47	55.10	56.67	66.70	73.36	75.46
A_2	87.03	89.33	87.27	82.91	84.95	83.36
A_3	91.09	94.16	93.47	89.38	87.92	88.89
A_4	93.78	96.06	95.31	94.54	92.50	92.67
A_5	95.63	96.92	97.21	97.08	94.71	94.34
A_6	96.85	97.75	98.27	98.14	96.96	96.18

direction $\theta = 180^\circ$ revealed in the histogram as a sharp peak with an angular width $\approx 0.5^\circ$. However, these photons constitute less than 1% of all the scattered photons, and their role is negligible in the overall radiation balance. This backscattering peak is omitted in Fig. 7 but is discussed in the following section.

Table 1 lists the cumulative probabilities $A_m = \sum_{n=0}^m a_n$ calculated for hexagonal ice columns and plates as well as for a droxtal (tapered column). The main conclusion following from Table 1 is that about 90% of photons exit a crystal after the zeroth, first, and second internal collisions. Higher orders of collisions $m \geq 3$ account for only $\sim 10\%$ of the scattered photons. The second conclusion is that the majority of the scattered photons undergo only one internal collision: the probability a_1 is $\sim 70\%$ for the majority of particle shapes, although it decreases to $\sim 50\%$ for hexagonal columns with an aspect ratio of ~ 1 . For hexagonal columns and plates, these one-internal-collision photons are represented by only three terms: b_0 corresponds to the passage through parallel plane facets ($\alpha_0 = 0^\circ$, $\theta_0 = 0^\circ$), while b_1 and b_2 correspond to traversing the $\alpha_1 = 60^\circ$ and $\alpha_2 = 90^\circ$ wedges causing the $\theta_1 \approx 22^\circ$ and $\theta_2 \approx 46^\circ$ halos, respectively.

Thus, our statistical approach [3] largely relies upon the fact that the main features of the phase function are caused by beams with one internal reflection. Then these features are modeled adequately. In particular, light scattering by randomly oriented ice crystals is reduced to the calculation or modeling of the coefficients a_m and b_i , while the functions $p_m(\theta)$ and $p_{1i}(\theta)$ are assumed to be known reasonably well (Fig. 8). As to the inverse scattering problem, it is expedient to retrieve just the coefficients a_m and b_i from the experimentally measured phase functions because of their transparent physical meaning.

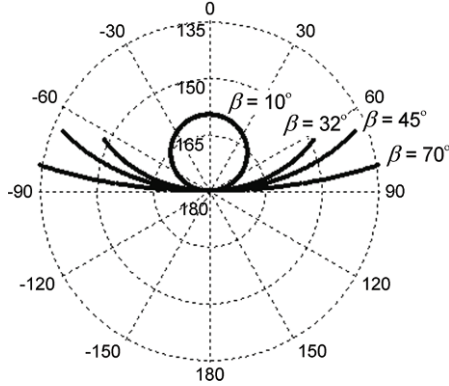


Fig. 9. Mapping of orientation coordinate lines $\beta = \text{const}$ onto the scattering-direction sphere ($\theta = 180^\circ$ is the exact backscattering direction).

3.2. Backscattering peak for hexagonal ice crystals

Many authors have calculated the Mueller matrix for hexagonal ice crystals using either ray-tracing or facet-tracing techniques and obtained a sharp phase-function peak in the backward direction. The reason for this is that a hexagonal crystal is equivalent to a 2D corner reflector owing to the 90° dihedral angle.

By use of our computer code, we have decomposed the backscatter into constituent ray trajectories and demonstrated that only four types of ray trajectories are responsible for the backscattering peak [2,4,6]. Two of these trajectories shown in Fig. 3 give the main contribution. Figure 9 depicts a mapping of such a trajectory on the scattering-direction sphere. Here, the angle γ is scanned continuously, while the angle β takes discrete values. We see that all curves converge at the exact backscattering direction forming a straight line $\varphi = \pm\pi/2$. Thus, in the vicinity of backscatter we obtain

$$\langle \mathbf{m}(\theta, \varphi) \rangle = \mathbf{C} \delta(\varphi \pm \pi/2) / \sin \theta. \quad (49)$$

Then the total Mueller matrix $\langle \mathbf{M}(\theta) \rangle$ is readily found after the integration over the orientation azimuth angle α according to Eq. (41):

$$\langle \mathbf{M}(\theta) \rangle = \frac{\mathbf{C}}{2\pi \sin \theta}. \quad (50)$$

Thus, only the constant matrix \mathbf{C} needs to be calculated numerically, and the requisite analytical description of the backscattering peak reduces to the singular function $1/\sin \theta$. Note that this singularity was also mentioned by Muinonen et al. [18].

The Mueller matrix of Eq. (50) describes the scattered light at an arbitrary observation point except that in the exact backward direction $\theta=0$, where the scattered intensity becomes infinite. Furthermore, there is an uncertainty in the orien-

tation of the scattering plane at this point. Nevertheless, these circumstances are not obstacles for the interpretation of experimental data because, in practice, the receiver of a lidar system detects a certain integral of the Mueller matrix of Eq. (50) over a cone $\Delta\theta$ around the backward direction. In this case, we have to replace the Mueller matrix of Eq. (50) by the average matrix

$$\boldsymbol{\mu}(\theta) = \frac{1}{2\pi \sin \theta} \int \frac{d\varphi}{2\pi} \mathbf{L}(\varphi) \mathbf{C} \mathbf{L}(\varphi) d\varphi = \frac{\mathbf{A}}{2\pi \sin \theta}, \quad (51)$$

where \mathbf{A} is constant diagonal matrix expressed in terms of the matrix \mathbf{C} as follows:

$$\mathbf{A} = \begin{bmatrix} C_{11} & 0 & 0 & 0 \\ 0 & \frac{1}{2}(C_{22} - C_{33}) & 0 & 0 \\ 0 & 0 & -\frac{1}{2}(C_{22} - C_{33}) & 0 \\ 0 & 0 & 0 & C_{44} \end{bmatrix}. \quad (52)$$

Thus, only three quantities must be calculated numerically.

In lidar measurements, the linear depolarization ratio $\Delta = I_{\perp} / I_{\parallel}$ is often used assuming that the incident light is fully linear polarized; then Δ is the ratio of the cross- and co-polarized components of the scattered intensity. For the conventional case of Eq. (52), we have:

$$\Delta = \frac{I_{\perp}}{I_{\parallel}} = \frac{A_{11} - A_{22}}{A_{11} + A_{22}}. \quad (53)$$

The depolarization ratio found numerically proves to be too large compared to experimental data. We explain this by the fact that the interference among the beams is not accounted for in the geometric-optics matrix. Ways to correct for this deficiency in the framework of the physical optics formalism are under consideration.

3.3. Diffraction of shadow-forming beams

At present, the physical-optics Mueller matrix of randomly oriented crystals cannot be calculated yet for all scattering directions because of excessive consumption of computer resources. Therefore, only scattering directions of practical interest should be considered. The simplest case that can be studied is forward scattering. Indeed, in the forward direction the shadow-forming beams dominate, the shape of each beam being a polygon. As is well known, the Fraunhofer diffraction on a polygon is reduced to the sum of rather simple functions over the polygon corners that can be calculated efficiently. Averaging over particle orientation takes into account only the changing shape and area of the particle projection. However, the computation of the average small-angle scattering phase function is also computer-intensive because of the large dynamic range of intensity values in the diffraction patterns.

To overcome this problem, we invoke the Fourier transform of the scattering phase function, hereinafter called the shadow function. As compared to the scat-

tering phase function, the shadow function has a number of advantages. First, it has an obvious geometrical meaning, which facilitates a visual evaluation of numerical results. Second, this function virtually vanishes outside a finite domain with a size equal to the maximal diameter of the particle shadow. Third, the shadow function, unlike the small-angle scattering phase function, is a smooth function with a small dynamic range of values. And fourth, this function is wavelength-independent because it is determined by the near-zone scattered field.

The shadow-forming beam propagating in the direction of the incident plane wave is described by Eqs. (3), (10), and (25), where $\boldsymbol{\rho} = (x, y)$ are coordinates in the plane normal to the incidence direction; the function $\eta(\boldsymbol{\rho})$ will be called the shadow indicator. At large distances from the particle, i.e., in the wave zone, this field is transformed into a diverging spherical wave distributed over scattering directions according to the so-called scattering amplitude $f(\mathbf{n})$ which, for the shadow-forming field, is described by the classical Fraunhofer diffraction equation (25). In this section, let us re-write this equation as follows:

$$f(\mathbf{v}) = \frac{k}{2\pi i} \int \eta(\boldsymbol{\rho}) \exp(-ik\mathbf{v} \cdot \boldsymbol{\rho}) d\boldsymbol{\rho}, \quad (54)$$

where \mathbf{v} is the projection of the scattering direction onto the plane normal to the incidence direction. The small-angle scattered field deviates from zero noticeably only in near-forward directions with $|\mathbf{v}| \ll 1$; therefore, the scattering amplitude (54) can be formally considered a function defined in the infinite plane of \mathbf{v} values. Then the energy conservation law is reduced to the Parseval theorem and has the following simple form:

$$\int |f(\mathbf{v})|^2 d\mathbf{v} = \int \eta^2(\boldsymbol{\rho}) d\boldsymbol{\rho} = s, \quad (55)$$

where s is the area of the particle shadow. The small-angle scattering amplitude $f(\mathbf{v})$ carries full information on the shadow indicator $\eta(\boldsymbol{\rho})$ which could be retrieved from the experimentally measured scattering amplitude by a trivial 2D Fourier transform. However, in real optical experiments, the real-valued quadratic value of the field,

$$I(\mathbf{v}) = |f(\mathbf{v})|^2, \quad (56)$$

is commonly measured instead of the complex scattering amplitude $f(\mathbf{v})$. The $I(\mathbf{v})$ is the standard Fraunhofer-diffraction pattern of a hole in a $\eta(\boldsymbol{\rho})$ -shaped black screen. In terms of the light scattering theory, $I(\mathbf{v})$ corresponds to the conventional scattering phase function $p(\mathbf{v})$, differing, as is evident from Eq. (55), only by the normalization factor

$$p(\mathbf{v}) = I(\mathbf{v})/s. \quad (57)$$

This poses a question of whether all the information on the shadow shape is retained in the scattering phase function. The answer follows from the Fourier trans-

form of the phase function or, more conveniently, the diffraction pattern $I(\mathbf{v})$:

$$S(\boldsymbol{\rho}) = \int I(\mathbf{v}) \exp(ik\mathbf{v} \cdot \boldsymbol{\rho}) d\mathbf{v}, \quad (58)$$

where the inverse Fourier transform yields the initial diffraction pattern:

$$I(\mathbf{v}) = \left(\frac{k}{2\pi}\right)^2 \int S(\boldsymbol{\rho}) \exp(-ik\mathbf{v} \cdot \boldsymbol{\rho}) d\boldsymbol{\rho}. \quad (59)$$

Substituting Eqs. (54) and (56) into Eq. (58), we obtain the following formula:

$$S(\boldsymbol{\rho}) = \int \eta(\boldsymbol{\rho}') \eta(\boldsymbol{\rho}' - \boldsymbol{\rho}) d\boldsymbol{\rho}', \quad (60)$$

which has a clear geometrical meaning of an autocorrelation of the shadow indicator. It is this function that will be referred to as the shadow function.

Thus, if what is measured experimentally is the scattering phase function rather than the scattering amplitude then we can retrieve only the smoother autocorrelation or shadow function but not the initial shadow indicator. Let us recapitulate the main properties of the shadow function (60). First, it reaches a maximum in the center, $\boldsymbol{\rho} = \mathbf{0}$, the maximum value being equal to the shadow area:

$$S(0) = s. \quad (61)$$

Then it decreases rapidly, vanishing at a distance from the center equal to the shadow diameter in the given direction on the $\boldsymbol{\rho}$ plane. If the shadow shape is convex, then the shadow function decreases monotonically in any direction. At $\boldsymbol{\rho} = \mathbf{0}$ it has a singularity in the form of a sharp peak. Here the directional derivative experiences a jump and a reversal of sign, since $S(\boldsymbol{\rho}) = S(-\boldsymbol{\rho})$. The integral of the shadow function over the plane is equal to the square of the shadow area:

$$\int S(\boldsymbol{\rho}) d\boldsymbol{\rho} = s^2. \quad (62)$$

Let us compare the properties of scattering phase functions with those of shadow functions. A small-angle scattering phase function is formally defined by Eqs. (54)–(57) in the unlimited plane of \mathbf{v} values. Although it deviates from zero only in the central spot with $|\mathbf{v}| \leq \lambda/d$, there are also side diffraction lobes rapidly decreasing in magnitude and carrying additional information on the shadow shape. Therefore, the proper choice of a limited area in the \mathbf{v} plane, where the phase function is calculated, and cell sizes is always complicated. Moreover, the choice depends on the wavelength, which must be fixed in the calculations. None of the above-mentioned problems appear in numerical calculations of shadow functions.

Moreover, note that the scattering phase function for nonspherical particles is a complex and oscillating function of two variables. Therefore, a particle with a complex shape is usually approximated by a simplified one, most often a sphere. In the domain of scattering phase functions, fitting phase-function parameters for different particle shapes numerically is a formal and ambiguous procedure. Thus,

another advantage of shadow functions over scattering phase functions is that the former makes the substitution of one particle shape for another (e.g., using a simpler shape or a superposition of shapes) a geometrically descriptive and physically justified procedure.

The above-introduced functions $\eta(\boldsymbol{\rho})$ and $S(\boldsymbol{\rho})$ depend on geometrical sizes and shapes of particles. It is convenient to exclude the trivial size dependence and consider only the shape dependence. To this end, we switch from the dimensional variable $\boldsymbol{\rho}$ to a dimensionless one \mathbf{R} defined as

$$\boldsymbol{\rho} = \mathbf{R} s^{1/2}, \quad (63)$$

where $d\boldsymbol{\rho} = dx dy = s d\mathbf{R}$. The shadow indicator, written in terms of these variables,

$$\eta_0(\mathbf{R}) = \eta(\boldsymbol{\rho} s^{-1/2}), \quad (64)$$

will be called the reduced shadow indicator. Since $\eta_0(\mathbf{R})$ identifies a unit area on the \mathbf{R} plane, it is the reduced shadow indicator $\eta_0(\mathbf{R})$ that characterizes the particle shape. The absolute particle size can be introduced easily by returning to the variable $\boldsymbol{\rho}$. The autocorrelation (60) of the shadow indicator (64) yields the reduced shadow function:

$$S_0(\mathbf{R}) = \int \eta_0(\mathbf{R}') \eta_0(\mathbf{R}' - \mathbf{R}) d\mathbf{R}' = S(\boldsymbol{\rho} s^{-1/2})/s, \quad (65)$$

which is defined only by the shadow shape and has the following properties:

$$S_0(0) = 1, \quad (66)$$

$$\int S_0(\mathbf{R}) d\mathbf{R} = 1. \quad (67)$$

In optics of scattering media, the usual subject of study is not a single particle but a statistical ensemble of particles having certain sizes, shapes, and spatial orientations. In this case the detector sums up the diffraction patterns (56) formed by different particles. If N is the number of particles then the detector reading is given by $N\langle I(\mathbf{v}) \rangle$, where $\langle I(\mathbf{v}) \rangle$ is the ensemble average of the single-particle diffraction pattern possessing the following properties:

$$\int \langle I(\mathbf{v}) \rangle d\mathbf{v} = \langle s \rangle, \quad (68)$$

$$\langle I(0) \rangle = \langle s^2 \rangle / \lambda^2, \quad (69)$$

$\langle s \rangle$ and $\langle s^2 \rangle$ being the mean area and the mean squared area of the shadow, respectively. The Fourier transform (58) of the diffraction pattern $\langle I(\mathbf{v}) \rangle$ yields the shadow function

$$\langle S(\boldsymbol{\rho}) \rangle = \int \langle I(\mathbf{v}) \rangle \exp(ik\mathbf{v} \cdot \boldsymbol{\rho}) d\mathbf{v} = \int \langle \eta(\boldsymbol{\rho}') \eta(\boldsymbol{\rho}' - \boldsymbol{\rho}) \rangle d\boldsymbol{\rho}' \quad (70)$$

possessing the following properties:

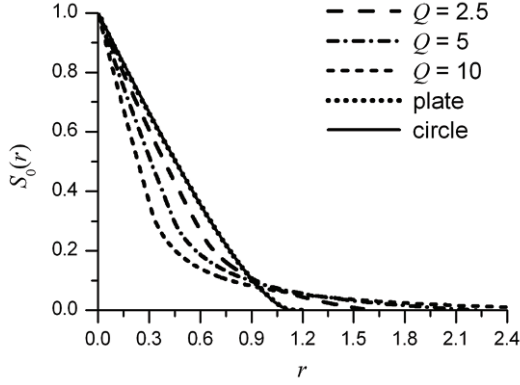


Fig. 10. Shadow functions for horizontally oriented hexagonal crystals.

$$\langle S(0) \rangle = \langle s \rangle, \quad (71)$$

$$\int \langle S(\boldsymbol{\rho}) \rangle d\boldsymbol{\rho} = \langle s^2 \rangle. \quad (72)$$

To exclude absolute particle sizes, one defines the specific shadow function for the statistical ensemble via Eq. (65), where the shadow area is replaced by the mean area

$$S_0(\mathbf{R}) = \langle S(\boldsymbol{\rho} \langle s \rangle^{-1/2}) \rangle / \langle s \rangle. \quad (73)$$

The normalization condition (66) still holds for this reduced function, while the condition (67) is replaced by the equation

$$\int S_0(\mathbf{R}) d\mathbf{R} = \langle s^2 \rangle / \langle s \rangle^2. \quad (74)$$

In certain cases, faceted particles can have preferential spatial orientations. For example, plate-like ice crystals are horizontally oriented, with a probable spread of deviations from the horizontal plane not exceeding 5° . The principal axes of ice columns also lie in the horizontal plane while having a random azimuth-angle distribution; if a pair of side faces of a column is also horizontally oriented then the column is called Parry-oriented. It is horizontally oriented ice crystals that cause a number of halos well known in atmospheric optics.

Consider the simplest case of light normally incident on a hexagonal Parry-oriented column. Its shadow indicator is a rectangle. For a random horizontal orientation of a column, the shadow function must be averaged over the azimuth angle, resulting in a one-argument function $S_0(r)$, where r is the distance from the center $\mathbf{R} = \mathbf{0}$. Figure 10 shows the shadow functions (73) calculated for different aspect ratios Q ; obviously, the circle shadow function, also shown in Fig. 10 for comparison, represents the limiting case.

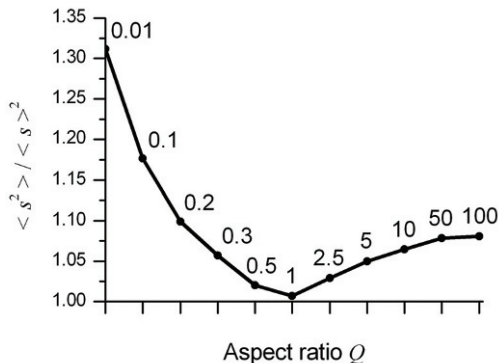


Fig. 11. Relative dispersion of the shadow area for randomly oriented hexagonal columns and plates.

The greater the aspect ratio of a column, the greater the deviation of its shadow function from that for a circle. As is seen from Fig. 10, the shadow function for a horizontally oriented hexagonal plate essentially coincides with that for a circle. Note that since the particle shadow area remains constant during the 2D orientation averaging, all functions in Fig. 10 satisfy Eqs. (66) and (67).

In the case of a uniform 3D orientation of a particle, the area of its projection changes with orientation, and the reduced shadow function satisfies Eq. (54). The values of the integral parameter Q of shadow functions calculated for hexagonal columns and plates are given in Fig. 11. One can see that the cube-like shape of crystals with $Q \approx 1$ represent an extremum. Here, the particle shape is most close to that of a sphere and, consequently, the parameter $\mu = \langle s^2 \rangle / \langle s \rangle^2$ is very close to unity. For plates ($Q < 1$), μ increases and approaches the limiting value $\mu_1 = 4/3$ corresponding to infinitesimally thin plates, which is easily demonstrated analytically. For elongated columns ($Q > 1$), μ increases and approaches its limiting value $\mu_2 \approx 1.08$, which can be also be shown analytically.

The shadow functions of randomly oriented hexagonal crystals are shown in Fig. 12. Note the qualitative difference between the shadow-function profiles for columns and plates. The shadow functions of hexagonal plates are close to that of a circle up to $r \approx 0.6$. The excess over the latter observed at $1.6 > r > 0.6$ fills a narrow gap between the shadow functions for a circle and an infinitesimally thin plate ($Q \rightarrow 0$). Note that this excess corresponds to the above-mentioned increase of the parameter μ .

For hexagonal columns with $Q \gg 1$, μ is not large. Therefore, a change in the shadow function profile with increasing Q causes first a decrease in $S_0(r)$ compared to the shadow function of a circle at $r \leq 1$, followed by an excess over the shadow function of a circle at $r > 1$.

In summary, small-angle phase functions have been calculated by other authors only for specific particle shapes and fixed wavelengths. The formalism of

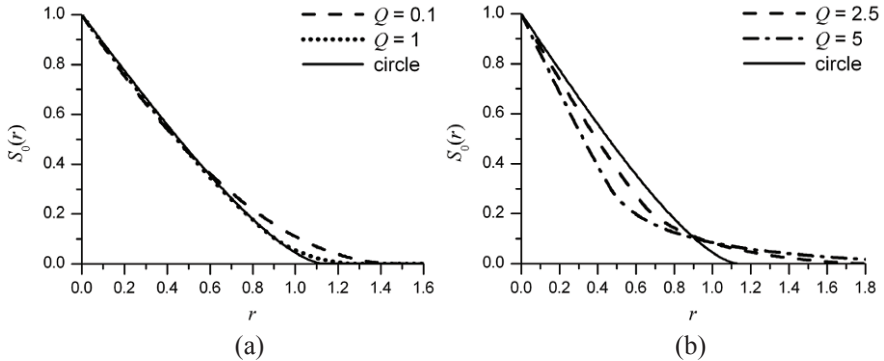


Fig. 12. Shadow functions for randomly oriented hexagonal crystals: (a) plates; (b) columns.

shadow functions outlined above has a number of important advantages over the direct calculation of diffraction patterns and, therefore, can be applied efficiently to various problems of small-angle scattering. Furthermore, the procedure of replacing a complex particle shape with a simpler one becomes physically justified and transparent. The formalism developed should be especially useful in solving practical problems of aureole scattering and laser sensing of cirrus clouds.

References

1. Borovoi, A. G., 2006: Multiple scattering of short waves by uncorrelated and correlated scatterers. *Light Scattering Rev.* **1**, 181–252.
2. Borovoi, A. G., and I. A. Grishin, 2003: Scattering matrices for large ice crystal particles. *J. Opt. Soc. Am. A* **20**, 2071–2080.
3. Borovoi, A., and N. Kustova, 2006: Statistical approach to light scattering by convex ice crystals. *Opt. Lett.* **31**, 1747–1749.
4. Borovoi, A., I. Grishin, E. Naats, and U. Oppel, 2000: Backscattering peak of hexagonal ice columns and plates. *Opt. Lett.* **25**, 1388–1390.
5. Borovoi, A., I. Grishin, E. Naats, and U. Oppel, 2002: Light backscattering by hexagonal ice crystals. *J. Quant. Spectrosc. Radiat. Transfer* **72**, 403–417.
6. Borovoi, A. G., N. V. Kustova, and U. G. Oppel, 2005: Light backscattering by hexagonal ice crystal particles in the geometrical optics approximation. *Opt. Eng.* **44**, 071208.
7. Cai, Q., and K. N. Liou, 1982: Polarized light scattering by hexagonal ice crystals: theory. *Appl. Opt.* **21**, 3569–3580.
8. Heffels, C., D. Heitzmann, E. D. Hirtleman, and B. Scarlett, 1995: Forward light scattering for arbitrary sharp-edged convex crystals in Fraunhofer and anomalous diffraction approximations. *Appl. Opt.* **34**, 6552–6560.
9. Hess, M., and M. Wiegner, 1994: COP: a data library of optical properties of hexagonal ice crystals. *Appl. Opt.* **33**, 7740–7746.
10. Jackson, J. D., 1998: *Classical Electrodynamics* (Wiley, New York).

11. Liou, K. N., 2002: *An Introduction to Atmospheric Radiation* (Academic Press, San Diego).
12. Macke, A., 1993: Scattering of light by polyhedral ice crystals. *Appl. Opt.* **32**, 2780–2788.
13. Macke, A., J. Mueller, and E. Raschke, 1996: Single scattering properties of atmospheric ice crystals. *J. Atmos. Sci.* **53**, 2813–2825.
14. Mishchenko, M. I., and A. Macke, 1998: Incorporation of physical optics effects and computation of the Legendre expansion for ray-tracing phase functions involving δ -function transmission. *J. Geophys. Res.* **103**, 1799–1805.
15. Mishchenko, M. I., J. W. Hovenier, and L. D. Travis, Eds., 2000: *Light Scattering by Nonspherical Particles: Theory, Measurements and Applications* (Academic Press, San Diego).
16. Mishchenko, M. I., L. D. Travis, and A. A. Lacis, 2002: *Scattering, Absorption, and Emission of Light by Small Particles* (Cambridge University Press, Cambridge).
17. Muinonen, K., 1989: Scattering of light by crystals: a modified Kirchhoff approximation. *Appl. Opt.* **28**, 3044–3050.
18. Muinonen, K., K. Lumme, J. Peltoniemi, and W. M. Irvine, 1989: Light scattering by randomly oriented crystals. *Appl. Opt.* **28**, 3051–3060.
19. Noel, V., G. Ledanois, H. Chepfer, and P. H. Flamant, 2001: Computation of a single-scattering matrix for nonspherical particles randomly or horizontally oriented in space. *Appl. Opt.* **40**, 4365–4375.
20. Rockwitz, K. D., 1989: Scattering properties of horizontally oriented ice crystal columns in cirrus clouds. *Appl. Opt.* **28**, 4103–4110.
21. Taflove, A., 1995: *Computational Electrodynamics: The Finite-Difference Time-Domain Method* (Artech House, Boston).
22. Takano, Y., and K. Jayaweera, 1985: Scattering phase matrix for hexagonal ice crystals computed from ray optics. *Appl. Opt.* **24**, 3254–3263.
23. Yang, P., and K. N. Liou, 1995: Light scattering by hexagonal ice crystals: comparison of finite-difference time domain and geometric optics models, *J. Opt. Soc. Am. A* **12**, 162–176.
24. Yang, P., and K. N. Liou, 1996: Geometric-optics–integral-equation method for light scattering by nonspherical ice crystals. *Appl. Opt.* **35**, 6568–6584.
25. Yang, P., and K. N. Liou, 1997: Light scattering by hexagonal ice crystals: solutions by a ray-by-ray integration algorithm. *J. Opt. Soc. Am. A* **14**, 2278–2289.

Regularization of inverse problems in atmospheric remote sensing

Adrian Doicu*, Thomas Trautmann, and Franz Schreier

German Aerospace Center, Remote Sensing Technology Institute, Oberpfaffenhofen, Germany

Abstract. In this chapter we present the basic concepts of numerical regularization theory. We analyze direct and iterative methods for solving nonlinear ill-posed problems in a general framework. From the category of direct methods we choose the method of Tikhonov regularization, while from the category of iterative methods we pay attention to the iteratively regularized Gauss–Newton method and the regularizing Levenberg–Marquardt method. For the Tikhonov regularization, we discuss the practical aspects of the regularization parameter choice methods, which are decisive for obtaining a reliable solution of the inverse problem. The efficiency of the regularization methods is analyzed from the numerical point of view by considering several test problems of atmospheric remote sensing.

Keywords: atmospheric remote sensing, inverse problems, regularization methods

1. Introduction

In atmospheric remote sensing, the relationship between the physical state parameters x and the collected observations making up some set of data y is described by a forward model F . This encapsulates a radiative transfer model and an instrument model, and formally, we write

$$F(x) = y.$$

The task of computing the data y given the physical state parameters x is called the forward problem, while the mathematical process of retrieving x given y is called the inverse problem. Atmospheric remote sensing deals with the inverse problem. In fact, inverse problems are ubiquitous challenges in almost any field of science and engineering, from astrophysics, helioseismology, geophysics, quantum-mechanical scattering theory, and materials science to medicine with its large diversity of imaging and tomographic techniques.

The forward model is closely connected with the actual observation being performed and mirrors the physics of the measurement process. In contrast the approach to solve the inverse problem is (to some extent) independent of the physi-

* Corresponding author. E-mail: adrian.doicu@dlr.de

cal process, and so the methods developed throughout this chapter can be applied to inverse problems in other fields as well.

In the general framework, the data y may be a function of frequency (or wavelength) or they may be a collection of discrete observations. In the former case, the problem is called a continuous problem, while in the latter case it is called a semi-discrete problem. When both x and y are discrete, the corresponding problem is a discrete problem. In order to avoid possible confusion, vectors will be denoted by bold upright letters, e.g., \mathbf{x} is a vector of state parameters or simply a state vector, while x is a state parameter function. As any measurement system can deliver only a discrete, finite set of data, the problems arising in atmospheric remote sensing are semi-discrete. Moreover, due to the complexity of radiative transfer, the forward model has to be computed by a numerical algorithm, which, in turn, requires a discretization of the state parameter function. For these reasons, the atmospheric inverse problems we are dealing with are discrete.

An important issue is that actual observations contain instrumental or measurement noise. We can thus envision data \mathbf{y}^δ as generally consisting of noiseless observations \mathbf{y} from a “perfect” instrument plus a noise component δ , i.e.,

$$\mathbf{y}^\delta = \mathbf{y} + \delta.$$

Due to the instrumental noise, the retrieval problems arising in atmospheric remote sensing belong to the class of so-called discrete ill-posed problems. These problems are unstable under data perturbations, and can be solved by numerical regularization methods whereby the solution is stabilized by taking additional information into account.

The goal of this chapter is to present the fundamentals of numerical regularization methods for solving inverse problems, and in particular of those arising in atmospheric remote sensing. The chapter is organized as follows. In Section 1 we clarify the concept of ill-posedness, while in Section 2 we discuss practical aspects of the method of Tikhonov regularization for solving linear inverse problems. These results are extended, in Section 3, to the case of nonlinear inverse problems. As an alternative to the method of Tikhonov regularization, we present in Section 4 two important iterative regularization methods, namely, the iteratively regularized Gauss–Newton method and the regularizing Levenberg–Marquardt method. For an exhaustive analysis of other regularization methods such as, for example, the Newton–CG method, the regularized total least-squares method, and mollifier methods, we refer the reader to the recent monograph by Doicu et al. [5].

Our presentation is mainly focused on practical aspects of numerical regularization theory and not on the theoretical aspects associated with the proof of convergence and convergence rate results. To interested readers we mention that the mathematical fundamentals of regularization theory can be found in the books by Engl et al. [6] and Rieder [29], the mathematical foundation of iterative regularization methods is the subject of the book by Kaltenbacher et al. [20], and the state of the art in numerical regularization can be found in the book by Hansen [16].

2. Ill-posedness of linear problems

Inverse problems typically involve the estimation of certain quantities based on indirect measurements of these quantities. The inversion process is often ill-posed in the sense that noise in the data gives rise to significant errors in the estimate. In this section we introduce the concept of ill-posedness and analyze the solvability and ill-posedness of linear discrete equations. Our analysis is focused on the classical example in atmospheric remote sensing, namely, the temperature retrieval by nadir sounding. In a continuous setting, this retrieval problem is modeled by a Fredholm integral equation of the first kind, which is the prototype of ill-posed problems.

2.1. Continuous, semi-discrete, and discrete problems

In a plane-parallel atmosphere and under the assumption that the background contribution from the surface can be neglected, the diffuse radiance at the top of the atmosphere $z = z_{\max}$ and in a direction with zero scan angle is given by the Schwarzschild equation

$$I(\nu) = \int_0^{z_{\max}} B[\nu, T(z)] \frac{\partial \mathbb{T}}{\partial z}(\nu, z) dz,$$

where B is the Planck function, T is the temperature, and \mathbb{T} is the transmission function of the atmosphere. In the infrared and microwave regions, the Rayleigh–Jeans approximation allows the following representation of the Planck function:

$$B[\nu, T(z)] = 2ck_B\nu^2T(z),$$

where c is the speed of light and k_B is the Boltzmann constant. Moreover, neglecting the temperature dependence of the transmission yields

$$I(\nu) = \int_0^{z_{\max}} k(\nu, z)T(z) dz, \quad (1)$$

with

$$k(\nu, z) = 2ck_B\nu^2 \frac{\partial \mathbb{T}}{\partial z}(\nu, z)$$

being the kernel function. Equation (1), which we rewrite in the generic form ($y = I$ and $x = T$)

$$y(\nu) = \int_0^{z_{\max}} k(\nu, z)x(z) dz, \quad \nu \in [\nu_{\min}, \nu_{\max}],$$

is a Fredholm integral equation of the first kind and represents the mathematical model of a continuous problem. We introduce the linear operator $K: X = L^2([0, z_{\max}]) \rightarrow Y = L^2([\nu_{\min}, \nu_{\max}])$ by the relation

$$Kx = \int_0^{\bar{z}_{\max}} k(\cdot, z)x(z) dz$$

and express the integral equation as

$$Kx = y,$$

K being a bounded and compact operator.

A spectral instrument cannot measure a continuous signal, and so the data are always a collection of discrete observations. The radiances

$$y(v_i) = I(v_i),$$

with $\{v_i\}_{i=1,m}$ being an equidistant set of points in the spectral interval $[v_{\min}, v_{\max}]$, represent the data, and the equation to be solved takes the form

$$y(v_i) = \int_0^{\bar{z}_{\max}} k(v_i, z)x(z) dz, \quad i = 1, \dots, m.$$

The semi-discrete equation is a mathematical model for discrete observations of a physical process and can be expressed in compact form as

$$K_m x = y_m.$$

The data y_m is a vector with entries $[y_m]_i = y(v_i)$, $i = 1, \dots, m$, and K_m is a linear operator acting between the Hilbert space X and the finite-dimensional Euclidean space \mathbb{R}^m :

$$[K_m x]_i = (Kx)(v_i) = \int_0^{\bar{z}_{\max}} k(v_i, z)x(z) dz.$$

The discretization approach which transforms a continuous equation into a semi-discrete equation is known as the collocation method.

For a complete discretization, we consider the subspace $X_n \subset X$ with the (not necessarily orthonormal) basis $\{\Phi_{nj}\}_{j=1,n}$ and define the interpolant $x_n \in X$ of x as the solution of the equation

$$K_m x_n = y_m.$$

Representing x_n as a linear combination of basis functions,

$$x_n = \sum_{j=1}^n \xi_j \Phi_{nj},$$

we obtain the system of equations

$$y(v_i) = \sum_{j=1}^n \left[\int_0^{\bar{z}_{\max}} k(v_i, z) \Phi_{nj}(z) dz \right] \xi_j, \quad i = 1, \dots, m. \quad (2)$$

In the matrix form, Eq. (2) can be written as

$$\mathbf{K}_{mn} \mathbf{x}_n = \mathbf{y}_m,$$

where $\mathbf{x}_n = [\xi_1, \dots, \xi_n]^T$ (T stands for “transposed”) is the coordinate vector and the matrix \mathbf{K}_{mn} , with entries

$$[\mathbf{K}_{mn}]_{ij} = [K_m \Phi_{nj}]_i = (K \Phi_{nj})(v_i) = \int_0^{z_{\max}} k(v_i, z) \Phi_{nj}(z) dz,$$

is a linear map between the finite-dimensional Euclidean spaces \mathbf{R}^n and \mathbf{R}^m . The discretization approach that transforms a continuous equation into a discrete equation is called a projection method.

2.2. Concept of ill-posedness

The mathematical formulation of inverse problems leads to equations that typically are ill-posed. According to Hadamard, the equation

$$Kx = y,$$

with K being a linear operator acting from the Hilbert space X into the Hilbert space Y , is called well-posed provided that

1. for any $y \in Y$, a solution x exists;
2. the solution is unique; and
3. the solution is stable with respect to perturbations in y in the sense that if $Kx_0 = y_0$ and $Kx = y$ then $x \rightarrow x_0$ whenever $y \rightarrow y_0$.

Equivalently, the linear equation is called well-posed if the operator K is bijective and the inverse operator K^{-1} is continuous. If one of the Hadamard’s conditions is violated, the problem is called ill-posed. Denoting by

$$R(K) = \{Kx \mid x \in X\}$$

the range space of K and by

$$N(K) = \{x \in X \mid Kx = 0\}$$

the null space of K , it is apparent that

1. if K is not surjective ($R(K) \neq Y$) then the linear equation is not solvable for all $y \in Y$ (non-existence);
2. if K is not injective ($N(K) \neq \{0\}$) then the linear equation may have more than one solution (non-uniqueness); and
3. if K^{-1} exists but is not continuous, then the solution x of the linear equation does not depend continuously on the data (instability).

Non-existence can occur in practice because the forward model is approximate or because the data contains noise. Non-uniqueness is introduced by the need for discretization and is a manifestation of so-called rank deficient problems characterized by a matrix \mathbf{K}_{mn} with a non-trivial null space. In particular, state vectors

\mathbf{x}_0 that lie in the null space of \mathbf{K}_{mn} solve the equation $\mathbf{K}_{mn}\mathbf{x}_0 = \mathbf{0}$, and by superposition, any linear combination of these null-space solutions can be added to a particular solution without changing the fit to the data. Violation of the third Hadamard's condition creates serious numerical problems because small errors in the data space can be dramatically amplified in the state space.

When a continuous ill-posed problem is discretized, the underlying discrete problem inherits this ill-posedness, and we say that we are dealing with a discrete ill-posed problem. The ill-posedness of a discrete linear problem written in the form of a linear system of equations results in a huge condition number of the corresponding coefficient matrix. In this regard, the term "ill-conditioned system of equations" is also used to describe instability. To stabilize the inversion process, one may impose additional constraints that bias the solution, a process that is generally referred to as *regularization*.

2.3. Solvability and ill-posedness

For an $m \times n$ matrix \mathbf{K} with $\text{rank}(\mathbf{K}) = r$, the representation

$$\mathbf{K} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T,$$

where $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_m]$ and $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_n]$ are orthogonal $m \times m$ and $n \times n$ matrices, respectively, and $\mathbf{\Sigma}$ is an $m \times n$ matrix of the form

$$\mathbf{\Sigma} = \begin{bmatrix} \text{diag}(\sigma_i)_{r \times r} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

with $\text{diag}(\sigma_i)_{r \times r}$ being an $r \times r$ diagonal matrix, is called the singular value decomposition (SVD) of the matrix \mathbf{K} . The sets $\{\mathbf{v}_i\}_{i=1,r}$ and $\{\mathbf{v}_i\}_{i=r+1,n}$ are orthonormal bases of $N(\mathbf{K})^\perp$ and $N(\mathbf{K})$, respectively, i.e.,

$$N(\mathbf{K})^\perp = \text{span}\{\mathbf{v}_i\}_{i=1,r}, \quad N(\mathbf{K}) = \text{span}\{\mathbf{v}_i\}_{i=r+1,n},$$

while the sets $\{\mathbf{u}_i\}_{i=1,r}$ and $\{\mathbf{u}_i\}_{i=r+1,m}$ are orthonormal bases of $R(\mathbf{K})$ and $R(\mathbf{K})^\perp$, respectively, i.e.,

$$R(\mathbf{K}) = \text{span}\{\mathbf{u}_i\}_{i=1,r}, \quad R(\mathbf{K})^\perp = \text{span}\{\mathbf{u}_i\}_{i=r+1,m}.$$

The solvability of linear equations is stated by the following result: the linear equation

$$\mathbf{K}\mathbf{x} = \mathbf{y} \tag{3}$$

with $\text{rank}(\mathbf{K}) = n$ is solvable if and only if $\mathbf{y} \in R(\mathbf{K})$, and the unique solution is given by

$$\mathbf{x}^\dagger = \sum_{i=1}^n \frac{1}{\sigma_i} (\mathbf{u}_i^T \mathbf{y}) \mathbf{v}_i.$$

The above equation defines a linear operator $\mathbf{K}^\dagger: \mathbf{R}^m \rightarrow \mathbf{R}^n$ by the relation

$$\mathbf{K}^\dagger \mathbf{y} = \sum_{i=1}^n \frac{1}{\sigma_i} (\mathbf{u}_i^T \mathbf{y}) \mathbf{v}_i, \quad \mathbf{y} \in \mathbb{R}^m,$$

which also allows a representation by an $n \times m$ matrix,

$$\mathbf{K}^\dagger = \sum_{i=1}^n \frac{1}{\sigma_i} \mathbf{v}_i \mathbf{u}_i^T.$$

This operator or matrix, which maps $\mathbf{y} \in R(\mathbf{K})$ onto the solution \mathbf{x}^\dagger of Eq. (3), i.e.,

$$\mathbf{x}^\dagger = \mathbf{K}^\dagger \mathbf{y},$$

is called the generalized inverse. By convention, the data vector $\mathbf{y} \in R(\mathbf{K})$ will be referred to as the exact data vector and $\mathbf{x}^\dagger = \mathbf{K}^\dagger \mathbf{y}$ will be called the exact solution.

In practice, a noisy data vector \mathbf{y}^δ is a perturbation of the exact data vector \mathbf{y} , and we have the representation

$$\mathbf{y}^\delta = \mathbf{y} + \boldsymbol{\delta},$$

where $\boldsymbol{\delta}$ is the instrumental noise. In general, $\mathbf{y}^\delta \in \mathbb{R}^m$, and there is no guarantee that $\mathbf{y}^\delta \in R(\mathbf{K})$. As a consequence, the linear equation is not solvable for arbitrary noisy data, and so one needs another concept of solution, namely, the least squares solution. For the noisy data vector, the least squares solution of the linear equation is defined by

$$\mathbf{x}^\delta = \sum_{i=1}^n \frac{1}{\sigma_i} (\mathbf{u}_i^T \mathbf{y}^\delta) \mathbf{v}_i.$$

The least squares solution can be characterized as follows:

1. the image of \mathbf{x}^δ under \mathbf{K} is the projection of \mathbf{y}^δ onto $R(\mathbf{K})$, i.e.,

$$\mathbf{K} \mathbf{x}^\delta = P_{R(\mathbf{K})} \mathbf{y}^\delta = \sum_{i=1}^n (\mathbf{u}_i^T \mathbf{y}^\delta) \mathbf{u}_i;$$

2. \mathbf{x}^δ has the optimality property

$$\mathbf{x}^\delta = \arg \min_{\mathbf{x}} \|\mathbf{y}^\delta - \mathbf{K} \mathbf{x}\|;$$

3. \mathbf{x}^δ solves the normal equation

$$\mathbf{K}^T \mathbf{K} \mathbf{x} = \mathbf{K}^T \mathbf{y}^\delta.$$

For discrete ill-posed problems, the following features of the singular values and vectors are relevant:

- as the dimension of \mathbf{K} increases, the number of small singular values also increases;

- as the singular values σ_i decrease, the corresponding singular vectors \mathbf{u}_i and \mathbf{v}_i have more sign changes in their components.

As a consequence of the oscillatory behavior of the high-order singular vectors, the norm of the least squares solution becomes extremely large and \mathbf{x}^δ is not a reliable approximation of \mathbf{x}^\dagger . To be more specific, let i^* be a large singular-value index and let us consider a perturbation of the exact data vector \mathbf{y} in the direction of the singular vector \mathbf{u}_{i^*} ,

$$\mathbf{y}^\delta = \mathbf{y} + \Delta \mathbf{u}_{i^*},$$

with $\Delta = \|\mathbf{y}^\delta - \mathbf{y}\|$ being the noise level. The least squares solution is then given by

$$\mathbf{x}^\delta = \mathbf{x}^\dagger + \frac{\Delta}{\sigma_{i^*}} \mathbf{v}_{i^*},$$

and the ratio

$$\frac{\|\mathbf{x}^\delta - \mathbf{x}^\dagger\|}{\|\mathbf{y}^\delta - \mathbf{y}\|} = \frac{1}{\sigma_{i^*}}$$

is very large if σ_{i^*} is very small.

3. Tikhonov regularization for linear problems

The Tikhonov regularization is the most widely used technique for regularizing ill-posed problems [31,32]. We begin our analysis by formulating the Tikhonov regularization for linear problems and by making some remarks on the selection of the regularization matrix. We then describe the errors in the state space and the data space and analyze regularization parameter choice methods.

3.1. Formulation

For the linear data model

$$\mathbf{y}^\delta = \mathbf{K}\mathbf{x} + \boldsymbol{\delta},$$

the Tikhonov solution \mathbf{x}_α^δ satisfies the regularized normal equation

$$(\mathbf{K}^T \mathbf{K} + \alpha \mathbf{L}^T \mathbf{L}) \mathbf{x} = \mathbf{K}^T \mathbf{y}^\delta, \quad (4)$$

and can be expressed as $\mathbf{x}_\alpha^\delta = \mathbf{K}_\alpha^\dagger \mathbf{y}^\delta$, where the matrix

$$\mathbf{K}_\alpha^\dagger = (\mathbf{K}^T \mathbf{K} + \alpha \mathbf{L}^T \mathbf{L})^{-1} \mathbf{K}^T$$

is the regularized generalized inverse. The parameter $\alpha \geq 0$ is called the regularization parameter and \mathbf{L} is known as the regularization matrix. Since $\mathbf{L}^T \mathbf{L}$ is positive semi-definite, the spectrum of the matrix $\mathbf{K}^T \mathbf{K} + \alpha \mathbf{L}^T \mathbf{L}$ is shifted in the positive direction, and the solution of the regularized normal equation is less sus-

ceptible to perturbations in the data.

The Tikhonov regularization can be interpreted as a penalized least squares problem, i.e., \mathbf{x}_α^δ minimizes the Tikhonov function

$$F_\alpha(\mathbf{x}) = \|\mathbf{y}^\delta - \mathbf{K}\mathbf{x}\|^2 + \alpha \|\mathbf{L}\mathbf{x}\|^2$$

and vice versa. Minimizing the Tikhonov function means to search for some \mathbf{x}_α^δ providing at the same time a small residual $\|\mathbf{y}^\delta - \mathbf{K}\mathbf{x}\|$ and a moderate value of the penalty term $\|\mathbf{L}\mathbf{x}\|$. If the regularization parameter is chosen too small, Eq. (4) is too close to the original problem and we must expect instabilities; if α is too large, the equation we solve has only little connection with the original problem.

Regularization methods for solving ill-posed problems can be analyzed in a deterministic or a semi-stochastic setting:

- In a deterministic setting, the solution \mathbf{x}^\dagger corresponding to the exact data vector \mathbf{y} is assumed to be deterministic, and only information on the noise level Δ , defined as $\|\mathbf{y}^\delta - \mathbf{y}\| \leq \Delta$, is available.
- In a semi-stochastic setting, the solution \mathbf{x}^\dagger is deterministic, while the instrumental noise $\boldsymbol{\delta}$ is assumed to be an m -dimensional random vector. Typically $\boldsymbol{\delta}$ is supposed to be a discrete white noise with zero mean and covariance $\mathbf{C}_\delta = \sigma^2 \mathbf{I}_m$, where σ^2 is the noise variance.

The noise level can be estimated by using the probability distribution of the noise, i.e., $\Delta = E\{\|\boldsymbol{\delta}\|\}$ or $\Delta^2 = E\{\|\boldsymbol{\delta}\|^2\}$, where E is the expected value operator. These estimates can be computed either numerically by generating randomly a sample of noise vectors followed by averaging, or analytically, if the explicit integrals of probability densities are available. In the case of white noise, the second criterion yields

$$\Delta^2 = m\sigma^2.$$

From a practical point of view, the Tikhonov solution does not depend on which setting the problem is treated; differences arise when proving convergence and convergence rate results for different regularization parameter choice methods.

3.2. Regularization matrices

The penalty term in the expression of the Tikhonov function is called the discrete smoothing norm. The discrete smoothing norm takes into account the additional information about the solution, and its role is to stabilize the problem and single out a useful and stable solution.

If we intend to control the magnitude of the solution then \mathbf{L} can be chosen as either the identity matrix ($\mathbf{L}_0 = \mathbf{I}_n$) or a diagonal matrix.

If the solution should be smooth then we have to use another measure of the solution, such as the discrete approximations to derivative operators

$$\mathbf{L}_1 = \begin{bmatrix} -1 & 1 & \dots & 0 & 0 \\ 0 & -1 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & -1 & 1 \end{bmatrix} \in \mathbf{R}^{(n-1) \times n}$$

and

$$\mathbf{L}_2 = \begin{bmatrix} 1 & -2 & 1 & \dots & 0 & 0 & 0 \\ 0 & 1 & -2 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -2 & 1 \end{bmatrix} \in \mathbf{R}^{(n-2) \times n}.$$

If we have some knowledge about the magnitude of the state vector and want to constrain the solution to be smooth, we can combine several derivative orders and determine the regularization matrix by the Cholesky factorization

$$\mathbf{L}^T \mathbf{L} = \omega_0 \mathbf{L}_0^T \mathbf{L}_0 + \omega_1 \mathbf{L}_1^T \mathbf{L}_1,$$

where ω_0 and ω_1 are positive weighting factors satisfying the normalization condition

$$\omega_0 + \omega_1 = 1.$$

The regularization matrix can also be constructed by means of statistical information, that is, \mathbf{L} can be the Cholesky factor of an a priori profile covariance matrix. The covariance matrix \mathbf{C}_x corresponding to an exponential correlation function is given by

$$[\mathbf{C}_x]_{ij} = \sigma_{xi} \sigma_{xj} [\mathbf{x}_a]_i [\mathbf{x}_a]_j \exp\left(-2 \frac{|z_i - z_j|}{l_i + l_j}\right), \quad i, j = 1, \dots, n,$$

where σ_{xi} are dimensionless profile standard deviations and l_i are the lengths which determine the correlation between the parameters at different altitudes z_i . To compute the regularization matrix \mathbf{L} we have two options:

1. If the profile standard deviations are known to a certain accuracy, we set

$$\mathbf{C}_x^{-1} = \mathbf{L}^T \mathbf{L}.$$

2. If the profile standard deviations are unknown, we assume that $\sigma_{xi} = \sigma_x$ for all $i = 1, \dots, n$. The covariance matrix can be expressed as $\mathbf{C}_x = \sigma_x^2 \mathbf{C}_{nx}$, and we set

$$\mathbf{C}_{nx}^{-1} = \mathbf{L}^T \mathbf{L}.$$

In the first case, α can be regarded as a scale factor, and a regularization parameter choice method will yield a value of α close to one. In the second case, α reproduces the profile standard deviation σ_x , and a regularization parameter choice method will yield an estimate for σ_x .

3.3. Solution representation

In the framework of Tikhonov regularization, the regularized solution can be expressed in terms of the generalized singular value decomposition (GSVD) of the matrix pair (\mathbf{K}, \mathbf{L}) . Essentially, the GSVD is a superior numerical tool which yields important insights into the regularization problem involving both the Jacobian and the regularization matrices. The use of SVD and GSVD in the analysis of discrete ill-posed problems goes back to Hanson [18] and Varah [34]. For easy reference, we review this canonical decomposition following the presentation by Hansen [16].

If \mathbf{K} is an $m \times n$ matrix and \mathbf{L} is a $p \times n$ matrix with $m > n \geq p$ and, further, if $\text{rank}(\mathbf{L}) = p$ then the GSVD of the matrix pair (\mathbf{K}, \mathbf{L}) is given by

$$\mathbf{K} = \mathbf{U}\mathbf{\Sigma}_1\mathbf{W}^{-1}, \quad \mathbf{L} = \mathbf{V}\mathbf{\Sigma}_2\mathbf{W}^{-1},$$

where $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are block matrices of the form

$$\mathbf{\Sigma}_1 = \begin{bmatrix} \text{diag}(\sigma_i)_{p \times p} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n-p} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \mathbf{\Sigma}_2 = [\text{diag}(\mu_i)_{p \times p} \quad \mathbf{0}],$$

the matrices $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_m] \in \mathbb{R}^{m \times m}$ and $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_p] \in \mathbb{R}^{p \times p}$ are orthogonal, and the matrix $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_n] \in \mathbb{R}^{n \times n}$ is nonsingular. Moreover, $\text{diag}(\sigma_i)_{p \times p}$ and $\text{diag}(\mu_i)_{p \times p}$ are $p \times p$ diagonal matrices whose elements are positive and normalized according to

$$\sigma_i^2 + \mu_i^2 = 1, \quad i = 1, \dots, p.$$

The generalized singular values of (\mathbf{K}, \mathbf{L}) are defined by

$$\gamma_i = \frac{\sigma_i}{\mu_i},$$

and we shall assume them to appear in the descending order:

$$\gamma_1 \geq \dots \geq \gamma_p > 0.$$

As for the ordinary method, the number of small generalized singular values γ_i increases as the dimension of \mathbf{K} increases, while the singular vectors \mathbf{u}_i , \mathbf{v}_i , and \mathbf{w}_i have more sign changes in their components as the corresponding generalized singular values γ_i decrease.

In terms of the generalized singular systems of the matrix pair (\mathbf{K}, \mathbf{L}) , the regularized solution \mathbf{x}_α^δ is given by

$$\mathbf{x}_\alpha^\delta = \mathbf{K}_\alpha^\dagger \mathbf{y}^\delta = \sum_{i=1}^p \frac{\gamma_i^2}{\gamma_i^2 + \alpha} \frac{1}{\sigma_i} (\mathbf{u}_i^T \mathbf{y}^\delta) \mathbf{w}_i + \sum_{i=p+1}^n (\mathbf{u}_i^T \mathbf{y}^\delta) \mathbf{w}_i,$$

where the second term,

$$\mathbf{x}_0^\delta = \sum_{i=p+1}^n (\mathbf{u}_i^T \mathbf{y}^\delta) \mathbf{w}_i,$$

is the component of the solution in the null space of \mathbf{L} . The regularized generalized inverse $\mathbf{K}_\alpha^\dagger$ can be factorized according to

$$\mathbf{K}_\alpha^\dagger = (\mathbf{K}^T \mathbf{K} + \alpha \mathbf{L}^T \mathbf{L})^{-1} \mathbf{K}^T = \mathbf{W} \Sigma_\alpha^\dagger \mathbf{U}^T,$$

with

$$\Sigma_\alpha^\dagger = \begin{bmatrix} \text{diag} \left(\frac{\gamma_i^2}{\gamma_i^2 + \alpha} \frac{1}{\sigma_i} \right)_{p \times p} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n-p} & \mathbf{0} \end{bmatrix}.$$

If $p = n$ then the expression of the regularized solution simplifies to

$$\mathbf{x}_\alpha^\delta = \mathbf{K}_\alpha^\dagger \mathbf{y}^\delta = \sum_{i=1}^n \frac{\gamma_i^2}{\gamma_i^2 + \alpha} \frac{1}{\sigma_i} (\mathbf{u}_i^T \mathbf{y}^\delta) \mathbf{w}_i.$$

3.4. Errors in state space and data space

An error analysis can be performed in the state space, by evaluating the solution error $\mathbf{x}^\dagger - \mathbf{x}_\alpha^\delta$, or in the data space, by estimating the predictive risk $\mathbf{y} - \mathbf{K} \mathbf{x}_\alpha^\delta$. In fact, an error analysis is not only a tool for characterizing the accuracy of the obtained solution; it is also the basis for selecting the optimal solution.

In the state space, the deviation of the regularized solution from the exact solution is given by

$$\mathbf{x}^\dagger - \mathbf{x}_\alpha^\delta = (\mathbf{x}^\dagger - \mathbf{x}_\alpha) + (\mathbf{x}_\alpha - \mathbf{x}_\alpha^\delta),$$

where \mathbf{x}_α is the regularized solution for the exact data vector \mathbf{y} , i.e.,

$$\mathbf{x}_\alpha = \mathbf{K}_\alpha^\dagger \mathbf{y}.$$

Defining the total error by $\mathbf{e}_\alpha^\delta = \mathbf{x}^\dagger - \mathbf{x}_\alpha^\delta$ and the smoothing and noise errors by $\mathbf{e}_{s\alpha}^\delta = \mathbf{x}^\dagger - \mathbf{x}_\alpha$ and $\mathbf{e}_{n\alpha}^\delta = \mathbf{x}_\alpha - \mathbf{x}_\alpha^\delta$, respectively, we obtain

$$\mathbf{e}_\alpha^\delta = \mathbf{e}_{s\alpha}^\delta + \mathbf{e}_{n\alpha}^\delta.$$

The smoothing error quantifies the loss of information due to the regularization, while the noise error quantifies the loss of information due to inaccurate data. The smoothing error can be expressed in terms of the exact data vector \mathbf{y} as

$$\mathbf{e}_{s\alpha}^\delta = (\mathbf{K}^\dagger - \mathbf{K}_\alpha^\dagger) \mathbf{y} = \sum_{i=1}^n \frac{\alpha}{\gamma_i^2 + \alpha} \frac{1}{\sigma_i} (\mathbf{u}_i^T \mathbf{y}) \mathbf{w}_i,$$

and in terms of the exact solution \mathbf{x}^\dagger as

$$\mathbf{e}_{s\alpha} = \mathbf{x}^\dagger - \mathbf{x}_\alpha = (\mathbf{I}_n - \mathbf{A}_\alpha) \mathbf{x}^\dagger.$$

Here, the $n \times n$ matrix $\mathbf{A}_\alpha = \mathbf{K}_\alpha^\dagger \mathbf{K}$ is called the resolution matrix or the averaging kernel matrix. The noise error can be expressed in terms of the noise vector $\boldsymbol{\delta}$, i.e.,

$$\mathbf{e}_{n\alpha}^\delta = \mathbf{x}_\alpha - \mathbf{x}_\alpha^\delta = -\mathbf{K}_\alpha^\dagger \boldsymbol{\delta} = -\sum_{i=1}^n \frac{\gamma_i^2}{\gamma_i^2 + \alpha} \frac{1}{\sigma_i} (\mathbf{u}_i^T \boldsymbol{\delta}) \mathbf{w}_i.$$

In a semi-stochastic setting and for white noise, the smoothing error is deterministic, while the noise error is stochastic with zero mean and covariance $\mathbf{C}_{\text{en}} = \sigma^2 \mathbf{K}_\alpha^\dagger \mathbf{K}_\alpha^{\dagger T}$. The expected value of the total error is given by

$$E\{\|\mathbf{e}_\alpha^\delta\|^2\} = \|\mathbf{e}_{s\alpha}\|^2 + E\{\|\mathbf{e}_{n\alpha}^\delta\|^2\},$$

whereas the expected value of the noise error is computed as

$$E\{\|\mathbf{e}_{n\alpha}^\delta\|^2\} = \sigma^2 \sum_{i=1}^n \left(\frac{\gamma_i^2}{\gamma_i^2 + \alpha} \frac{1}{\sigma_i} \right)^2 \|\mathbf{w}_i\|^2.$$

The smoothing error $\|\mathbf{e}_{s\alpha}\|^2$ is an increasing function of α , while the expected value of the noise error $E\{\|\mathbf{e}_{n\alpha}^\delta\|^2\}$ is a decreasing function of α . Consequently, $E\{\|\mathbf{e}_\alpha^\delta\|^2\}$ has a minimum for an optimal value of the regularization parameter.

In the data space, the accuracy of the regularized solution can be characterized in terms of the predictive error or the predictive risk, defined as

$$\mathbf{p}_\alpha^\delta = \mathbf{K} \mathbf{e}_\alpha^\delta = \mathbf{p}_{s\alpha} + \mathbf{p}_{n\alpha}^\delta.$$

The predictive smoothing error is given by

$$\mathbf{p}_{s\alpha} = \mathbf{K} \mathbf{e}_{s\alpha} = (\mathbf{I}_m - \mathbf{K} \mathbf{K}_\alpha^\dagger) \mathbf{y} = (\mathbf{I}_m - \hat{\mathbf{A}}_\alpha) \mathbf{y},$$

where the $m \times m$ matrix $\hat{\mathbf{A}}_\alpha = \mathbf{K} \mathbf{K}_\alpha^\dagger$ is called the influence matrix. We have the expansion

$$\mathbf{p}_{s\alpha} = \sum_{i=1}^n \frac{\alpha}{\gamma_i^2 + \alpha} (\mathbf{u}_i^T \mathbf{y}) \mathbf{u}_i,$$

while for the predictive noise error,

$$\mathbf{p}_{n\alpha}^\delta = \mathbf{K} \mathbf{e}_{n\alpha}^\delta = -\mathbf{K} \mathbf{K}_\alpha^\dagger \boldsymbol{\delta} = -\hat{\mathbf{A}}_\alpha \boldsymbol{\delta} = -\sum_{i=1}^n \frac{\gamma_i^2}{\gamma_i^2 + \alpha} (\mathbf{u}_i^T \boldsymbol{\delta}) \mathbf{u}_i.$$

The expected value of the predictive error is given by

$$E\{\|\mathbf{p}_\alpha^\delta\|^2\} = \|\mathbf{p}_{s\alpha}\|^2 + E\{\|\mathbf{p}_{n\alpha}^\delta\|^2\},$$

with

$$\|\mathbf{p}_{s\alpha}\|^2 = \|(\mathbf{I}_m - \hat{\mathbf{A}}_\alpha)\mathbf{y}\|^2 = \sum_{i=1}^n \left(\frac{\alpha}{\gamma_i^2 + \alpha} \right)^2 (\mathbf{u}_i^T \mathbf{y})^2$$

and

$$E\{\|\mathbf{p}_{n\alpha}^\delta\|^2\} = \sigma^2 \text{trace}(\hat{\mathbf{A}}_\alpha^T \hat{\mathbf{A}}_\alpha) = \sigma^2 \sum_{i=1}^n \left(\frac{\gamma_i^2}{\gamma_i^2 + \alpha} \right)^2.$$

As in the state space, the predictive smoothing error $\|\mathbf{p}_{s\alpha}\|^2$ is an increasing function of α , while the expected value of the noise error $E\{\|\mathbf{p}_{n\alpha}^\delta\|^2\}$ is a decreasing function of α .

3.5. Regularization parameter choice methods

With too little regularization, reconstructions have highly oscillatory artifacts due to noise amplification. With too much regularization, the reconstructions are too smooth. Ideally, we would like to select a regularization parameter so that the corresponding regularized solution minimizes some indicator of solution fidelity, e.g., some measure of the magnitude of the solution error.

The formulations of regularization parameter choice methods in deterministic and semi-stochastic settings are very similar. The reason is that the noise level Δ can be estimated as $\Delta^2 = m\sigma^2$, and this estimate can be used to reformulate a “deterministic” parameter choice method in a semi-stochastic setting. According to the standard deterministic classification,

1. a regularization parameter choice method depending only on Δ , $\alpha = \alpha(\Delta)$, is called an *a priori* parameter choice method;
2. a regularization parameter choice method depending on Δ and \mathbf{y}^δ , $\alpha = \alpha(\Delta, \mathbf{y}^\delta)$, is called an *a posteriori* parameter choice method; and
3. a regularization parameter choice method depending only on \mathbf{y}^δ , $\alpha = \alpha(\mathbf{y}^\delta)$, is called an error-free parameter choice method.

4. *A priori* parameter choice methods

In a deterministic setting, an *a priori* parameter choice method is of the form $\alpha \propto \Delta^p$, while in a semi-stochastic setting, this selection rule translates into the choice $\alpha \propto \sigma^p$ [6,36,29]. In a semi-stochastic setting, we define the optimal regularization parameter for error estimation as the minimizer of the expected error,

$$\bar{\alpha}_{\text{opt}} = \arg \min_{\alpha} E\{\|\mathbf{e}_\alpha^\delta\|^2\}.$$

The optimal regularization parameter is not a computable quantity because the exact solution is unknown, but we may design an *a priori* parameter choice method by combining this selection criterion with a Monte Carlo technique. The resulting algorithm involves the following steps:

1. perform a random exploration of a domain in which the exact solution is supposed to lie by considering a set of state vector realizations $\{\mathbf{x}_i^\dagger\}_{i=1,N}$, where N is the sample size;
2. for each \mathbf{x}_i^\dagger , compute the optimal regularization parameter for error estimation

$$\bar{\alpha}_{\text{opt}i} = \arg \min_{\alpha} E\{\|\mathbf{e}_{\alpha}^{\delta}(\mathbf{x}_i^\dagger)\|^2\}$$

and determine the exponent

$$p_i = \frac{\log \bar{\alpha}_{\text{opt}i}}{\log \sigma};$$

3. compute the sample mean exponent

$$\bar{p} = \frac{1}{N} \sum_{i=1}^N p_i;$$

4. choose the regularization parameter as $\alpha_c = \sigma^{\bar{p}}$.

5. Discrepancy principle

The most popular *a posteriori* parameter choice method is the discrepancy principle due to Morozov [25]. In this method, the regularization parameter is chosen via a comparison between the residual norm (discrepancy)

$$\|\mathbf{r}_{\alpha}^{\delta}\|^2 = \|\mathbf{y}^{\delta} - \mathbf{K}\mathbf{x}_{\alpha}^{\delta}\|^2$$

and the assumed noise level Δ ,

$$\|\mathbf{r}_{\alpha}^{\delta}\|^2 = \tau \Delta^2, \quad \tau > 1.$$

A heuristic motivation for this method is that as long as we have only the noisy data vector \mathbf{y}^{δ} and know that $\|\mathbf{y}^{\delta} - \mathbf{y}\| \leq \Delta$, it does not make sense to look for an approximate solution $\mathbf{x}_{\alpha}^{\delta}$ with a discrepancy $\|\mathbf{y}^{\delta} - \mathbf{K}\mathbf{x}_{\alpha}^{\delta}\| < \Delta$; a residual norm of the order of Δ is the best one should ask for. In a semi-stochastic setting, we have the equation

$$\|\mathbf{r}_{\alpha}^{\delta}\|^2 = \tau m \sigma^2,$$

which, in terms of a generalized singular system of (\mathbf{K}, \mathbf{L}) , takes the form

$$\sum_{i=1}^m \left(\frac{\alpha}{\gamma_i^2 + \alpha} \right)^2 (\mathbf{u}_i^T \mathbf{y}^{\delta})^2 = \tau m \sigma^2.$$

In Fig. 1 we plot the expected value of the residual norm,

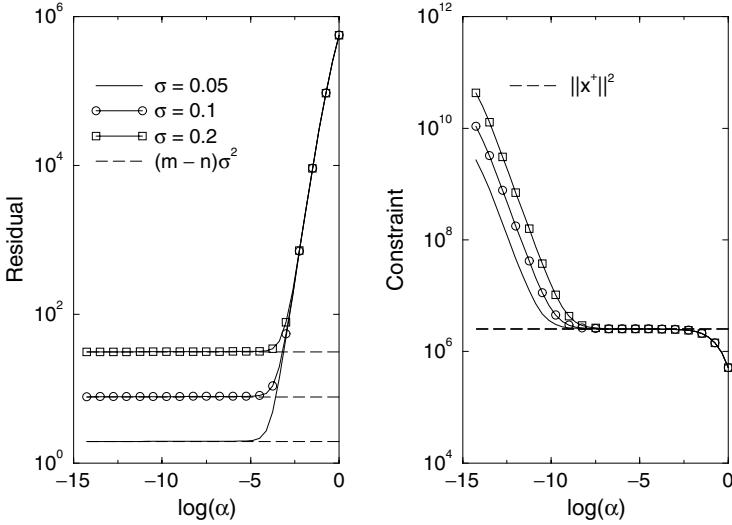


Fig. 1. Expected residual (left) and constraint (right) curves for different noise variances.

$$E\{\|\mathbf{r}_\alpha^\delta\|^2\} = (m-n)\sigma^2 + \sum_{i=1}^n \left(\frac{\alpha}{\gamma_i^2 + \alpha} \right)^2 [(\mathbf{u}_i^T \mathbf{y})^2 + \sigma^2]$$

and, for comparison, the expected value of the constraint norm

$$E\{\|\mathbf{c}_\alpha^\delta\|^2\} = \sum_{i=1}^n \left(\frac{\gamma_i}{\gamma_i^2 + \alpha} \right)^2 [(\mathbf{u}_i^T \mathbf{y})^2 + \sigma^2]$$

in the case $\mathbf{L} = \mathbf{I}_n$. The expected residual is an increasing function of the regularization parameter and has a plateau at $(m-n)\sigma^2$, while the expected constraint is a decreasing function with a plateau in the neighborhood of $\|\mathbf{x}^\dagger\|^2$. Roughly speaking, the discrepancy principle seeks to locate that point on the curve where the residual norm changes to a rapidly increasing function of the regularization parameter.

6. Generalized discrepancy principle

In some applications, the discrepancy principle gives too small a regularization parameter, and the solution is undersmoothed. In the generalized version of the discrepancy principle [9,28,23], α is the solution of the equation

$$\|\mathbf{r}_\alpha^\delta\|^2 - \mathbf{r}_\alpha^{\delta T} \hat{\mathbf{A}}_\alpha \mathbf{r}_\alpha^\delta = \tau \Delta^2, \quad \tau > 1.$$

As $\hat{\mathbf{A}}_\alpha$ is positive definite, the left-hand side of this equation is smaller than the residual $\|\mathbf{r}_\alpha^\delta\|^2$ and, therefore, the regularization parameter computed by the gen-

eralized discrepancy principle is larger than that corresponding to the ordinary method. In a semi-stochastic setting, we have

$$\|\mathbf{r}_\alpha^\delta\|^2 - \mathbf{r}_\alpha^{\delta T} \hat{\mathbf{A}}_\alpha \mathbf{r}_\alpha^\delta = \tau m \sigma^2$$

or, explicitly,

$$\sum_{i=1}^m \left(\frac{\alpha}{\gamma_i^2 + \alpha} \right)^3 (\mathbf{u}_i^T \mathbf{y}^\delta)^2 = \tau m \sigma^2.$$

7. Unbiased predictive risk estimator method

The computation of the regularization parameter by analyzing the solution error is not practical, since the exact solution is unknown. Instead, the predictive error can be used as an indicator of the solution fidelity, because it can be accurately estimated in the framework of the unbiased predictive risk estimator method. This approach is also known as the C_L -method or the predictive mean square error method, and was originally developed by Mallows [24]. The predictive risk estimator is defined through the relation

$$\pi_\alpha^\delta = \|\mathbf{r}_\alpha^\delta\|^2 + 2\sigma^2 \text{trace}(\hat{\mathbf{A}}_\alpha) - m\sigma^2,$$

and it can be shown that π_α^δ is an unbiased estimator for the expected value of the predictive error, i.e.,

$$E\{\pi_\alpha^\delta\} = E\{\|\mathbf{p}_\alpha^\delta\|^2\}.$$

The unbiased predictive risk estimator method chooses the regularization parameter as

$$\alpha_{\text{pr}} = \arg \min_{\alpha} \pi_\alpha^\delta,$$

and a computable expression for π_α^δ reads as

$$\pi_\alpha^\delta = \sum_{i=1}^m \left(\frac{\alpha}{\gamma_i^2 + \alpha} \right)^2 (\mathbf{u}_i^T \mathbf{y}^\delta)^2 + 2\sigma^2 \sum_{i=1}^n \frac{\gamma_i^2}{\gamma_i^2 + \alpha} - m\sigma^2.$$

Although both have the same expected values, it does not necessarily follow that π_α^δ and $\|\mathbf{p}_\alpha^\delta\|^2$ have the same minimizers. The analysis performed by Lukas [23] has shown that these minimizers are close provided that these functions do not have flat minima.

The predictive risk estimator possesses a minimum since $\|\mathbf{r}_\alpha^\delta\|^2$ is an increasing function of α and $\text{trace}(\hat{\mathbf{A}}_\alpha)$ is a decreasing function of α . However, this minimum can be very flat, especially when the trace term is very small as compared to the residual term. Note that $\text{trace}(\hat{\mathbf{A}}_\alpha) \rightarrow 0$ as $\alpha \rightarrow \infty$, and in the asymptotic limit of large α , the predictive risk curve is similar to the residual curve:

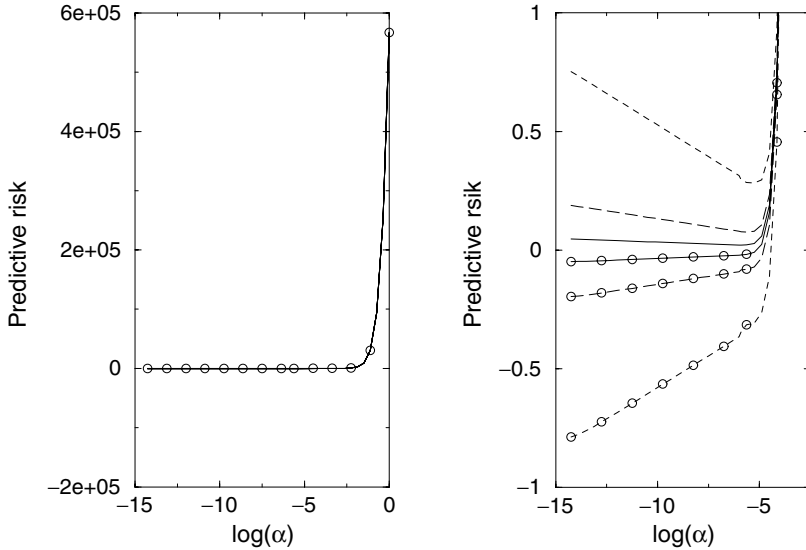


Fig. 2. Expected predictive risk and its approximation. In the left-hand panel, the curve is plotted over the entire domain of variation of the regularization parameter, while in the right-hand panel a reduced domain of variation is considered. The curves correspond to three values of the noise standard deviation: 0.05 (solid), 0.1 (long dashed) and 0.2 (dashed). The approximations are marked with circles.

$$E\{\|\mathbf{p}_\alpha^\delta\|^2\} \approx E\{\|\mathbf{r}_\alpha^\delta\|^2\} - m\sigma^2.$$

In Fig. 2 we show the expected value of the predictive risk along with its asymptotical approximation. The plots illustrate that $E\{\|\mathbf{r}_\alpha^\delta\|^2\} - m\sigma^2$ is a good approximation of $E\{\|\mathbf{p}_\alpha^\delta\|^2\}$ for large values of α (in the rapidly increasing portion of the curve). The expected predictive risk has a flat minimum which moves toward large α with increasing the noise standard deviation, while from the right-hand panel of Fig. 2 it is also apparent that the flatness of the curves becomes more pronounced as σ decreases.

8. Generalized cross-validation

The generalized cross-validation method is an alternative to the unbiased predictive risk estimator method that does not require the knowledge of σ^2 . This method was developed by Wahba [37,38] and is a very popular error-free method for choosing the regularization parameter. The generalized cross-validation function can be derived from the “leave one out” principle. In the “leave one out” cross validation, models that are obtained by leaving one of the m data points out of the inversion process are considered. The regularization parameter is computed as

$$\alpha_{\text{gcv}} = \arg \min_{\alpha} v_{\alpha}^{\delta},$$

where the generalized cross-validation function v_{α}^{δ} is given by

$$v_{\alpha}^{\delta} = \frac{\|\mathbf{r}_{\alpha}^{\delta}\|^2}{[\text{trace}(\mathbf{I}_m - \hat{\mathbf{A}}_{\alpha})]^2}.$$

The generalized cross-validation method seeks to locate the transition point where the residual norm changes from a very slowly varying function of α to a rapidly increasing function of α . But instead of working with the residual norm, the generalized cross-validation method uses the ratio of the residual norm and the degree of freedom for noise, which is a monotonically increasing function of α .

The expected value of the cross-validation function can be expressed as

$$E\{v_{\alpha}^{\delta}\} = \frac{E\{\|\mathbf{p}_{\alpha}^{\delta}\|^2\} - 2\sigma^2 \text{trace}(\hat{\mathbf{A}}_{\alpha}) + m\sigma^2}{[m - \text{trace}(\hat{\mathbf{A}}_{\alpha})]^2},$$

and, since the trace term is small for $m \gg n$, we obtain the approximation

$$E\{v_{\alpha}^{\delta}\} \approx \frac{1}{m^2} E\{\|\mathbf{p}_{\alpha}^{\delta}\|^2\} + \frac{\sigma^2}{m}.$$

Thus, the minimizer of the expected value of the generalized cross-validation function is close to the minimizer of the expected value of the predictive risk. In view of this equivalence, the generalized cross validation may suffer from the same drawback as the unbiased predictive risk estimator method: the unique minimum can be very flat, thus leading to numerical difficulties in computing the regularization parameter. In Fig. 3 we plot the expected generalized cross-validation curve and its approximation. The agreement between the curves is acceptable over the entire domain of variation of α .

9. Maximum likelihood estimation

Based on a Monte Carlo analysis by Thompson et al. [30], it has been observed that the generalized cross-validation function may not have a unique minimum and that the unbiased predictive risk estimator may result in severe undersmoothing. The maximum likelihood estimation is equivalent to the maximization of the marginal likelihood function when Gaussian densities are assumed, and computes the regularization parameter as [3,21]

$$\alpha_{\text{mle}} = \arg \min_{\alpha} \lambda_{\alpha}^{\delta},$$

where $\lambda_{\alpha}^{\delta}$ is the maximum likelihood function defined by

$$\lambda_{\alpha}^{\delta} = \frac{\mathbf{y}^{\delta T} (\mathbf{I}_m - \hat{\mathbf{A}}_{\alpha}) \mathbf{y}^{\delta}}{m \sqrt{\det(\mathbf{I}_m - \hat{\mathbf{A}}_{\alpha})}} = \sum_{i=1}^m \frac{(\mathbf{u}_i^T \mathbf{y}^{\delta})^2}{\gamma_i^2 + \alpha} \bigg/ m \sqrt{\prod_{i=1}^m \frac{1}{\gamma_i^2 + \alpha}}.$$

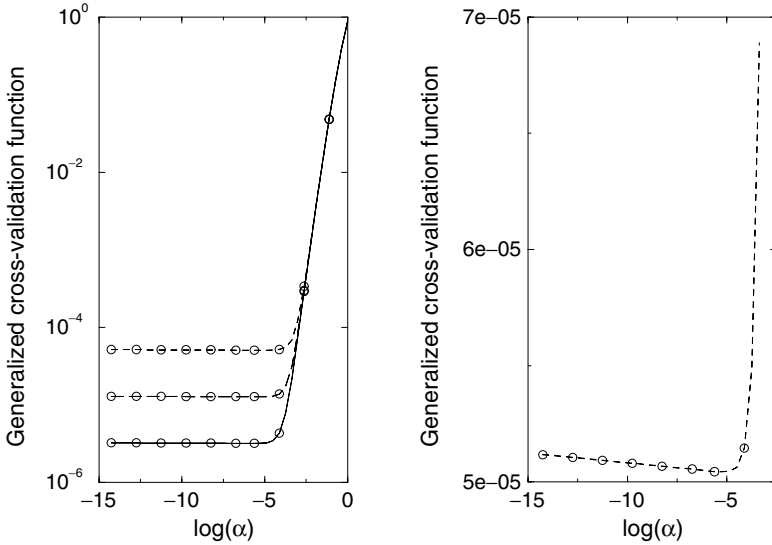


Fig. 3. Expected generalized cross-validation curve and its approximation. In the left-hand panel, the curve is plotted over the entire domain of variation of the regularization parameter, while in the right-hand panel a reduced domain of variation is considered. The curves correspond to three values of the noise standard deviation: 0.05 (solid), 0.1 (long dashed) and 0.2 (dashed). The approximations are marked with circles.

The plots in Fig. 4 show that the maximum likelihood function has a more pronounced minimum than the predictive risk and the generalized cross-validation functions.

10. Quasi-optimality criterion

The quasi-optimality criterion is based on the hypothesis of the existence of a plateau of $\|\mathbf{x}_\alpha^\delta - \mathbf{x}^\dagger\|$ near the optimal regularization parameter, so that α_{qo} is chosen as the minimizer of the function

$$\zeta_\alpha^\delta = \left\| \alpha \frac{d\mathbf{x}_\alpha^\delta}{d\alpha} \right\|^2.$$

This method was pioneered by Tikhonov and Glasko [33] in a slightly different form, and has been considered by numerous authors, especially in the Russian literature [26]. Note that the derivative of the regularized solution can be computed by using the relation

$$\alpha \frac{d\mathbf{x}_\alpha^\delta}{d\alpha} = - \sum_{i=1}^n \frac{\alpha \gamma_i^2}{(\gamma_i^2 + \alpha)^2} \frac{1}{\sigma_i} (\mathbf{u}_i^T \mathbf{y}^\delta) \mathbf{w}_i.$$

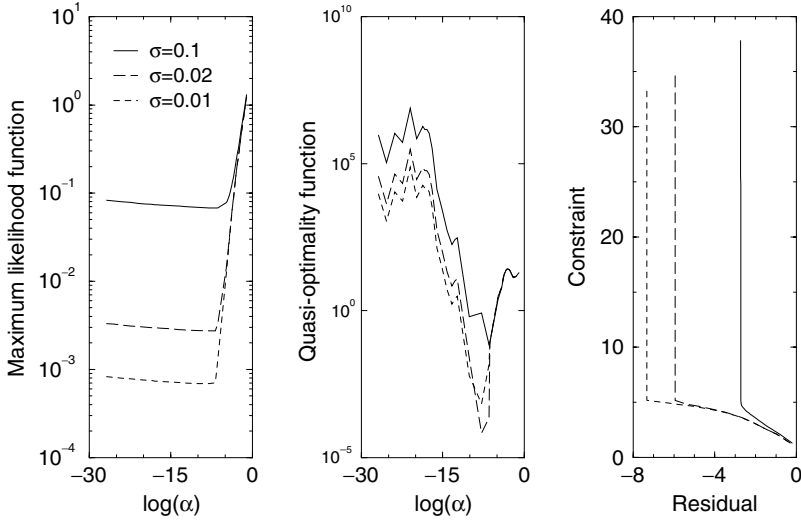


Fig. 4. The maximum likelihood function (left), the quasi-optimality function (middle), and the L-curve (right) for three values of the noise standard deviation.

In practice, the minimization of the quasi-optimality function is complicated because this function has many local minima (see the middle panel of Fig. 4).

11. L-curve method

The L-curve criterion advocated by Hansen [15] is based on the L-curve, which is a parametric plot of the constraint $\|\mathbf{c}_\alpha^\delta\|^2$ against $\|\mathbf{r}_\alpha^\delta\|^2$ on the log-log scale. The corner of the L-curve appears for regularization parameters close to the optimal parameter that balances the smoothing and noise errors. The notion of a corner originates from a purely visual impression, and it is not at all obvious how to translate this impression into mathematics. According to Hansen and O’Leary [17], the corner of the L-curve is the point of maximum curvature. Defining the L-curve components by

$$x(\alpha) = \log(\|\mathbf{r}_\alpha^\delta\|^2), \quad y(\alpha) = \log(\|\mathbf{c}_\alpha^\delta\|^2),$$

we select that value of α that maximizes the curvature function

$$\alpha_{lc} = \arg \max_{\alpha} k_{lc\alpha}^\delta,$$

where

$$k_{lc\alpha}^\delta = \frac{x''(\alpha)y'(\alpha) - x'(\alpha)y''(\alpha)}{[x'(\alpha)^2 + y'(\alpha)^2]^{3/2}}.$$

Table 1. Average values of the relative solution errors (in %) for different regularization parameter choice methods. The noise standard deviation is 0.1.

Regularization parameter choice method	Relative solution error
Optimal regularization parameter	0.14
Expected error estimation method	1.23
Maximum likelihood estimation	1.20
Discrepancy principle	2.19
Generalized discrepancy principle	2.01
Generalized cross-validation	2.68
Quasi-optimality criterion	3.59
L-curve method	3.17

The L-curve is illustrated in the right-hand panel of Fig. 4. The plots show that the corner of the L-curve is more pronounced when the noise standard deviation is small.

11.1. Ozone retrieval test problem

In this section we analyze the performance of the regularization parameter choice methods for an ozone retrieval test problem. The atmospheric ozone profile is retrieved from a sequence of simulated limb spectra in a spectral interval ranging from 323 to 333 nm. The number of limb scans is 11 and the limb tangent height varies between 14 and 49 km in steps of 3.5 km. The atmosphere is discretized with a step of 3.5 km between 0 and 70 km and a step of 5 km between 70 and 100 km. The problem is assumed to be nearly linear in the sense that a linearization of the forward model about the a priori state is appropriate to find a solution. The forward model assumes piecewise constant interpolation for profile representation, while the regularization matrix is the Cholesky factor of a normalized covariance matrix with an altitude-independent correlation length of $l = 3.5$ km.

The exact state vector is chosen as a translated and a scaled version of a climatological profile with a translation distance of 3 km and a scaling factor of 1.3. For a fixed value of the noise standard deviation σ , we compute the exact data vector \mathbf{y} and generate the noisy data vector $\mathbf{y}_i^\delta = \mathbf{y} + \boldsymbol{\delta}_i$, $\boldsymbol{\delta}_i$ being a random sample of the white noise with the normal distribution $\mathbf{N}(\mathbf{0}, \sigma^2 \mathbf{I}_m)$. The number of noisy data realizations is 100, and for each \mathbf{y}_i^δ , we determine the regularization parameter α_i by a particular parameter choice method.

The average values of the solution errors over noisy data realizations are given in Table 1, where the optimal regularization parameter for the noisy data vector \mathbf{y}_i^δ is computed according to

$$\alpha_{\text{opt}i} = \arg \min_{\alpha} \|\mathbf{K}_{\alpha}^{\dagger} \mathbf{y}_i^{\delta} - \mathbf{x}^{\dagger}\|.$$

The results show that for this application, the smallest reconstruction errors correspond to the maximum likelihood estimation and the expected error estimation methods.

12. Tikhonov regularization for nonlinear problems

The majority of inverse problems arising in atmospheric remote sensing are nonlinear. In this section, we discuss the practical aspects of the Tikhonov regularization for solving the nonlinear problem

$$\mathbf{F}(\mathbf{x}) = \mathbf{y},$$

where $\mathbf{F}(\mathbf{x})$ is the forward model. As in the linear case, the nonlinear system of equations is called a discrete ill-posed problem because the underlying continuous problem is ill-posed.

Nonlinear problems are treated in the same framework as linear problems. The right-hand side \mathbf{y} is supposed to be contaminated by measurements errors: $\mathbf{y}^{\delta} = \mathbf{y} + \boldsymbol{\delta}$. In a deterministic setting, the data error is characterized by the noise level Δ , while in a semi-stochastic setting, $\boldsymbol{\delta}$ is assumed to be a discrete white noise with the covariance matrix $\mathbf{C}_{\delta} = \sigma^2 \mathbf{I}_m$.

The formulation of the Tikhonov regularization for nonlinear problems is straightforward: we minimize the objective function

$$F_{\alpha}(\mathbf{x}) = \frac{1}{2} [\|\mathbf{y}^{\delta} - \mathbf{F}(\mathbf{x})\|^2 + \alpha \|\mathbf{L}(\mathbf{x} - \mathbf{x}_a)\|^2],$$

where \mathbf{x}_a is the *a priori* state vector, i.e., the best beforehand estimate of \mathbf{x}^{\dagger} . For a positive regularization parameter, minimizers of the Tikhonov function always exist, but are not unique.

12.1. Minimization of the Tikhonov function

The minimization of the Tikhonov function can be formulated as a least squares problem, and the regularized solution can be computed by using optimization methods for unconstrained minimization problems. For this purpose, “step-length-based-methods” or “trust-region methods” can be employed [4,10]. Nonlinear optimization methods are iterative methods which compute the new iterate $\mathbf{x}_{\alpha k+1}^{\delta}$ by approximating the objective function around the actual iterate $\mathbf{x}_{\alpha k}^{\delta}$ by a quadratic model and by imposing the so-called descent condition

$$F_{\alpha}(\mathbf{x}_{\alpha k+1}^{\delta}) < F_{\alpha}(\mathbf{x}_{\alpha k}^{\delta}).$$

The Gauss–Newton method for least squares problems belongs to the category of step-length methods and has an important practical interpretation, which we now discuss. At the iteration step k , let us consider a linearization of the forward model about the actual iterate $\mathbf{x}_{\alpha k}^{\delta}$, that is,

$$\mathbf{F}(\mathbf{x}) \approx \mathbf{F}(\mathbf{x}_{ak}^\delta) + \mathbf{K}_{ak}(\mathbf{x} - \mathbf{x}_{ak}^\delta),$$

and let us replace the nonlinear equation $\mathbf{F}(\mathbf{x}) = \mathbf{y}^\delta$ by its linearization

$$\mathbf{K}_{ak}(\mathbf{x} - \mathbf{x}_a) = \mathbf{y}_k^\delta,$$

where

$$\mathbf{y}_k^\delta = \mathbf{y}^\delta - \mathbf{F}(\mathbf{x}_{ak}^\delta) + \mathbf{K}_{ak}(\mathbf{x}_{ak}^\delta - \mathbf{x}).$$

Because the nonlinear problem is ill-posed, its linearization is also ill-posed, and we solve the linearized equation by means of the Tikhonov regularization. The Tikhonov function for the linearized equation takes the form

$$F_{l\alpha k}(\mathbf{x}) = \frac{1}{2} [\|\mathbf{y}_k^\delta - \mathbf{K}_{ak}(\mathbf{x} - \mathbf{x}_a)\|^2 + \alpha \|\mathbf{L}(\mathbf{x} - \mathbf{x}_a)\|^2],$$

and its minimizer

$$\mathbf{x}_{ak+1}^\delta = \mathbf{x}_a + \mathbf{K}_{ak}^\dagger \mathbf{y}_k^\delta,$$

with

$$\mathbf{K}_{ak}^\dagger = (\mathbf{K}_{ak}^T \mathbf{K}_{ak} + \alpha \mathbf{L}^T \mathbf{L})^{-1} \mathbf{K}_{ak}^T$$

being the generalized inverse at the iteration step k , is exactly the new iterate computed in the framework of the Gauss–Newton method (with a unit step length). Thus, the solution of a nonlinear ill-posed problem by means of the Tikhonov regularization is equivalent to the solution of a sequence of ill-posed linearizations of the forward model about the current iterate.

12.2. Parameter choice methods

As for linear problems, the choice of the regularization parameter plays an important role in computing a reliable approximation of the solution. In this section, we extend the expected error estimation method to the nonlinear case and present selection criteria with variable and constant regularization parameters. In the first case, the regularization parameter is estimated at each iteration step, while in the second case, the minimization of the Tikhonov function is repeated a few times with different regularization parameters.

As we will see, the regularization parameter choice methods for nonlinear problems are natural extensions of the selection criteria for linear problems. This fact is a consequence of the equivalence between the error representations in the nonlinear and linear cases. Indeed, assuming that the sequence of iterates $\{\mathbf{x}_{ak}^\delta\}$ converges to \mathbf{x}_α^δ and that \mathbf{F} is continuously differentiable implies that

$$\mathbf{e}_\alpha^\delta = \mathbf{e}_{s\alpha} + \mathbf{e}_{n\alpha} + \mathbf{e}_{l\alpha}^\delta,$$

where $\mathbf{e}_\alpha^\delta = \mathbf{x}^\dagger - \mathbf{x}_\alpha^\delta$ is the total error in the solution, $\mathbf{e}_{s\alpha} = (\mathbf{I}_n - \mathbf{A}_\alpha)(\mathbf{x}^\dagger - \mathbf{x}_\alpha)$ is the smoothing error, $\mathbf{e}_{n\alpha}^\delta = -\mathbf{K}_\alpha^\dagger \boldsymbol{\delta}$ is the noise error, and $\mathbf{e}_{l\alpha}^\delta = -\mathbf{K}_\alpha^\dagger \mathbf{R}(\mathbf{x}^\dagger, \mathbf{x}_\alpha^\delta)$ is

the nonlinearity error. In the above relations, $\mathbf{R}(\mathbf{x}^\dagger, \mathbf{x}_\alpha^\delta)$ is the remainder term of the first-order Taylor expansion of the forward model about \mathbf{x}_α^δ , while $\mathbf{K}_\alpha^\dagger$ and \mathbf{A}_α are the generalized inverse and the averaging kernel matrix evaluated at \mathbf{x}_α^δ . Thus, the error representations in the nonlinear and linear cases are similar, except for an additional term representing the nonlinearity error. If the minimizer \mathbf{x}_α^δ is sufficiently close to the exact solution \mathbf{x}^\dagger , the nonlinearity error can be neglected, and the agreement is complete.

13. *A priori* parameter choice methods

In the linear case, the expected error estimation method has been formulated as an *a priori* parameter selection criterion. The idea was to perform a random exploration of a domain in which the solution is supposed to lie and, for each state vector realization \mathbf{x}_i^\dagger , compute the optimal regularization parameter for error estimation $\bar{\alpha}_{\text{opt}i} = \arg \min_\alpha E\{\|\mathbf{e}_\alpha^\delta(\mathbf{x}_i^\dagger)\|^2\}$ and the exponent $p_i = \log \bar{\alpha}_{\text{opt}i} / \log \sigma$. The regularization parameter is then chosen as $\alpha_c = \sigma^{\bar{p}}$, where $\bar{p} = (1/N) \sum_{i=1}^N p_i$ is the sample mean exponent.

The expected error estimation method for nonlinear problems can be formulated by representing the expected error at the solution as

$$E\{\|\mathbf{e}_\alpha^\delta\|^2\} = \|\mathbf{e}_{s\alpha}\|^2 + E\{\|\mathbf{e}_{n\alpha}^\delta\|^2\},$$

where $\|\mathbf{e}_{s\alpha}\|^2$ and $E\{\|\mathbf{e}_{n\alpha}^\delta\|^2\}$ are expressed in terms of the generalized singular values of the matrix pair $(\mathbf{K}_\alpha, \mathbf{L})$. The Jacobian matrix \mathbf{K}_α is evaluated at the solution (the generalized singular system depends on α), and, as a consequence, the optimal regularization parameter for error estimation has to be computed for each state vector realization by repeatedly solving the nonlinear minimization problem. The resulting algorithm is extremely computationally expensive, and in order to ameliorate this drawback we approximate the Jacobian matrix at the solution by the Jacobian matrix at the *a priori* state. The *a priori* parameter choice method is then equivalent to the expected error estimation method applied to a linearization of the forward model about the *a priori* state.

14. Selection criteria with variable regularization parameters

As the solution of a nonlinear ill-posed problem by means of the Tikhonov regularization is equivalent to the solution of a sequence of ill-posed linearizations of the forward model, parameter choice methods for linear problems can be used to compute the regularization parameter at each iteration step.

The errors in the right-hand side of the linearized equation are due to the instrumental noise and the linearization error. At an iteration step k , the noise level can be estimated by the minimum value of the linearized residual corresponding to the smallest singular value $\|\mathbf{r}_{1\text{min}k}^\delta\|$. A heuristic version of the discrepancy principle then takes the following form: at the iteration step k , compute the regularization parameter as the solution of the equation

$$\|\mathbf{r}_{lck}^\delta\|^2 = \tau \|\mathbf{r}_{lmin k}^\delta\|^2, \quad \tau > 1,$$

where $\mathbf{r}_{lck}^\delta = (\mathbf{I}_m - \hat{\mathbf{A}}_{ck}) \mathbf{y}_k^\delta$ is the linearized residual vector and $\hat{\mathbf{A}}_{ck}$ is the influence matrix.

Due to the difficulties associated with the data error estimation, error-free parameter choice methods (based only on information about the noisy data) are more attractive. In this regard, we mention that the generalized cross validation has been applied to the linearized problem by Haber [12], Haber and Oldenburg [13], and Farquharson and Oldenburg [8]. Selection of the regularization parameter by using the L-curve criterion has been reported by Li and Oldenburg [22], and Hasekamp and Landgraf [19]. For the linearized equation, the following parameter choice methods can be considered:

1. the generalized cross-validation method, $\alpha_{gcvk} = \arg \min_{\alpha} v_{ck}^\delta$, with

$$v_{ck}^\delta = \frac{\|\mathbf{r}_{lck}^\delta\|^2}{[\text{trace}(\mathbf{I}_m - \hat{\mathbf{A}}_{ck})]^2};$$

2. the maximum likelihood estimation, $\alpha_{mlek} = \arg \min_{\alpha} \lambda_{ck}^\delta$, with

$$\lambda_{ck}^\delta = \frac{\mathbf{y}_k^{\delta T} (\mathbf{I}_m - \hat{\mathbf{A}}_{ck}) \mathbf{y}_k^\delta}{m \sqrt{\det(\mathbf{I}_m - \hat{\mathbf{A}}_{ck})}};$$

3. the L-curve method, $\alpha_{lck} = \arg \max_{\alpha} \mathbf{k}_{lck}^\delta$, with

$$\mathbf{k}_{lck}^\delta = \frac{x_k''(\alpha) y_k'(\alpha) - x_k'(\alpha) y_k''(\alpha)}{[x_k'(\alpha)^2 + y_k'(\alpha)^2]^{3/2}},$$

$$x_k(\alpha) = \log(\|\mathbf{r}_{lck}^\delta\|^2), \quad y_k(\alpha) = \log(\|\mathbf{c}_{ck}^\delta\|^2), \quad \text{and} \quad \mathbf{c}_{ck}^\delta = \mathbf{L} \mathbf{K}_{ck}^\dagger \mathbf{y}_k^\delta.$$

In practice, the following recommendations for choosing the regularization parameter should be taken into account:

1. in the beginning of the iterative process, large α -values should be used to avoid local minima and to get well-conditioned least squares problems to solve;
2. during the iteration, the regularization parameter should be decreased slowly in order to achieve a stable solution.

Numerical experiments have shown that a brute-force use of the regularization parameter computed with one of the above selection criteria may lead to an oscillatory sequence of α -values. A heuristic formula that addresses this problem has been proposed by Eriksson [7]: at the iteration step k , the regularization parameter α_k is the weighted sum of the previous regularization parameter α_{k-1} and the proposed regularization parameter α , that is,

$$\alpha_k = \begin{cases} \xi \alpha_{k-1} + (1 - \xi) \alpha, & \alpha < \alpha_{k-1}, \\ \alpha_{k-1}, & \alpha \geq \alpha_{k-1}. \end{cases}$$

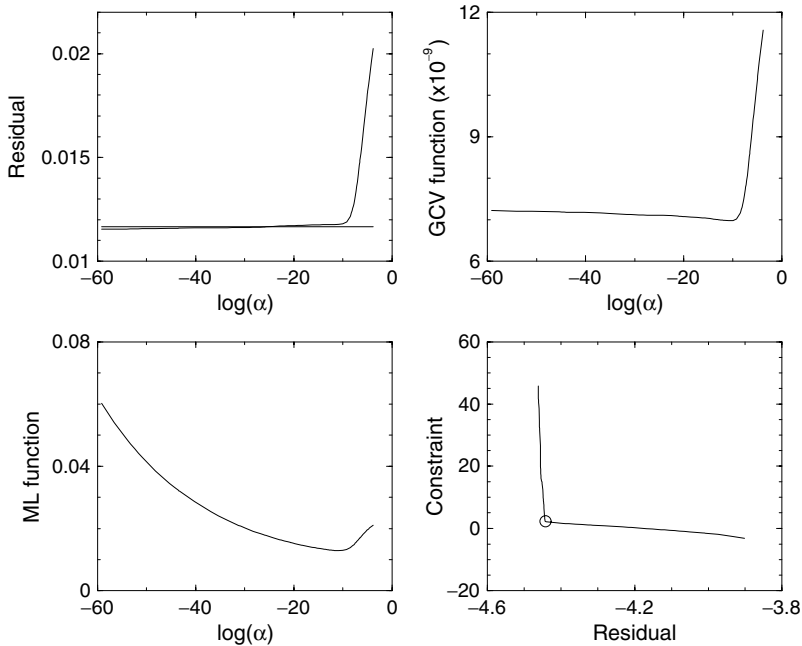


Fig. 5. The residual curve, the generalized cross-validation function, the maximum likelihood function, and the L-curve for the ozone retrieval test problem. The curves are computed at the first iteration.

The residual and the L-curves as well as the generalized cross-validation and maximum likelihood functions are shown in Fig. 5. The curves exhibit the same behaviors as in the linear case: the generalized cross-validation function has a flat minimum, the maximum likelihood function has a distinct minimum and the L-curve has a sharp corner.

To examine the efficiency of the Tikhonov regularization, we consider two limb retrieval test problems: the ozone retrieval in a spectral interval ranging from 520 to 580 nm and the BrO retrieval in a spectral interval ranging from 337 to 357 nm. The limb tangent height varies between 13.6 and 49.9 km in steps of 3.3 km. The atmosphere is discretized in 1.75-km steps between 0 and 42 km, a step of 3.5 km between 42 and 70 km, and a step of 10 km between 70 and 100 km. The total number of levels is 36, and the spectral resolution is 0.25 nm.

The solution errors listed in Table 2 show that the Tikhonov regularization with variable regularization parameter yields accurate results, and that the maximum likelihood estimation is superior to the other regularization parameter choice methods.

Table 2. Relative solution errors (in %) for Tikhonov regularization with variable regularization parameters corresponding to the following selection criteria: the discrepancy principle (DP), the maximum likelihood estimation (MLE), generalized cross-validation (GCV), and the L-curve (LC) method.

Ozone				BrO			
DP	MLE	GCV	LC	DP	MLE	GCV	LC
6.01	5.24	5.37	5.64	6.31	6.26	6.28	6.22

15. Selection criteria with constant regularization parameters

The numerical realization of these parameter choice methods necessitates the solution of the nonlinear minimization problem several times for different regularization parameters. Each minimization is solved with a regularization parameter α yielding a solution \mathbf{x}_α^δ .

In the framework of the discrepancy principle, the regularization parameter α solves the equation

$$\|\mathbf{y}^\delta - \mathbf{F}(\mathbf{x}_\alpha^\delta)\|^2 = \tau \Delta^2, \quad \tau > 1.$$

Because, for nonlinear problems, the discrepancy principle equation only has a solution under very strong restrictive assumptions, we use a simplified version of this selection criterion, as follows: if $\{\alpha_i\}$ is a decreasing sequence of regularization parameters, we choose the largest α_{i^*} such that the residual norm is below the noise level, i.e.,

$$\|\mathbf{y}^\delta - \mathbf{F}(\mathbf{x}_{\alpha_{i^*}}^\delta)\|^2 \leq \tau \Delta^2 < \|\mathbf{y}^\delta - \mathbf{F}(\mathbf{x}_{\alpha_i}^\delta)\|^2, \quad 0 \leq i < i^*.$$

Error-free methods with constant regularization parameter are natural extensions of the corresponding selection criteria for linear problems. For example, the generalized cross-validation and the maximum likelihood functions take the forms

$$v_\alpha^\delta = \frac{\|\mathbf{r}_\alpha^\delta\|^2}{[\text{trace}(\mathbf{I}_m - \hat{\mathbf{A}}_\alpha)]^2}$$

and

$$\lambda_\alpha^\delta = \frac{\mathbf{y}_\alpha^{\delta T} (\mathbf{I}_m - \hat{\mathbf{A}}_\alpha) \mathbf{y}_\alpha^\delta}{m \sqrt{\det(\mathbf{I}_m - \hat{\mathbf{A}}_\alpha)}},$$

with $\mathbf{y}_\alpha^\delta = \mathbf{y}^\delta - \mathbf{F}(\mathbf{x}_\alpha^\delta) + \mathbf{K}_\alpha(\mathbf{x}_\alpha^\delta - \mathbf{x}_a)$, respectively. Application of the generalized cross validation in conjunction with the Tikhonov regularization method for solving a temperature retrieval problem and an inverse scattering problem have been reported by O'Sullivan and Wahba [27] and Vogel [35], respectively.

Table 3. Relative solution errors (in %) for Tikhonov regularization with a constant regularization parameter corresponding to the discrepancy principle (DP), the maximum likelihood estimation (MLE), generalized cross-validation (GCV), and the L-curve (LC) method.

Ozone				BrO			
DP	MLE	GCV	LC	DP	MLE	GCV	LC
5.99	5.29	5.36	17.32	6.61	6.61	6.61	6.64

The use of the L-curve for nonlinear problems has been suggested by Eriksson [7]. The nonlinear L-curve is the plot of the constraint $\|\mathbf{c}_\alpha^\delta\| = \|\mathbf{L}(\mathbf{x}_\alpha^\delta - \mathbf{x}_a)\|$ against the residual $\|\mathbf{r}_\alpha^\delta\| = \|\mathbf{y}^\delta - \mathbf{F}(\mathbf{x}_\alpha^\delta)\|$ for a range of values of the regularization parameter α . This curve is monotonically decreasing and convex [11]. In a computational sense, the nonlinear L-curve consists of a number of discrete points corresponding to the different values of the regularization parameter, and in practice, the following techniques can be used for choosing the regularization parameter:

1. fit a cubic spline curve to the discrete points of the L-curve $\{x(\alpha_i), y(\alpha_i)\}$, with $x(\alpha) = \log(\|\mathbf{r}_\alpha^\delta\|^2)$ and $y(\alpha) = \log(\|\mathbf{c}_\alpha^\delta\|^2)$, and determine the point on the original discrete curve that is closest to the spline curve's corner;
2. in the framework of the minimum distance function approach [2], compute $\alpha_{lc} = \arg \min_i d(\alpha_i)^2$ for the distance function

$$d(\alpha)^2 = [x(\alpha) - x_0]^2 + [y(\alpha) - y_0]^2,$$

with $x_0 = \min_i x(\alpha_i)$ and $y_0 = \min_i y(\alpha_i)$;

3. detect the minimum of the logarithmic L-curve rotated by $\pi/4$ radians,

$$\alpha_{lc} = \arg \min_i [x(\alpha_i) + y(\alpha_i)].$$

The curves corresponding to the nonlinear parameter choice methods with a constant regularization parameter are illustrated in Fig. 6. The plots show that the maximum likelihood function has a sharper minimum than the generalized cross-validation function, and that the L-curve corner is not distinctive.

The solution errors listed in Table 3 indicate that the best results correspond to the maximum likelihood estimation, while the worst results correspond to the L-curve method. Especially noteworthy is the failure of the L-curve method in application to the ozone retrieval test problem: the predicted value of the regularization parameter is considerably larger than the optimal value, and the retrieved profile is close to the *a priori* profile.

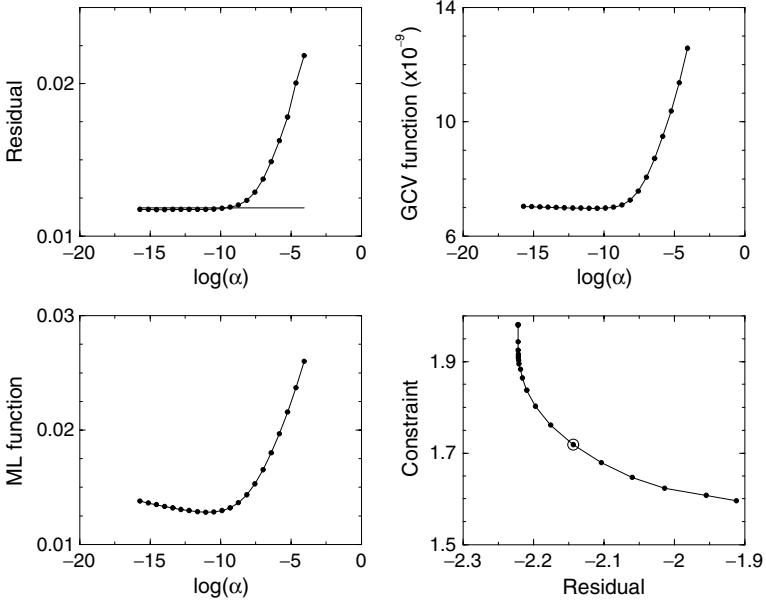


Fig. 6. The nonlinear residual curve, the generalized cross-validation function, the maximum likelihood function, and the L-curve for the ozone retrieval test problem.

16. Iterative regularization methods

Numerical experience shows that the Tikhonov function usually has many local minima, and a descent method for solving the optimization problem may tend to get stuck, especially for severely ill-posed problems.

For iterative regularization methods, the number of iteration steps k plays the role of the regularization parameter, and the iterative process has to be stopped after an appropriate number of steps k^* in order to avoid uncontrollable amplification of the noise error. A widely used *a posteriori* choice for the stopping index k^* depending on the noise level and noisy data vector is the discrepancy principle, i.e., the iterative process is stopped after k^* steps such that

$$\|\mathbf{y}^\delta - \mathbf{F}(\mathbf{x}_{k^*}^\delta)\|^2 \leq \tau \Delta^2 < \|\mathbf{y}^\delta - \mathbf{F}(\mathbf{x}_k^\delta)\|^2, \quad 0 \leq k < k^*,$$

with $\tau > 1$ chosen sufficiently large.

In this section, we address the practical aspects of Newton-type methods, such as the iteratively regularized Gauss–Newton method [1] and the regularizing Levenberg–Marquardt method [14].

16.1. Iteratively regularized Gauss–Newton method

The iteratively regularized Gauss–Newton method relies on the solution of the linearized equation

$$\mathbf{K}_k(\mathbf{x} - \mathbf{x}_a) = \mathbf{y}_k^\delta$$

by means of the Tikhonov regularization with the penalty term $\|\mathbf{L}(\mathbf{x} - \mathbf{x}_a)\|^2$ and the regularization parameter α_k . The new iterate minimizes the function

$$F_k(\mathbf{x}) = \frac{1}{2} [\|\mathbf{y}_k^\delta - \mathbf{K}_k(\mathbf{x} - \mathbf{x}_a)\|^2 + \alpha_k \|\mathbf{L}(\mathbf{x} - \mathbf{x}_a)\|^2]$$

and is given by

$$\mathbf{x}_{k+1}^\delta = \mathbf{x}_a + \mathbf{K}_k^\dagger \mathbf{y}_k^\delta,$$

where $\mathbf{K}_k^\dagger = (\mathbf{K}_k^T \mathbf{K}_k + \alpha_k \mathbf{L}^T \mathbf{L})^{-1} \mathbf{K}_k^T$ is the regularized generalized inverse at the iteration step k .

At first glance, this method seems to be identical to the Tikhonov regularization method with a variable regularization parameter, but the following differences do exist:

- the regularization parameters are the terms of a decreasing sequence satisfying the requirements

$$\alpha_k > 0, \quad 1 < \frac{\alpha_k}{\alpha_{k+1}} \leq c, \quad \lim_{k \rightarrow \infty} \alpha_k = 0;$$

- the iterative process is stopped according to the discrepancy principle instead of requiring the convergence of iterates and employing the discrepancy principle as an *a posteriori* parameter choice method.

Several strategies for selecting the regularization parameters α_k can be considered. One may use the selection criterion

$$\alpha_k = q_k \alpha_{k-1},$$

where q_k can be chosen as the ratio of a geometric sequence, i.e., $q_k = q < 1$ is constant, or as

$$q_k = \frac{\tau \Delta^2}{\|\mathbf{r}_k^\delta\|^2} \quad (5)$$

and

$$q_k = 1 - \frac{\tau \Delta^2}{\|\mathbf{r}_k^\delta\|^2}. \quad (6)$$

With the choice (5) the regularization parameter decreases very quickly at the beginning of iteration, while the scheme (6) permits sufficient regularization to be applied at the beginning of iteration and then to be gradually decreased.

Any iterative method using the discrepancy principle as stopping rule requires the knowledge of the noise level or its statistical estimate. Because in many practical problems arising in atmospheric remote sensing data errors cannot be estimated (due to the forward model errors), we propose the following stopping rules:

1. For a geometric sequence of regularization parameters, store all iterates \mathbf{x}_k^δ and require the convergence of the nonlinear residuals $\|\mathbf{r}_k^\delta\|$ within a prescribed tolerance. If $\|\mathbf{r}^\delta\|$ is the residual at the last iteration step, choose the solution $\mathbf{x}_{k^*}^\delta$, with k^* being given by

$$\|\mathbf{r}_{k^*}^\delta\|^2 \leq \tau \|\mathbf{r}^\delta\|^2 < \|\mathbf{r}_k^\delta\|^2, \quad 0 \leq k < k^*, \quad \tau > 1.$$

2. For the selection rules (5) and (6), first estimate the noise level. For this purpose, minimize the sum of squares

$$F(\mathbf{x}) = \frac{1}{2} \|\mathbf{y}^\delta - \mathbf{F}(\mathbf{x})\|^2$$

by requiring relative function convergence, compute the equivalent noise variance

$$\sigma_c^2 = \frac{1}{m-n} \|\mathbf{r}^\delta\|^2,$$

where $\|\mathbf{r}^\delta\|$ is the residual at the last iteration step, and then set $\Delta^2 = m\sigma_c^2$.

The above heuristic stopping rules do not have any mathematical justification but work sufficiently well in practice.

In Fig. 7 we illustrate the solution errors for the iteratively regularized Gauss–Newton method and the Tikhonov regularization. In the iteratively regularized Gauss–Newton method, the exponent p characterizes the initial value of the regularization parameter, $\alpha_0 = \sigma^p$, while at all subsequent iteration steps, the regularization parameters are the terms of a geometric sequence with the ratio $q = 0.8$. For the Tikhonov regularization, the error curves possess a minimum, and the minimizers of the solution errors are the optimal values of the regularization parameters. The plots show that the iteratively regularized Gauss–Newton method still yields reliable results for small values of the exponent p or, equivalently, for large initial values of the regularization parameter. Evidently, stronger regularization at the beginning of the iterative process requires a larger number of iteration steps, as can be seen in the right panels of Fig. 7. The main conclusion that one can derive from this numerical simulation is that the iteratively regularized Gauss–Newton method is more stable than the Tikhonov regularization with respect to overestimations of the regularization parameter.

16.2. Regularizing Levenberg–Marquardt method

In the regularizing Levenberg–Marquardt method, the linearized equation

$$\mathbf{K}_k \mathbf{p} = \mathbf{r}_k^\delta,$$

where $\mathbf{p} = \mathbf{x} - \mathbf{x}_k^\delta$ is the step or the search direction and $\mathbf{r}_k^\delta = \mathbf{y}^\delta - \mathbf{F}(\mathbf{x}_k^\delta)$ is the nonlinear residual, is solved by means of the Tikhonov regularization with the

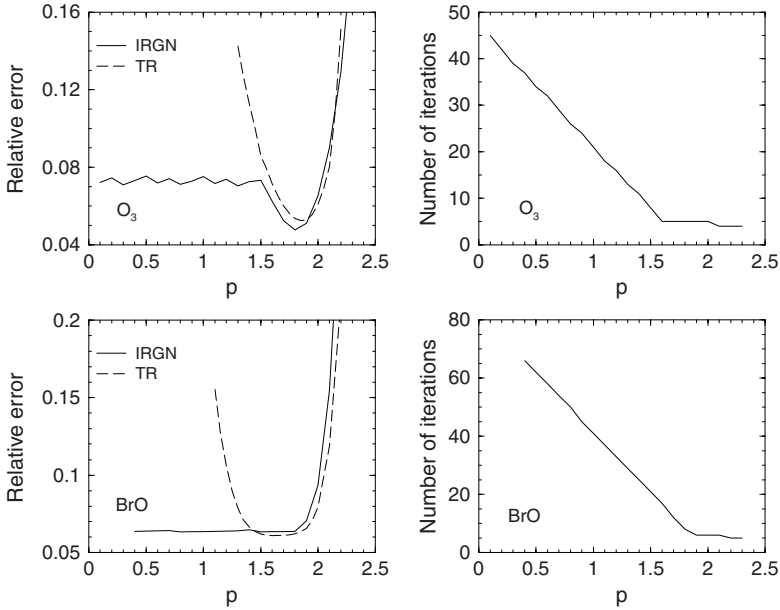


Fig. 7. Relative solution errors and the number of iteration steps for different values of the exponent p . The results are computed with the iteratively regularized Gauss–Newton (IRGN) method and Tikhonov regularization (TR).

penalty term $\|\mathbf{Lp}\|^2$ and the regularization parameter α_k . The Newton step minimizing the Tikhonov function

$$F_{lk}(\mathbf{p}) = \frac{1}{2} [\|\mathbf{r}_k^\delta - \mathbf{K}_k \mathbf{p}\|^2 + \alpha_k \|\mathbf{Lp}\|^2]$$

is given by

$$\mathbf{p}_k^\delta = \mathbf{K}_k^\dagger \mathbf{r}_k^\delta,$$

and the new iterate is taken as

$$\mathbf{x}_{k+1}^\delta = \mathbf{x}_k^\delta + \mathbf{p}_k^\delta.$$

Evidently, the difference from the iteratively regularized Gauss–Newton method consists of the penalty term which now depends on the previous iterate rather than on the *a priori* state.

The parameter choice rule $\alpha_k = q_k \alpha_{k-1}$ with $q_k < 1$, designed for the iteratively regularized Gauss–Newton method, can be used for the regularizing Levenberg–Marquardt method as well. Otherwise, the regularization parameter can be selected by applying the discrepancy principle to the linearized equation: if $\mathbf{p}_{\alpha k}^\delta = \mathbf{K}_{\alpha k}^\dagger \mathbf{r}_k^\delta$, with

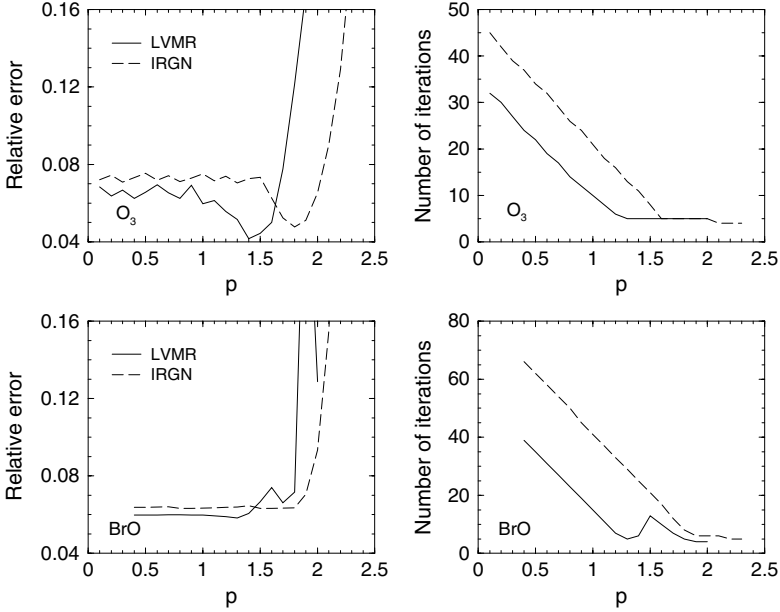


Fig. 8. Relative solution errors and the number of iteration steps for different values of the exponent p . The results correspond to the regularizing Levenberg–Marquardt (LVMR) method and the iteratively regularized Gauss–Newton (IRGN) method.

$$\mathbf{K}_{\alpha k}^{\dagger} = (\mathbf{K}_k^T \mathbf{K}_k + \alpha \mathbf{L}^T \mathbf{L})^{-1} \mathbf{K}_k^T$$

denoting the minimizer of the Tikhonov function for an arbitrary α , then the Levenberg–Marquardt parameter α_k is chosen as the solution of the “discrepancy principle” equation

$$\|\mathbf{r}_k^{\delta} - \mathbf{K}_k \mathbf{p}_{\alpha k}^{\delta}\|^2 = \theta \|\mathbf{r}_k^{\delta}\|^2, \quad 0 < \theta < 1,$$

and the Newton step is given by $\mathbf{p}_k^{\delta} = \mathbf{p}_{\alpha_k k}^{\delta}$.

The regularizing Levenberg–Marquardt method is also insensitive to overestimations of the regularization parameter. The results in Fig. 8 show that the regularizing Levenberg–Marquardt method is superior to the iteratively regularized Gauss–Newton method: for large initial values of the regularization parameter, the number of iteration steps as well as the solution errors are smaller.

17. Conclusions

In this chapter we discussed numerical regularization methods for solving ill-posed problems. The efficiency of the Tikhonov regularization method is affected by the selection of the regularization parameter. When reliable information about

the instrumental noise is available, it is important to make use of this information, and this is the gist of the discrepancy principle and related methods. When no specific information about the instrumental noise is available or when forward model errors are present, the error-free parameter choice methods are a viable alternative. In actuality, there is no infallible regularization parameter choice method. This is because:

- the expected error estimation method requires the knowledge of a physically meaningful solution domain and is time-consuming;
- the discrepancy principle and its generalized version are sensitive to the selection of the control parameter τ ;
- the predictive risk, the generalized cross-validation, and sometimes the maximum likelihood functions may have very flat minima;
- the quasi-optimality function has several local minima and sometimes does not have a global minimum at all;
- the L-curve may lose its L-shape.

As compared to the Tikhonov regularization method, iterative methods are insensitive to overestimations of the regularization parameter. However, they require the knowledge of the noise level or of an appropriate estimate of the systematic and instrumental errors. The latter requirement necessitates an additional computational step or the storage of all iterates.

In this regard, it is advantageous to design a regularization tool incorporating various methods and parameter selection criteria and to select an appropriate solution based on the output of all these strategies. The regularization tool DRACULA (aDvanced Retrieval of the Atmosphere with Constrained and Unconstrained Least squares Algorithms) which we developed at the German Aerospace Center addresses this purpose. With this software package we have been able to retrieve atmospheric state parameters, e.g., temperature or constituent concentration, from a variety of atmospheric sounding instruments such as, for example, the instruments SCIAMACHY and MIPAS on ESA's environmental remote sensing satellite ENVISAT, and the spectrometer instruments GOME-2 and IASI on EUMETSAT's MetOp operational meteorological satellite.

References

1. Bakushinsky, A. B., 1992: The problem of the convergence of the iteratively regularized Gauss–Newton method. *Comput. Math. Phys.* **32**, 1353–1359.
2. Belge, M., M. E. Kilmer, and E. L. Miller, 2002: Efficient determination of multiple regularization parameters in a generalized L-curve framework. *Inverse Probl.* **18**, 1161–1183.
3. Demoment, G., 1989: Image reconstruction and restoration: Overview of common estimation problems. *IEEE Trans. Acoust. Speech Signal Process.* **37**, 2024–2036.
4. Dennis, J. E. Jr., and R. B. Schnabel, 1996: *Numerical Methods for Unconstrained Optimization and Nonlinear Equations* (Prentice-Hall, Englewood Cliffs, NJ).

5. Doicu, A., T. Trautmann, and F. Schreier, 2010: *Numerical Regularization for Atmospheric Inverse Problems* (Springer, Heidelberg).
6. Engl, H. W., M. Hanke, and A. Neubauer, 2000: *Regularization of Inverse Problems* (Kluwer, Dordrecht).
7. Eriksson, J., 1996: Optimization and regularization of nonlinear least square problems. Ph.D. dissertation (Umea University, Sweden).
8. Farquharson, C. G., and D. W. Oldenburg, 2004: A comparison of automatic techniques for estimating the regularization parameter in nonlinear inverse problems. *Geophys. J. Int.* **156**, 411–425.
9. Gfrerer, H., 1987: An a posteriori parameter choice method for ordinary and iterated Tikhonov regularization of ill-posed problems leading to optimal convergence rates. *Math. Comput.* **49**, 507–522.
10. Gill, P. E., W. Murray, and M. H. Wright, 1981: *Practical Optimization* (Academic Press, London).
11. Gulliksson, M. E., and P. A. Wedin, 1999: Optimization tools for inverse problems using the nonlinear L-curve and A-curve. In *Proceedings of the 3rd ASME International Conference on Inverse Problems in Engineering* (Port Ludlow, Washington), pp. 1–6.
12. Haber, E., 1997: Numerical Strategies for the Solution of Inverse Problems. Ph.D. Thesis (The University of British Columbia, Vancouver, Canada).
13. Haber, E., and D. Oldenburg, 2000: A GCV based method for nonlinear ill-posed problems. *Comput. Geosci.* **4**, 41–63.
14. Hanke, M., 1997: A regularizing Levenberg–Marquardt scheme, with applications to inverse groundwater filtration problems. *Inverse Probl.* **13**, 79–95.
15. Hansen, P. C., 1992: Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Rev.* **34**, 561–580.
16. Hansen, P. C., 1998: *Rank Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion* (SIAM, Philadelphia, PA).
17. Hansen, P. C., and D. P. O’Leary, 1993: The use of the L-curve in the regularization of discrete ill-posed problems. *SIAM J. Sci. Comput.* **14**, 1487–1503.
18. Hanson, R. J., 1971: A numerical method for solving Fredholm integral equations of the first kind using singular values. *SIAM J. Numer. Anal.* **8**, 616–622.
19. Hasekamp, O., and J. Landgraf, 2001: Ozone profile retrieval from backscattered ultraviolet radiances: the inverse problem solved by regularization. *J. Geophys. Res.* **106**, 8077–8088.
20. Kaltenbacher, B., A. Neubauer, and O. Scherzer, 2008: *Iterative Regularization Methods for Nonlinear Ill-Posed Problems* (Walter de Gruyter, Berlin).
21. Kitagawa, G., and W. Gersch, 1985: A smoothness priors long AR model method for spectral estimation. *IEEE Trans. Autom. Contr.* **30**, 57–65.
22. Li, Y., and D. W. Oldenburg, 1999: 3-D inversion of DC resistivity data using an L-curve criterion. In *Extended Abstracts of 69th SEG Meeting* (Houston, TX), pp. 251–254.
23. Lukas, M. A., 1998: Asymptotic behaviour of the minimum bound method for choosing the regularization parameter. *Inverse Probl.* **14**, 149–159.
24. Mallows, C. L., 1973: Some comments on Cp. *Technometrics* **15**, 661–676.
25. Morozov, V. A., 1966: On the solution of functional equations by the method of regularization. *Soviet Math. Dokl.* **7**, 414–417.
26. Morozov, V. A., 1984: *Methods for Solving Incorrectly Posed Problems* (Springer, New York).

27. O'Sullivan, F., 1990: Convergence characteristics of a method of regularization estimators for nonlinear operator equations. *SIAM J. Numer. Anal.* **27**, 1635–1649.
28. Raus, T., 1985: The principle of the residual in the solution of ill-posed problems with non-selfadjoint operators. *Acta Comment. Univ. Tartuensis* **715**, 12–20.
29. Rieder, A., 2003: *Keine Probleme mit Inversen Problemen* (Vieweg, Wiesbaden).
30. Thompson, A. M., J. W. Kay, and D. M. Titterton, 1989: A cautionary note about the crossvalidatory choice. *J. Statist. Comput. Simul.* **33**, 199–216.
31. Tikhonov, A. N., 1963: Solution of incorrectly formulated problems and the regularization method. *Soviet Math. Dokl.* **4**, 1035–1038.
32. Tikhonov, A. N., and V. Y. Arsenin, 1977: *Solutions of Ill-Posed Problems* (Wiley, New York).
33. Tikhonov, A. N., and V. B. Glasko, 1965: Use of the regularization method in nonlinear problems. *USSR Comp. Math. Math. Phys.* **5**, 93–107.
34. Varah, J. M., 1973: On the numerical solutions of ill-conditioned linear systems with applications to ill-posed problems. *SIAM J. Numer. Anal.* **10**, 257–267.
35. Vogel, C. R., 1985: Numerical solution of a non-linear ill-posed problem arising in inverse scattering. *Inverse Probl.* **1**, 393–403.
36. Vogel, C. R., 2002: *Computational Methods for Inverse Problems* (SIAM, Philadelphia, PA).
37. Wahba, G., 1977: Practical approximate solutions to linear operator equations when the data are noisy. *SIAM J. Numer. Anal.* **14**, 651–667.
38. Wahba, G., 1990: *Spline Models for Observational Data* (SIAM, Philadelphia, PA).



Kristan Gurton, Konstantin Gilev, and Anton Lopatin (from left to right) enjoying the picnic at the Main Astronomical Observatory.



Field trip to the Main Astronomical Observatory. From left to right: Mikhail Alexandrov, Olga Kalashnikova, Igor Geogdzhayev, and Larissa Nazarenko.



The last night of the ASI. From left to right: Gorden Videen, Vera Rosenbush, Tamara Bul'ba, Anatoli Borovoi, Irina Kulyk, and Nikolai Kiselev.

Light scattering resonances in small particles with electric and magnetic optical properties

Braulio García-Cámara^{1*}, Francisco González¹, Fernando Moreno,¹
and Gorden Videen²

¹ Grupo de Óptica, Departamento de Física Aplicada, Universidad de Cantabria,
Avda. de los Castros s/n 39005, Santander, Spain

² Army Research Laboratory, 2800 Powder Mill Road, Adelphi, Maryland, 20783, USA

Abstract. With the recent progress of metamaterials and nanoscience providing encouragement, new avenues of light-scattering research is focusing on particles having unconventional optical properties. Light scattered by such particles can have interesting features, visible in their resonances or directional scattering, for instance. In this chapter we focus on the simple Lorenz–Mie resonances that appear in light scattering of small particles. By considering a large range of values of the electric permittivity and the magnetic permeability, either positive or negative, unexpected resonant behavior may be observed. Special attention has been paid to the double-negative or left-handed materials.

Keywords: resonances, left-handed materials, Mie coefficients, nanoparticles, size effects, zero-backscattering, zero-forward scattering

1. Introduction

Advances in nanophotonics offer new technical possibilities like intra-chip optical communications (Almeida et al. 2004; Engheta 2007; Miller 2010) or biosensors (Anker et al. 2008; Yan et al. 2008, Lin and Chung 2009; Ladd et al. 2009). To achieve these goals, smaller components with specialized behavior are required, such as high optical response or directional control of the light scattering, i.e., waveguiding. This brought researchers to concentrate on systems much smaller in size than those considered a few years ago. Technological advances have allowed them to produce well-defined structures and particles in the nanometric range (Sun and Xia 2002; Krasheninnikov and Banhart 2007; Wang et al. 2007). For such small systems, the scattering cross-section is usually very small, which is inconvenient for structures which are supposed to affect or alter the propagation of light. This problem can be overcome if scatterers can sustain resonances that enhance the light-matter interaction. In this context, metallic nanostructures have received a lot of attention due to their ability to confine elec-

* Corresponding author. E-mail: braugarcia@gmail.com

tromagnetic waves to regions in space much smaller than the wavelength by means of localized surface plasmons (Prasad 2004). These surface plasmons are collective oscillation of the electron plasma at the interface between a metal and a dielectric, or in other words, on the surface of a metallic nanostructures.

Light scattering resonances appear in non-metallic structures and have been studied for a long time (Barber and Chang 1988). By considering unusual optical properties, Videen and Bickel (1990) were able to derive analytical expressions for very small spheres. They analyzed Lorenz–Mie resonances for very small dielectric ($\epsilon > 0$, $\mu = 1$) spherical particles and studied resonances as a function of particle size and refractive index. As particles are much smaller than the incident wavelength, some approximations to Mie theory can be used. The authors showed that Mie coefficients of order higher than 2 can be neglected. Furthermore, they proposed approximate expressions for the first four Mie coefficients a_1 , a_2 , b_1 , and b_2 .

On the experimental front, new engineered materials or metamaterials (Smith et al. 2002) have extended the range of values that the electric permittivity (ϵ) and the magnetic permeability (μ) can reach. Special attention has been dedicated to the case of negative electric permittivity ($\epsilon < 0$) and negative magnetic permeability ($\mu < 0$). Such materials are referred to as double-negative or left-handed, for which light propagation takes on interesting characteristics (Smith et al. 2004).

Such new materials provide the opportunity to consider arbitrary values of the refractive index and hence, different and new scattering behaviors, especially in resonant features. In this chapter, we have generalized the analysis of Videen and Bickel (1990), removing all restrictions on the particle's optical properties and we study the evolution of these resonances, in particular their position and width, as a function of the optical properties and size.

2. Theoretical model

2.1. The Lorenz–Mie theory: scattering coefficients

The problem of the electromagnetic scattering from a homogeneous sphere of arbitrary size was solved by Lorenz and Mie many years ago (Mie 1908; Lorenz 1890, 1898). The Lorenz–Mie theory considers an isotropic sphere characterized by a dielectric constant that could be complex and frequency dependent, immersed in a homogeneous and isotropic medium, as shown in Fig. 1

By means of this theory, the scattered field (\mathbf{E}_s , \mathbf{H}_s) that depends on the incident field (\mathbf{E}_i , \mathbf{H}_i), can be expressed as an expansion of the *Vector Spherical Harmonics* (VSH) and one may obtain the expansion coefficients using the boundary conditions that relate the fields as follows:

$$\mathbf{E}_s = \sum_{n=1}^{\infty} E_n (i a_n \mathbf{N}_{e1n}^{(3)} - b_n \mathbf{M}_{o1n}^{(3)}), \quad (1a)$$

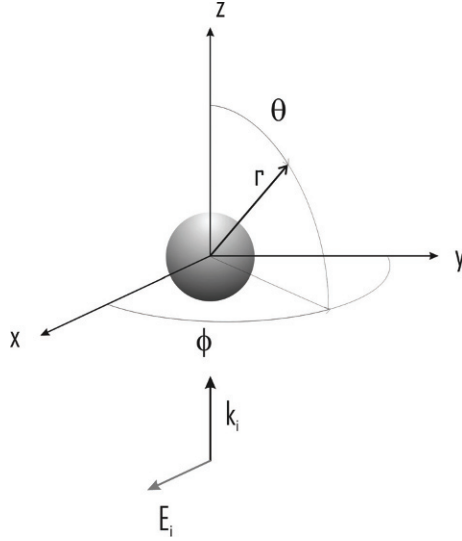


Fig. 1. Scattering geometry.

$$\mathbf{H}_s = \frac{k}{\omega\mu} \sum_{n=1}^{\infty} E_n (ib_n \mathbf{N}_{oln}^{(3)} + a_n \mathbf{M}_{eln}^{(3)}), \quad (1b)$$

where $i = (-1)^{1/2}$, and the superscript (3) means that the radial dependence of the generating function is given by $h_n^{(1)}$. The coefficients a_n and b_n are the Lorenz–Mie coefficients for the scattered field that can be expressed as (Bohren and Huffman 1983):

$$a_n = \frac{\mu m^2 j_n(mx) [xj_n(x)]' - \mu_l j_n(x) [mxj_n(mx)]'}{\mu m^2 j_n(mx) [xh_n^{(1)}(x)]' - \mu_l h_n^{(1)}(x) [mxj_n(mx)]'}, \quad (2a)$$

$$b_n = \frac{\mu_l m^2 j_n(mx) [xj_n(x)]' - \mu j_n(x) [mxj_n(mx)]'}{\mu_l m^2 j_n(mx) [xh_n^{(1)}(x)]' - \mu h_n^{(1)}(x) [mxj_n(mx)]'}, \quad (2b)$$

where μ_l and μ are the magnetic permeability of the sphere and the surrounding medium respectively, x is the size parameter and m the relative refractive index defined, respectively, as

$$x = kR = \frac{2\pi R}{\lambda}, \quad m = \frac{n_l}{n}. \quad (3)$$

Here, R is the radius of the sphere, λ is the incident wavelength, and n_l and n are the refractive index of the sphere and the surrounding medium, respectively.

In order to simplify the expressions of the Mie coefficients we introduce the *Riccati–Bessel functions* expressed as

$$\psi_n(\rho) = \rho j_n(\rho), \quad \xi_n(\rho) = \rho h_n^{(1)}(\rho). \quad (4)$$

Then the scattering Lorenz–Mie coefficients can be written as

$$a_n = \frac{\tilde{m} \psi'_n(x) \psi_n(mx) - \psi'_n(mx) \psi_n(x)}{\tilde{m} \xi_n(x) \psi_n(mx) - \psi'_n(mx) \xi_n(x)}, \quad (5a)$$

$$b_n = \frac{\psi'_n(x) \psi_n(mx) - \tilde{m} \psi'_n(mx) \psi_n(x)}{\xi_n(x) \psi_n(mx) - \tilde{m} \psi'_n(mx) \xi_n(x)}, \quad (5b)$$

where $\tilde{m} = m/\mu_i$ and $\mu = 1$.

2.2. Scattering, absorption, and extinction cross sections

The scattering and absorption cross sections are defined as the ratio of the rate at which the electromagnetic energy is scattered (W_s) or absorbed (W_a) by an imaginary sphere of radius $r \geq R$ around the particle with respect to the incident irradiance (I_i):

$$C_{abs} = \frac{W_a}{I_i}, \quad C_{sca} = \frac{W_s}{I_i}. \quad (6)$$

The extinction cross section is defined as the sum of the scattering and absorption cross sections:

$$C_{ext} = C_{abs} + C_{sca}. \quad (7)$$

Using the scattering Lorenz–Mie coefficients we can express these parameters as

$$\begin{aligned} C_{sca} &= \frac{W_s}{I_i} = \frac{2\pi}{k^2} \sum_{n=1}^{\infty} (2n+1) (|a_n|^2 + |b_n|^2), \\ C_{ext} &= \frac{W_{ext}}{I_i} = \frac{2\pi}{k^2} \sum_{n=1}^{\infty} (2n+1) \operatorname{Re}(a_n + b_n). \end{aligned} \quad (8)$$

Commonly, the scattering or extinction efficiencies are used. These are defined as

$$\begin{aligned} Q_{ext} &= \frac{C_{ext}}{G} = \frac{2}{x^2} \sum_{n=1}^{\infty} (2n+1) \operatorname{Re}(a_n + b_n), \\ Q_{sca} &= \frac{C_{sca}}{G} = \frac{2}{x^2} \sum_{n=1}^{\infty} (2n+1) (|a_n|^2 + |b_n|^2), \end{aligned} \quad (9)$$

where G is the particle cross-sectional area projected onto a plane perpendicular to the incident beam. In the case of a spherical particle of radius R , $G = \pi R^2$.

2.3. Light scattering by small particles: Rayleigh approximation

Light scattering from particles of size much smaller than the incident wavelength is often called Rayleigh scattering and expressions that also refers to the

approximations made under such condition. Rayleigh scattering approximation or Dipolar approximation requires the following two conditions:

- the size parameter x must be much smaller than one; and
- the relative refractive index of the particle m must be small, i.e., $|m|x \ll 1$.

Under these conditions, the electromagnetic scattering is dominated by the first term of the Lorenz–Mie expansion. This means that only the first two scattering coefficients (a_1 and b_1) are not negligible and these can be reduced to the following expressions:

$$\begin{aligned} a_1 &= i \frac{2x^3}{3} \frac{(\varepsilon - 1)}{(\varepsilon + 2)} + O(x^5), \\ b_1 &= i \frac{2x^3}{3} \frac{(\mu - 1)}{(\mu + 2)} + O(x^5), \\ a_n &\approx b_n \approx 0. \end{aligned} \quad (10)$$

The light scattered by a particle under these conditions is similar to that emitted by an electric or a magnetic dipole. Depending of the values of the electric permittivity ε and the magnetic permeability μ one or the other dominates the scattered radiation. In this situation the scattering and extinction efficiencies can be expressed as

$$\begin{aligned} Q_{ext} &= \frac{C_{ext}}{G} = \frac{6}{x^2} \operatorname{Re}(a_1 + b_1), \\ Q_{sca} &= \frac{C_{sca}}{G} = \frac{6}{x^2} (|a_1|^2 + |b_1|^2). \end{aligned} \quad (11)$$

2.4. Second-order dipolar approximation

In a previous work, Videen and Bickel (1990) considered light scattering when the second Rayleigh condition is relaxed and there is no restriction on the refractive index. They found new resonances and developed approximate expressions for the first four Mie coefficients that could be used for studying the behaviour of such resonances as a function of system parameters like the refractive index and the radius of the sphere.

For the case in which the isolated particles show a magnetic response to the incident field ($\mu \neq 1$), the scattering properties are also of interest (Kerker et al. 1983; Merchiers et al. 2007). With the recent appearance of engineered materials whose optical properties can be controlled, known as *metamaterials*, light scattering by small particles having exceptional properties, like left-handed properties, is receiving significant attention from the scientific community (Soukoulis et al. 2007; Shalaev 2007; Zheludev 2010). One suggested application is the construction of optical nanocircuits based on metamaterials for optical communication and computing applications (Engheta 2007).

We consider a generalized formulation of the Lorenz–Mie coefficients when the particles are small, analogous to the study performed by Videen and Bickel (1990). Particles for which $R \ll \lambda$, only the first four scattering coefficients (a_1 , b_1 , a_2 , and b_2) are significant and higher orders are neglected. In this limit, the Riccati–Bessel functions can be expanded and the first few terms are as follows:

$$\begin{aligned} \psi_1(x) &= \frac{\sin x}{x} - \cos x, & \psi'_1(x) &= \frac{\cos x}{x} - \frac{\sin x}{x^2} + \sin x, \\ \xi_1(x) &\approx e^{ix} \left(\frac{i}{x} + 1 \right), & \xi'_1(x) &\approx e^{ix} \left(\frac{i}{x^2} + \frac{1}{x} - i \right), \\ \psi_2(x) &\approx \left(\frac{3}{x^2} - 1 \right) \sin x - \frac{3}{x} \cos x, & & (12) \\ \psi'_2(x) &\approx \frac{1}{x^3} (6 \sin x - 6x \cos x + x^3 \cos x - 3x^2 \sin x), \\ \xi_2(x) &\approx e^{ix} \left(\frac{3i}{x^2} - \frac{3}{x} - i \right), & \xi'_2(x) &\approx e^{ix} \left(\frac{6i}{x^3} + \frac{6}{x^2} - \frac{3i}{x} - 1 \right). \end{aligned}$$

As we consider $x \ll 1$, the sine, cosine and exponential functions whose argument is x can be substituted for its power expansion. We have checked that for the analysis of the resonance, which is our purpose, only the first two terms of the Taylor expansion are sufficient, this is:

$$\begin{aligned} \cos x &= 1 - \frac{1}{2}x^2, \\ \sin x &= x - \frac{1}{6}x^3, & & (13) \\ e^{ix} &= \cos x + i \sin x = \left(1 - \frac{1}{2}x^2 \right) + i \left(x - \frac{1}{6}x^3 \right). \end{aligned}$$

Substituting these into the Ricatti–Bessel function expansions (12), and then those into the general expressions of the scattering coefficients (5) we obtain approximate expressions for the first four Lorenz–Mie coefficients that are valid for any pair of the optical constants (ε, μ):

$$a_1 = \frac{a_1^{num}}{a_1^{den}},$$

where

$$\begin{aligned} a_1^{num} &= \tilde{m}mx^3 [mx \cos(mx) - \sin(mx)], \\ a_1^{den} &= \cos(mx) [-\tilde{m}m^2x^2 - i\tilde{m}m^2x^2 + imx^3 + imx] \\ &\quad + \sin(mx) [\tilde{m}mx^3 + i\tilde{m}m - ix^2 + im^2x^4 - i + im^2x^2], \end{aligned}$$

$$\begin{aligned}
a_2 &= \frac{a_2^{num}}{a_2^{den}}, \\
a_2^{num} &= \cos(mx)[6\tilde{m}m^2x^2 - 6mx^2 + m^3x^4] \\
&\quad + \sin(mx)[6x - 3m^2x^3 - 6\tilde{m}mx - 2\tilde{m}m^3x^3] \\
a_2^{den} &= \cos(mx)[3i\tilde{m}m^3x^2 + 2\tilde{m}m^3x^3 - i\tilde{m}m^3x^4 + 6i\tilde{m}m^3 - 9i\tilde{m}m \\
&\quad - 6\tilde{m}mx + 3i\tilde{m}mx^2 + 18i\tilde{m}mx^{-2} - (-3ix^{-2} - 2i - x)(6 - 3m^2x^2)], \\
&\quad + \sin(mx)[9i\tilde{m}m^2x + 6\tilde{m}m^2x^2 - 3i\tilde{m}m^2x^3 \\
&\quad + 18i\tilde{m}m^2x^{-1} - (-3ix^{-2} - 2i - x)(6mx + m^3x^3)]. \tag{14}
\end{aligned}$$

Due to the spherical symmetry, a_n and b_n are related in the following way

$$b_n(1/\tilde{m}, m, x) = a_n(\tilde{m}, m, x). \tag{15}$$

Under this approach, the efficiencies of extinction and scattering are given by

$$\begin{aligned}
Q_{ext} &= \frac{C_{ext}}{G} = \frac{2}{x^2} [3\text{Re}(a_1 + b_1) + 5\text{Re}(a_2 + b_2)], \\
Q_{sca} &= \frac{C_{sca}}{G} = \frac{2}{x^2} [3(|a_1|^2 + |b_1|^2) + 5(|a_2|^2 + |b_2|^2)]. \tag{16}
\end{aligned}$$

In Fig. 2 we show the extinction efficiency Q_{ext} in a semi-logarithmic scale as function of the electric permittivity for a particle of radius $R = 0.01\lambda$ with a magnetic permeability equal to one ($\mu = 1$). The extinction efficiency is calculated using three different methods: (i) the exact expressions of Mie coefficients (exact); (ii) the approximate coefficients given by Eqs. (14) and (AC1); and (iii) another approximation using the first four terms of the Taylor expansion of sine and cosine functions.

3. Light scattering resonances

3.1. Mie resonances

Light scattering by a particle depends on the size, shape and optical properties of the scatterer as well as the frequency of the incident wave. In the same way, resonances and their spectral properties (peak width and position) depend also on particle size and optical properties. From a mathematical point of view, resonances are excited when the denominator of the Mie coefficients, Eq. (5a) or Eq. (5b), are zero. For particles in the Rayleigh limit ($x \rightarrow 0$) the zeros of the denominator of a_n and b_n appear when (Bohren and Huffman 1983)

$$m^2 = -\frac{n+1}{n}, \quad n = 1, 2, \dots, \tag{17}$$

where m is the relative refractive index of the particle. For a non-magnetic particle

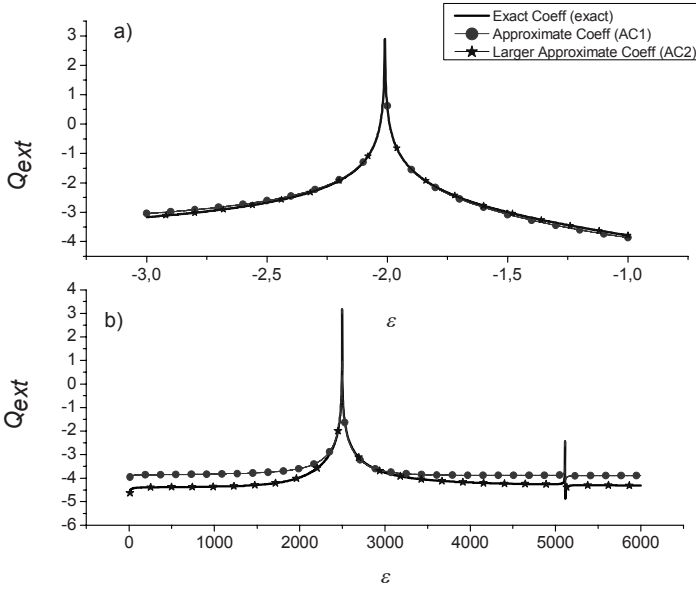


Fig. 2. Comparative plot of Q_{ext} for three different expressions of the Lorenz–Mie coefficients: the exact one (solid curve), the approximation using Eq. (14) (labeled AC1), and the approximation using more coefficients in the sin and cos expansions (labeled AC2) for $\mu = 1$. Panels (a) and (b) show the metallic and the dielectric case, respectively.

($\mu = 1$) with an electric permittivity $\varepsilon = \varepsilon' + i\varepsilon''$, immersed in a medium with a nonabsorbing electric permittivity ε_m , the last condition can be rewritten as

$$\varepsilon' = -2\varepsilon_m, \quad \varepsilon'' = 0. \quad (18)$$

This resonance is known as the *Fröhlich resonance* or *mode of uniform polarization* (Bohren and Huffman 1983).

A wide variety of resonances can be excited in a particle depending on its size and optical properties as was shown by Videen and Bickel (1990). In that work, the authors limited their analysis to the case of dielectric and non-magnetic particles ($\varepsilon > 0$ and $\mu = 1$). Using the scattering coefficients given by Eq. (14), this study can be generalized to describe resonant behaviours for almost every combination of optical properties (García-Cámara et al. 2008a).

For this study we have calculated the evolution of the extinction efficiency as a function of the optical constants and particle size to analyze the main characteristics (position and width) of the resonant modes.

Particle sizes have been considered in the range $R \in [0.01-0.1]\lambda$ such that particles are finite but still smaller than the incident wavelength. This allows us to consider only the first four terms of the Lorenz–Mie coefficients in such a way that the extinction efficiency can be written as

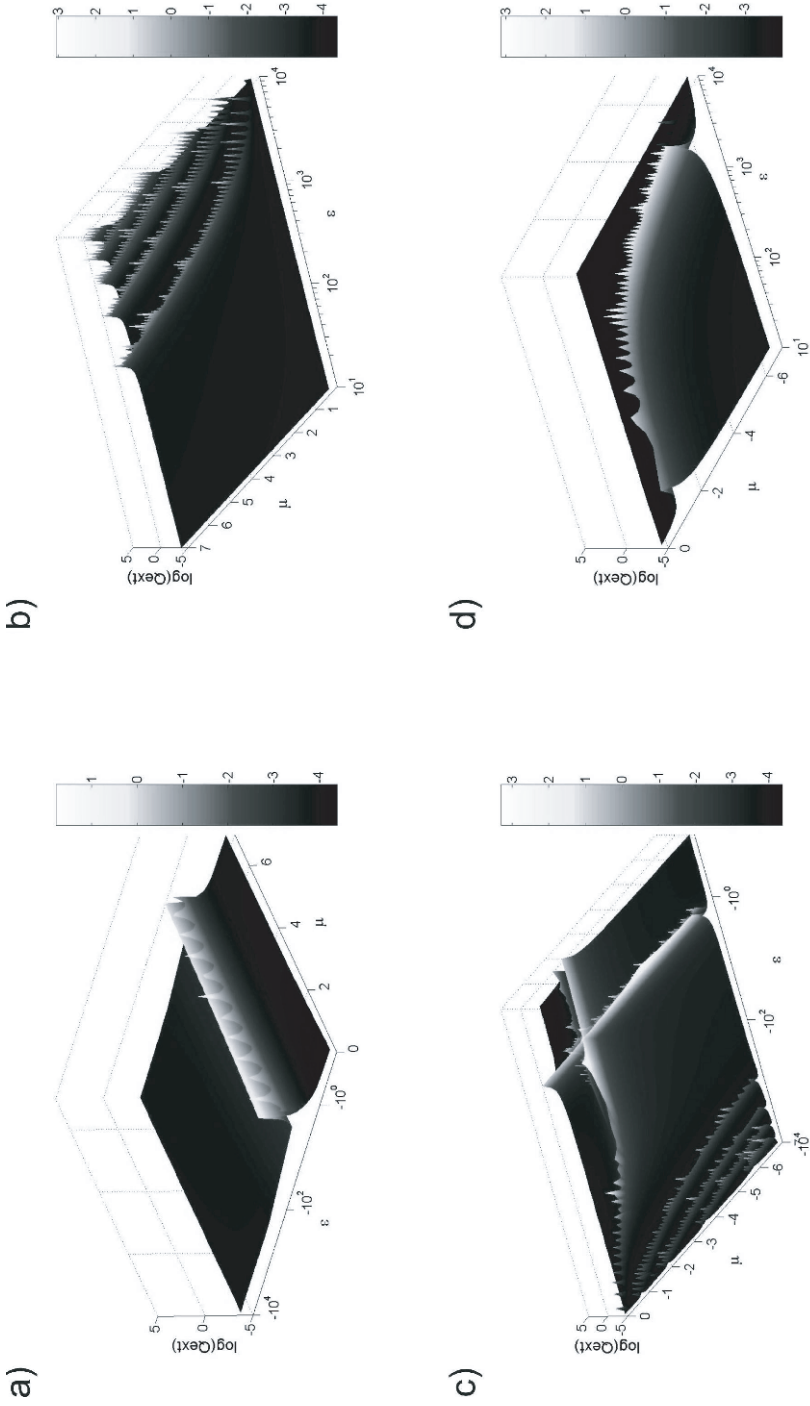


Fig. 3. 3D plots of $\log(Q_{ext})$ as a function of the optical properties (ϵ and μ) for a spherical particle with $R = 0.01\lambda$.

$$Q_{ext} = \frac{2}{x^2} [3 \operatorname{Re}(a_1 + b_1) + 5 \operatorname{Re}(a_2 + b_2)]. \quad (19)$$

The expressions used for the scattering coefficients are those described by Eq. (14). In what follows, we refer to a_1 and a_2 as electric terms (dipolar and quadrupolar, respectively) and b_1 and b_2 as magnetic terms (dipolar and quadrupolar, respectively). Resonant modes appearing due to zeros in the denominator of one of these coefficients also can be referred in the same way. For instance, a mode due to a zero in the denominator of a_1 will be called dipolar electric resonance.

3.2. Dependence of Mie resonances on the optical properties

The purpose of this section is to analyze the dependence of the resonances of a small particles on its optical properties. We consider arbitrary values for the electric permittivity and the magnetic permeability. In Fig. 3 we plot Q_{ext} (semi-logarithmic scale) for spherical particles as a function of the optical properties (ε, μ). The particle size is $R = 0.01\lambda$ and the expressions (14) for the scattering coefficients reproduce accurately the cross-sections. The values for the electric permittivity were chosen in order to be comparable with those in Videen and Bickel (1990).

For a detailed analysis, we have considered separately four different cases that coincide with the four quadrants in the $\varepsilon - \mu$ representation: (i) $\varepsilon > 0$ and $\mu > 0$; (ii) $\varepsilon < 0$ and $\mu > 0$; (iii) $\varepsilon < 0$ and $\mu < 0$; and (iv) $\varepsilon > 0$ and $\mu < 0$. In the first case, the conventional dielectric materials for the visible with $\varepsilon > 0$ and $\mu = 1$ are included. The second case includes metallic materials ($\varepsilon < 0$ and $\mu = 1$) while the third one represents left-handed materials ($\varepsilon < 0$, $\mu < 0$) and finally the refractive index $m < 0$. This last case does not have any physical interest at present, but we include it for completion.

Case $\varepsilon > 0$ and $\mu > 0$ (Fig. 3b). First, if we restrict the plot to $\mu = 1$ we reproduce the typical Lorenz–Mie resonances shown in Fig. 1 of Videen and Bickel (1990). In the full surface plot we consider the evolution of the position of these resonances as the magnetic permeability μ changes. We see a series of concentric curves that evolve with ε and μ . In Fig. 4a we can see that the resonances are organized in branches, each corresponding to a particular Lorenz–Mie coefficient being on resonance. These corresponding coefficients are labelled in the figure. The evolution of the resonance position is such that as μ increases, the resonance appears for smaller values of $|\varepsilon|$.

Case $\varepsilon < 0$ and $\mu > 0$ (Fig. 3a). In this domain, only the Mie coefficients a_1 and a_2 may take high values, which imply that only electric resonances (dipolar and quadrupolar) can be excited. The dipolar resonance, associated with maximum values of a_1 , appears around $\varepsilon \sim -2$. Only when the particle size tends to zero and the particle becomes dipole-like, is the resonance position located at exactly $\varepsilon = -2$. This resonance is the Fröhlich resonance described in Eq. (18), which is slightly shifted due to the quadrupolar effects (a_2). Although the particle size is

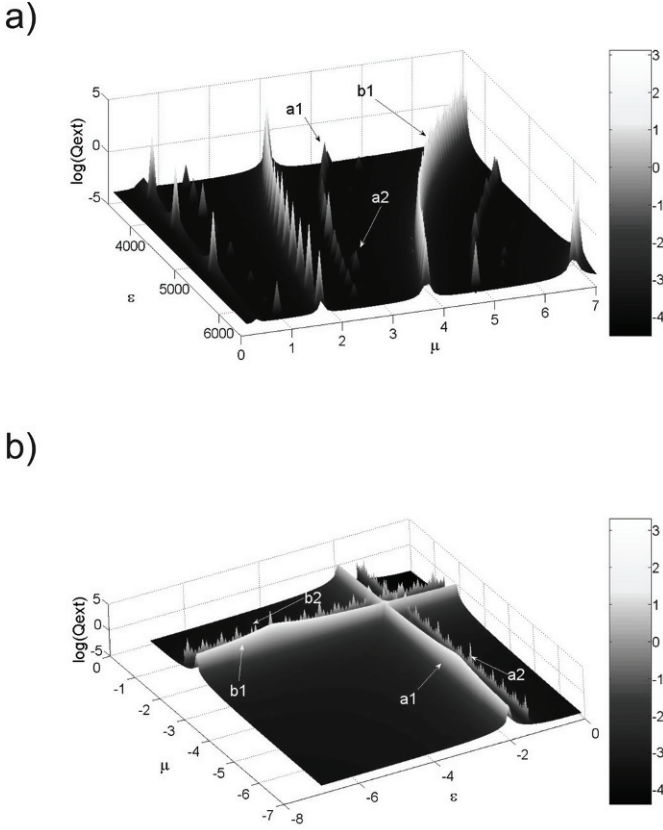


Fig. 4. Enlargement of two interesting zones of Fig. 3: (a) region with $\epsilon > 0$, $\mu > 0$; (b) region with $\epsilon < 0$, $\mu < 0$. The resonant modes are indicated.

small, quadrupolar effects can be observed. In this domain an electric quadrupolar resonance can be observed at $\epsilon \sim -1.5$, and it is associated with high values of the coefficient a_2 . Resonances in this domain do not change, or change only slightly, their spectral position as the magnetic permeability changes.

Case $\epsilon < 0$ and $\mu < 0$ (Fig. 3c). This region has the greatest number of features and also corresponds to that of left-handed, or negative-refractive-index, materials (Veselago 1968). While these values of the electric permittivity and the magnetic permeability have not been found in naturally occurring materials, recent developments have found their construction to be possible (Smith et al. 2004; Valentine et al. 2009; Boltasseva and Shalaev 2008).

We can distinguish two different behaviors in this quadrant: for low values of the modulus of the electric permittivity ($|\epsilon| < 10^2$) and for larger values ($|\epsilon| > 10^2$). The first case is enlarged in Fig. 4b. Here, both electric and magnetic resonances can be observed. The electric ones are a continuation of those de-

scribed in the previous case. The dipolar electric resonance appears at $\varepsilon \sim -2$ for every value of the magnetic permeability, while the quadrupolar electric mode is located at $\varepsilon \sim -1.5$, and remains fixed as μ changes. Two additional resonances can be observed. These are associated with the magnetic coefficients, b_1 and b_2 . The dipolar magnetic mode that corresponds with high values of b_1 appears at $\mu \sim -2$ and the quadrupolar one, associated with b_2 , at $\mu \sim -1.5$. Only when $R \rightarrow 0$ do resonances appear at the exact values predicted. As happens for the electric resonances, these modes remain fixed for every value of the electric permittivity in the considered range. An interesting $\varepsilon - \mu$ symmetry is observed in this domain: the values of ε at which the electric resonances appear, are equal to those values of μ at which the magnetic resonances are observed. This symmetry is related to that observed for the Mie coefficients under the Rayleigh approximation (10). Here, we see that this symmetry is still present when the Rayleigh approximation is not strictly valid, and that it can be extended to the quadrupolar terms. The dipolar resonances, both electric and magnetic, present a particular structure with an intense background and a peak that appears at $\mu = -5$ for the electric case and at $\varepsilon = -5$ for the magnetic case, as depicted in Fig. 4b.

For larger values of the electric permittivity ($|\varepsilon| > 10^2$), resonances similar to those observed for $\varepsilon > 0$, $\mu > 0$ are found. Also, their behaviour as a function of magnetic permeability is also very similar to that of the modes in the first case ($\varepsilon > 0$, $\mu > 0$): increasing the modulus of the magnetic permeability shifts the position of each resonance to lower values of the modulus of the electric permittivity $|\varepsilon|$.

Case $\varepsilon > 0$ and $\mu < 0$ (Fig. 3d). This range is the most unusual, since it is not related with any natural or engineered material. Here, only magnetic resonances are excited: the magnetic dipolar resonance, associated with b_1 and the magnetic quadrupolar resonance associated with b_2 . These resonances appear as a continuation of the magnetic modes appearing in the previous case. They are located at $\mu \sim -2$ for the dipolar resonances and at $\mu \sim -1.5$ for the quadrupolar resonance. However, for high values of the electric permittivity, the behavior of the resonances' positions is more complex.

Again, a $\varepsilon - \mu$ symmetry is observed for the electric and magnetic resonances in the range ($\varepsilon < 0$, $\mu < 0$) according to the symmetry of the Mie coefficients (10). We have already commented that the electric resonances in the second quadrant ($\varepsilon < 0$, $\mu > 0$) and the magnetic resonances in this quadrant ($\varepsilon > 0$, $\mu < 0$), are a continuation of the symmetric resonances in the negative-refractive-index quadrant ($\varepsilon < 0$, $\mu < 0$). A question that has to be raised is whether the electric and magnetic resonances of these ranges present a similar $\varepsilon - \mu$ symmetry. To analyze this, we plot in Fig. 5 the extinction efficiency Q_{ext} of a spherical particle ($R = 0.01\lambda$) as a function of the optical properties in these two regions ($\varepsilon < 0$, $\mu > 0$ and $\varepsilon > 0$, $\mu < 0$) when over the same range of ε and μ . It can be seen that the behavior is symmetric: the electric resonances in Fig. 5a depend on ε in the same way as the magnetic resonances depends on μ in Fig. 5b.

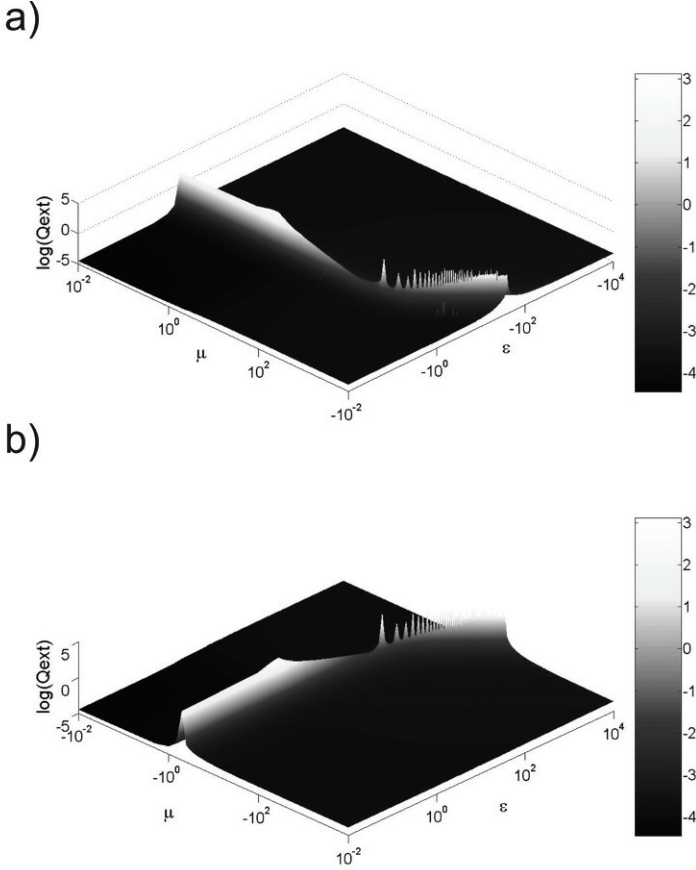


Fig. 5. 3D plots of $\log(Q_{ext})$ as a function of the optical properties (ε and μ) for two different ranges: (a) $\varepsilon < 0$ and $\mu > 0$, (b) $\varepsilon > 0$ and $\mu < 0$ when the range of values of ε and μ are the same.

3.3. Dependence of Mie resonances on particles size

In the previous section, we have considered particle sizes that are much smaller than the wavelength of the incident radiation. It is evident that the spectral behavior of the cross-sections depends on the particle size. The simplicity of the expressions introduced in this work for scattering coefficients, a_1 , a_2 , b_1 and b_2 , allow us to extend our former analysis to other particle sizes and to observe the changes in resonances as the particle grows. In Fig. 6 the logarithm of Q_{ext} is plotted for several particle sizes and for two different cases:

1. As a function of the electric permittivity and $\mu = -2$. In this case, the electric resonances are excited and the peak associated with the dipolar magnetic term b_1 is also shown.

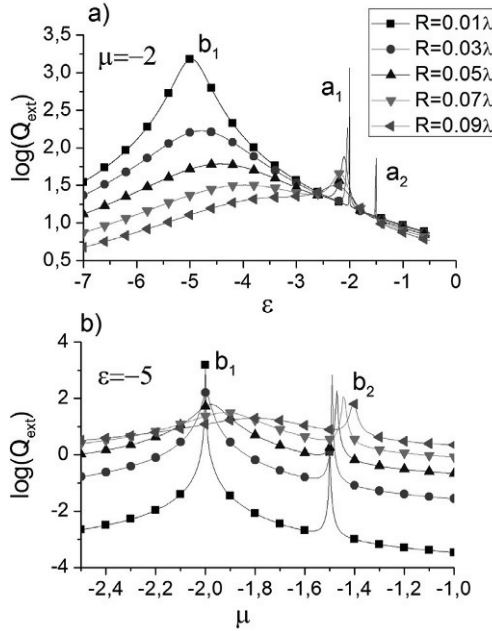


Fig. 6. Evolution of the extinction efficiency as a function of the electric permittivity in the range $\epsilon < 0$ for different values of particle radius R . The value of the magnetic permeability is indicated in the bottom right corner. The resonance modes are indicated.

2. As a function of the magnetic permeability and $\epsilon = -5$. In this case, the magnetic dipolar and quadrupolar resonance terms are labeled.

In Fig. 6a and for very small particles ($R = 0.01\lambda$), only dipolar resonances appear: the electric one is located at $\epsilon \approx -2$ where high values of a_1 are reached, and another peak due to a magnetic contribution is also observed at $\epsilon \approx -5$. As R increases, these two peaks evolve, becoming broader and shifted, Q_{ext} takes lower values, and new resonances of higher orders appear. In particular, an electric quadrupolar resonance, related to high values of a_2 , is observable at $\epsilon \approx -1.5$. For higher values of the particle size, the dipolar resonances tend to disappear, becoming less sharp until only the quadrupolar resonance remains as a well-defined peak for the highest value of R . Similar behavior can be observed for the pure magnetic resonances (b_1 and b_2) in Fig. 6b. As R increases, the modes' position shifts while the shape becomes less sharp and broader. For the highest value of R , the dipolar resonance has almost disappeared and only the quadrupolar resonance is still present. Here, in contrast to the behaviour seen in Fig. 6a, the quadrupolar mode is still observed for the lowest value of particles size. Also, as R increases, Q_{ext} takes higher values in the background, while at the peak it is lower.

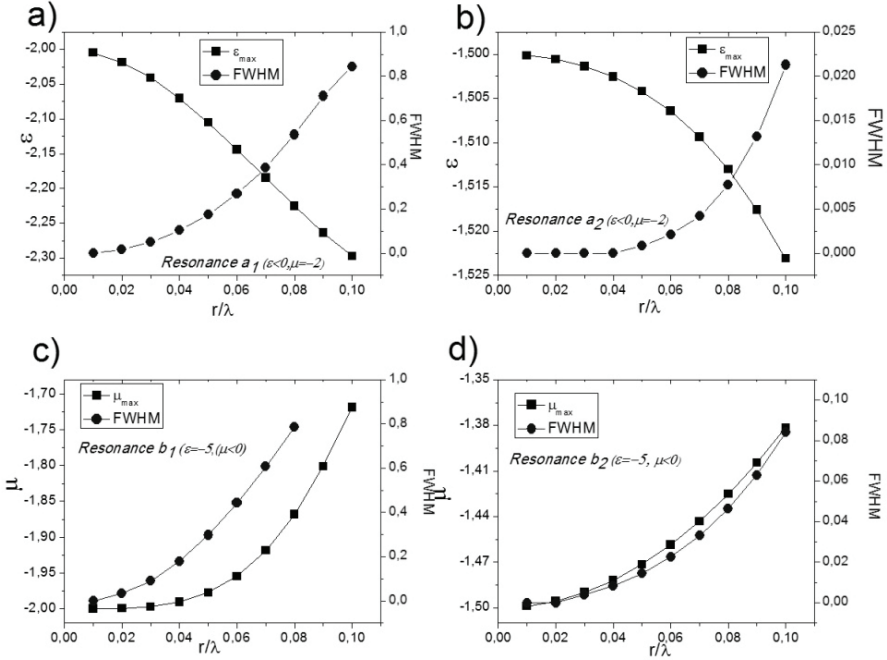


Fig. 7. Evolution of FWHM and position of the resonances as a function of particle size.

This brief analysis can be expanded using the simple expressions for the Lorenz–Mie coefficients introduced in Eq. (14) for a wide range for particle sizes. Resonances are often quantified by their position of maximum extinction Q_{ext} and by their full width at half maximum (FWHM). In Fig. 7, the position and the width of the previous resonances are presented as a function of particle size. The position of the resonances was analyzed for a permeability $\mu = -2$ and as a function of the permittivity ϵ (Figs. 7a and 7b) while that of the magnetic modes was studied as a function of the permeability μ at a fixed permittivity $\epsilon = -5$ (Figs. 7c and 7d). As R increases, the position of the peaks shifts to higher values of $|\epsilon|$ (red shift). The shift of the dipolar resonance, Fig. 7a, is from $\epsilon = -2$ to $\epsilon = -2.3$ while that of the quadrupolar resonance, Fig. 7b, is shifted from $\epsilon = -1.5$ to $\epsilon = -1.525$. The dipolar resonance with a 15% shift is much more sensitive to particle size than the quadrupolar mode with a 1.6% shift.

These results are quite similar to those presented by Meier and Wokaum (1983), who explained that the red-shift of the resonances is produced by the dynamic depolarization; whereas, the decrease in magnitude and broadening are due to radiation damping effects.

The evolution of the position for the magnetic resonances as the particle size increases differs from that of the electric ones. In both cases, for the dipolar mag-

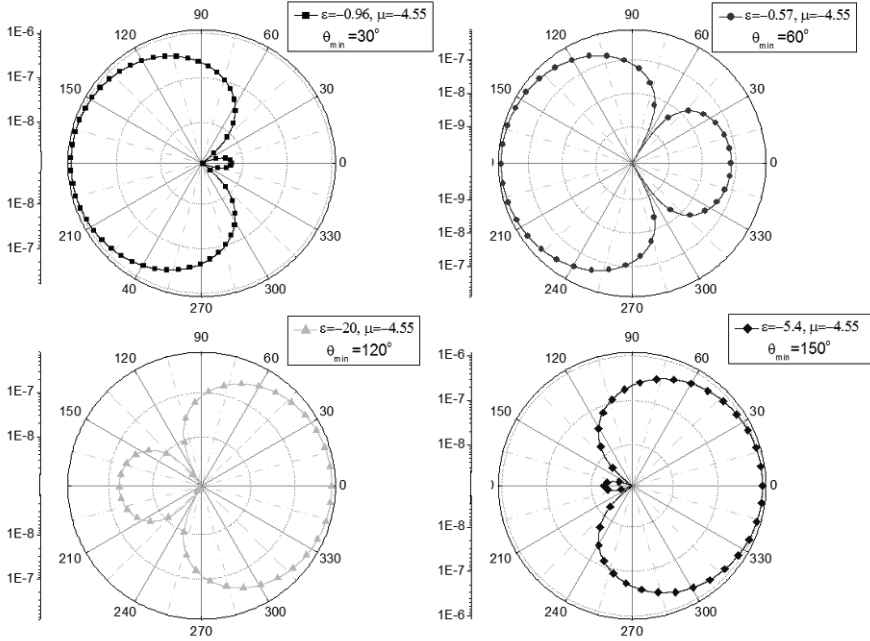


Fig. 8. Polar diagrams of the scattered intensity for a spherical particles with $R = 0.01\lambda$ and optical constants in the negative–negative range that produce a minimum intensity at certain scattering angles. The particle is illuminated with a linearly polarized plane wave with the electric field perpendicular to the scattering plane (TE polarization).

netic (Fig. 7c) and the quadrupolar magnetic (Fig. 7d) resonances, as R increases the peak tends to lower values of the modulus of the magnetic permeability $|\mu|$. The magnetic resonances evolve in the opposite direction as that of the electric resonances as the particle size changes. This fact also can be observed in Fig. 6a, where the peak associated with the dipolar magnetic term b_1 shows a shift opposite to that seen for the electric resonances. As in the previous case, the dipolar resonance suffers a larger shift (from $\mu = -2.0$ to $\mu = -1.75$) than the quadrupolar one (from $\mu = -1.5$ to $\mu = -1.38$) as particle size changes from $R = 0.01\lambda$ to $R = 0.1\lambda$. However, the difference between the dipolar (12.5%) and the quadrupolar (8%) shifts is much smaller than for the electric modes.

Concerning the FWHM in Figs. 6 and 7, it is clear that for every case we analyzed, the smaller the particle size, the narrower the resonance peak. In other words, the slope of the evolution of the width with particle size is positive for the four cases in Fig. 7. It is also interesting to remark from Figs. 6 and 7 that the dipolar resonances are broader than the quadrupolar ones. The electric quadrupolar mode, related to a_2 , is the narrowest. It is even difficult to observe for particle sizes lower than 0.05λ because the FWHM of this resonance is almost zero for these values of R .

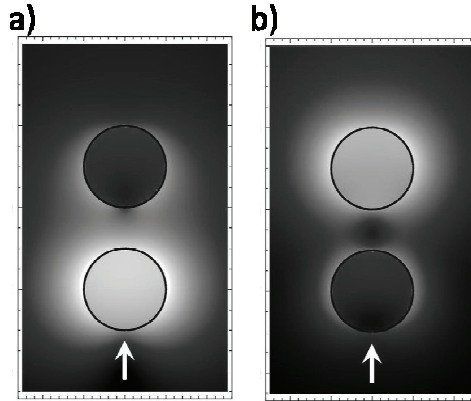


Fig. 9. Spatial distribution of the local field of a dimer with $R = 0.01\lambda$ and optical constants such that one particle scatters mainly in the forward direction [light particle, $(\epsilon, \mu) = (-2.01 + 0.1i, -2.01 + 0.1i)$] and the other particle scatters mainly in the backward direction [dark particle, $(\epsilon, \mu) = (-5 + 0.1i, -1.01 + 0.1i)$]. Bright areas correspond to high intensities of the local electric field, and dark areas correspond to low values. The arrow indicates the direction of the incident beam.

4. Applications of particles with arbitrary optical constants

Light scattering by particles with arbitrary optical properties, especially in the negative-negative range, has been the focus of several recent studies (Engheta 2007; Alù and Engheta 2009; Smith et al. 2004; Mirin and Halas 2009). Interest in such particles with unconventional optical properties lies mainly in the unconventional phenomena they suggest: negative refraction, invisibility cloak, etc. (Hess 2009; Pendry et al. 2006; Chen et al. 2010).

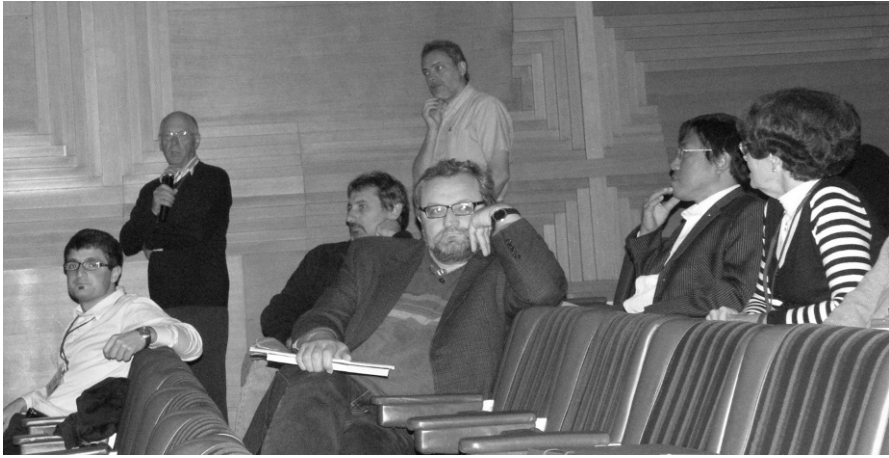
The possibility to tune the optical properties of the particle allows also a tuning of their scattering characteristics. In particular, the studies of Kerker et al. (1983) to consider directionally scattered light has been generalized and expanded (García-Cámara et al. 2010a; García-Cámara et al. 2008b). These directional effects, that were first proposed for the forward and backward directions and for dipolar particles, also have been found for finite particles and other scattering angles (García-Cámara et al. 2010b), as demonstrated in Fig. 8.

Aggregates of such particles with directional effects could prove useful for futuristic applications in the field of optical communications (Silveirinha et al. 2008; Gaylord et al. 2008; Miller 2010) or even to improve the characteristics of applications in the fields of nanolithography, microscopy or sensing (Acimovic et al. 2009; Ueno et al. 2010). For instance, through the interaction between these particles with unconventional optical properties, enhanced maxima or deep minima can be obtained, and resonators or anti-resonators can be designed (Fig. 9).

References

- Acimovic, S., M. P. Kreuzer, M. U. González, and R. Quidant, 2009: Plasmon near-field coupling in metal dimers as a step towards single-molecule sensing. *ACS Nano* **3**, 1231–1237.
- Almeida, V. R., C. A. Barrios, and R.R. Panepucci, 2004: All-optical control of light on a silicon chip. *Nature* **431**, 1081–1084.
- Alù, A., and N. Engheta, 2009: The quest for magnetic plasmons at optical frequencies. *Opt. Express* **17**, 5723–5730.
- Anker, J. N., W. P. Hall, O. Lyandres, et al., 2008: Biosensing with plasmonic nanosensors. *Nature Mater.* **7**, 442–453.
- Barber, P., and R. K. Chang, 1988: *Optical Effects Associated with Small Particles* (World Scientific, Singapore).
- Bohren, C., and D. Huffman, 1983: *Absorption and Scattering of Light by Small Particles* (Wiley, New York).
- Boltasseva, A., and V. M. Shalaev, 2008: Fabrication of optical negative-index metamaterials: recent advances and outlook. *Metamaterials* **2**, 1–17.
- Chen, H., C. T. Chan, and P. Sheng, 2010: Transformation optics and metamaterials. *Nature Mater.* **9**, 387–396.
- Engheta, N., 2007: Circuits with light at nanoscales: optical nanocircuits inspired by metamaterials. *Science* **317**, 1698–1702.
- García-Cámara, B., F. Moreno, F. González, et al., 2008a: Light scattering resonances in small particles with electric and magnetic properties. *J. Opt. Soc. Am. A* **25**, 327–334.
- García-Cámara, B., F. González, F. Moreno, and J. M. Saiz, 2008b: Exception for the zero-forward-scattering theory. *J. Opt. Soc. Am. A* **25**, 2875–2878.
- García-Cámara, B., J. M. Saiz, F. González, and F. Moreno, 2010a: Distance limit of the directionality conditions for the scattering of nanoparticles. *Metamaterials* **4**, 15–23.
- García-Cámara, B., J. M. Saiz, F. González, and F. Moreno, 2010b: Nanoparticles with unconventional scattering properties: size effects. *Opt. Commun.* **283**, 490–496.
- Gaylord, T. K., J. L. Stay, and J. D. Meindl, 2008: Optical interconnect devices and structures based on metamaterials. US patent No. 2008/0212921.
- Hess, O., 2008: Farewell to flatland. *Nature* **455**, 299–300.
- Krasheninnikov, A. V., and F. Banhart, 2007: Engineering of nanostructures carbon materials with electron or ion beams. *Nature Mater.* **6**, 723–733.
- Kerker, M., D. S. Wang, and C. L. Giles, 1983: Electromagnetic scattering by magnetic spheres. *J. Opt. Soc. Am.* **73**, 765–767.
- Ladd, J., A. D. Taylor, M. Piliarik, et al., 2009: Label-free detection of cancer biomarker candidates using surface plasmon resonance imaging. *Anal. Bional. Chem.* **393**, 1157–1163.
- Lin, T. -J., and M. -F. Chung, 2009: Detection of cadmium by a fiber-optic biosensor based on localized surface plasmon resonance. *Biosens. Bioelectron.* **24**, 1213–1218.
- Lorenz, L., 1890: Lysbevægelsen i og uden for en af plane Lysbølger belyst Kugle. *K. Dan. Vidensk. Selsk. Skr.* **6**, 1–62.
- Lorenz, L., 1898: Sur la lumière réfléchie et réfractée par une sphère (surface) transparente. In *Oeuvres scientifiques de L. Lorenz*, Tome Premier (Libraire Lehmann & Stage, Copenhagen), 403–529.
- Meier, M., and A. Wokaum, 1983: Enhanced fields on large metal particles: dynamic depolarization. *Opt. Lett.* **8**, 581–583.

- Merchiers, O., F. Moreno, F. González, et al., 2007: Electromagnetic wave scattering from two interacting small spherical particles. Influence of their optical constants, ϵ and μ . *Opt. Commun.* **269**, 1–7.
- Mie, G., 1908: Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen, *Ann. Phys.* **330**, 377–445.
- Miller, D. A. B., 2010: Are the optical transistors the logical step? *Nature Photon.* **4**, 3–5.
- Mirin, N. A., and N. J. Halas, 2009: Light-bending nanoparticles. *Nano Lett.* **9**, 1255–1259.
- Pendry, J. B., D. Schuring, and D. R. Smith, 2006: Controlling electromagnetic fields. *Science* **312**, 1780–1782.
- Prasad, P. N., 2004: *Nanophotonics* (Wiley, New York).
- Shalaev, V., 2007: Optical negative-index metamaterials. *Nature Photon.* **1**, 41–48.
- Silveirinha, M. G., A. Alù, J. Li, and N. Engheta, 2008: Nanoinsulators and nanoconnectors for optical nanocircuits. *J. Appl. Phys.* **103**, 064305.
- Smith, D. R., S. Schultz, P. Markos, and C. M. Soukoulis, 2002: Determination of effective permittivity and permeability of metamaterials from reflection and transmission coefficients. *Phys. Rev. B* **65**, 195104.
- Smith, D. R., J. B. Pendry, and M. C. Wiltshire, 2004: Metamaterials and negative refractive index. *Science* **305**, 788–792.
- Soukoulis, C. M., S. Linden, and M. Wegener, 2007: Negative refractive index at optical wavelength. *Science* **315**, 47–49.
- Sun, Y., and Y. Xia, 2002: Shape-controlled synthesis of gold and silver nanoparticles. *Science* **298**, 2176–2179.
- Ueno, K., S. Takabatake, Y. Nishijima, et al., 2010: Nanogap-assisted surface plasmon nanolithography. *J. Phys. Chem. Lett.* **1**, 657–662.
- Valentine, J., S. Zhang, T. Zentgraf, et al., 2009: An optical cloak made of dielectrics. *Nature Mater.* **8**, 568–571.
- Veselago, V. G., 1968: The electrodynamics of substrates with simultaneously negative values of ϵ and μ . *Phys. Usp.* **10**, 509–514.
- Videen, G., and W. S. Bickel, 1990: Light scattering resonances in small spheres. *Phys. Rev. A* **45**, 6008–6012.
- Wang, H., D. W. Brandi, P. Norlander, and N. J. Halas, 2007: Plasmonic nanostructures: artificial molecules. *Acc. Chem. Res.* **40**, 53–62.
- Yan, W., X. Feng, X. Chen, et al., 2008: A super highly sensitive glucose biosensor based on Au nanoparticles-AgCl@polyaniline hybrid material. *Biosens. Bioelectron.* **23**, 925–931.
- Zheludev, N. I., 2010: The road ahead for metamaterials. *Science* **328**, 582–583.



Working session of the NATO ASI. From left to right: Gergely Dolgos, Jim Hough, Stefano Bagnulo, Alexander Kokhanovsky, Michael Mishchenko, Wenbo Sun, and Zhanna Dlugach.



Nikolai Khlebtsov (left) and Victor Tishkovets at the first poster session.

Laboratory measurements of light scattered by clouds and layers of solid particles using an imaging technique

E. Hadamcik^{1*}, J.-B. Renard², A. C. Levasseur-Regourd¹, and J. Lasue^{3,4}

¹UPMC University Paris 06/CNRS, LATMOS-IPSL, 11 Bld D'Alembert, 78280 Guyancourt, France

²LPC2E/CNRS, 3A Avenue de la Recherche Scientifique, 45071 Orleans Cedex 2, France

³LANL, Space Science and Applications, ISR-1, Mail Stop D-466, Los Alamos, NM 87545, USA

⁴Lunar and Planetary Institute, 3600 Bay Area Blvd., Houston, TX 77058, USA

Abstract. The Optical PROPERTIES of Astronomical and Atmospheric Grains (PROGRA²) experiment was developed in 1993 to study the light scattered by dust clouds, with an emphasis on its linear polarization. The instrument was progressively improved. A short description of the PROGRA²-vis instrument operating in the visible is given in this chapter. In the framework of the experiment, solid particles are lifted using two different means: either by reduced gravity conditions during parabolic flight campaigns, or by a nitrogen draught in a ground-based laboratory setting. The second instrument, PROGRA²-surf, is used to compare the measurements, obtained with the PROGRA²-vis, of the light scattered by the same grains but deposited in layers on the ground. The latter configuration allows one to study planetary regolith analogs.

Various samples studied by each technique are presented and discussed. They are characterized by different types of constituent particles, including compact ones, either regularly or irregularly shaped, as well as aggregates of micrometer- and submicrometer-sized grains. We perform polarimetric comparisons between particle clouds and deposited layers and identify certain light-scattering tendencies *versus* particle size and absorption. Numerical simulations of the light-scattering processes allow us to cross-validate the experimental and numerical procedures.

We compare our laboratory results to those obtained by remote or *in situ* observations in order to provide a better understanding of the physical properties of solid particles in the Earth's atmosphere and in the solar system (e.g., cometary coma). Examples of such applications are also given.

Keywords: light scattering, linear polarization, comets, asteroids, solar system, aerosols, dust, experiment

* Corresponding author. E-mail: edith.hadamcik@latmos.ipsl.fr

1. Introduction

1.1. Context

Our knowledge of the physical properties of the solar system dust is limited to:

- Particles captured in space and studied on the ground, such as the Stardust particles collected in the coma of comet 81P/Wild 2 (see *Science* **314** (2006) for first results), the Hayabusa particles collected on the surface of asteroid 25143 Itokawa and delivered in the summer of 2010, and the interplanetary dust particles captured in the Earth's atmosphere at 17–19-km altitudes (e.g., Bradley and Brownlee 1986). This latter category contains particles ejected by comets and asteroids undergoing breakup, e.g., during collisions or disruptions (Jewitt et al. 2010). Compact particles have been assumed to be of asteroidal origin, but in the Stardust tracks, compact and fluffy aggregates were present (Hörtz et al. 2006; Burchell et al. 2008);
- *In situ* observations during encounters. The relatively low resolution of observations does not allow for the complete determination of the structure of the particles. Most results of the past missions came from chemical analyses, e.g., using mass spectrometers, like CHON measured by PUMA/Vega and PIA/Giotto (Kissel et al. 1986a,b), while the physical characteristics of the particles were determined indirectly by impact experiments [see, e.g., a summary of SP2/Vega by Mazets et al. (1986) and a summary of DIDS/Giotto by McDonnell et al. (1991)] and light-scattering measurements in the OPE/Giotto experiment (Levasseur-Regourd et al. 1986, 1999; Fulle et al. 2000). The Rosetta Mission is expected to help improve upon our current knowledge via results from the MIDAS instrument providing 3D imaging of particles from the nm to μm size range as well as information on the particle texture, shape, and size (Riedler et al. 2007), and from the Giada instrument yielding size distributions (Colangeli et al. 2007);
- Remote observations using different measurement techniques and numerical or experimental models to interpret the results. Space missions are expensive, rare, and limited to specific objects (e.g., periodic comets), while remote observations can be carried out on a large sample of objects in a systematic manner (e.g., sky surveys such as LINEAR or CATALINA).

The light scattered by dust particles is partially linearly polarized. The degree of linear polarization depends on the physical properties of the medium as well as on the wavelength and geometry of the observations. We are mainly interested in surface properties of, e.g., the Moon, asteroids, and cometary nuclei. Such properties are important in the development of space missions, wherein spacecraft and humans may be in contact with very fine dust materials, e.g., surface regoliths of the Moon, Mars, and asteroids.

The surfaces can be made of rocks, gravels, or regoliths having low packing densities. We are also interested in the properties of the dust particles in clouds

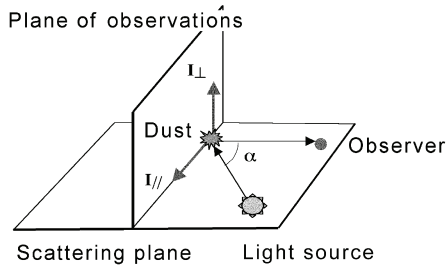


Fig. 1. Geometry of observations.

such as cometary comae, the interplanetary dust cloud, and aerosols in the Earth's and Titan's atmospheres. The particles can be compact as, e.g., sand particles in the terrestrial atmosphere (Renard et al. 2010), fragments of asteroids in the interplanetary dust (Grün et al. 1985), and compact particles in cometary comae (Burchell et al. 2008). They also can be fluffy aggregates made of submicrometer- or micrometer-sized grains, e.g., atmospheric soot (Renard et al. 2005; Francis et al. 2011), solid aerosols in the atmosphere of Titan (Cabane et al. 1993; Tomasko et al. 2008), and aggregates in cometary comae (Hörtz et al. 2006). The chemical composition varies, but silicates and carbonaceous compounds are generally found in varying ratios (Greenberg and Hage 1990; Jessberger et al. 2001; Ehrenfreund and Charley 2000). The existence of such compounds is confirmed by laboratory analyses of IDPs (Engrand and Maurette 1998; Matrajt et al. 2008; Flynn et al. 2006) and of the Stardust sample (Zolensky et al. 2006).

The goal of the work reviewed in this chapter (covering two decades of light-scattering studies) is to combine light-scattering observations, numerical simulations, and laboratory measurements in order to link the specific *in situ* analyses of solar-system objects to systematic, remotely observable properties and encompass a global view of the light-scattering properties of materials present in the Solar System.

1.2. Geometry of measurements and relevant equations

1.2.1. Geometry

All the measurements are made in the scattering plane (Fig. 1). The phase angle α is the angle between the direction toward the light source and that toward the detector (or observer) as seen from the scattering object.

1.2.2. Equations

If I_{\parallel} and I_{\perp} are the polarized components parallel and perpendicular to the scattering plane, respectively, then the degree of polarization P is the ratio of the difference to the sum of these components. By convention, P is negative when the

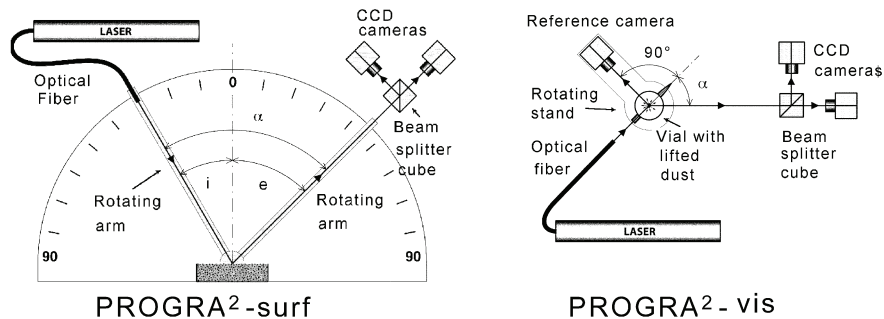


Fig. 2. Principles of the instruments: the SURF is projected onto the vertical plane, and the VIS is projected onto the horizontal plane.

parallel component is greater than the perpendicular component. The sum of the two components is the total scattered intensity in the direction of observation:

$$P = \frac{I_{\perp} - I_{\parallel}}{I_{\parallel} + I_{\perp}}, \quad (1)$$

$$I = I_{\parallel} + I_{\perp}. \quad (2)$$

2. Instruments

Two instruments are currently used (Fig. 2). PROGRA²-surf instrument (hereinafter SURF), is intended to study the light scattered by particles in layers deposited on a plane surface. The PROGRA²-vis instrument (hereinafter VIS), is designed to study the light scattered by clouds of particles levitated in reduced-gravity conditions or by a nitrogen draught (air draught before 2006). The imaging system was first described in Renard et al. (2002, 2005); details on the VIS can be found in Hadamcik et al. (2009a). The principles of the two instruments, along with supplementary descriptions, are explained below with an emphasis on various recent improvements.

2.1. Instrument principles

The light sources are randomly polarized He–Ne lasers operating at two wavelengths: 543.5 nm (green) and 632.8 nm (red). Optical fibers carry the light to the sample through collimator lenses. The laser-beam diameter is about 3 mm. A polarizing beam-splitter cube divides the light scattered by the particles into two components polarized perpendicularly and parallel to the scattering plane. Two cameras serve as detectors; photodiodes had been used in the SURF until the autumn of 2008. The two polarized components are recorded simultaneously, which is mainly important for the study of particle clouds in which the particles captured by the field of view change with time while floating through the cloud, thereby causing variations in the particle number, positions, and sizes.

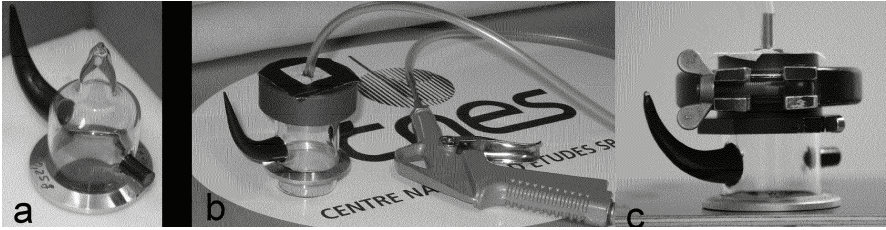


Fig. 3. Vials used for the study of lifted particles: (a) microgravity; (b, c) nitrogen-draught [(b) the previous vial; (c) the new vial with a new cap to better control the injection]. The light trap is on the left-hand side of the vials.

2.1.1. Camera characteristics

All five cameras are JAI/CV-A50's, with a sensitivity of about 0.03 lux. The SURF sensitivity is slightly better than that of the VIS cameras. A third camera is used with the VIS to normalize the intensities (see Section 2.3.2). As compared to the previous cameras in the VIS used between 2002 and 2008, the sensitivity is greatly increased, thereby enabling measurements for darker and smaller grains (e.g., for soot or very dark samples).

Telephoto lenses focus the cameras on the axis of the vial (VIS) or on the axes of the rotating arms (SURF). The phase angle ranges between $\sim 5^\circ$ and $\sim 160^\circ$. The resolution of the images is $10\ \mu\text{m}$, and the sensor of the cameras has 752×582 pixels. At a 90° phase angle, the horizontal field of view is $\sim 6.5\ \text{mm}$ and the vertical one corresponds to the diameter of the beam (2–3 mm). As the phase angle changes, the field of view varies as $6.5/\sin \alpha$ but is limited by the depth of the field. At $\alpha < 40^\circ$ and $\alpha > 140^\circ$, the maximum field of view is 10 mm. The error bar on α is 1° .

2.1.2. Dust containers

For cloud studies, the sample is placed in a cylindrical vial (Fig. 3). When the particles are levitated in reduced-gravity conditions, the vial is sealed with a low pressure inside (about 10^{-3} hPa), which helps minimize problems caused by humidity. When the particles are lifted by a nitrogen draught, a cap closes the vial. The nitrogen is dry and thus also reduces humidity-caused problems. It is injected at a controlled, low speed all around the cap and leaves the vial by the middle of the cap. The measurements are made after the injection, when the particles float in the vial before slowly settling down (see Section 2.2.2).

2.1.3. Phase-angle variations

For clouds levitated either in reduced-gravity conditions or by a draught, the vial is placed on a rotating tray to scan a range of phase angles. The light enters the vial through a plane surface fixed on an entrance tube sealed to the side of the

vial. The optical fiber injects the light perpendicularly to the plane surface and rotates with the vial. The detector system has a fixed position in the container box of the VIS.

For the SURF, the optical fiber is fixed on a rotating arm and the detector system is placed on another arm with a common axis. This allows one to choose the direction of the incident beam (i.e., the incidence angle i) and to change independently the direction of observation (i.e., the emergence angle e). Usually a mirror configuration is used (implying the phase angle $\alpha = 2i = 2e$), but other configurations are also possible. The sample surface is at the level of the rotation axis.

2.2. Advantages and limitations of the two levitation techniques

2.2.1. *Particles levitating in reduced gravity*

The successive phases of a typical profile of one parabolic flight on the dedicated A300-zeroG aircraft are as follows:

- the entry pull-up phase, during which the gravity is 1.8 g for 20 s;
- the injection phase, during which the gravity falls to about 0 g for less than 5 s;
- the reduced-gravity phase ($\pm 5 \times 10^{-2}$ g) lasting 22 s; and
- the pull-out phase, which is almost symmetrical to the pull-up phase.

A 2 min period at 1 g occurs between two parabolas. The pull up and pull down facilitate the adhesion of fluffy aggregates and small grains to the bottom of the vial. To loosen them, a mechanical device is used. Nevertheless, fluffy particles and often small grains form into large aggregates when levitating.

The number of parabolas in a flight is limited to 31, and 3 flights occur during three successive days of a campaign. The number of possible campaigns in a year is limited to between 1 and 3. This constrains severely the number of samples that can be analyzed in reduced gravity over a year and makes it difficult to perform redundant measurements in order to improve the statistics and decrease the error bars. During the reduced-gravity phase, the particles of any size, structure, and density float freely in the vial. The only restriction to their movement is the presence of the vial walls. However cumbersome it may be, this method is the only effective way to measure light scattering by large dust particles that would quickly sediment under terrestrial gravity conditions.

2.2.2. *Particles levitated by a nitrogen draught*

The use of a nitrogen draught to lift the small and/or low-density particles is preferable to air, which has a high level of oxygen and can have a high humidity content. Another reason to use nitrogen is that it allows one to use containers with controlled gas injection that reproduces similar conditions at all phase angles. When levitating, the particles float by convection and gravity-driven buoyancy with random speeds for several seconds (typically ~ 10 s) before settling down. The smaller and fluffier particles float for a longer time, allowing them to be stud-

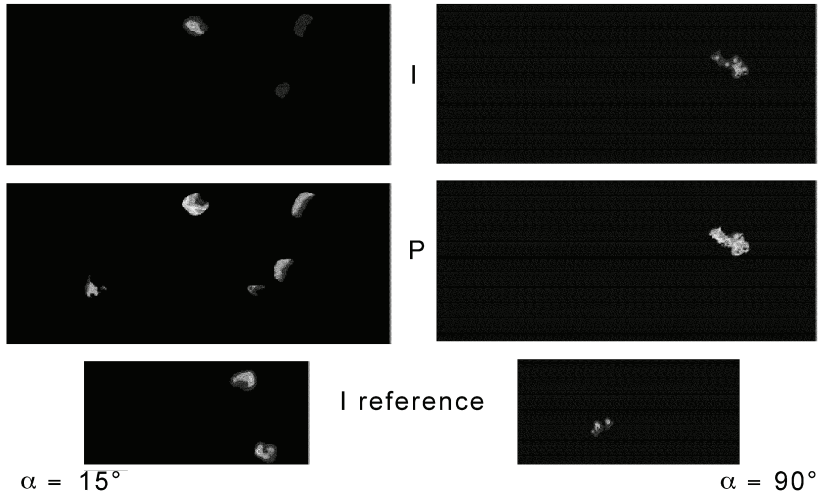


Fig. 4. Intensity (I), polarization (P), and reference intensity (I reference) maps for NaCl rounded compact crystals (after Mikrenska et al. 2006) and fluffy agglomerates of HCN polymers observed at two different phase angles.

ied separately from particles with diameters smaller than the resolution limit ($10\ \mu\text{m}$). With the nitrogen-draught technique, measurements can be repeated precisely as often as required, but large compact particles cannot be studied.

2.3. Advantages of the imaging technique

2.3.1. Polarization and intensity maps

The number of particles in the field of view is controlled on the images and allows one to select specific images with a small number of particles to study single scattering (thin medium) or multiple scattering (thick medium).

Intensity and polarization maps are built by using Eqs. (1) and (2) for each pixel of the polarized components. Such maps are shown in Fig. 4 for compact particles and grain agglomerates. For each series of measurements, a dark map without particles is built, which contains the offset and the eventual faint stray light, and is subtracted from all the images. In reduced-gravity measurements, the dark image is measured at each phase angle before the reduced-gravity phase. With the two instruments, the imaging technique allows one to reduce the field when contaminated by stray light. Saturated pixels (e.g., by specular reflection from flat facets) are removed from the analysis.

When the samples are made of large compact particles with diameters in the $20\text{--}300\ \mu\text{m}$ range, they appear as single particles, but when levitating in reduced-gravity, they can be agglomerated as grains in contact over small surface areas. When levitating in a nitrogen draught, the size distribution of the agglomerates is approximately Gaussian with a maximum in the $50\text{--}100\ \mu\text{m}$ range. When mi-

rometer- or submicrometer-sized grains are studied, they are lifted by the nitrogen-draught technique. Some large agglomerates may be observed in the field of view, but they fall out within a few seconds. Individual grains and aggregates smaller than $5\ \mu\text{m}$ still float and scatter the incoming light.

The imaging technique allows for the measurement of the polarization of the light scattered by single particles. The polarization maps for phase angles between 30° and 150° are automatically scanned, and the projected surface S of each particle is calculated in number of pixels independently of its shape. The equivalent diameter (in pixels) is

$$d = \sqrt{\frac{4S}{\pi}}. \quad (3)$$

Size ranges are defined in steps of some tens of micrometers, and a size distribution is given. For each size range, a polarization and a surface-normalized intensity value are calculated. To detect a significant variation of the polarization as a function of the particle size, it is necessary to measure the polarization for a large number of particles, usually more than a thousand. P values for clouds of particles selected by size can be obtained. Variations of these values can be indicative of the variation of P as a function of the surface-normalized intensity, which in turn varies with the porosity of the agglomerates (Hadamcik et al. 2002).

2.3.2. Polarization and intensity values

Since the degree of polarization is a ratio, it is not necessary to perform any normalization with the number of particles. The polarized intensities are added, and the polarization degree can be calculated. To obtain good statistics for the polarization values of a sample, numerous measurements are necessary (typically more than 20, and for some samples more than 100).

The number and size of the levitating particles change in the field of view with time, and, as a consequence, it is necessary to normalize the intensity. To do this, a camera is fixed to the tray at a 90° phase angle. The three cameras of the VIS are synchronized. The intensity function is given by I_p/I_r , where I_p is the surface normalized components as in Eq. (2) and I_r is the surface normalized intensity recorded by the reference camera.

3. Samples

It was already mentioned that the main motivation for the laboratory measurements is the study of the light-scattering properties of dust in the solar system. Therefore, samples that range in structure from compact single particles to very fluffy particles, with a variety of compositions analogous to what could be found in terrestrial and extra-terrestrial environments, are studied. Some samples can be used to study the polarization as a function of different physical properties of particles, such as size, porosity, refractive index, and shape. Spheres are used to improve the accuracy of the measurements and for comparison with numerical mod-

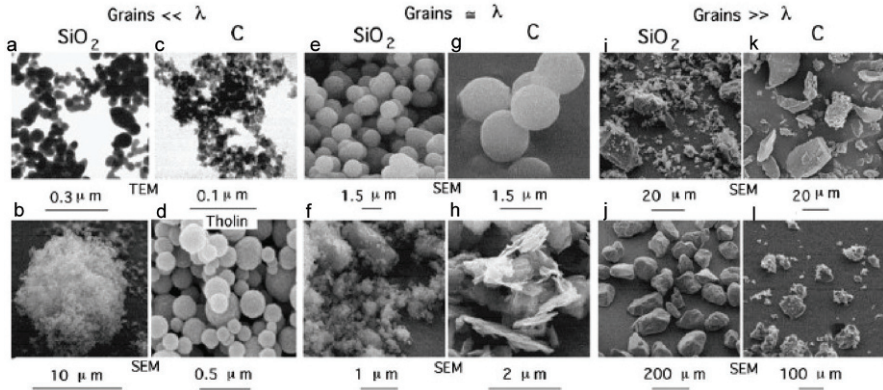


Fig. 5. Electron microscope images of grains in different size ranges. All samples are made of silica or carbon except image (d) for a carbonaceous compound (adapted from Hadamcik et al. 2009c).

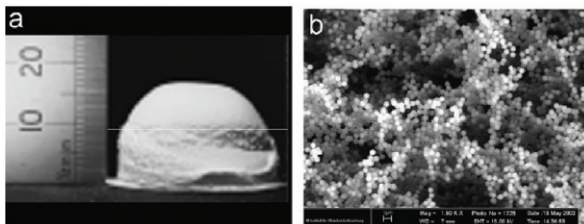


Fig. 6. Huge agglomerates and their surface produced by random ballistic deposition (adapted from Blum and Schröpfer 2004).

els. The size range of the grains is 10 nm – 500 μm . The sizes are measured using SEM or TEM images; more than 1000 data points are necessary to provide good statistics. Some grains are spherical, bare or coated; others are irregular with facets, and their edges can be rounded, like those of Gaussian particles (Muinonen et al. 1996), or sharp (e.g., for ground materials). Submicrometer-sized grains generally form micrometer-sized aggregates. Silica and carbon-black particles are produced in flames under different conditions, which defines their size and the structure of the aggregates that are of fractal type, like, e.g., soot. Some electron microscope images (TEM and SEM) of grains are presented in Fig. 5 in different size ranges. Some of the micrometer-sized grains form huge agglomerates produced by random ballistic deposition (Fig. 6 and Blum and Schröpfer 2004).

Layers of particles deposited on a horizontal surface may be sifted with a relatively low volume-filling factor (an average of 0.4 ± 0.1 for 1- μm spheres) and an irregular surface, or packed to increase the volume-filling factor of the grains (an average of 0.6–0.7 for 1- μm spheres). In that case the surface is relatively smooth

(mainly for grains smaller than 10 μm). For huge agglomerates, the volume-filling factor is between 0.12 and 0.20 for micrometer-sized spheres, and the surface is smoother than that for a sifted sample.

The silica particles are transparent, while carbon and carbonaceous compounds are highly absorbing. Silica and carbon are grey materials in the visible spectral domain, while the absorption of tholin and HCN polymers decreases with wavelength. Tables 1 and 2 list some characteristics of particles related to the present chapter.

4. Results

4.1. Spheres

Here we give examples of measurements performed for samples having different size ranges of spheres. Some samples are coated with an organic black compound to simulate the Greenberg model of interstellar grains (Greenberg and Li 1996). The intensity and polarization phase curves of light scattered by single spheres, with sizes close to or larger than the wavelength, exhibit typical Mie oscillations. Submicrometer-sized spheres form large aggregates which can be considered as irregular particles. For all the results, the fits take into account the aperture of the detectors (which causes a 1° uncertainty in the phase-angle values) and the size distribution of the spheres determined from SEM images. All the sizes are diameters. For grains larger than $\sim 50 \mu\text{m}$, the size distribution is controlled on the polarization maps. Single and small aggregates can be easily sorted out. The outputs of the best fits serve as a confirmation of the size or the size distribution of the levitating grains, and are also indicative of the values of the complex refractive index at each wavelength.

4.1.1. Spheres large as compared to the wavelength

Levitating 100- μm glass spheres: single scattering. Glass beads with a diameter of $100 \pm 25 \mu\text{m}$ are studied. The measurement data are compared to the results of Mie calculations (Fig. 7). The discrepancies between the fits and the data are mainly due to the sphere sizes and surface irregularities. The refractive indices implied by the best fits are $m_R = 1.52 + i8.5 \times 10^{-4}$ at 632.8 nm and $m_G = 1.52 + i6 \times 10^{-4}$ at 543.5 nm, where $i = (-1)^{1/2}$.

Levitating 100- μm glass spheres: multiple scattering. Polarization phase curves obtained for small agglomerates of 3 to 5 spheres (detected on the images) and for layers of spheres are compared (Renard et al. 2002; Hadamcik et al. 2003). The phase curves are presented in Fig. 8. The amplitudes of oscillations decrease as the number of spheres increases, thereby increasing the amount of multiple scattering between them. At phase angles greater than 130° , the increase of polarization is higher for the deposited sample than for the other ones, probably due to the effect of a smoother surface.

Table 1. Compact particles

Samples	Main components (%)	Mean diameter (μm)	Shape	$\text{Re}(m) + i \text{Im}(m)$
Glass beads	SiO_2 (71), Na_2O (14)	100 ± 25	Spheres < 1% spheroids	$1.52 + i7 \times 10^{-4}$
Coated glass beads	SiO_2 (71), Na_2O (14)+ Graphite C	Core 100 ± 25 Mantle 0.2 ± 0.1	Spheres < 1% spheroids Irregular	Core $1.52 + i7 \times 10^{-4}$ Mantle $2 + i3 \times 10^{-2}$
Silica	SiO_2	1.45 ± 0.06	Spheres < 2% spheroids	$1.48 + i10^{-4}$
Coated silica	SiO_2 + @Sicatar-black	Core 1.45 ± 0.06 Mantle 0.04 ± 0.01	Spheres < 2% spheroids	Core $1.48 + i10^{-4}$ Mantle $2.4 + i0.4$
Lunar analogue (JSC-1)	SiO_2 (43.5), Al_2O_3 (23.3), Fe_2O_3 (15.6)	100 ± 50	Rounded edges, some concave parts	$\cong 1.6$
Boron carbide	B_4C (crystals)	88 ± 5 13 ± 2 9 ± 1	Polyhedral, sharp edges <i>idem</i> <i>idem</i>	$2 + i0.04$ <i>idem</i> <i>idem</i>
Silicon carbide	SiC (99.4)	88 ± 8	Polyhedral, sharp edges	2.67 ± 0.3
Basalt	SiO_2 (50), Al_2O_3 (16), FeO (10), CaO (10),	110 ± 15	Polyhedral	$\cong 1.6$
Quartz	SiO_2	200 ± 100	Stochastically rough	$1.4 - 1.5$
Quartz	SiO_2	30 ± 20	Some sharp edges	$1.4 - 1.5$
Silica	SiO_2	$0.1 - 6$	Some sharp edges	$1.4 - 1.5$
Carbon	C (amorphous)	150 ± 100	Stochastically rough	$\cong 2$
Carbon	C (amorphous)	20 ± 10	Sharp edges	$\cong 2$
Graphite	C (100)	$(1 - 10) \times (0.1 - 2)$	Plates	$(2 \pm 0.5) + i(1 \pm 0.8)$
Mg-silicates	MgSiO_3 ($\cong 50$ crystals) Mg_2SiO_4 ($\cong 50$ crystals)	300 ± 150 (aggregates with 15 ± 12 crystals)	Polyhedral and aggregates of μm -sized crystals	$\cong 1$

Spheres

Irregular compact

Table 2. Fluffy particles

Samples	Main components (%)	Mean diameter (μm)	Shape	$\text{Re}(m) + i \text{Im}(m)$
Silica	SiO_2	0.040 ± 0.020 (grains) 8 \pm 6 (aggregates)	High porosity (HP) aggregates, amorphous	$\cong 1.45$
Carbon-black	C	0.095 ± 0.020 (grains) 5 \pm 4 (aggregates)	HP aggregates, amorphous	$\cong 2$
Carbon-black	C	0.014 ± 0.005 (grains) 7 \pm 3 (aggregates)	HP aggregates, amorphous	$\cong 2$
Mg-Silicates (5 \neq samples)	$\text{Mg}_x\text{Si}_y\text{O}_z$ mixtures	0.01 – 100 grains Average 0.05	HP aggregates, crystals, and amorphous grains	$\cong 1.6$
Fe-Silicates	$\text{Fe}_x\text{Si}_y\text{O}_z$ mixture	0.040 \pm 0.02	HP aggregates	$\cong 1.6$
Titan's analogues	$\text{C}_x\text{H}_y\text{N}_z$ 6 \neq samples	1.27 \pm 0.5, 0.68 \pm 0.35 0.27 \pm 0.15, 0.30 \pm 0.18 0.39 \pm 0.1, 0.19 \pm 0.06	HP aggregates, amorphous Radial growing	$\cong (1.6 - 1.7)$

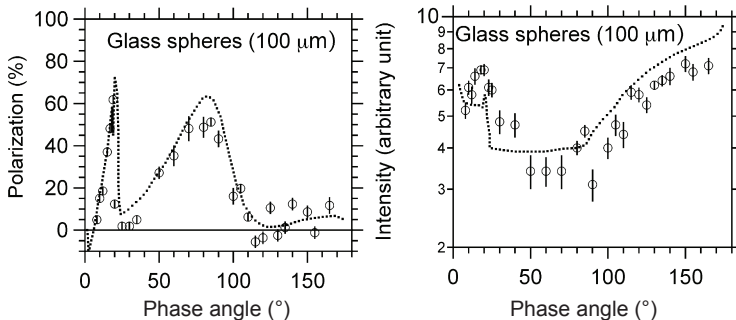


Fig. 7. Polarization and intensity for glass spheres as compared to Mie curves at $\lambda = 543.5$ nm (after Renard et al. 2002 and Hadamcik et al. 2003).

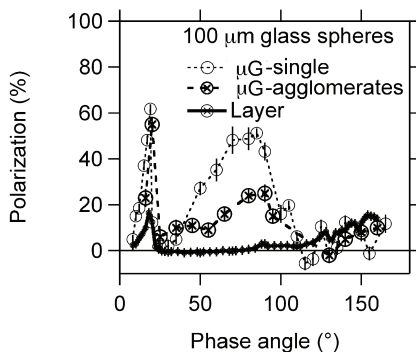


Fig. 8. Polarization phase curves for 100- μm glass spheres with and without interactions between them (after Hadamcik et al. 2003).

Levitating 100- μm glass spheres coated with graphite. Lasue et al. (2007) used a stratified sphere model (Toon et al. 1981) to fit the measured intensity and polarization from glass spheres coated with graphite. The model was adapted to take into account the size distribution measured from SEM images, as well as the specific parameters of the experiment. In order to estimate the confidence level of the numerical model fit to the data points, a χ^2 fit is calculated over the free parameters corresponding to the refractive index and thickness of the graphite mantle.

In Fig. 9, the phase curves are presented with the best fits in the red and green wavelength domains. The refractive indices of the core are those found for the bare spheres. The coating refractive index is close to $2 + i0.03$, which does not correspond to a highly absorbing material. The mantle thickness is $0.2 \pm 0.1 \mu\text{m}$. The graphite was deposited from a colloidal graphite solution, and the coating exhibits surface irregularities as seen in the SEM images (Fig. 10).

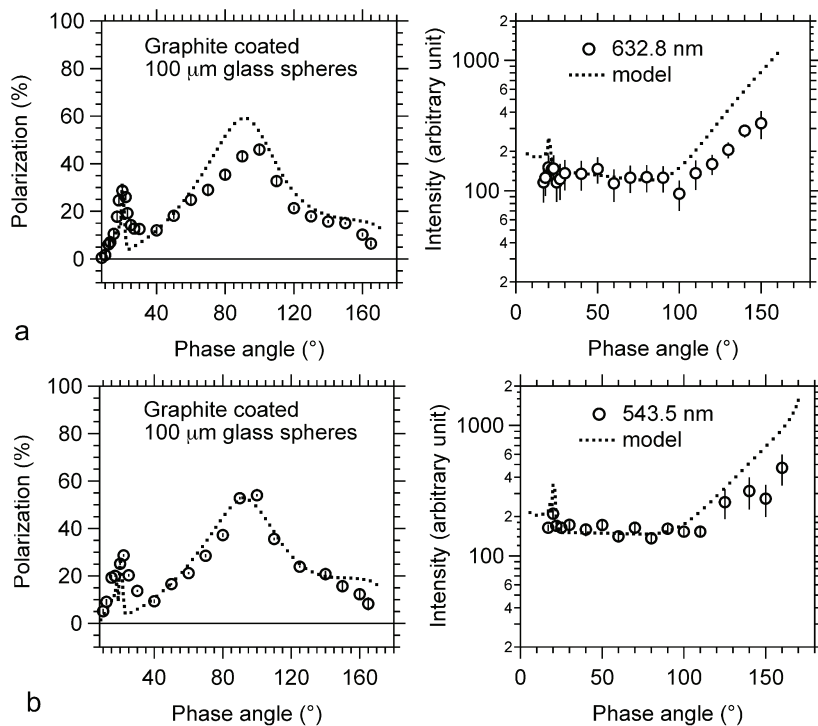


Fig. 9. Glass spheres with a graphite coating: (a) $\lambda = 632.8$ nm, core $d = 100 \pm 25$ μm , $m_c = 1.52 + i8.5 \times 10^{-4}$, mantle thickness = 0.2 ± 0.1 μm , and $m_m = 2 + i3 \times 10^{-3}$ (after Lasue et al. 2007); (b) $\lambda = 543.5$ nm, $m_c = 1.52 + i6 \times 10^{-4}$, mantle thickness = 0.2 ± 0.1 μm , and $m_m = 2 + i3 \times 10^{-3}$.

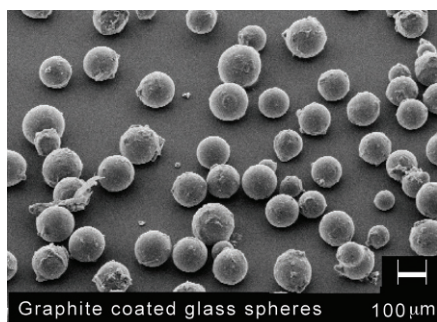


Fig. 10. SEM images of glass spheres with a graphite mantle.

4.1.2. Micrometer-sized bare silica spheres and spheres coated by a black mantle

Spheres with diameters between 1 and 2 μm have been studied to compare the light scattered by single spheres, small agglomerates of spheres, large agglomer-

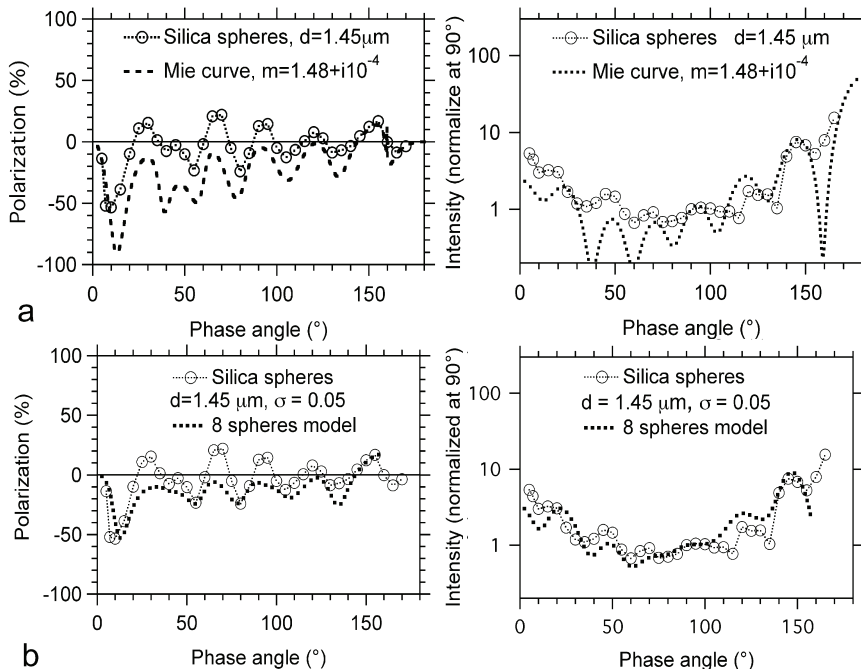


Fig. 11. Polarization and intensity phase curves at 632.8 nm for bare silica spheres. Best fits: (a) Mie curves for single spheres with $\sigma = 0.06$ for the size distribution; (b) a model of agglomerates with 8 constituent spheres (after Hadamcik et al. 2007a).

ates of spheres, deposited spheres, and huge agglomerates of spheres produced by random ballistic deposition. The purpose of these studies is to model the light scattering by spheres in a variety of general cases. Silica and silica spheres coated with a Sicastar-black® carbonaceous paint are studied.

Single scattering and multiple scattering by small agglomerates of bare silica spheres. Phase curves for silica spheres with a diameter of $1.45 \pm 0.15 \mu\text{m}$ lifted by a nitrogen draught are compared to Mie results (Fig. 11a). The polarization best fit obtained for single spheres with a size distribution measured from the SEM images is slightly below the data points, with negative values for $\alpha < 140^\circ$. The measured polarization oscillates around the mean value close to zero (at $\alpha > 30^\circ$). The phase-angle locations of the extrema on the polarization and intensity curves correspond to those of the measurements. In Hadamcik et al. (2007a), different size distributions are studied. When the size distribution is wider, the polarization is found to be less negative and the amplitude of the oscillations decreases. The fit also constrains the refractive index of the silica to $m = 1.48 + i10^{-4}$.

The experimental data are obtained with particles smaller than $5 \mu\text{m}$ made of the silica spheres, but the ratio and size of the agglomerates made of spheres is not known; it is statistically the same for each phase angle. The best fits in polariza-

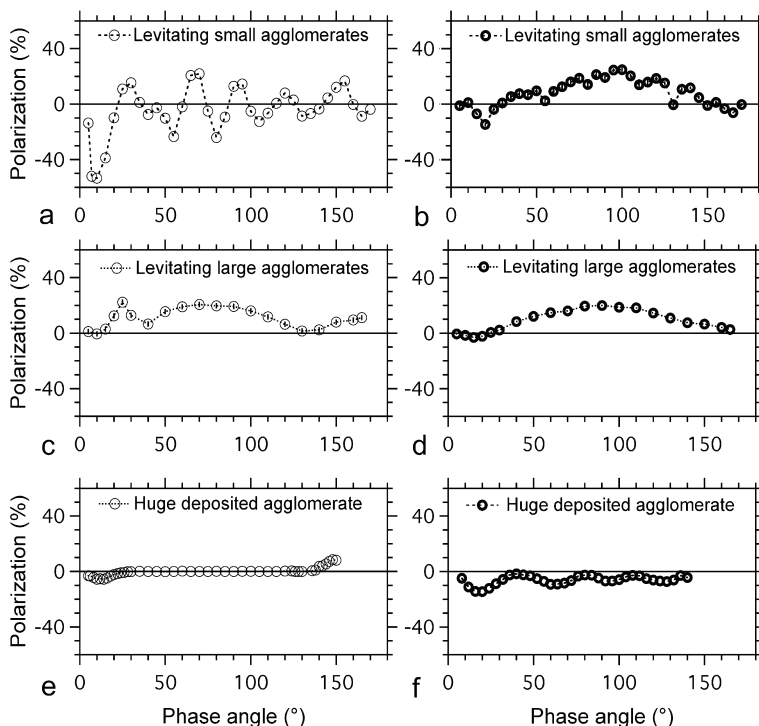


Fig. 12. Comparison between phase curves for small micrometer-sized agglomerates, large 50–100 μm sized agglomerates, and huge 2 cm \times 3 mm agglomerates deposited on a surface. (a, c, e) Bare silica spheres; (b, d, f) silica spheres coated with an organic black mantle (after Hadamcik et al. 2009c).

tion and intensity are obtained for 8 spheres and the size distribution with a dispersion of $\sigma = 0.05$ corresponding to an error bar on the size of 0.12 μm (Fig. 11b).

Multiple scattering in agglomerates of different sizes consisting of bare silica spheres and silica spheres coated by an absorbing mantle. A comparison of the polarization phase curves of agglomerates of different sizes are presented in Figs. 12a,c,e for bare silica spheres and in Figs. 12b,d,f for the same spheres but coated by a black mantle. The diameter of the silica spheres, as described in Section 4.2.1, is $1.45 \pm 0.15 \mu\text{m}$. The cores of the black-coated spheres have the same diameters and refractive indices.

Silica spheres. For transparent silica spheres, as the agglomerate size increases, the amount of multiple scattering increases while the amplitude of the oscillations decreases significantly at intermediate phase angles. For small agglomerates the Mie oscillations are present (Fig. 12a). At small phase angles, the phase curve for large levitating agglomerates (Fig. 12c) exhibits a very weak negative polarization

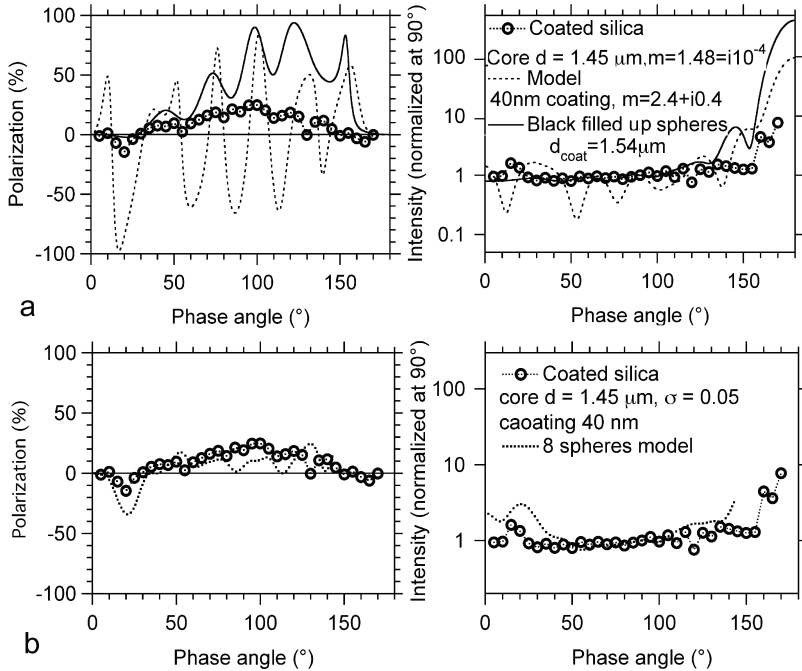


Fig. 13. Polarization and intensity phase curves at 632.8 nm for silica spheres coated by a Sicastar-black® mantle. Best fits: (a) model for single spheres with $\sigma = 0.05$ for the size distribution; (b) a model of 8 agglomerated coated spheres (after Hadamcik et al. 2007a).

branch, a spike at about 30° , a bell-shaped curve between 50° and 130° , and, finally, an increase of polarization for phase angles greater than 130° . For the huge deposited agglomerates (Fig. 12e), a small negative branch is observed at $\alpha < 30^\circ$; the polarization is close to zero between 30° and 130° and increases at $\alpha > 130^\circ$.

Coated spheres (Sicastar-black®). For coated spheres forming small agglomerates, a negative branch at $\alpha < 30^\circ$ and a positive branch between 30° and 140° with an overall bell shape and faint oscillations are present (Fig. 12b). For large levitating agglomerates (Fig. 12d) the polarization phase curve is similar to those measured for irregular particles, with a shallow negative branch at $\alpha < 30^\circ$ and a smooth positive bell-shaped branch at $30^\circ < \alpha < 130^\circ$. For huge agglomerates (Fig. 12f), the polarization is always negative, with oscillations typical of spherical grains. The presence of the absorbing black mantle decreases the amount of multiple scattering between the grains. Nevertheless the decrease of the amplitude of the oscillations corresponds to an increase of the amount of light scattered by the sample when the phase angle increases. Some interactions certainly exist between the spheres.

In Fig. 13a the best fits are obtained with a modified stratified-sphere model (Lasue et al. 2007) using single coated spheres; that in Fig. 13b is obtained for ag-

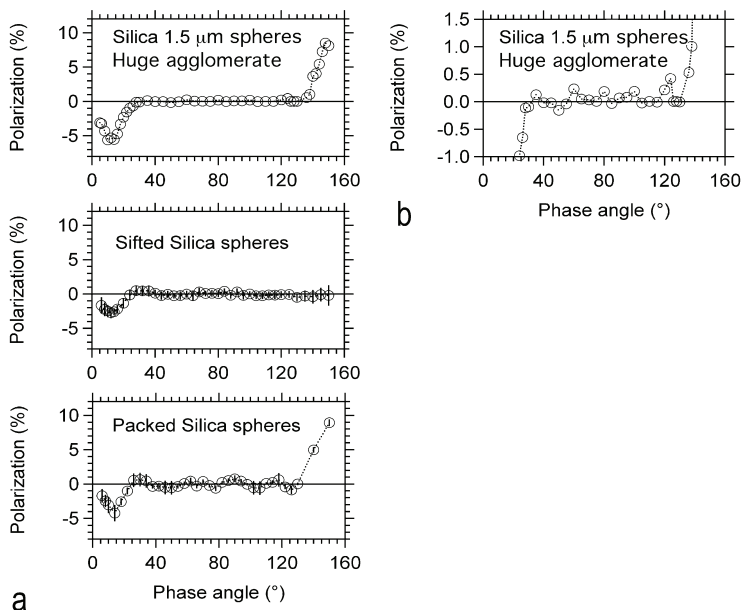


Fig. 14. Comparison between deposited particles. (a) Top panel: huge agglomerates; middle panel: sifted sample; bottom panel: packed sample. (b) Huge agglomerates but with a vertical scale increased to emphasize small oscillations at intermediate phase angles.

glomerates of 8 coated spheres with a size distribution for the core ($\sigma=0.05$). From the model, a width of 40 nm was found for the mantle with a refractive index $m=2.4+i0.4$. The best fits are obtained for the agglomerates of 8 spheres. The amplitude of the oscillations decreases with the number of constituent spheres in the agglomerates (not shown). When this number is very large, new oscillations appear. The positions of the extrema appear to be independent of the refractive index and mainly depend on the size of the spheres. In Fig. 13a, the Mie curve for plain spheres having the same diameter of the coated spheres of sicastar-black® material are also depicted. The positions of the extrema for the coated spheres correspond to the Mie curve of the plain sicastar-black® material.

Increasing volume-filling factor for deposited bare silica spheres. The purpose of this study is to demonstrate the influence of the filling-factor, or porosity, of a regolith on the light scattered by dust-covered surfaces, e.g., surfaces of asteroids. Is it necessary to build the sample by a random deposition in order to obtain an analog for the astrophysical surfaces, or is it possible to consider powdered samples? As a first approximation of the random deposition of dust particles, powdered samples deposited through a sieve are used. These require less material than samples built by random ballistic deposition, which are limited to micrometer-sized grains (Blum and Schröpfer 2004). The filling factor increases from huge agglom-

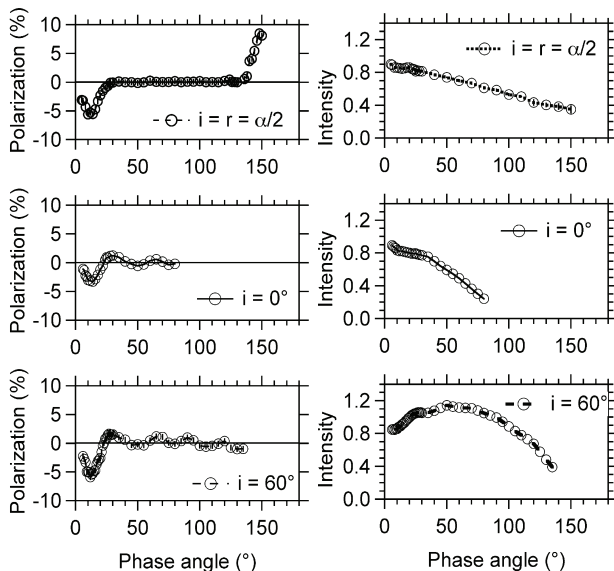


Fig. 15. Polarization and intensity phase curves of deposited huge agglomerates measured for different incidence angles at $\lambda = 632.8$ nm.

erates to sifted spheres and to packed spheres. As the volume-filling factor increases (the values are listed in Section 3), the amplitude of the polarization oscillations increases, while the extrema seem to occur at the same phase angles (Fig. 14a). In Fig. 14b, the vertical axis for huge agglomerates is stretched to show the presence of small-amplitude oscillations centered on the value zero at intermediate phase angles. The extrema positions are close to the previous ones, except perhaps between 95° and 125° . These oscillations are at the limit of the error bars, 0.1% to 0.2%. For each phase angle, about 200 measurements were necessary. Powdered, sifted samples represent a first-approximation analog to surfaces of asteroids.

Different incidence angles of light illuminating huge deposited agglomerates of bare silica spheres. To facilitate the eventual modeling of the light scattered by huge agglomerates, the incidence angle is fixed at two values $i = 0^\circ$ (normal incidence) and $i = 60^\circ$ (Fig. 15) to allow for the study of a larger domain of phase angles. When the emergence angle increases, the light path of the refracted light lengthens, and the amount of multiple scattering increases. The resulting differences are more pronounced in the intensity curves. In each curve a difference is noticed at $\alpha < 30^\circ$ (implying a change in the scattering regime). For $i = 60^\circ$, a maximum occurs when the emergence is normal to the surface.

4.1.3. Submicrometer-sized spheres in aggregates

The submicrometer-sized spheres form fluffy-structured aggregates similar to the fractal aggregates found, e.g., in soot. The size of the spheres is well defined

and ranges from 7 nm up to 100 nm with a relatively narrow size distribution, allowing one to study the dependence on the size of the constituent spheres and eventually on their packing density, or fractal dimension, in the aggregates. The typical size range of the aggregates is 1–10 μm . Two material components were mainly studied, silica and carbon, but also carbonaceous compounds such as tholins. When lifted, the aggregates are included in agglomerates with irregular structures, and the results are presented in the next section. The phase curves are smooth, with a shallow negative branch and a bell-shaped positive branch. The oscillations due to single spheres disappear completely for an agglomerate composed of thousands of submicrometer-sized grains.

4.2. Irregular particles

Remote observations of small solar-system bodies reveal phase curves typical of irregular particles. As shown by the different extra-terrestrial material samples collected *in situ*, the particles have different sizes, shapes, and structures and are made of transparent and absorbing materials. Series of measurements have been made on similar materials, like silica and carbon samples, and various minerals and organic materials. By compiling all these measurements, it is possible to determine the rate of change of the phase-curve parameters (like the amplitude of the positive branch and the depth of the negative branch) with the physical properties of the particles. In this section, some of these relationships are presented in the form of common rules.

4.2.1. Polarization phase curves

The different polarization phase curves for transparent silica and absorbing carbon particles of different size ranges are presented in Fig. 16a for submicrometer- and micrometer-sized grains and in Fig. 16b for grains large compared to the wavelength. For each size range, levitated and deposited silica and carbon grains are studied (Hadamcik et al. 2009c).

Amplitude of the positive branch P_{max} . For levitating particles composed of submicrometer-sized grains, the difference between transparent and absorbing materials is very small and depends on the average grain size. As the size of the transparent grains increases, the amplitude P_{max} decreases significantly and finally, for very large grains, increases slightly. The same behavior is observed for transparent deposited grains with a more significant decrease due to multiple scattering between the grains. For levitated and deposited absorbing particles, the decrease is less significant while the increase for very large particles is strong. The resulting value can even be greater than that for submicrometer-sized grains.

Phase angle of maximum polarization α_{max} . In the same size range, the phase angle α_{max} is smaller for transparent levitating particles than for absorbing ones. The opposite is found for deposited particles. For absorbing particles, α_{max} also increases with the grain size.

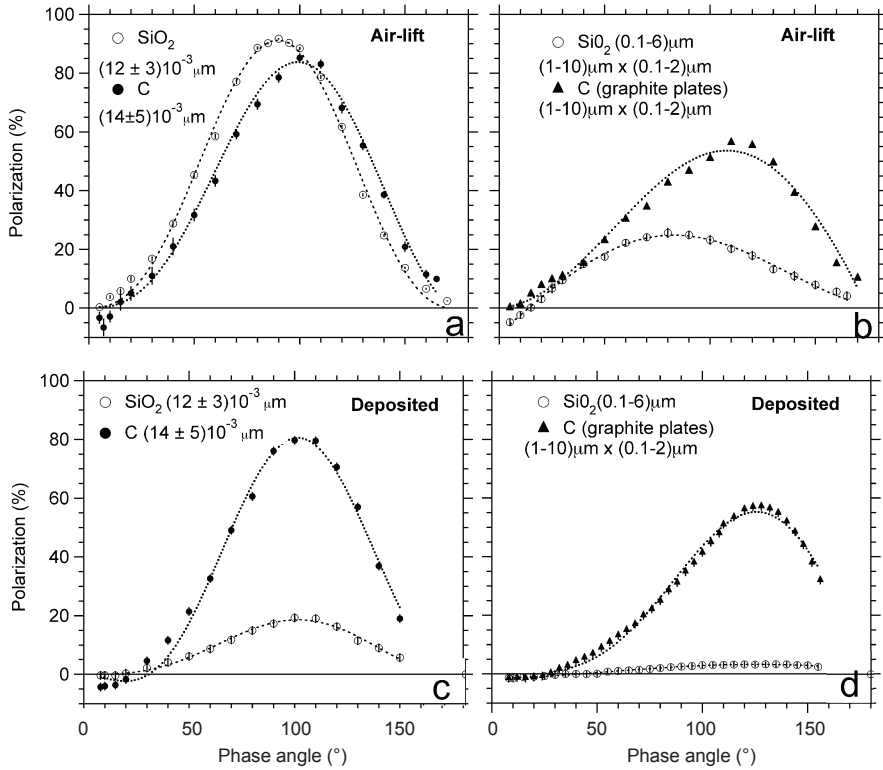


Fig. 16a. Polarization phase curves for (a–b) levitated and (c–d) deposited submicrometer- and micrometer-sized silica and carbon grains at $\lambda = 632.8$ nm.

Inversion angle α_0 . Deposited particles always exhibit a negative branch at small phase angles. This is not always the case for levitating particles. Except for submicrometer-sized grains, α_0 is smaller for levitating transparent particles than for absorbing particles. The opposite is observed for deposited samples. The inversion angle for levitated and deposited absorbing particles is between 10° and 30° when an inversion exists. The inversion angle for transparent levitating particles seems to be in the same range. For deposited transparent particles, α_0 may be shifted to values as large as 50° (e.g., for large $200\text{-}\mu\text{m}$ quartz particles).

4.2.2. Amplitude of the positive branch P_{max} as a function of the particle physical properties

To study the variation of P_{max} with size for both grains and agglomerates, we use the size parameter X defined as

$$X = \frac{\pi d}{\lambda}, \quad (4)$$

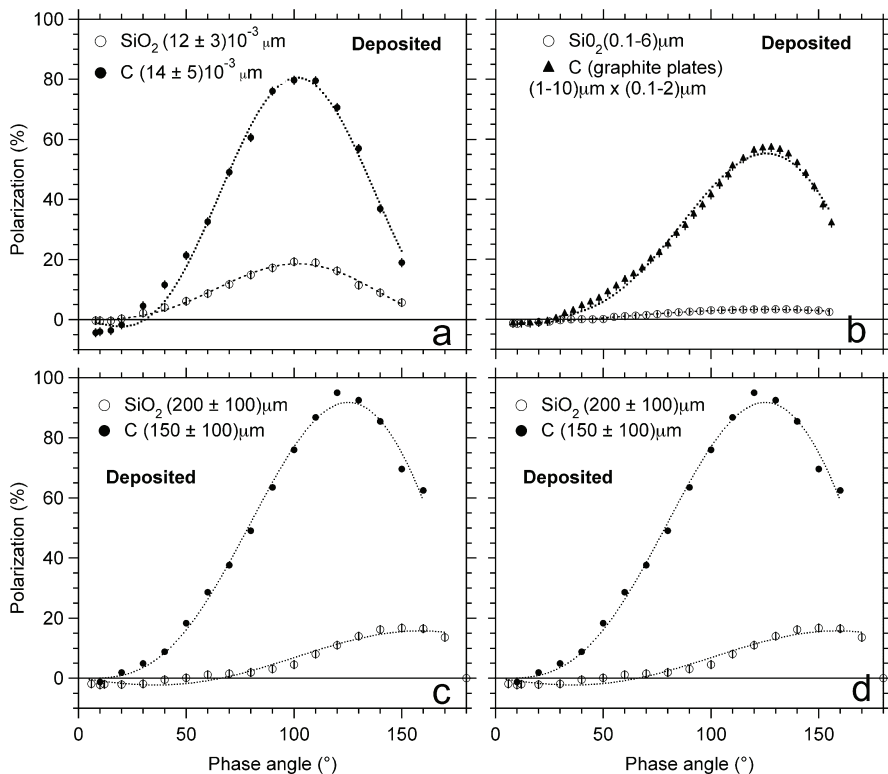


Fig. 16b. As in Fig. 16a, but for super-wavelength-sized grains.

where d is the equivalent diameter of a grain or a particle and λ is the wavelength of the incident light.

P_{max} as a function of grain size. Figure 17 shows variations of P_{max} with the equivalent size parameter of the grains for silica and carbon black (Fig. 17a) and for a dark brown tholin (Fig. 17b). The submicrometer-sized grains form agglomerates having approximately Gaussian size distributions and an average equivalent diameter of 50–60 μm . In Fig. 17a, the decrease of P_{max} is similar in green and red light; thus, the wavelength effect manifests itself only as a size effect for grey materials with refractive indices independent of the wavelength. For size parameters smaller than 0.1 (i.e., for diameters between 0.01 and 0.02 μm in the visible spectral domain), the amplitude of the linear polarization is similar for the transparent silica and the absorbing carbon black. For larger grains made of an absorbing material, the decrease of P_{max} with grain size tends toward a lower limit of 40%. A limit for the silica grains in fluffy aggregates is unknown owing to the lack of relevant samples. The fit of the data is quasi-linear for the silica grains and quasi-exponential for the carbon grains in fluffy agglomerates.

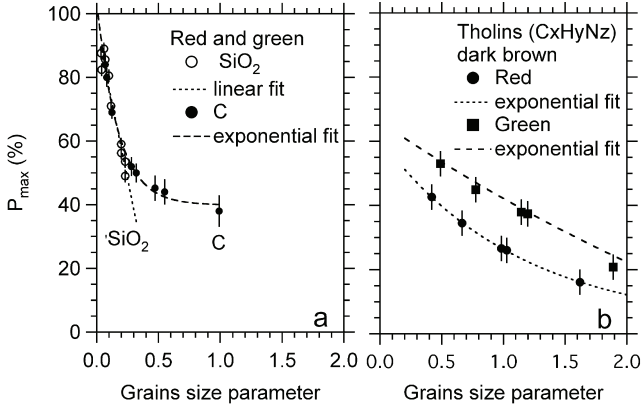


Fig. 17. P_{max} versus size parameter of submicrometer-sized grains. A size parameter of one is equivalent to a diameter of about 200 nm at 632.8 nm and 160 nm at 543.5 nm. Grains are in large levitated agglomerates: (a) grey materials; (b) an organic material with lower absorption at 632.8 nm than at 543.5 nm (updated from Hadamcik et al. 2009c).

Figure 17b shows the P_{max} for fluffy agglomerates of a dark brown absorbing tholin sample. P_{max} is greater in green light than in red light, thereby confirming the larger absorption at 543.5 nm than at 632.8 nm for this sample (Hadamcik et al. 2009b). The decrease of P_{max} with increasing size parameter for grains made of absorbing materials is in agreement with previous results (West et al. 1997; Petrova et al. 2004). The decrease of polarization with wavelength seems to be in contradiction with the usual variation for submicrometer-sized grains. This fact will be discussed in the following section.

Large compact particles can be observed individually with the VIS. The polarization variations observed for large particles, such as individual grains, are compared in Fig. 18 to the polarization produced by smaller grains (Fig. 18a for levitating grains and Fig. 18b for deposited grains). A logarithmic scale is used for the sizes. Up to a size parameter ~ 1 , the difference in polarization between levitating silica and carbon grains is small. For carbon grains with size parameters larger than ~ 1 , the amplitude of the positive branch increases up to values close to those measured for very small grains having size parameters smaller than 0.1; nevertheless, the rate of the increase is small for size parameters larger than ~ 250 (i.e., grain sizes $\sim 40 \mu\text{m}$ in the visible spectral domain), as also seen in Fig. 17a for carbon. For silica grains, the amplitude of the positive branch P_{max} continues to decrease for size parameters up to ~ 300 and then increases slightly for larger size parameters.

In Fig. 18b, deposited grains are considered. For silica grains, due to multiple scattering between the grains, the values of P_{max} are smaller than those found for particle clouds. For absorbing carbon grains, the values of P_{max} measured for deposited grains are similar to those found for clouds of carbon grains; i.e., they are slightly smaller for submicrometer-sized grains and slightly higher for large grains. The polarization decreases as the size parameter increases up to ~ 10 for

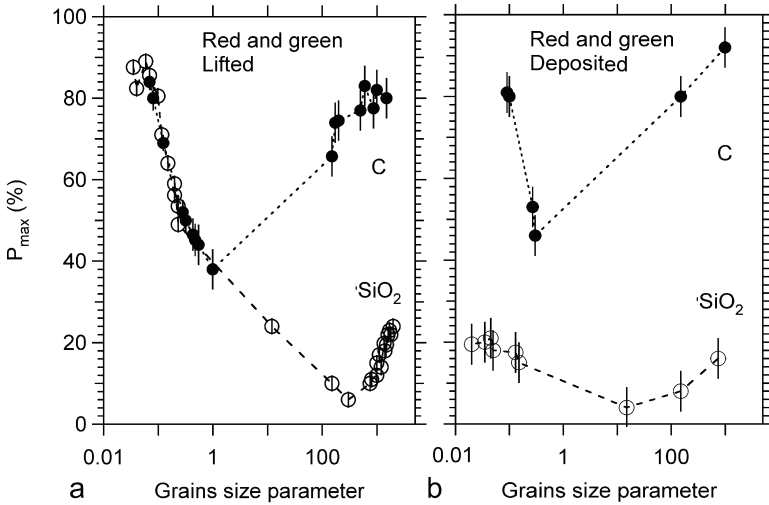


Fig. 18. P_{max} versus size parameter of grains (small grains in agglomerates, large grains as individual particles) made of transparent silica and absorbing carbon black (updated from Hadamcik et al. 2009c).

silica grains and ~ 1 for carbon grains and then increases for larger size parameters.

For grey materials, increasing the wavelength is equivalent to decreasing the grain size; i.e., the results is a decrease of P_{max} for size parameters smaller than the value corresponding to the minimum on the curves and an increase of P_{max} for greater size parameters.

P_{max} as a function of the size parameter of levitating particles. For agglomerates of grains and for compact particles (individual grains), the behavior of P_{max} depends not only on the grain sizes but also on the agglomerate sizes. For fluffy agglomerates, as seen in Section 4.2.2, the size distribution is usually Gaussian with a maximum between 50 and 60 μm . The smaller particles were not detected on the screen with the previous cameras but they may have had an influence. With the new cameras, smaller particles are detected and are taken into account in the polarization calculations. If these small particles exist in large numbers, the polarization may change, and comparisons between recent samples and previous samples must be done with caution by considering the two size distributions and the variation of polarization phase curves for different particle sizes. Nevertheless, this effect is usually small as compared to the grain-size influence.

Figure 19 shows the variation of P_{max} for transparent materials having different grain sizes. For fluffy particles (with porosities greater than 90%) composed of submicrometer-sized grains, P_{max} decreases with the size parameter of the agglomerates. The opposite behavior is observed for Mg-silicates with constituent grain sizes between 1 μm and 10 μm and compact quartz grains displaying an increase of P_{max} followed by near-constant values.

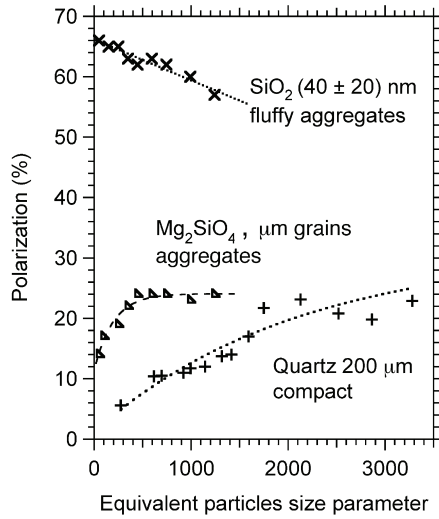


Fig. 19. P_{max} versus size parameter of levitated transparent particles with different structures and sizes of the constituent grains (after Hadamcik et al. 2006).

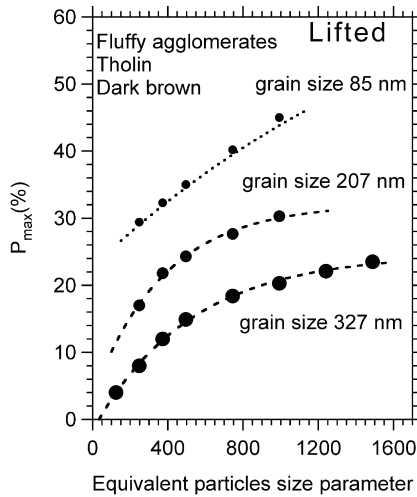


Fig. 20. P_{max} versus size parameter of levitated fluffy aggregates made of a carbonaceous material and having different size distributions of the constituent grains ($\lambda = 632.8$ nm) (after Hadamcik et al. 2009b).

Generally for absorbing materials, fluffy or not, the positive polarization branch increases with the particle size. Figure 20 shows the variation of P_{max} for an absorbing carbonaceous compound in fluffy agglomerates. Tholins are considered analogs of Titan’s solid aerosols. Three different samples are made of the

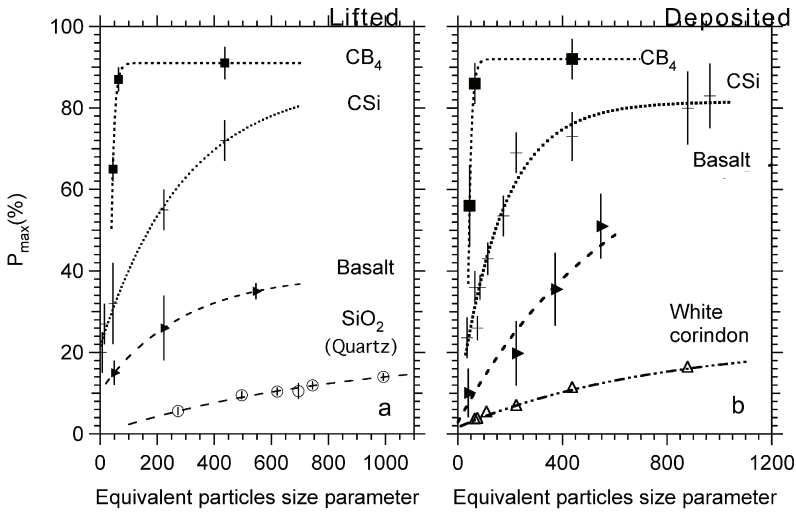


Fig. 21. P_{max} versus size parameter of levitated and deposited compact particles. Absorption increases from quartz or corindon to basalt, silicon carbide, and boron carbide.

same materials with different size distributions of the constituent grains: 80 ± 30 nm, 207 ± 30 nm, and 315 ± 185 nm. As expected, the polarization increases when the constituent grain size parameter decreases. The increase in polarization with the agglomerate size parameter seems to stabilize when the size parameter is sufficiently large; this stabilization is reached for smaller sizes when the size of the constituent grains increases. Figure 21 shows the amplitude of the positive branch for compact particles with increasing absorption as a function of the size parameter of the particles. Figure 21a depicts the results for levitating particles in reduced gravity and Fig. 21b for deposited particles. Except for the two transparent samples (i.e., levitating quartz particles and deposited white corindon particles), the absorbing samples are made of the same materials for the two series of studies. The overall variations with size are similar for levitating and deposited samples, with an increase of P_{max} with increasing size of the particles and near-constant values for large particles. The maximum for each sample is obtained for smaller particles when the material is more absorbing and the P_{max} values are greater.

Hadamcik et al. (2009c) suggested tentative interpretations of the observed behavior of the positive polarization when the size of large compact or fluffy absorbing particles increases. For increasing size parameters of individual grains or of constituent grains in agglomerates, the amplitude of the positive branch increases up to a maximum value for transparent and absorbing materials. When the size of compact particles increases, more refracted light is absorbed, and the polarization increases. When all the refracted light is completely absorbed, P_{max} remains constant. A similar trend is observed for levitated and deposited samples. As the size of large fluffy agglomerates of absorbing materials increases, more

light is absorbed. The scattered light is dominated by single scattering on grains in the external regions of the particles.

5. Applications to atmospheric physics and astrophysics

5.1. Atmospheric dust

Aerosols in the Earth's atmosphere are mainly droplets, but solid particles are also present. They can be, e.g., residues of combustions, such as soot from natural fires or industrial activities with a concentration of a few particles per cm^3 (Renard et al. 2008). Volcanic ashes are also present at different atmospheric levels. For example, the eruption of the Eyjafjöll volcano in Iceland in the spring of 2010 disturbed the air-traffic in Europe with ash in the troposphere. The eruption in 1991 of Mt. Pinatubo was the most violent and destructive volcanic event of the 20th century that sent ash into the stratosphere that remained there for several years (Russell et al. 1996). Interplanetary micrometer-sized dust particles and meteorites are also found, but their concentration is about 10^{-4} cm^{-3} . Mineral dust, such as sands, can be transported over large distances (Peyridieu et al. 2010), and polarization can help characterize these solid components. A database has been built with the PROGRA² experiment and progressively enlarged to include new samples. The database is located at <http://www.icare.univ-lille.fr/progra2>. Results for deposited particles are also incorporated in the database for comparison.

5.1.1. Sands

Experiments on different sands with large (100–250 μm on average) levitating grains in microgravity as well as deposited on a surface were integrated into the database to help interpret observations from, e.g., balloon-borne experiment such as MicroRadibal (Brogniez et al. 2003; Renard et al. 2008, 2010). Some micrometer-sized ground sands were also studied and compared to clay. The color of the sands is beige to reddish, except for one black sample.

The shapes of the polarization phase curves are typical for deposited samples with large grains, with a negative branch at relatively small phase angles. The inversion angles range from 30° to 60° . The amplitudes of the positive branch are between 16% and 26%. Smaller inversion angles correspond to smaller amplitudes. The phase angle of maximum polarization ranges between 140° and 160° . The amplitude of the positive branch is smaller for levitating particles, with values between 5% and 15%. The phase angle of maximum polarization is $\sim 90^\circ$, while the inversion angle is $\sim 10^\circ$.

Average phase curves have been deduced from the experimental results to define typical phase curves for deposited and levitated sands with hundreds of micrometer-sized particles (Fig. 22a). In this figure, error bars indicate the scatter of the data between the different samples. The difference between the red and green spectral domains is generally negligible. The amplitude of the positive branch is slightly greater for deposited samples with a phase-angle shift toward larger values (150° – 160°) for deposited particles. For micrometer-sized particles, the main

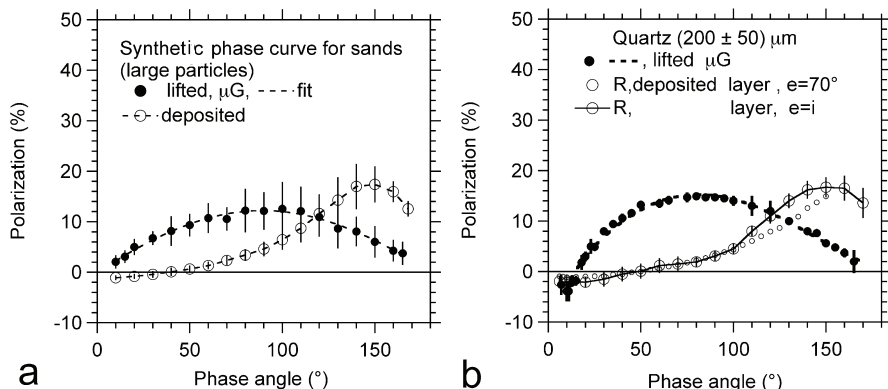


Fig. 22. Comparison of polarization phase curves for levitated sand particles (single scattering) and deposited particles (multiple scattering). (a) Averaged phase curves for different sands (error bars corresponds to the scatter of the data between sands); (b) phase curves for white sand (quartz). Particle sizes range from 100 to 300 μm .

difference is the amplitude of the positive branch, which decreases with increasing amount of multiple scattering between the particles. The phase angles of maximum polarization are close to 90° for both levitated and deposited particles.

Figure 22b depicts polarization phase curves for white quartz sand with particles sizes ~ 200 μm . They are remarkably similar to the average curve, except that the phase angle of maximum polarization (14%) is $\sim 80^\circ$ for levitating particles. Two other curves are also shown for deposited particles with an emergence angle of 70° to enable a comparison with the measurements by Grynko and Shkuratov (2008) for the same sand. The three phase curves are similar at phase angles smaller than 110° , where the slope of increase of polarization is steeper in the mirror configuration up to $P_{max} = 17\% \pm 1\%$ at a phase angle of $155^\circ \pm 5^\circ$. Shkuratov et al. (2007) compare single scattering by different particles levitated by an air jet to light scattering by the same but deposited particles. They find similar differences between the phase curves and their parameters. Multiple scattering between the particles decreases the depth of the negative branch and the amplitude of the positive branch mainly for transparent particles.

5.1.2. Soots

Soot particles are fluffy aggregates with a fractal structure. Depending on their origin and formation conditions, they are made of aggregated carbon grains with different carbonaceous compounds. Preliminary studies can be found in Renard et al. (2005). Carbon black is composed of aggregated carbon grains. They result from well-controlled burning processes intended to produce, for example, toner for printers. Their content of carbonaceous compounds is smaller than 1% or 2%.

Like other types of soot, they can be found in the Earth's atmosphere. Soot or carbon black are absorbing materials, but their albedo depends on the size of the constituent grains and on their number density in the aggregate which is related to their fractal dimension. A study of four soot and two carbon black samples is described in Francis et al. (2011). Depending on their production conditions, the size of the constituent grains and the fractal dimension of the aggregates can vary, which is expected to facilitate the eventual development of optical characterization techniques based on polarization phase curves. The increasing amplitude of the positive branch for absorbing materials with increasing size of the agglomerates is confirmed (see Fig. 20). Francis et al. also confirm that P_{max} for deposited absorbing samples can be greater than for the same but levitating grains (cf. Figs. 16a and 18).

5.2. Solar-system dust

To help interpret remote observations of solar-system bodies it is necessary to use laboratory analogs. Some meteorites are also available in sufficient quantities. Measurements with the VIS need relatively small amounts of material (0.1 to 0.3 g); larger quantities (1 or 2 g for surface studies) are required with the SURF. Phase curves of brightness and polarization as well as polarization variations with the wavelength can be compared to results obtained by remote observations.

5.2.1. *Analogs for deposited dust on planetary surfaces*

Deposited in layers, the JSC1 Martian- and Lunar-analog polarization phase curves are compatible with the phase curves obtained by remote observations (Worms et al. 2000; Hadamcik et al. 2003).

Recently, Hadamcik et al. (2011) compared the polarization phase curves and the spectral behavior of CV3 and CO3 powdered meteorites in different size distributions to observations of asteroid 21 Lutetia. From the light-scattering results, the powdered CV3 meteorite (Allende) seems to be a good analog. The powdered CO3 (NWA 4868) Saharan meteorite phase curve is also comparable, but exhibits an inverse wavelength variation of polarization. From the experimental measurements, an average size distribution ($< 50 \mu\text{m}$) has also been derived, which is compatible with infrared spectral estimates. In an international coordinated study (McFadden et al. 2009), the light scattered by an aubrite meteorite from Antarctica was compared to the remotely observed properties of 2867 Steins and, more generally, of E-type asteroids. The polarization phase curves are compatible for the smaller grain sizes (Hadamcik et al. 2010; Levasseur-Regourd et al. 2011). The variation of polarization as a function of size for levitating particles and the similarities of variation for the same but deposited particles (Fig. 23) allow one to estimate the size when the amplitude of the positive branch is known. As for 21 Lutetia, an average size of $< 40 \mu\text{m}$ is derived when considering the amplitude of the positive branch exhibited by the synthetic phase curve. This result is compatible with retrievals based on infrared measurements.

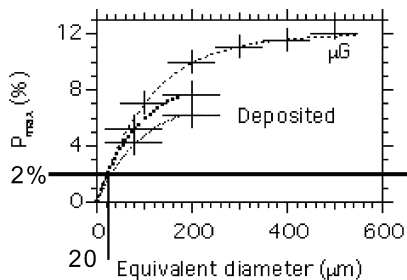


Fig. 23. Polarization *versus* size for levitated aubrite grains ($\lambda=632.8$ nm) and deposited grains with two size distributions ($\lambda=543.5$ nm in the G band and 632.8 nm in the R band).

Very large agglomerates produced by random ballistic deposition may be considered as regolith analogs when their constituent grains are irregular or as proto-planetary analogs (Blum and Schr apler 2004; Blum et al. 2006). Their volume-filling factor is in the 0.1–0.2 range, depending on the sample. A relatively large amount of material is necessary to produce usable samples, and such amounts are usually not available from meteorites. To evaluate the optical interactions between the constituent grains in deposited layers, the samples also are levitated in reduced-gravity or nitrogen-draught conditions to derive the phase curves of single particles and of small or large agglomerates of grains (Worms et al. 2000; Hadamcik et al. 2007a, 2009c, 2011).

5.2.2. *Analogs of cometary dust particles*

The variations of polarization with the physical properties of the dust, such as size, porosity, or albedo for the different kinds of particles (compact or fluffy, transparent or absorbing) are used to interpret remote-sensing observations.

Cometary comae are extended objects. The dust and gas emissions are observed in the same coma regions. To retrieve the dust properties, it is necessary to use specific filters in order to avoid contamination by the gaseous species in the continuum. Numerous observations have been made by integrating polarized intensities over apertures of different sizes. Observations conducted at different phase angles and wavelengths yield averaged dust properties. With the imaging technique, polarization maps can be built. The differences in polarization between coma regions indicate differences in physical properties of the dust in these regions. The imaging technique allows for the study of the coma radial variations in specific directions and also the evolution of the polarization and intensities.

For example, jets observed in comet C/1995 O1 (Hale–Bopp) and in active comets exhibit stronger polarization than other regions (Hadamcik and Levasseur-Regourd 2003a, b). The jets and arcs are rich in submicrometer-sized grains compared to the background. This interpretation has been confirmed by other diagnostic observations, such as more pronounced infrared emission features owing to the presence of silicates in these regions (Hayward et al. 2000) or the bluer color of

the intensity in the jets (Furusho et al. 1999). Such particles were also found in the dust ejected into the coma of comet 9P/Tempel 1 by the Deep Impact event (Harker et al. 2007; Hadamcik et al. 2007b). Before the event, the polarization evolution was correlated to the nucleus rotation. High polarization was also measured in the inner regions before and some days after the event. The results obtained by PROGRA² on large absorbing particles (see Section 4.2 and Fig. 19) correlate with the extension of that region on the images and with the intensity decrease that still allowed for large, relatively compact absorbing particles (and eventually aggregated grains) covered by carbon or carbonaceous compounds. Farnham et al. (2007) suggested the existence of large, slowly moving particles in that region.

When particles break up, the polarization and intensity evolution in a coma may be indicative of the particle structure. Large, relatively compact fragmenting particles are suggested by the significant decrease in polarization with increasing aperture around the nucleus of comet C/2000 WM1 (LINEAR) (Hadamcik et al. 2003) and around the main debris in comet C/1999 S4 (LINEAR) (Hadamcik and Levasseur-Regourd 2003c). In the coma of comet 67P/Churyumov–Gerasimenko, the target of the Rosetta mission, Hadamcik et al. (2010a) found very large particles before perihelion with no evident structures in the polarization maps, and fine grains in fluffy aggregates in the month following perihelion. This last region appeared on the maps with a greater polarization around the nucleus.

When the aperture is increased, the polarization varies, depending on the structures within the field. For sufficiently large apertures including the main features, the polarization remains constant; the coma inside this aperture is called the “whole coma” and a polarization value can be attributed to the comet at the defined phase angle. From the whole-coma studies of various comets at different wavelengths, typical polarization phase curves have been found. The particles inside originate from different regions and at different times; they have different physical properties, and averaged characteristics can be deduced. The samples brought by Stardust have confirmed that a significant fraction of the dust particles are fluffy aggregates made of submicrometer- or micrometer-sized constituent grains, but compact particles are also found (Hörtz et al. 2006). The particles analyzed consist of a mixture of silicates (mainly Mg-silicates) and carbonaceous compounds; some iron components are also present. This composition is consistent with the composition measured in comet Halley (Jessberger et al. 1988). When a comet exhibits significant jet activity, the size distribution of the grains is shifted toward smaller grains with the presence of Mg-silicates crystals (Hanner and Bradley 2004).

Cometary analogs made of Mg-silicates and Fe-silicates were provided to us by the NASA Goddard Space Flight Center (Nuth et al. 2002). The average size of the constituent grains was 0.05 μm . The polarization phase curves of agglomerates in the 50–100 μm size range were studied in different mixtures at two wavelengths (543.5 and 632.8 nm) with the VIS. Carbon black with submicrometer-sized grains was mixed with silicates. Some of the phase curves thus obtained are

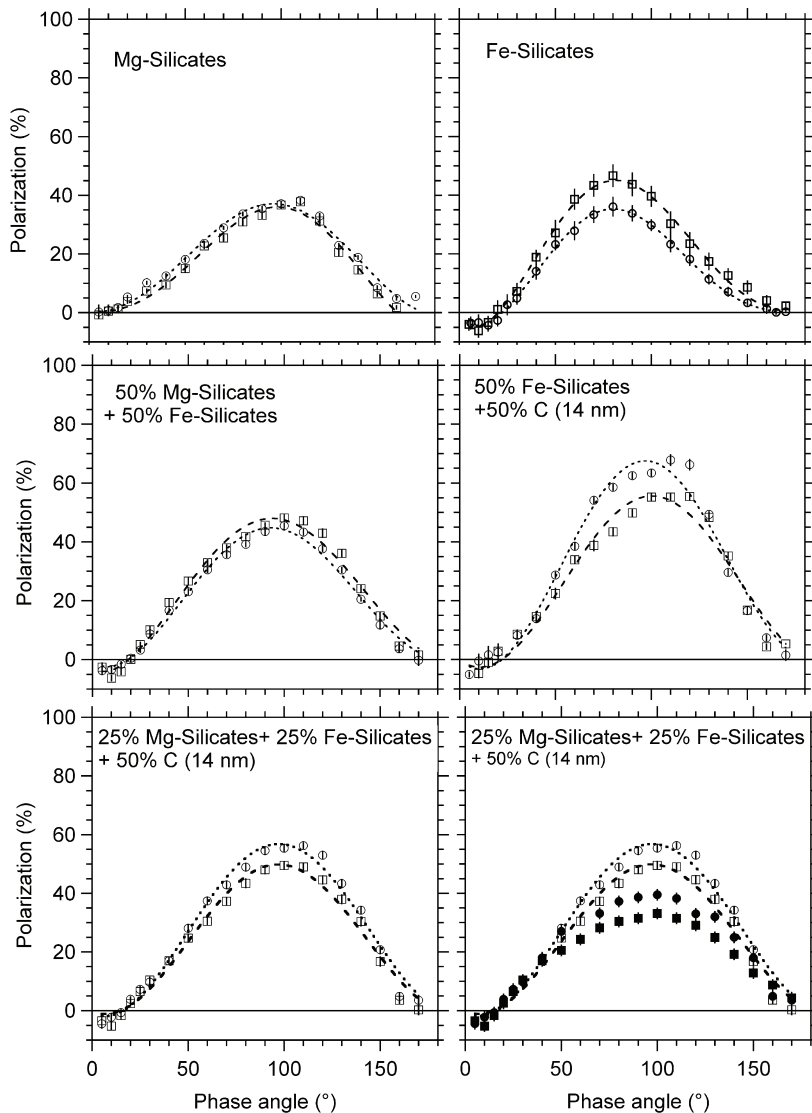


Fig. 24. Polarization phase curves for cometary analogs (open symbols for mixtures of fluffy aggregates, filled symbols for mixtures with compact particles, circles at $\lambda = 543.5$ nm, squares at $\lambda = 632.8$ nm).

depicted in Fig. 24. For samples without carbon, the wavelength effect is approximately neutral (Mg-silicates) or manifests itself as a decrease of polarization with wavelength (Fe-silicates). For mixtures with carbon black, the polarization increases with wavelength, as is usually observed for comets.

With fluffy aggregates of submicrometer-sized constituent grains only, the amplitude of the positive branch was found to be too high in comparison with cometary phase curves. Some compact Mg-silicate particles were added to the mixture to yield polarization compatible with the observations. Then the amplitude of the positive branch is $\sim 30\%$ and increases with wavelength, the inversion angle is $\sim 20^\circ$, and there is a shallow negative branch with a depth smaller or equal to 2%. The increase in polarization with wavelength is only observed for mixtures with carbon (Figs. 24d,e,f). P_{max} increases when the average size of submicrometer-sized grains in aggregates decreases (Fig. 24d) and decreases when large compact particles are added (see the lower curves in Fig. 24f).

Numerical simulations with mixtures of aggregates and compact particles yield similar results (Lasue and Levasseur-Regourd 2006; Levasseur-Regourd et al. 2007). Lasue et al. (2009) compared the whole-coma phase curves in different wavelengths for comets 1P/Halley and C/1995 O1 (Hale–Bopp). The numerical simulations with mixtures of aggregates and compact particles of silicates and carbonaceous compounds give results comparable to the observations with a 50%–50% mass mixture of silicates and carbonaceous absorbing compounds having volume-equivalent diameters between 0.2 and 40 μm . More precisely, there were fewer small particles in the 1P/Halley coma than in the Hale–Bopp coma. The larger number of small particles in Hale–Bopp could be owing to the large number of jets and strong activity of this comet.

An anomalous region around the nucleus (the so-called circum-nucleus halo) was detected in different comets: *in situ* in the coma of 1P/Halley by Levasseur-Regourd et al. (1986); through a small aperture by Dollfus and Suchail (1987); by imaging polarimetry in the coma of comet C/1990 K1 (Levy) by Renard et al. (1992) and comet 47P/Ashbrook–Jackson by Renard et al. (1996); or in comet Hale–Bopp by numerous observers (e.g., Hadamcik et al. 2003; Jockers et al. 1997). A synthetic phase curve can be found in Hadamcik and Levasseur-Regourd (2003 a) showing the increased depth of the negative branch (up to 6%), an inversion angle of $\sim 30^\circ$, and a decrease of the amplitude of the positive branch as compared to the whole coma. Zubko et al. (2007) suggested that this region could be characterized by a material that is not very absorbing and could be modeled by debris of fluffy particles. In the PAMPRE experiment tholins were generated in a radio-frequency plasma reactor (Szopa et al. 2006); tholins produced from specific gas mixtures have a clear brown color. The size distribution of the grains is 595 ± 390 nm (Hadamcik et al. 2009b). They form agglomerates linked by electrostatic forces outside the reactor. Figure 25 shows the polarization phase curves for agglomerates having 50–100- μm sizes. The phase curves are compatible with the synthetic phase curves derived from remote cometary observations. P_{max} is higher at 543.5 nm than at 632.8 nm, as was observed *in situ* for Halley (Levasseur-Regourd et al. 2005). Grains in agglomerates can be covered by such a material, increasing the compactness in the aggregates and then disappearing at larger distances from the nucleus.

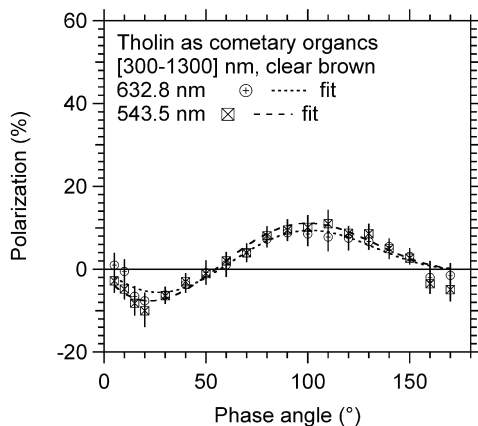


Fig. 25. Cometary particle analog ($C_xH_yN_z$) in the halo region.

6. Conclusions and perspectives

Combining numerical and experimental simulations creates a synergy benefiting both types of study. Experimental results are usually more qualitative than numerical ones. It is relatively easy to separate the effects of different parameters in numerical studies, but it is also important to work with real particles. To perform a comprehensive analysis, it is essential to complement the laboratory work with numerical simulations and to establish the limitations of each method.

With the PROGRA² experiment, we have studied the light-scattering (mainly the linear-polarization) behavior of transparent and absorbing samples in reduced-gravity and ground-based conditions. The levitated particles and the particles deposited in layers had different sizes, shapes, and filling factors. Specifically:

- Spheres and aggregates of spheres have been used to compare experimental results with numerical simulations, thereby allowing us to vary each parameter separately and validate both analyses; and
- The use of irregular particles (either compact or in aggregates) has allowed us to identify effects of size variations from a fraction of a micrometer to hundreds of micrometers.

Our study has revealed:

- The influence of absorbing mantles on the light-scattering behavior of the particles, even with very thin mantles as compared to the diameter of the core;
- The influence of aggregation and filling factors through multiple scattering for transparent materials and absorption for dark materials; and
- The influence of interconnection between size and absorption on the amplitude of the positive polarization branch as a function of wavelength.

We have shown that PROGRA² polarimetric studies are complementary to other diagnostic techniques in determining adequate analogs for particles forming asteroidal regoliths as well as comets. These results are of interest for the interpretation of remote atmospheric and astronomical observations in terms of the formation, evolution, and composition of dust particles.

Complementary results are expected from the Interactions in Cosmic and Atmospheric Particles Systems (ICAPS) microgravity experiment and its precursor (IPE) onboard the International Space Station, which will tentatively simulate the processes of accretion of cometary nuclei and planetesimals in the protosolar nebula.

Acknowledgements

This research was supported by the French Space Agency (CNES), the French National Program of Atmospheric Chemistry (PNCA), and the Programme National of Planetology (PNP). The continuation of the experimental program was recently funded by a contract with the Environment-SA company. The reduced-gravity flights have been operated by Novespace and funded by CNES and ESA.

The authors thank S. Borensztajn (UPMC Univ. Paris 06, LISE) for the SEM images, B. Couté, and M. Jeannot for the improvements of the instruments, and V. Guerrini for the vials. We thank all those who have provided particulate samples and cosmic dust analogs. We are grateful to the organizers of the NATO Advance Study Institute in Kyiv and thank M. Mishchenko, V. Rosenbush, and G. Videen for their help in improving the final version of this chapter.

References

- Blum, J., and R. Schräpler, 2004: Structure and mechanical properties of high-porosity macroscopic agglomerates formed by random ballistic deposition. *Phys. Rev. Lett.* **93**, 115503.
- Blum, J., R. Schräpler, B. J. R. Davidsson, and J. M. Trigo-Rodriguez, 2006: The physics of protoplanetary dust agglomerates. I. Mechanical properties and relations to primitive bodies in the Solar System. *Astrophys. J.* **652**, 1768–1781.
- Bradley, J. P., and D. E. Brownlee, 1986: Cometary particles – thin sectioning and electron beam analysis. *Science* **231**, 1542–1544.
- Brogniez, C., N. Huret, S. Eckermann, et al., 2003: Polar stratospheric cloud microphysical properties measured by the microRADIBAL instrument on 25 January 2000 above Esrange and modeling interpretation. *J. Geophys. Res.* **108**, 8332.
- Burchell, M. J., S. A. J. Fairey, P. Wozniakiewicz, et al., 2008: Characteristics of cometary dust tracks in Stardust aerogel and laboratory calibrations. *Meteorit. Planet. Sci.* **43**, 23–40.
- Cabane, M., P. Rannou, E. Chassefiere, and G. Israel, 1993: Fractal aggregates in Titan's atmosphere. *Planet. Space Sci.* **41**, 257–267.
- Colangeli, L., J. J. Lopez-Moreno, P. Palumbo, et al., 2007: The grain impact analyser and dust accumulator (GIADA) experiment for the Rosetta mission: design, performances and first results. *Space Sci. Rev.* **128**, 803–821.

- Dollfus, A., and J.-L. Suchail, 1987: Polarimetry of grains in the coma of P/Halley. I. Observations. *Astron. Astrophys.* **187**, 669–688.
- Ehrenfreund, P., and S. B. Charnley, 2000: Organic molecules in the interstellar medium, comets, and meteorites: a voyage from dark clouds to the early Earth. *Annu. Rev. Astron. Astrophys.* **38**, 427–483.
- Engrand, C., and M. Maurette, 1998: Carbonaceous micrometeorites from Antarctica. *Meteorit. Planet. Sci.* **33**, 565–580.
- Farnham, T. L., D. D. Wellnitz, D. L. Hampton, et al., 2007: Dust coma morphology in the Deep Impact images of Comet 9P/Tempel 1. *Icarus* **187**, 26–40.
- Flynn, G. J., P. Bleuett, J. Borg, et al., 2006: Elemental composition of Comet 81P/Wild 2 samples collected by Stardust. *Science* **314**, 1731–1735.
- Francis, M., J.-B. Renard, E. Hadamcik, et al., 2011: New studies on scattering properties of different kinds of soot and carbon-black. *J. Quant. Spectrosc. Radiat. Transfer* **112**, in press.
- Fulle, M., A. C. Levasseur-Regourd, N. McBride, and E. Hadamcik, 2000: *In situ* dust measurements from within the coma of 1P/Halley: first-order approximation with a dust dynamical model. *Astrophys. J.* **119**, 1968–1977.
- Furusho, R., B. Suzuki, N. Yamamoto, et al., 1999: Imaging polarimetry and color of the inner coma of comet Hale–Bopp (C/1995 O1). *Publ. Astron. Soc. Japan* **51**, 367–373.
- Greenberg, J. M., and J. I. Hage, 1990: From interstellar dust to comets – a unification of observational constraints. *Astrophys. J.* **361**, 260–274.
- Greenberg, J. M., and A. Li, 1996: What are the true astronomical silicates? *Astron. Astrophys.* **309**, 258–266.
- Grün, E., H. A. Zook, H. Fechtig, and R. H. Giese, 1985: Collisional balance of the meteoritic complex. *Icarus* **62**, 244–272.
- Grynko, Ye., and Yu. G. Shkuratov, 2008: Light scattering from particulate surfaces in geometrical optics approximation. *Light Scattering Rev.* **3**, 329–382.
- Hadamcik, E., and A. C. Levasseur-Regourd, 2003a: Imaging polarimetry of cometary dust: different comets and phase angles. *J. Quant. Spectrosc. Radiat. Transfer* **79–80**, 661–678.
- Hadamcik, E., and A. C. Levasseur-Regourd, 2003b: Dust evolution of comet C/1995 O1 (Hale–Bopp) by imaging polarimetric observations. *Astron. Astrophys.* **403**, 757–768.
- Hadamcik, E., and A. C. Levasseur-Regourd, 2003c: Dust coma of Comet C/1999 S4 (LINEAR): imaging polarimetry during nucleus disruption. *Icarus* **166**, 188–194.
- Hadamcik, E., J. B. Renard, A. C. Levasseur-Regourd, and J. C. Worms, 2002: Polarimetric study of levitating dust aggregates with the PROGRA² experiment. *Planet. Space Sci.* **50**, 895–901.
- Hadamcik, E., J. B. Renard, A. C. Levasseur-Regourd, and J. C. Worms, 2003: Laboratory light scattering measurements on “natural” particles with the PROGRA² experiment: an overview. *J. Quant. Spectrosc. Radiat. Transfer* **79–80**, 679–693.
- Hadamcik, E., J.-B. Renard, A. C. Levasseur-Regourd, and J. C. Worms, 2006: PROGRA2 experiment: new results for dust clouds and regoliths analogs. *Adv. Space Res.* **38**, 2006–2012.
- Hadamcik, E., J.-B. Renard, J. Lasue, et al., 2007a: Light scattering by low-density agglomerates of micron-sized grains with the PROGRA² experiment. *J. Quant. Spectrosc. Radiat. Transfer* **106**, 74–89.
- Hadamcik, E., J.-B. Renard, F. J. M. Rietmeijer, et al., 2007b: Light scattering by fluffy Mg–Fe–SiO and C mixtures as cometary analogs (PROGRA² experiment). *Icarus* **190**, 660–671.

- Hadamcik, E., J.-B. Renard, A. C. Levasseur-Regourd, and J. C. Worms, 2009a: Laboratory measurements of the light scattered by clouds of solid particles by imaging technique. *Light Scattering Rev.* **4**, 31–70.
- Hadamcik, E., J.-B. Renard, G. Alcouffe, et al., 2009b: Laboratory light-scattering measurements with Titan's aerosols analogues produced by a dusty plasma. *Planet. Space Sci.* **57**, 1631–1641.
- Hadamcik, E., J.-B. Renard, A. C. Levasseur-Regourd, et al., 2009c: Light scattering by agglomerates: interconnecting size and absorption effects (PROGRA² experiment). *J. Quant. Spectrosc. Radiat. Transfer* **110**, 1755–1770.
- Hadamcik, E., A. K. Sen, A. C. Levasseur-Regourd, et al., 2010a: Polarimetric observations of comet 67P/Churyumov–Gerasimenko during its 2008–2009 apparition. *Astron. Astrophys.* **517**, A86.
- Hadamcik, E., A. C. Levasseur-Regourd, J.-B. Renard, et al., 2010b: Observations and laboratory simulations of asteroids by polarization measurements. In: K. Muinonen, A. Penttilä, H. Lindqvist, et al., Eds., *Electromagnetic and Light Scattering XII* (University of Helsinki, Helsinki), pp. 70–73.
- Hadamcik, E., A. C. Levasseur-Regourd, J.-B. Renard, et al., 2011: Polarimetric observations and laboratory simulations of asteroidal surfaces: the case of 21 Lutetia. *J. Quant. Spectrosc. Radiat. Transfer* **112**, in press.
- Hanner, M. S., and J. P. Bradley, 2004: Composition and mineralogy of cometary dust. In M. C. Festou, H. U. Keller, and H. A. Weaver, Eds., *Comets II* (University of Arizona Press, Tucson, AZ), pp. 555–564.
- Harker, D. E., C. E. Woodward, D. H. Wooden, et al., 2007: Gemini-N mid-IR observations of the dust properties of the ejecta excavated from Comet 9P/Tempel 1 during Deep Impact. *Icarus* **190**, 432–453.
- Hayward, T. L., M. S. Hanner, and Z. Sekanina, 2000: Thermal infrared imaging and spectroscopy of comet Hale–Bopp (C/1995 O1). *Astrophys. J.* **538**, 428–455.
- Hörz, F., R. Bastien, J. Borg, et al., 2006: Impact features on Stardust: implications for comet 81P/Wild2 dust. *Science* **314**, 1716–1719.
- Jessberger, E. K., T. Stephan, D. Rost, et al., 2001: Properties of interplanetary dust: information from collected samples. In E. Grün, B. Å. S. Gustafson, S. Dermott, and H. Fechtig, Eds., *Interplanetary Dust* (Springer, Berlin), pp. 253–294.
- Jewitt, D., H. Weaver, J. Agarwal, et al., 2010: A recent disruption of the main-belt asteroid P/2010A2. *Nature* **467**, 817–819.
- Jockers, K., V. K. Rosenbush, T. Bonev, and T. Credner, 1997: Images of polarization and colour in the inner coma of comet Hale–Bopp. *Earth Moon Planets* **78**, 373–379.
- Kissel, J., R. Z. Sagdeev, J. L. Bertaux, et al., 1986a: Composition of comet Halley dust particles from VEGA observations. *Nature* **321**, 280–282.
- Kissel, J., D. E. Brownlee, K. Buchler, et al., 1986b: Composition of comet Halley dust particles from Giotto observations. *Nature* **321**, 336–337.
- Lasue, J., and A. C. Levasseur-Regourd, 2006: Porous irregular aggregates of sub-micron sized grains to reproduce cometary dust light scattering observations. *J. Quant. Spectrosc. Radiat. Transfer* **100**, 220–236.
- Lasue, J., A. C. Levasseur-Regourd, E. Hadamcik, and J.-B. Renard, 2007: Light scattering by coated spheres: experimental results and numerical simulations. *J. Quant. Spectrosc. Radiat. Transfer* **106**, 212–224.
- Lasue, J., A. C. Levasseur-Regourd, E. Hadamcik, and G. Alcouffe, 2009: Cometary dust properties retrieved from polarization observations: application to C/1995 O1 Hale–Bopp and 1P/Halley. *Icarus* **199**, 129–144.

- Levasseur-Regourd, A. C., J. L. Bertaux, R. Dumont, et al., 1986: Optical probing of comet Halley from the Giotto spacecraft. *Nature* **321**, 341–344.
- Levasseur-Regourd, A. C., N. McBride, E. Hadamcik, and M. Fulle, 1999: Similarities between *in situ* measurements of local dust light scattering and dust flux impact data within the coma of 1P/Halley. *Astron Astrophys* **348**, 636–641.
- Levasseur-Regourd, A. C., E. Hadamcik, and J. Lasue, 2005: Light scattering as a clue to cometary dust structure. In O. Engold, Ed., *Highlights of Astronomy 13* (Astronomical Society of the Pacific, San Francisco), pp. 498–500.
- Levasseur-Regourd, A. C., T. Mukai, J. Lasue, and Y. Okada, 2007: Physical properties of cometary and interplanetary dust. *Planet. Space Sci.* **55**, 1010–1020.
- Levasseur-Regourd, A. C., E. Hadamcik, J.-B. Renard, and L. A. McFadden, 2011: Clues to links between aubrites and (2867) Steins, from comparison between polarimetric laboratory measurements on meteoritic samples and observations. *Planet. Space Sci.*, in press.
- Matrajt, G., M. Ito, S. Wirick, et al., 2008: Carbon investigation of two Stardust particles: a TEM, NanoSIMS, and XANES study. *Meteorit. Planet. Sci.* **43**, 315–334.
- Mazets, E. P., R. L. Aptekar, S. V. Golenetskii, et al., 1986: Comet Halley dust environment from SP-2 detector measurements. *Nature* **321**, 276–278.
- McDonnell, J. A. M., P. L. Lamy, and G. S. Pankiewicz, 1991: Physical properties of cometary dust. In R. L. Newburn, M. Neugebauer, and J. Rahe, Eds., *Comets in the post-Halley Era*, Vol. 2 (Kluwer, Dordrecht), pp. 1043–1073.
- McFadden, L. A., E. Ammonito, E. A. Cloutis, et al., 2009: Coordinated laboratory studies of meteorites supporting Rosetta mission's asteroid flyby target: 2867 Steins. In *40th Lunar and Planetary Science Conference* (Woodlands, TX), p. 2887.
- Mikrenska, M., P. Koulev, J.-B. Renard, et al., 2006: Direct simulation Monte Carlo ray tracing model of light scattering by a class of real particles and comparison with PROGRA² experimental results. *J. Quant. Spectrosc. Radiat. Transfer* **100**, 256–267.
- Muinsonen, K., T. Nousiainen, P. Fast, et al., 1996: Light scattering by Gaussian random particles: ray optics approximation. *J. Quant. Spectrosc. Radiat. Transfer* **55**, 577–601.
- Nuth, J. A. III, F. J. M. Rietmeijer, and H. G. M. Hill, 2002: Condensation processes in astrophysical environments: the composition and structure of cometary grains. *Meteorit. Planet. Sci.* **37**, 1579–1590.
- Petrova, E. V., V. P. Tishkovets, and K. Jockers, 2004: Polarization of light scattered by solar system bodies and the aggregate model of dust particles. *Solar Syst. Res.* **38**, 309–324.
- Peyridieu, S., A. Chédin, D. Tanré, et al., 2010: Saharan dust infrared optical depth and altitude retrieved from AIRS: a focus over North Atlantic – comparison to MODIS and CALIPSO. *Atmos. Chem. Phys.* **10**, 1953–1967.
- Renard, J. B., A. C. Levasseur-Regourd, and A. Dollfus, 1992: Polarimetric CCD imaging of Comet Levy (1990c). *Ann. Geophys.* **10**, 288–292.
- Renard, J.-B., E. Hadamcik, and A. C. Levasseur-Regourd, 1996: Polarimetric CCD imaging of comet 47P/Ashbrook–Jackson and variability of polarization in the inner coma of comets. *Astron. Astrophys.* **316**, 263–269.
- Renard, J.-B., J.-C. Worms, T. Lemaire, et al., 2002: Light scattering by dust particles in microgravity: polarization and brightness imaging with the new version of the PROGRA² instrument. *Appl. Opt.* **41**, 609–618.
- Renard, J.-B., D. Augeron, P. Personne, et al., 2005: Optical properties of randomly distributed soot: improved polarimetric and intensity scattering functions. *Appl. Opt.* **44**, 591–596.

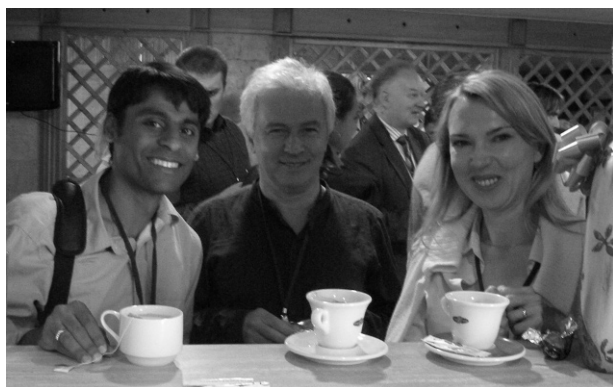
- Renard, J.-B., C. Brogniez, G. Berthet, et al., 2008: Vertical distribution of the different types of aerosols in the stratosphere: detection of solid particles and analysis of their spatial variability. *J. Geophys. Res.* **113**, D21303.
- Renard, J.-B., M. Francis, E. Hadamcik, et al., 2010: Scattering properties of sands. 2. Results for sands from different origins, *Appl. Opt.* **49**, 3552–3559.
- Riedler, W., K. Torkar, H. Jeszenszky, et al., 2007: MIDAS: the Micro-Imaging Dust Analysis System for the Rosetta Mission. *Space Sci. Rev.* **128**, 869–904.
- Russell, P. B., J. M. Livingston, R. F. Pueschel, et al., 1996: Global to microscale evolution of the Pinatubo volcanic aerosol derived from diverse measurements and analyses. *J. Geophys. Res.* **1010**, 18 745–18 764.
- Shkuratov, Yu., S. Bondarenko, V. Kaydash, et al., 2007: Photometry and polarimetry of particulate surfaces and aerosol particles over a wide range of phase angles. *J. Quant. Spectrosc. Radiat. Transfer* **106**, 487–508.
- Shkuratov, Yu., A. A. Ovcharenko, V. A. Psarev, and S. Y. Bondarenko, 2008: Laboratory measurements of reflected light intensity and polarization for selected surfaces. *Light Scattering Rev.* **3**, 383–401.
- Szopa, C., G. Cernogora, L. Boufendi, et al., 2006: PAMPRE: a dusty plasma experiment for Titan's tholins production and study. *Planet. Space Sci.* **54**, 394–404.
- Tomasko, M. G., L. Doose, S. Engel, et al., 2008: A model of Titan's aerosols based on measurements made inside the atmosphere. *Planet. Space Sci.* **56**, 669–707.
- Toon, O. B., and T. P. Ackerman, 1981: Algorithms for the calculation of scattering by stratified spheres. *Appl. Opt.* **20**, 3657–3660.
- West, R. A., L. R. Doose, A. M. Eibl, et al., 1997: Laboratory measurements of mineral dust scattering phase function and linear polarization. *J. Geophys. Res.* **102**, 16 871–16 882.
- Worms, J.-C., J.-B. Renard, E. Hadamcik, et al., 2000: Light scattering by dust particles with the PROGRA² instrument – comparative measurements between clouds under microgravity and layers on the ground. *Planet. Space Sci.* **48**, 493–505.
- Zolensky, M. E., T. J. Zega, H. Yano, et al., 2006: Mineralogy and petrology of comet 81P/Wild 2 nucleus samples. *Science* **314**, 1735–1739.
- Zubko, E., H. Kimura, T. Yamamoto, and H. Kobayashi, 2007: Differences in polarimetric properties of cometary jets and circumnucleus halos. In G. Videen, M. Mishchenko, M. P. Mengüç, and N. Zakharova, Eds., *Peer-Reviewed Abstracts of the Tenth Conference on Electromagnetic & Light Scattering* (Bodrum, Turkey), pp. 257–260.



Edith Hadamcik



Picnic at the Main Astronomical Observatory.



From left to right: Vijay Natraj, Anthony Davis, and Olga Kalashnikova.

High sensitivity polarimetry: techniques and applications

J. Hough*

Centre for Astrophysics Research, University of Hertfordshire, College Lane,
Hatfield AL10 9AB, UK

Abstract. High sensitivity polarimetry will refer to fractional polarizations of below, and often significantly below, 10^{-3} . Measuring low fractional polarization requires a large number of photons and also special techniques are needed to remove systematic effects. The highest sensitivities have been achieved using very fast modulation with single-element detectors, giving sensitivities of 10^{-6} , or better, in fractional polarization. However, it is often important to have good spatial and/or spectral information and to provide this efficiently requires the use of area detectors which bring their own problems, particularly for the highest sensitivities.

Applications of highly sensitive polarimetry described here, include observing the reflected light from exoplanets, the properties of dust in the Earth's atmosphere, and the remote sensing of biomarkers.

Keywords: Stokes parameters, temporal and spatial modulation, exoplanets, atmospheric dust, biomarkers

1. Introduction

Sensitivity is a measure of how well a polarimeter can detect small degrees of polarization, and is not to be confused with accuracy which is the absolute error in the fractional polarization, although ideally a polarimeter should have both high sensitivity and high accuracy.

There are no strict definitions of levels of sensitivity, but high sensitivity here refers to fractional polarizations of $\sim 10^{-3}$, or often significantly better. It will, however, be useful to look at other systems so as to understand better what techniques are required to achieve the highest sensitivity.

As degrees of polarization can often be tens of percent it is not immediately obvious why there is any need to build highly sensitive polarimeters. One fairly common example is when the field of view of the polarimeter contains a large amount of diluting unpolarized flux, leading to small polarizations, as is the case when trying to observe the reflected light from a spatially unresolved exoplanet (Section 3.1). The planet may be polarized at a level of several percent but this is

* Corresponding author. E-mail: j.h.hough@herts.ac.uk

then diluted by the largely unpolarized flux from the very much brighter central star.

There are penalties with high-sensitivity polarimetry, as a very large number of photons are required; for example, to detect a fractional polarization of 10^{-6} requires $\sim 10^{12}$ photons. Depending on the application this may be a major problem and achieving such a high sensitivity will only be possible with very bright sources, or, in astronomy, by using relatively large and hence expensive telescopes.

It is usual to use the Stokes vector, made up of four parameters $\{I, Q, U, V\}$, to describe polarized light (either completely polarized, partially polarized or unpolarized). Each parameter can easily be related to the properties of the polarization ellipse, and, most importantly, the parameters Q , U , and V correspond to the sum or difference of *measurable* intensities: I is the total intensity, Q (the difference in intensities between North-South and East-West linearly polarized components, U (the difference in intensities between linearly polarized components oriented at $+45^\circ$ and -45° from North), and V (the difference in intensities between right and left circularly polarized components). The Stokes parameters are often normalized by the intensity I , giving the dimensionless quantities: $q = Q/I$, $u = U/I$, and $v = V/I$. The degree of linear polarization $p = (q^2 + u^2)^{1/2}$, and v is the degree of circular polarization. The position angle of polarization, usually measured East from North, $\theta = \frac{1}{2} \arctan(u/q)$, with $u = p \sin 2\theta$, and $q = p \cos 2\theta$. Note that

$$\begin{aligned} (Q^2 + U^2 + V^2)^{1/2} &\leq I, \\ (q^2 + u^2 + v^2)^{1/2} &\leq 1, \end{aligned}$$

and pI is the linearly polarized flux.

Ideally a polarimeter should measure all four Stokes parameters, although in practice most measure either linear or circular polarization, and not both. Measurements of circular polarization are less common. This is because in many situations degrees of circular polarization are far less than degrees of linear polarization; some exceptions are the AM Her binary stars (Cropper 1988) or in high mass star forming regions (Chrysostomou et al. 2000). Particular care has to be taken when measuring circular polarization as some of the (usually) much higher linear polarization can be measured erroneously as circular polarization. One of the most straight-forward ways of eliminating this is to continuously rotate a half-wave plate in front of the modulator as this rotates the position angle of polarization, effectively smearing out the linear polarization, but is often not included through lack of space. Typically rejection factors of a few thousand can be obtained (Hough 2005). Another technique is to take measurements with the polarimeter rotated through 90 degrees, reversing the sign of the cross-talk circular polarization without affecting the true circular polarization, but this leads to calibration problems for extended sources with area detectors.

Before discussing high-sensitivity polarimeters it is useful to briefly look at some of the more common mechanisms for producing polarized radiation and

some of the key diagnostics of polarimetry measurements:

Intrinsic mechanisms: these include synchrotron (linear polarization) and cyclotron radiation (circular polarization). In addition, aligned dust grains, common in astronomy, and also occurring in the Earth's atmosphere (Section 3.2), produce linearly polarized radiation with the **E**-vector parallel to the long axis of the grain. The wavelengths at which this polarized emission can be observed depend on the temperature of the dust. The polarization position angle can provide the orientation axis of the dust and from this it is often possible to determine the mechanism responsible for the dust alignment.

Secondary mechanisms: scattering of unpolarized radiation produces linear polarization, but if the radiation is already polarized then circular polarization can be produced, as will occur when there is multiple scattering. Circular polarization can also be produced by scattering of unpolarized light off aligned grains. Linear polarization can be produced by the passage of radiation through a medium of aligned grains (linear dichroic absorption), and if the alignment of the grains twists along the line of sight then circular polarization can be produced through the linear birefringence of the grains (Martin 1972).

The polarization state of radiation contains a wealth of information on the nature of radiation sources, on the geometrical and velocity relationship between a radiation source, scatterer and observer – often on spatial scales that are unattainable by other techniques, and on the chemical and physical properties of an intervening medium between the source and observer when the medium produces polarization by dichroic absorption. As polarized flux is independent of dilution by unpolarized radiation it can be used to determine the spectral index of a source and even when there are two polarized components, these can be separated, provided they have different spectral slopes and a different position angle of polarization. Polarized flux can be used for enhancing the contrast between two close sources, when only one is significantly polarized. Hough and Aitken (2003) give a more complete description of the diagnostic power of polarimetry, particularly at infrared wavelengths.

2. Measuring polarization

There are two basic techniques, each relying on modulating in some way the polarized component of the light (Fig. 1).

The modulation can either be *temporal*, with an active modulator switching the amount of light falling onto a detector as a function of time (zero polarization ideally produces a constant intensity), or the *spatial* splitting of the light, split according to its polarization state (zero polarization ideally produces an exactly equal split of the light). Many polarimeters include both types of modulation.

2.1. Temporal modulation

In principle, degrees of polarization can be very easily measured using a po-

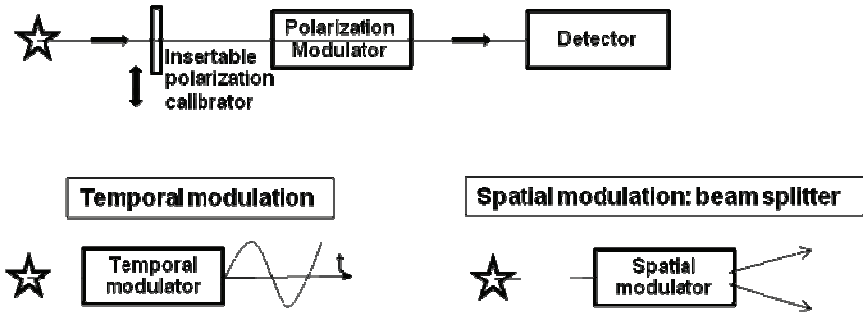


Fig. 1. Upper panel shows the general principles for measuring polarization; lower left-hand panel shows temporal modulation and lower right-hand panel shows spatial modulation.

larizer, either by stepping to a set of fixed angles, or by continuously rotating it, in front of a suitable detector (Fig. 2). As the polarizer transmission axis rotates, relative to the \mathbf{E} -vector of the radiation, the transmitted intensity changes from I_{\max} to I_{\min} , and the degree of polarization is given by $p = (I_{\max} - I_{\min}) / (I_{\max} + I_{\min})$. At optical and near infrared wavelengths (below $\sim 2 \mu\text{m}$) various types of sheet Polaroid can be used, and at longer wavelengths wiregrid polarizers.

In practice, however, there are a number of problems: (i) the detector may be polarization sensitive and will produce a different signal as the \mathbf{E} -vector of the radiation transmitted by the polarizer rotates. It is difficult to distinguish this from a change in signal resulting from the incident radiation being polarized; (ii) a polarizer will transmit no more than 50% of the incident radiation and this is a problem for low degrees of polarization which require very large number of photons; (iii) if the light source varies in intensity on a short timescale (or moves relative to the detector), or the atmosphere produces fast fluctuations, for example atmospheric seeing which can be at a few hundred Hz, then the polarizer has to be rotated faster than the frequency of the fluctuations.

Sparks and Axon (1999) give a detailed analysis of processing imaging polarimetry data taken through a set of polarizers, and Batcheldor et al. (2009) show that for the NICMOS instrument on the Hubble Space Telescope, a polarization

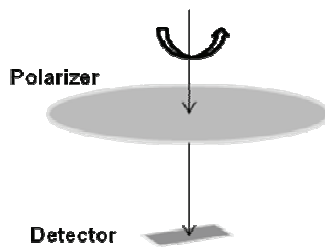


Fig. 2. Stepped or continuously rotating polarizer in front of a detector.

accuracy of $\sim 0.5\%$ can be achieved taking exposures through a set of fixed angle polarizers. The HST has the advantage of no atmospheric fluctuations but has an instrumental polarization of $\sim 1.3\%$ at $2.2 \mu\text{m}$ which can limit sensitivity (Section 2.2.3).

In general, therefore, to achieve high sensitivity it is important to detect as many photons as possible (not throwing away 50% using a polarizer), to have the detector see only a fixed plane of polarization and, for ground-based polarimeters, to modulate the polarized component of the light as quickly as possible, ideally at a few hundred Hz, so as to eliminate atmospheric effects. The latter requirement imposes restrictions on the type of detector that can be used. In particular, it is not easy to use area detectors such as CCDs where it is usually important to integrate on chip for as long as possible to maximize signal to noise (Section 2.2.2).

2.1.1. Achieving sensitivities of at least 10^{-6}

Kemp et al. (1987) showed that the linear polarization of the integrated light from the Sun had an upper limit of 3×10^{-7} . His polarimeter viewed the Sun directly, thus avoiding the problem of any telescope polarization which can be a real problem for very high sensitivity polarimetry (Section 2.2.3). Kemp pioneered the use of photoelastic modulators (PEM) in astronomy, in which a bar of non-birefringent material is stressed at the natural frequency of the slab using a quartz piezoelectric transducer attached to the end of the bar. By using the natural Q of the slab the power required to set up the standing wave is very low. A stress-induced birefringence is produced (with the refractive index along the bar different to that orthogonal to the bar) given by $\delta(x, t) = A_0 \sin \omega t \sin(\pi x/L)$, where A_0 is the amplitude of the modulation, determined by the driving force ($A_0 = \pi$ for half-wave and $\pi/2$ for quarter-wave retardation), ω is the resonant angular frequency, and L is the length of the slab. The birefringence changes with distance x along the slab, being a maximum at the centre of the slab ($x = L/2$). A polarimeter is produced by simply adding a polarizer after the PEM, with its transmission axis at 45° to the slab (Fig. 3). Kemp (1987) (also see Hough et al. 2006), derived the equations for the intensity of light incident on the detector as

$$I = \frac{1}{2}[1 - \cos(A_0 \cos \omega t)]$$

and

$$I = \frac{1}{2}[1 - J_0(A_0) + 2J_2(A_0) \cos(2\omega t) + 2J_4(A_0) \cos(4\omega t) - \dots],$$

where J_n is a Bessel function of order n . It can be shown that the V Stokes parameter is modulated at frequency ω , and the Stokes U parameter is modulated at 2ω , and these can be easily measured using lock-in amplifiers set to either the fundamental frequency or the second harmonic respectively.

With a single PEM only one of the two linear Stokes parameters can be measured, specifically the component polarized at 45° to the PEM axis. The second lin-

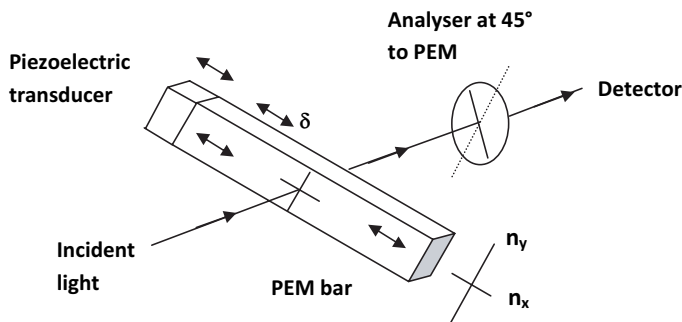


Fig. 3. Basics of a polarimeter using a PEM modulator. The piezo transducer sets up a longitudinal vibration producing a time varying birefringence $n_x - n_y$. The optical phase shift between orthogonal wave components of the transmitted wave is given by $2\pi d(n_x - n_y)/\lambda$, where d is the bar thickness and λ the wavelength of the light. The analyzer (polarizer) has its transmission axis at 45° to the bar axis.

ear Stokes parameter is measured by rotating the instrument through 45° . It is possible to measure both linear Stokes parameters simultaneously by using a pair of PEMs, operated at different frequencies, and angled at 45° to each other, although the modulation efficiency is only half that using a single PEM.

Various materials can be used (CaF_2 , fused silica, ZnSe, silicon), covering wavelengths from $0.2 \mu\text{m}$ to the far-IR. Resonant frequencies range from 20–80 kHz (depending on the size of the bar) and the very fast modulation eliminates atmospheric effects such as seeing or scintillation fluctuations produced by turbulence in the Earth's atmosphere. This, together with phase sensitive detection of the modulated signal (using lock-in-amplifiers) enables very high sensitivities (better than 10^{-6}) to be obtained. One of the major problems with such high sensitivities is eliminating systematic effects. Some of these can be reduced by rotating the analyzer by 90 degrees (from -45° to $+45^\circ$ relative to the PEM axis) thereby changing the polarity of the modulated signals and eliminating any offsets in the output of the phase sensitive detectors. This so-called “secondary chopping” is described by Kemp and Barbour (1981).

Not only are PEMs capable of very high sensitivity, but as they are true zero-order retarders, they have high acceptance angles (as large as ± 40 degrees), and so can be used with fast beams without the need for any prior optics which could alter the polarization state of any incident radiation, and with no rotating elements there is no beam wander, arising from any wedge angles, nor any periodic fluctuations in signals caused by dust on the modulator. PEMs do, however, have some disadvantages: (i) there is a relatively steep fall off in modulation efficiency with wavelength, λ , relative to the tuned wavelength, λ_c , given by $J_2(\pi\lambda_c/\lambda)/2^{1/2}$, which is a steeper fall-off in efficiency than the $\sin(\frac{1}{2}\pi\lambda_c/\lambda)$ factor of crystal retarders (Section 2.2.1); (ii) it is not possible to construct achromatic PEMs as can be done for crystal waveplates and ferroelectric liquid crystals

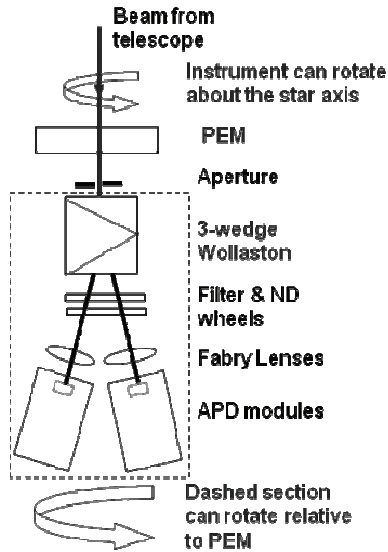


Fig. 4. Schematic of PlanetPol; there is a second identical channel for measuring sky polarization (after Hough et al. 2006).

(Section 2.3.2); (iii) the retardance of PEMs is spatially dependent, limiting the useful aperture; and (iv) the modulation is not square wave, reducing the polarization efficiency. Hough et al. (2006) show that for small fractional polarizations, where the total intensity is dominated by unpolarized flux, the modulation efficiency is only 0.687.

More recently, a polarimeter using techniques pioneered by Kemp was constructed at the University of Hertfordshire, specifically designed to detect the reflected light from exoplanets (known as PlanetPol; Hough et al. 2006). It achieves sensitivities of at least 10^{-6} , with an absolute accuracy $\sim 1\%$ when used on the William Herschel Telescope, La Palma. Although this is a lower sensitivity than achieved by Kemp, PlanetPol is mounted on a conventional telescope and allowance has to be made for the telescope polarization. In order to minimize telescope polarization PlanetPol is mounted on the unfolded Cassegrain focus of the telescope, where symmetry about the optical axis should lead, in the case of mirrors with perfectly uniform reflectivity, to zero telescope polarization. The problem of telescope polarization can be very significant for high sensitivity polarimetry (Section 2.2.3).

PlanetPol (Fig. 4) uses a 20 kHz fused silica PEM, a 3-wedge calcite Wollaston prism which provides better image quality for a given divergence of e - and o -beams, and Avalanche Photodiode Detectors (APD) which can have a very high dynamic range and can be used with very bright sources. With this very fast modulation there is, in principle, no requirement to use a 2-beam analyzer with simultaneous measurements of the e - and o -beam intensities (Section 2.2.1), how-

ever, it does provide double the throughput, very important in measuring very small fractional polarizations. The telescope aperture is imaged on to the detector using a Fabry lens, so that the illuminated patch is fixed, even if there is image motion in the telescope focal plane. Most measurements were made with a very broadband red filter covering the wavelength range from 590 nm to the detector cutoff at about 1000 nm. With such a broad passband the effective wavelength and the polarization efficiency depends on the colour of the object observed.

Removing the telescope and instrument polarization, when measuring very small polarizations, presents a number of problems. Foremost is the lack of appropriate unpolarized standards. Astronomers have used a number of such standards for many years but they are only known to a level of a few $\times 10^{-4}$ (e.g. Gil-Hutton and Penavidez 2003). Hough et al. (2006) determined the telescope polarization (TP) by observing bright nearby stars (typically within 25 pc), with the telescope de-rotator enabled causing the TP to rotate while any interstellar polarization and instrument polarization are fixed. Measuring the polarization of these nearby stars as a function of parallactic angle will produce, in the absence of any intrinsic stellar polarization and interstellar polarization, a sinusoidal curve in Q and in U with an amplitude equal to the TP, phase shifted by 45 degrees. A Gauss–Newton method (Hough et al. 2006) was used to separate the TP from any (small) intrinsic and/or interstellar polarization, and was found to be $(16.4 \pm 0.3) \times 10^{-6}$, showing that the TP can be measured to an accuracy of a few parts in 10^7 . The value of TP can vary with time as the uniformity of the mirror surface changes though dust and other contaminants, and needs to be measured for every observing run. The change can be as much as 50% over several months, or immediately before and after re-aluminizing of the mirror. Figure 5 shows the polarization for stars with very small interstellar and/or intrinsic polarization used to determine the TP for one particular observing run. As part of the observing programme, PlanetPol has been used to determine the linear polarization of ~ 50 nearby stars measured at parts per million (Bailey et al. 2010).

Hough et al. (2006) report that the instrument itself has a polarization (IP) of $\sim 2 \times 10^{-6}$, measured by observing the same low polarization stars at instrument rotation angles of 0° and 90° or 45° and -45° . This measures Stokes Q and $-Q$ or U and $-U$ respectively, so the results should change sign. The sum of such a pair of readings (after correcting for any change in TP between the readings) is equal to 2IP. It is not easy to identify the reasons for producing the IP, although Kemp et al. (1987) showed that a misalignment of \mathcal{G} between the PEM and the telescope axis produces a linear polarization of $\sim 0.1\mathcal{G}^2$, which is $\sim 1 \times 10^{-6}$ for only a 0.18° misalignment.

2.2. Spatial modulation

The introduction of high quantum efficiency area detectors enabled spatial modulation to be used for polarimetry. In order to limit the effect of readout noise with CCDs, it is necessary to fill the wells, provided the linear range of the detec-

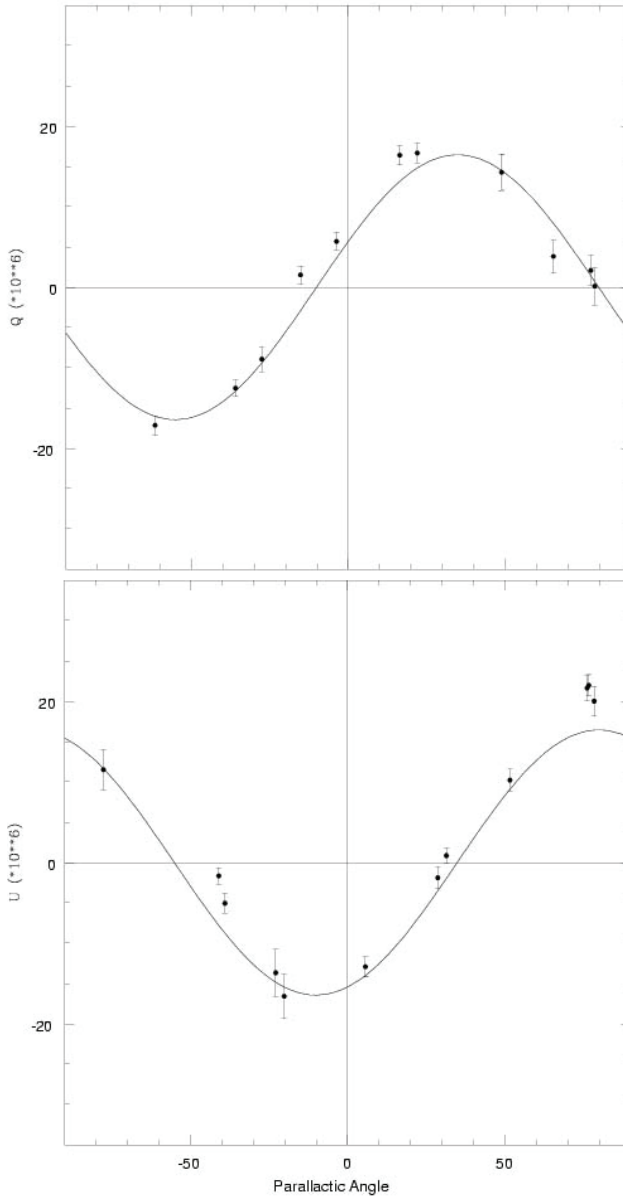


Fig. 5. Directly measured q (top) and u (bottom) polarizations for stars that have very small interstellar and/or intrinsic polarization (after Hough et al. 2006). Note the expected phase shift of 45° between q and u (see text).

tor is not exceeded. This, together with the time taken for readout, means that fast temporal modulation is not possible and detector exposure times might typically

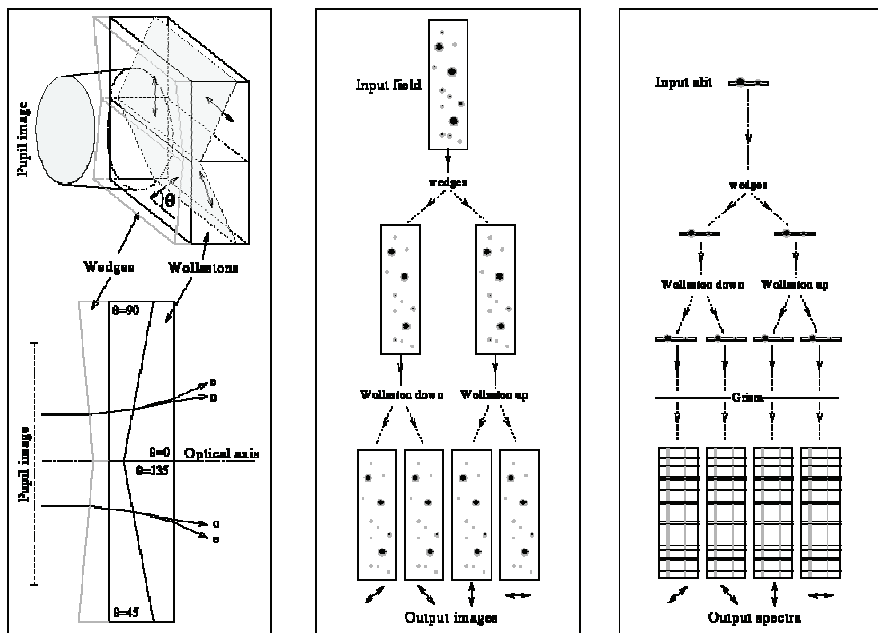


Fig. 6. The wedged double-Wollaston for simultaneous measurement of I , Q , and U (after Oliva 1997).

be seconds. It is possible, however, to measure polarization by spatially separating the two orthogonal polarization components, for example using a Wollaston prism. The polarization then being the difference between the two polarizations divided by the sum of the two (i.e. the intensity). A novel device is the double-Wollaston (Geyer et al. 1996), and the wedged double-Wollaston (Oliva 1997; Pernechele et al. 2003; Kawabata et al. 2008). For the latter, the two wedges (Fig. 6) split the pupil image, with the rays then entering two Wollaston prisms which have crystal axes at 45° , emerging at four different angles. The wedges, by deviating the rays from the optical axis, prevent vignetting at the interface between the two sets of prisms.

The combination of the two Wollaston prisms, in the pupil plane of a camera/spectrometer, enables simultaneous measurements of polarized flux at 0° , 45° , 90° , and 135° enabling I , Q , and U to be determined. An obvious attraction is that there are no moving parts and the device, being relatively compact, can often be included in an existing instrument. As I , Q , and U are determined simultaneously the polarimeter is less prone to tracking and guiding errors and is also well suited to objects that vary on short timescales. Unfortunately it is not possible to get high sensitivity, at best a few times 10^{-3} , as calibrating the relative efficiency of the four channels is difficult. A technique that can be used for point sources is to rotate the instrument through 90° , as this swaps the 0° and 90° beams and the 45°

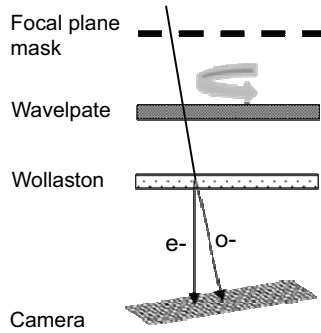


Fig. 7. Improving spatial modulation by including a variable retarder in front of a Wollaston prism; in this case a rotating waveplate.

and 135° beams. With extended sources the images end up on different pixels making the calibration depend on the flat-fielding of the detector, usually limited to a few tenths of a percent accuracy. Another potential problem is the presence of multiple reflections between the large number of surfaces. As with any polarimeter, in which multiple beams are imaged, a focal plane mask is needed to ensure there is no overlap on the detector of the different images, then requiring more than one exposure to cover the whole of an extended field. Another problem that can occur with Wollaston prisms is the lateral chromatism of the Wollaston material; that is the wavelength dependence of birefringence, producing elongated images in the dispersion direction when using wide band filters. Particular care has to be taken when highly birefringent materials are used so as to make the Wollaston(s) as compact as possible, as many of these have high lateral chromatism. Two exceptions are Lithium Yttrium Fluoride and α -Barium Borate but neither material is easily obtained.

2.2.1. Spatial modulation with improved sensitivity

The sensitivity of spatial modulation can be improved considerably by including a variable retarder in front of a single Wollaston prism (Fig. 7).

There is no loss of efficiency in using a single Wollaston as exposure times are half those when using a double Wollaston. Crystal waveplates have been the most commonly used retarders although liquid crystal variable retarders (LCVR) and ferroelectric liquid crystals (FLC) are increasingly used (Section 2.3). The two light intensities emerging from a Wollaston prism, which follows a waveplate with retardance δ and whose optical axis makes an angle ψ relative to the principal axis of the Wollaston, are given (Serkowski 1974) by (the lower sign has the Wollaston rotated by 90°):

$$I' = \frac{1}{2} \left\{ I \pm \frac{1}{2} Q [(1 + \cos \delta) + (1 - \cos \delta) \cos 4\psi] \pm \frac{1}{2} U (1 - \cos \delta) \sin 4\psi \mp V \sin \delta \sin 2\psi \right\};$$

for a $\lambda/4$ plate ($\delta = 90^\circ$):

$$I' = \frac{1}{2} \left\{ I \pm \frac{1}{2} Q \pm \frac{1}{2} Q \cos 4\psi \pm \frac{1}{2} U \sin 4\psi \mp V \sin 2\psi \right\};$$

for a $\lambda/2$ plate ($\delta = 180^\circ$):

$$I' = \frac{1}{2} \{ I \pm Q \cos 4\psi \pm U \sin 4\psi \}.$$

The Q , U , and V can be measured using a $\lambda/4$ plate (angle ψ_1) followed by a $\lambda/2$ plate (angle ψ_2):

$$I' = \frac{1}{2} \left\{ I \pm \frac{1}{2} Q [\cos 4(\psi_1 - \psi_2) + \cos 4\psi_2] \pm \frac{1}{2} U [\sin 4(\psi_1 - \psi_2) + \sin 4\psi_2] \mp V \sin(2\psi_1 - 4\psi_2) \right\}.$$

Crystal waveplates can be used in step and stare mode or rotated continuously; relatively large plates can be made (diameter of 90 mm, or more if mosaiced), and can be used from 0.2 to 24 μm (using quartz, MgF_2 , CdSe) and in the submillimetre (using sapphire). They can be made achromatic, using a combination of two materials with different wavelength-dependence of birefringence or even superachromatic, using three pairs of achromats in a prescription developed by Pancharatnam (1955a, b), where the two outer plates have parallel fast axes, and the inner plate is rotated by 60° . Quartz and magnesium fluoride plates are commonly used, giving excellent efficiency from 0.3 μm to 1.1 μm , or even extending to 2.2 μm , albeit with reduced efficiency. One disadvantage of the Pancharatnam design is that the fast axis depends on wavelength, varying by a few degrees, but this can readily be calibrated.

Disadvantages of crystal waveplates are that non-parallelism can lead to image motion; and dust on the waveplates can lead to periodic modulation. Changes in source intensity and in (atmospheric) transparency has no effect on the polarization as the e - and o -beams are imaged simultaneously. This demonstrates a very important advantage of polarimetry; being a differential measurement accurate polarimetry can be carried out even under varying conditions, although a longer time is needed to obtain a particular accuracy because of the potential loss of photons. Also, variations in sky background between exposures affect both states of polarization equally, and so can be eliminated.

In principle (for I , Q , and U) only exposures with the waveplate at 0° and 22.5° are required. However, the two channels will have different gains (e.g. different pixel sensitivity), which will introduce systematic errors, producing a non-zero signal for unpolarized light. Although flat-fielding can take account of different gains at an accuracy of a few tenths of a per cent, more complete compensation for the different gains can be made if the modulator reverses the polarization states, i.e. switching the e - and o -beams between the two detectors/pixels. Taking exposures at 0° and 45° and 22.5° and 67.5° switches the beams and allows calibration of the different gains between the two beams: $I + Q$ and $I - Q$ for angles 0°

and 45° , and $I+U$ and $I-U$ for 22.5° and 67.5° . Each of these intensity pairs occurs in each beam of the analyzer allowing the effects of the different channel gains to be eliminated.

Astronomers tend to use the so-called “ratio method” to reduce the data (Tinbergen 1996), eliminating common-mode noise such as atmospheric scintillation (subscripts 0, 22.5, 45, and 67.5 refer to e - and o - intensities measured at those waveplate angles):

$$R_Q^2 = [(e/o)_0 / (e/o)_{45}], \quad R_U^2 = [(e/o)_{22.5} / (e/o)_{67.5}],$$

$$q = Q/I = (R_Q - 1) / (R_Q + 1), \quad u = U/I = (R_U - 1) / (R_U + 1).$$

This breaks down for fractional polarizations close to unity and when there is very poor signal to noise, when negative square roots can be encountered, and then the “difference method” is used:

$$F_Q = (e_0 + e_{45}) / (o_0 + o_{45}), \quad F_U = (e_{22.5} + e_{67.5}) / (o_{22.5} + o_{67.5}),$$

$$q = Q/I = [e_0 - e_{45} - F_Q(o_0 - o_{45})] / [(e_0 + e_{45} + F_Q(o_0 + o_{45}))],$$

$$u = U/I = [e_{22.5} - e_{67.5} - F_U(o_{22.5} - o_{67.5})] / [e_{22.5} + e_{67.5} + F_U(o_{22.5} + o_{67.5})].$$

For extended sources, a focal plane mask has to be used to prevent the e - and the o -beams overlapping and only half the field is imaged at any one exposure

With this type of polarimeter sensitivities of $\sim 10^{-4}$ can be achieved. The limitation is set by rapid changes in, for example, seeing and that there are now two optical paths, which are likely to have different aberrations. Also, it is not easy to produce a high sensitivity polarimeter (e.g., 10^{-6}) using this technique unless the sensitivities of the two beams are calibrated with the same precision (10^{-6}), which is extremely difficult.

When non-zero-order waveplates (e.g., those consisting of more than one waveplate), are used at moderate to high spectral resolution, the linear polarization and position angle spectrum can both show ripples. This is caused by multiple reflections within the waveplates and historically was removed from the data by identifying the separate Fourier components of the ripple in the Q and U spectra and interpolating through them (e.g., Adamson and Whittet 1995). Such a procedure can easily result in some loss of spectral information. Aitken and Hough (2001) showed that a simple modification in the data taking process can eliminate much of the ripple. The four sets of observations used to define Q and U should be repeated from 90° to 157.5° and the results from observations at each angle and that angle plus 90° are averaged.

2.2.2. High-sensitivity polarimetry with area detectors

As already noted, CCDs do not lend themselves to high-sensitivity polarimetry as long exposures are needed to minimize readout noise, hence requiring slow modulation. A way of overcoming this problem is to avoid reading

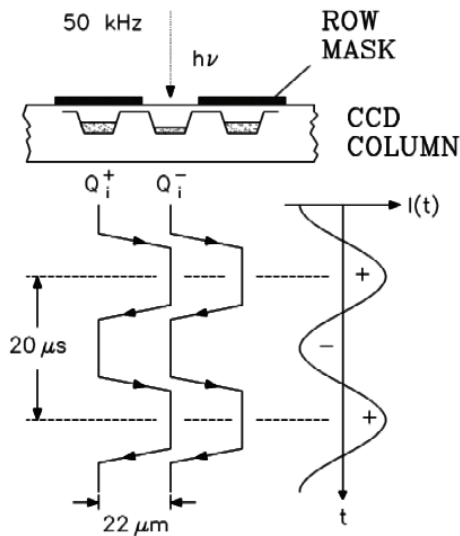


Fig. 8. CCD array used as a photosensitive and storage device (after Keller 2002); see text for further information.

out the array after each polarization switch and to shuffle the charge on the array, with the array then acting as both a detector and storage device. Povel (1995) first proposed that the CCD array detector is alternately divided into photo-sensitive rows and storage rows with the latter shielded from light by a mask (Fig. 8).

Charges generated in the photo-sensitive rows during the first modulation half-cycle are shifted into the left storage row at the transition to the second modulation half-cycle. The charges generated during the second modulation half-cycle are shifted into the right storage row at the transition to the first modulation half-cycle. This is repeated over many modulation cycles, and the respective charges are integrated alternately and synchronously to the modulation. Whenever the desired amount of charge has been accumulated, the charges are read out. There are two key requirements: the transfer time must be about a factor of 100 faster than the modulation frequency, and the charge transfer efficiency must be high to minimize the loss of charge and hence signal.

A series of polarimeters based on this principle, called ZIMPOLs (Zurich Imaging Polarimeters), were developed for solar studies. One such instrument (Keller 2002) used two PEMs at $\pm 22.5^\circ$ to an analyzer with three cameras, used to measure Q , U , and V . They have the advantage of very fast modulation (kHz), there are no flat fielding problems and they are capable of sensitivities $\sim 10^{-5}$, not quite as good as when PEMs are used with single element detectors. Specific disadvantages are (i) the photomask reduces the efficiency by a factor of two; (ii) restriction of a single CCD demodulator to only two states means losing a factor of three for measuring all three parameters simultaneously (this was overcome in a later

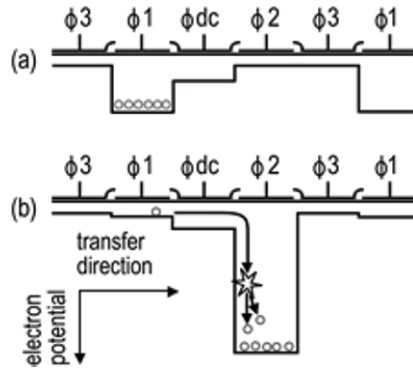


Fig. 9. Electron multiplying CCD (<http://www.emccd.com>).

version (ZIMPOL II) with three out of four CCDs rows masked for storage); (iii) it is difficult to synchronize the exposures of each camera; and (iv) they use front-sided illuminated CCDs which have modest quantum efficiency (QE).

2.2.3. Fast cameras

The development of electron-multiplying CCD cameras (EMCCD, Fig. 9) has led to new opportunities for polarimetry. An additional section to the CCD, the Gain Register, provides internal gain by charge being repeatedly accelerated and producing secondary ionization (similar to APDs). This signal amplification makes readout noise negligible.

One disadvantage, as with APDs, is that the statistical nature of the amplification process leads to additional noise ($\sim 40\%$) but this can be effectively eliminated by photon counting. Such cameras can be used with polarization modulators at speeds of tens of Hz, possibly ~ 100 Hz with limited pixel numbers. Although faster rates would be advantageous in overcoming all atmospheric fluctuations, the system is overall more efficient and simpler than the ZIMPOL devices.

The Astronomical Institute at the University of Utrecht is developing a polarimeter, called ExPo (Extreme Polarimeter; Rodenhuis et al. 2008, Fig. 10), using an Andor EMCCD camera, synchronized with an achromatic FLC modulator (Section 2.3.2). The two orthogonally polarized beams will be imaged simultaneously onto the camera which will operate at frame rates up to ~ 35 Hz with frame transfers < 0.15 ms. The camera has 512×512 active pixels and will be back illuminated, giving a maximum QE of over 90%. The aim is to achieve sensitivities of 10^{-5} , and to do this at the Nasmyth focus at the William Herschel Telescope where instrumental polarization from the telescope will be a few per cent at optical wavelengths. The presence of instrument polarization is a particular problem in the presence of any detector non-linearity. Keller (1996) shows that for a detector with non-linearity of 1%, the instrument polarization needs to be $< 1 \times 10^{-3}$ in or-

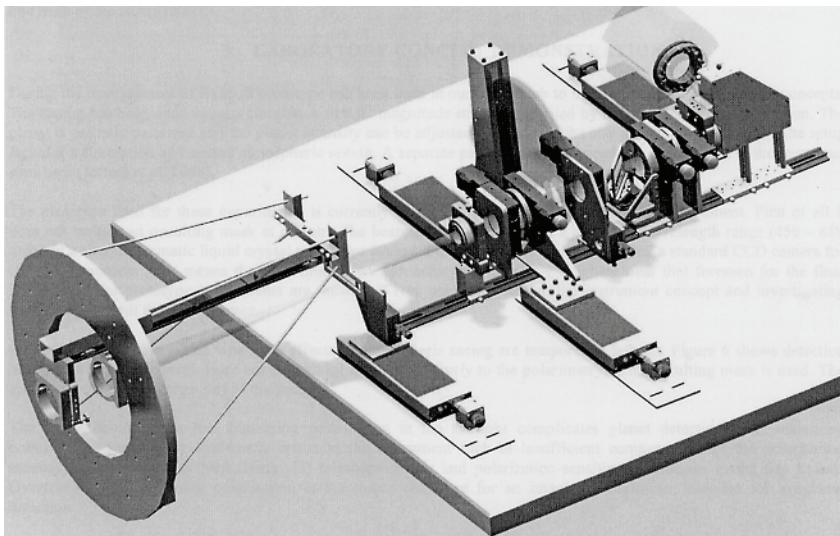


Fig. 10. The Extreme Polarimeter (ExPo) used on the William Herschel Telescope, La Palma (after Rodenhuis et al. 2008).

der to achieve a fractional polarization sensitivity of 10^{-5} . The Utrecht group will use compensating optics (reflectors and retarders) at the Nasmyth focus to significantly reduce the telescope polarization, which otherwise would be a few per cent. One problem that can arise when there is significant telescope polarization, is a coupling between that and any detector non-linearity. This coupling is proportional to intensity and can lead to strongly polarized spectral features for (unpolarized) strong lines (Keller 1996).

2.3. Liquid crystal modulators

2.3.1. Nematic liquid crystals

The orientation of rod-like nematic liquid crystals, sandwiched between electrically conducting fused silica windows which are spaced a few microns apart, can be changed with an applied voltage (typically up to $\sim 20\text{V}$). They can have a large birefringence, which is a maximum when the field is zero and the long axes of the molecules are parallel to the window faces (Fig. 11). They act as electrically variable zero-order retarders with a fixed fast axis (Jochum et al. 2003), and can be used in the visible and near-infrared, with clear apertures of $\sim 40\text{ mm}$. The modulation frequency is about 50 Hz , too low to eliminate all atmospheric effects. They have other disadvantages: (i) a residual retardance of $\sim 30\text{ nm}$ even at high voltages, caused by the inability of crystals at the substrate boundary to rotate freely, although this can be compensated by the addition of a fixed retarder; (ii) the retardance changes by about -0.4% per $^{\circ}\text{C}$, so the crystals have to be main-

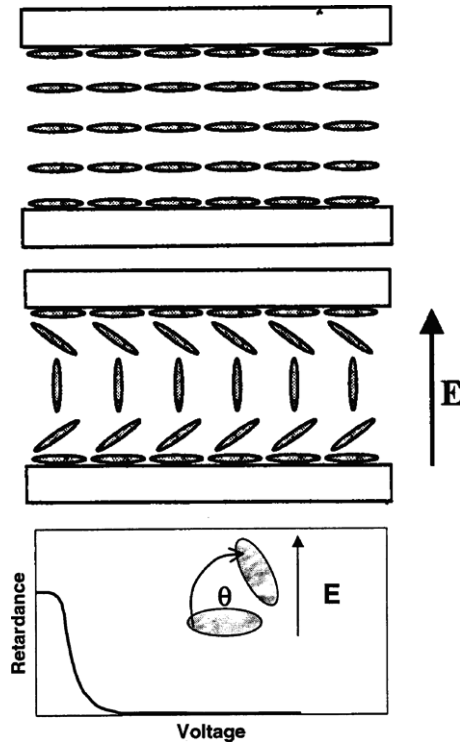


Fig. 11. Schematic of a nematic liquid crystal (after Jochum et al. 2003).

tained at a constant temperature; (iii) achromatic devices cannot be produced using the Pancharatnam design (Section 2.2.1).

2.3.2. Ferroelectric liquid crystals

These are composed of chiral molecules that can tilt away from the layer normal (the so-called smectic C^* phase). When constrained between bounding plates with a close gap, the individual molecular polarizations line up to give an overall permanent polarization with two stable states for the direction of the fast axis that can be switched by about 45° with an externally applied field. The switching speed is $\sim 150 \mu\text{s}$, much faster than with the nematic liquid crystals and fast enough to eliminate most atmospheric effects for polarimetry measurements. The switching angle is temperature dependent, but the (zero-order) retardance, set by the thickness of the crystal layer, is less so. As they have a fixed retardance, combinations of FECs can be made achromatic using the same Pancharatnam scheme (Section 2.2.1) as used with crystal waveplates, and they are useable from 400 to 750 nm (Gisler et al. 2003).

3. Applications

3.1. Detecting reflected light from exoplanets

One of the most exciting developments in astronomy, in recent years, has been the discovery of several hundred exoplanets, planets orbiting other stars (~500 as of November 2010; Schneider 2010). Most of these are discovered indirectly through radial velocity measurements, with the orbiting planet producing a wobble in the central star which is then measured spectroscopically. An increasing number of exoplanets are discovered through primary transits when the planet passes in front of the star, producing a decrease in brightness. Even with transits, confirmation that there is one, or more, exoplanet requires verification through radial velocity measurements. A few exoplanets have been imaged directly but these are young planets that are still sufficiently warm that they are observed through their intrinsic radiation, rather than through light reflected from the central star. A review of exoplanet detection methods can be found in Doyle (2008).

Detecting the reflected light from an exoplanet is very difficult and to date has not been achieved by any technique. The basic problem is not the intrinsic brightness of the planets but rather their proximity to the central star which makes spatially resolving them extremely difficult. The so-called hot-Jupiters, exoplanets with a mass typically that of Jupiter (M_{Jup}) but with an orbit of 0.1 AU or less, may never be spatially resolved in the foreseeable future. For such a system, the planet may be at least 10^4 times fainter than the star and detecting the small change in brightness with orbital phase is not possible, at least from the ground. In contrast, polarimetry, being a differential measurement, can achieve sensitivities of 10^{-6} or better from the ground (Section 2.1.1). If a planet has a typical maximum orbital polarization of ~10%, the observed fractional polarization will be $\sim 10^{-5}$, for a dilution factor of 10^4 . The polarization will vary in a characteristic way with the orbital phase angle of the planet, being zero for an inclination of 90° , and at a phase angle of 0° . More detailed calculations, using different models for the planetary atmosphere, suggest typical maximum orbital polarization of a few times 10^{-6} (Seager et al. 2000).

Measuring the polarization of light scattered from a planet, as a function of orbital phase, can provide a great deal of information: (i) the inclination of the orbit i can be determined from the variation of polarization position angle with orbital phase (removing the $m \sin i$ uncertainty that results from radial velocity measurements); (ii) the fraction of light coming from the planet can be determined from the magnitude of peak polarization (giving information on the albedo and planet radius); and (iii) the size and nature of the scattering particles can be determined from the phase of peak polarization.

The very high sensitivity polarimeter, PlanetPol, has been used to observe two of the earliest discovered exoplanets, 55Cnc and τ Boo (Lucas et al. 2009). The 55Cnc system (HR 3522) is a wide stellar binary with the primary ($I = 5.1$ mag) orbited by five planets. The innermost planet 55Cnc e, is thought to be a “hot Neptune” ($m \sin i = 0.034 M_{\text{Jup}}$), with an orbital radius 0.038 AU (period 2.817 days),

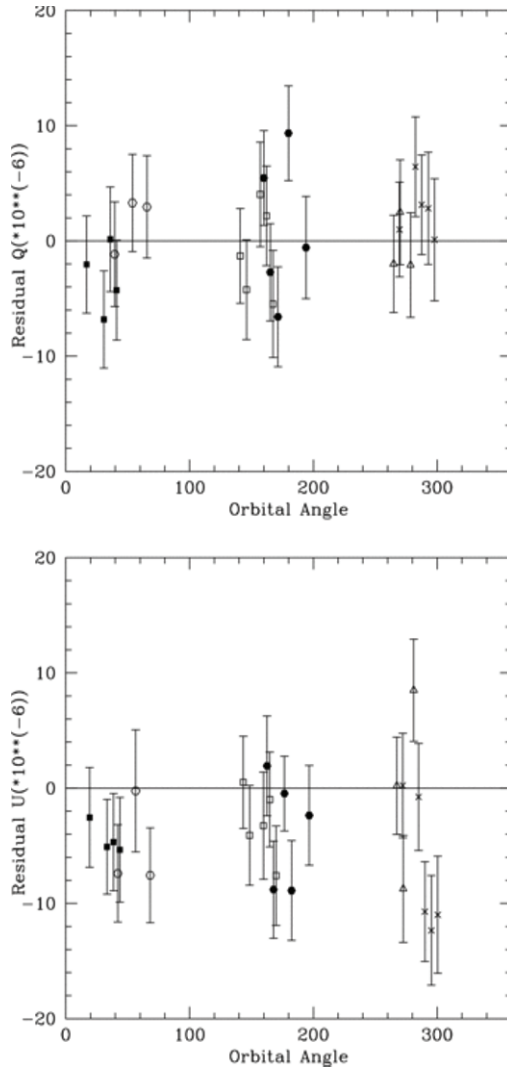


Fig. 12. Polarization of the 55Cnc system as a function of orbital phase angle of 55Cnc e. Minimum illumination at phase angle 180° (after Lucas et al. 2009). Telescope and instrument polarization have been removed.

and the next planet out, 55Cnc b is a hot Jupiter ($m \sin i = 0.824 M_{\text{Jup}}$), with orbital radius 0.115 AU (period 14.65 days). The much larger orbit of 55Cnc b makes it less likely to be detected, although this will depend on the relative planet radii.

Figure 12 shows the polarization of the 55Cnc system as a function of orbital phase angle of 55Cnc e. Lucas et al. (2009) saw no discernable orbital signal despite the standard deviation in the nightly averaged Q/I and U/I being as low as

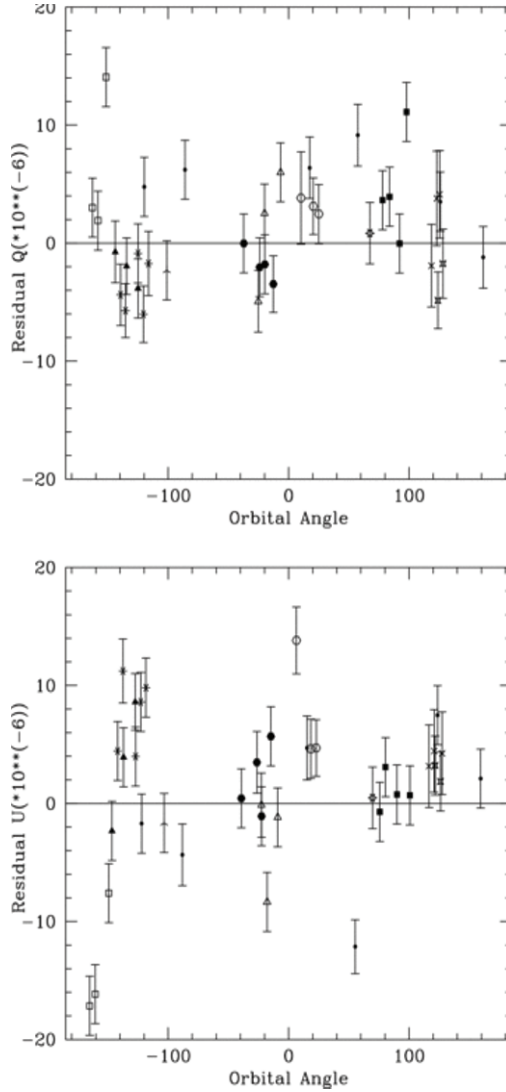


Fig. 13. Polarization of the τ Boo system as a function of orbital phase angle τ Boo b. Minimum illumination at phase angle 180° (after Lucas et al. 2009). Telescope and instrument polarization have been removed.

2.2×10^{-6} . They calculated a 4σ upper limit for the geometric albedo to be $0.13[R/(1.2R_{\text{Jup}})]^{-2} p_m^{-1}$, where p_m is the maximum fractional polarization produced by scattering of unpolarised light through 90° .

The τ Boo system is also a wide stellar binary. The primary ($I=4.0$ mag), is orbited by one of the most massive known hot-Jupiters, τ Boo b ($m \sin i = 4.13$

M_{Jup} , with a most likely mass of $6-7 M_{\text{Jup}}$). The orbital radius is 0.0481 AU (period 3.31 days). [Figure 13](#) shows the polarization of τBoo as a function of orbital phase angle of τBoo b. Again, no orbital signature was observed, and, despite τBoo being brighter than 55Cnc, the standard deviation in the nightly averaged Q/I , U/I is 5.1×10^{-6} , over twice the value for 55Cnc.

Lucas et al. (2009) calculated a 4σ upper limit for the geometric albedo of, $0.37[R/(1.2R_{\text{Jup}})]^{-2} p_{\text{m}}^{-1}$. The large standard deviation in the polarization data is consistent with the space-based photometry by the MOST satellite which shows variability in the stellar flux at the relatively large millimagnitude level (Walker et al. 2008). A likely explanation is magnetic activity on the stellar surface that is linked to the massive planet which has tidally spun up at the least the outer layers of the star to have the same rotation period as the planet's orbit.

3.2. Saharan dust event

When using PlanetPol to measure polarization at sensitivities of $\sim 10^{-6}$, it is necessary to very accurately measure the telescope polarization. This is achieved by observing a number of stars with low intrinsic and interstellar polarization (Section 2.1.1). Typically, many of these stars have polarizations which are at the level of a few times 10^{-6} . For a few days, during observations covering over two weeks in April and May 2005, Bailey et al. (2008) observed that these stars had polarizations which had increased by up to a factor of twenty and that the increase was a function of the zenith angle at which the stars were observed. These increased polarizations coincided with a Saharan dust event in which desert dust from the African continent is observed in the atmosphere over the islands. [Figure 14](#) shows the polarization of a number of low polarization stars measured in clear skies, as a function of zenith angle, and the same stars measured during the dust event. The increased polarization correlated well with the dust optical depth measured by the Carlsberg Meridian Telescope, sited close to the William Herschel Telescope, and the optical depths of the dust layer measured by the Cimel sun-photometer in Santa Cruz on the nearby island of Tenerife (<http://aeronet.gsfc.nasa.gov>).

The very small aperture of PlanetPol (5 arcsec in diameter), the lack of moonlight and simultaneous measurements made in the sky aperture of PlanetPol, enabled Bailey et al. (2008) to rule out scattered light as the cause of the increased polarization. The observations were, however, consistent with dichroic extinction of the starlight by Saharan dust in the atmosphere. The implication is that the dust must be aligned and, from the position angle of polarization the dust must be aligned vertically; this is an excellent demonstration of the diagnostic power of polarimetry. The effect is analogous to that of interstellar polarization produced by the passage of starlight through aligned grains in the interstellar medium. [Figure 15](#) shows a schematic of the passage of light through vertically aligned particles in the atmosphere.

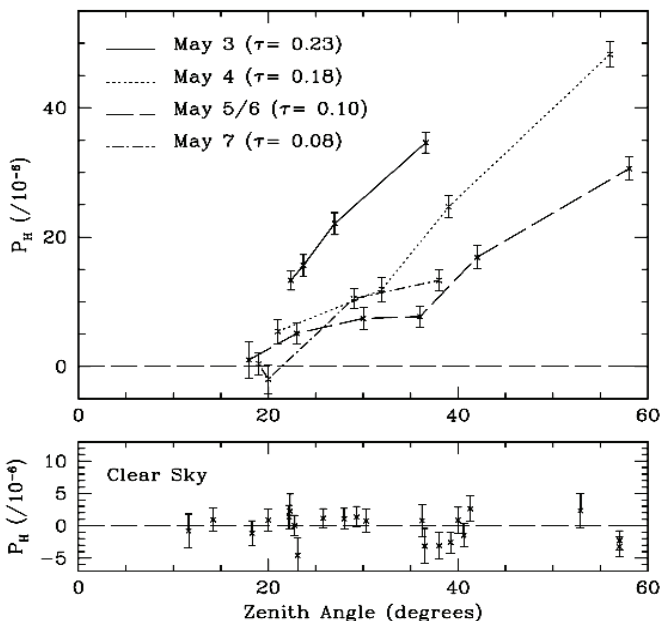


Fig. 14. Clear sky (lower panel) and dust affected (upper panel) linear polarization as a function of zenith angle. The dust optical depth (τ) measured by the Carlsberg Meridian Telescope (CMT) is indicated for each night (after Bailey et al. 2008).

The presence of vertically aligned particles, was unexpected (Ulanowski et al. 2007) as atmospheric particles are typically oriented horizontally due to aerodynamic forces. An exception occurs near electrically active storm clouds when ice crystals are thought to align vertically. Ulanowski et al. proposed that electric fields might be responsible for the alignment of mineral dust, although these aerosols are much smaller than atmospheric ice crystals and there was no evidence of storm clouds, and hence very high electric fields, associated with the Saharan dust event over La Palma.

Ulanowski et al. calculated the probability of a prolate ellipsoidal particle, density 2.6 g cm^{-3} , making an angle θ with the vertical, subject to aerodynamic drag and rotational Brownian motion, in the presence of an electric field. They show that particles with maximum dimension less than $\sim 5 \text{ }\mu\text{m}$ are randomly oriented, and those with maximum dimension $> 25 \text{ }\mu\text{m}$ are oriented horizontally. Particles with maximum dimension between 5 and $25 \text{ }\mu\text{m}$ align vertically with the strength of the alignment depending on the strength of the electric field (Fig. 16)

The sunphotometer measurements from Santa Cruz, Tenerife, during the dust event, showed the majority of particles had a diameter of the volume equivalent sphere up to about $10 \text{ }\mu\text{m}$. Using the T -matrix method (Mishchenko 2000), Ulanowski et al. (see also Bailey et al. 2008) calculated the polarization extinction

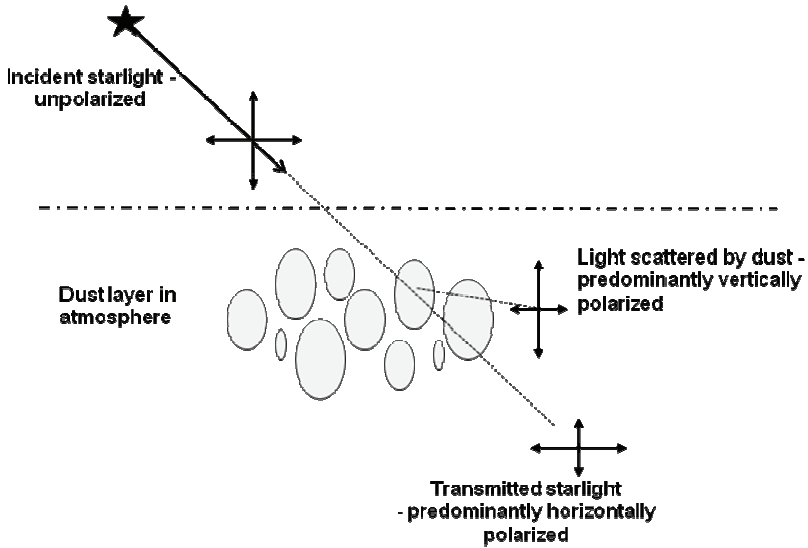


Fig. 15. Schematic of the passage of light through vertically aligned particles in the atmosphere (courtesy of Z. Ulanowski).

for prolate spheroids, aspect ratio 1.5 having the Aeronet particle size distribution, and the calculated particle orientation distribution, and showed that the observed polarization during the dust event could be reproduced for field strengths of $1.5\text{--}2\text{ kVm}^{-1}$ (Fig. 17).

Such field strengths are thought to occur in aerosol dust layers although the mechanism for generating the electric field is not yet clear. Ulanowski et al (2007)

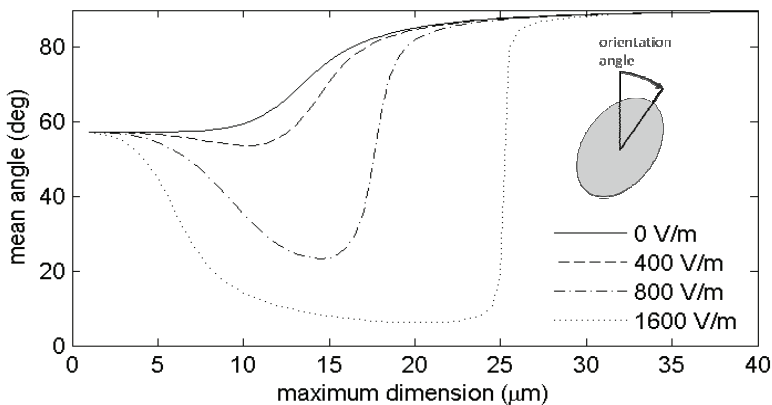


Fig. 16. Mean orientation angle of prolate ellipsoidal particles with aspect ratio 1.5, falling under gravity and subjected to a vertical electric field strength shown in the figure (after Ulanowski et al. 2007).

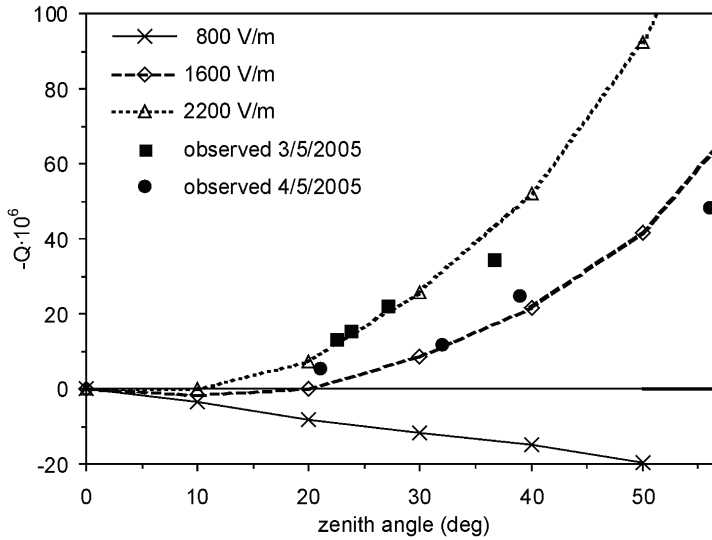


Fig. 17. Calculated polarization as a function of zenith angle for three different field strengths (after Ulanowski et al. 2007). The symbols represent polarization data taken during the dust event.

cite two possibilities: (i) small ion scavenging, producing larger ions with reduced mobility – leading to decreased conductivity and higher field strengths for relatively thin dust layers, although this was thought to be unlikely for this particular event; (ii) charging of the aerosols through triboelectric effects, enhanced by gravitational separation of particles with opposite charge and different aerodynamic size.

Although some details, particularly the mechanism for generating the electric fields are uncertain, it is clear that high-sensitivity optical polarimetry has shown, for the first time, that mineral aerosols can be vertically aligned in the atmosphere. Ulanowski et al. (2008) calculate that this can lead to a change in (i) the transmitted flux of $\sim 10\%$ (“Venetian blind effect”); (ii) the measured phase function by $\sim 20\%$, particularly at large angles; and (iii) a change in the measured degree of linear polarization that would affect retrievals from satellite polarimeters such as PARASOL on A-Train, and the Aerosol Polarimetry Sensor (Mishchenko et al. 2007). The event reported by Ulanowski et al was not at all unusual and La Palma was on the periphery of the dust event, hence the observed effect could have been much larger elsewhere or at other times.

Although it has been long assumed that the atmosphere does not affect polarimetry measurements (air is not birefringent) it is now seen that the presence of atmospheric dust can produce polarization through dichroic absorption, important for high sensitivity polarimetry.

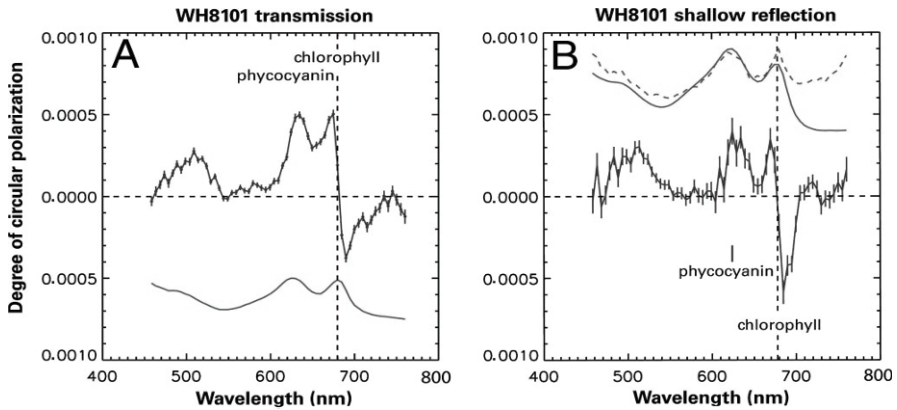


Fig. 18. Circular polarization spectra of cyanobacteria WH8101 (after Sparks et al. 2009). (a) Transmission circular polarization spectrum of *Synechococcus* WH8101 (upper curve); scaled version of the absorbance spectrum (lower curve). (b) Reflection circular polarization spectrum as in (a) (lower curve); scaled version of the $-\log_{10}(\text{Reflectance})$ (upper solid curve); scaled plot of linear polarization degree (upper dashed curve).

3.3. Biomarkers

One of the holy grails of science today is the detection of life elsewhere. A number of possible biomarkers (Gledhill et al. 2007) have been proposed, including the detection of oxygen, ozone, methane and chlorophyll. One unique feature of all life on Earth, and presumably life elsewhere, is its homochirality. All amino acids, except glycine, are chiral and are the building blocks for proteins, and for all life forms they are left-handed. Although the origin of this homochirality remains a mystery (see Bailey et al. 1998; Bailey 2001; Lucas et al. 2005) it is potentially an excellent biomarker. Circular polarization provides a potential method to remotely detect homochirality, and, in transmission, circular dichroism (the differential absorption of left and right-handed circularly polarized light) is a very common technique used to study protein structure. Sparks et al. (2009) were able to show that the circular polarization signature can be seen in reflected light, important for remote observations. Such a signature can be seen either with incident unpolarized light or by the difference between incident left and right-handed circularly polarized light (Sparks et al. 2009; Martin et al. 2010).

Sparks et al. (2009), using incident unpolarized light, as would occur for a planet illuminated by the central star, show that the circular polarization signature associated with light scattered from photosynthetic microbial organisms, is typically a few $\times 10^{-4}$, for example associated with the chlorophyll feature at 680 nm in cyanobacteria WH8101 (Fig. 18). Martin et al. (2010), using a tuneable laser, and with a far-more constrained geometry than used by Sparks et al., find much larger degrees of circular polarization, however, detection of such a signal would require a lander on the planetary surface.

4. Conclusions

Polarimetry is a technique that is capable of very high sensitivity reaching parts per million in fractional polarization. Using single element detectors such sensitivities can be readily achieved, using fast temporal modulation, although care has to be taken to eliminate systematic effects. When used on telescopes, polarimeters need to account for telescope polarization. At the unfolded Cassegrain focus the telescope polarization is typically very low, but still needs to be determined with high accuracy, however, at the Nasymth focus it is essential to reduce the telescope polarization using compensating optics in order to achieve sensitivities of better than 10^{-4} . Using array detectors presents additional problems as fast modulation is difficult. One way of overcoming this is to use the CDD as both a detector and storage device (as with the ZIMPOL polarimeters) when sensitivities of 10^{-5} can be achieved, using very fast modulation, although such devices are both relatively complex and have low efficiency. The introduction of CCDs with internal gain (EMCCDs) has opened up the possibility of efficient polarimeters with relatively fast modulation (tens of Hz using FLCs), with potentially sensitivities of 10^{-5} .

High-sensitivity polarimetry has many applications, either when the intrinsic polarization is low, for example the polarization produced by the passage of light through aligned dust aerosols in the Earth's atmosphere, or when the polarimeter's field of view contains a large amount of unpolarized flux as occurs when measuring the reflected light from exoplanets that are unresolved from the central star.

References

- Adamson, A. J., and D. C. B. Whittet, 1995: A search for polarization in diffuse interstellar bands toward HD 197770. *Astrophys. J.* **448**, L49–L52.
- Aitken, D. K., and J. H. Hough, 2001: Spectral modulation, or ripple, in retardation plates for linear and circular polarization. *Publ. Astron. Soc. Pac.* **113**, 1300–1305.
- Bailey, J. 2001: Astronomical sources of circularly polarized light and the origin of homochirality. *Origins Life Evol. B* **31**, 167–183.
- Bailey, J., A. Chrysostomou, J. H. Hough, et al., 1998: Circular polarization in star-formation regions: implications for biomolecular homochirality. *Science* **281**, 672–674.
- Bailey, J., P. W. Lucas, and J. H. Hough, 2010: The linear polarization of nearby bright stars measured at the parts per million level. *Mon. Not. R. Astron. Soc.* **405**, 2570–2578.
- Bailey, J., Z. Ulanowski, P. W. Lucas, et al., 2008: The effect of airborne dust on astronomical polarization measurements. *Mon. Not. R. Astron. Soc.* **386**, 1016–1022.
- Batcheldor, D., G. Schneider, D. C. Hines, et al., 2009: High-accuracy near-infrared imaging polarimetry with NICMOS. *Publ. Astron. Soc. Pac.* **121**, 153–166.
- Chrysostomou, A., T. M. Gledhill, F. Menard, et al., 2000: Polarimetry of young stellar objects – III. Circular polarimetry of OMC-1. *Mon. Not. R. Astron. Soc.* **312**, 103–115.
- Cropper, M., 1988: Optical properties of AM Her stars. *Adv. Space Res.* **8**, 273–281.
- Doyle, L. R., 2008: Overview of extrasolar planet detection methods. In H. Deeg, J. A. Belmonte, and A. Aparicio, Eds., *Extrasolar Planets: XVI Canary Islands Winter School of Astrophysics* (Cambridge University Press, Cambridge, UK), pp. 1–23.

- Geyer, E. H., K. Jockers, N. N. Kiselev, and G. P. Chernova, 1996: A novel quadruple beam imaging polarimeter and its application to Comet Tanaka–Machholz 1992X. *Astrophys. Space Sci.* **239**, 259–274.
- Gil-Hutton, R., and P. Benavidez, 2003: Southern stars that can be used as unpolarized standards. *Mon. Not. R. Astron. Soc.* **345**, 97–99.
- Gisler, D., A. Feller, and A. M. Gandorfer, 2003: Achromatic liquid crystal polarisation modulator. *Proc. SPIE* **4843**, 45–54.
- Gledhill, T. M., W. B. Sparks, Z. Ulanowski, et al., 2007: Astro-biological signatures. In A. Hoekstra, V. Maltsev, and G. Videen, Eds., *Optics of Biological Particles* (Springer, Berlin), pp. 193–210.
- Hough, J. H., 2005: Polarimetry techniques at optical and infrared wavelengths. *ASP Conf. Ser.* **343**, 3–13.
- Hough, J. H., and D. K. Aitken, 2003: Polarimetry in the infrared: what can be learned? *J. Quant. Spectrosc. Radiat. Transfer* **79–80**, 733–740.
- Hough, J. H., P. W. Lucas, J. A. Bailey, et al., 2006: PlanetPol: a very high sensitivity polarimeter. *Publ. Astron. Soc. Pac.* **118**, 1302–1318.
- Jochum, L., P. Herrero, M. Collados, et al., 2003: Liquid crystal optical retarders for IMAx to fly with SUNRISE. *Proc. SPIE* **4843**, 30–38.
- Kawabata, K. S., O. Nagae, S. Chiyonobu et al., 2008: Wide-field one-shot optical polarimeter: HOWPol. *Proc. SPIE* **7014**, 70144L.
- Keller, C. U., 1996: Recent progress in imaging polarimetry. *Sol. Phys.* **164**, 243–252.
- Keller, C. U., 2002: Instrumentation for astrophysical spectropolarimetry. In J. Trujillo-Bueno, F. Moreno-Insertis, and F. Sánchez, Eds., *Astrophysical Spectropolarimetry: XII Canary Islands Winter School of Astrophysics* (Cambridge University Press, Cambridge, UK), pp. 303–354.
- Kemp, J. C., 1987: Polarized light and its interaction with modulating devices (Hinds International, Inc., Hillsboro, OR).
- Kemp, J. C., and M. S. Barbour, 1981: A photoelastic-modulator polarimeter at Pine Mountain Observatory. *Publ. Astron. Soc. Pac.* **93**, 521–525.
- Kemp, J. C., G. D. Henson, C. T. Steiner, and E. R. Powell, 1987: The optical polarization of the sun measured at a sensitivity of parts in ten million. *Nature* **326**, 270–273.
- Lucas, P. W., J. H. Hough, J. A. Bailey, et al., 2005: UV circular polarisation in star formation regions: the origin of homochirality? *Origins Life Evol. B* **35**, 29–60.
- Lucas, P. W., J. H. Hough, J. A. Bailey, et al., 2009: Planetpol polarimetry of the exoplanet systems 55Cnc and τ Boo. *Mon. Not. R. Astron. Soc.* **393**, 229–244.
- Martin, P. G., 1972: Interstellar circular polarization. *Mon. Not. R. Astron. Soc.* **159**, 179–190.
- Martin, W. E., E. Hesse, J. H. Hough, et al., 2010: Polarized optical scattering signatures from biological materials. *J. Quant. Spectrosc. Radiat. Transfer* **111**, 2444–2459.
- Mishchenko, M. I., 2000: Calculation of the amplitude matrix for a nonspherical particle in a fixed orientation. *Appl. Opt.* **39**, 1026–1031.
- Mishchenko, M. I., B. Cairns, G. Kopp, et al., 2007: Accurate monitoring of terrestrial aerosols and total solar irradiance: introducing the Glory mission. *Bull. Amer. Meteorol. Soc.* **88**, 677–691.
- Oliva, E., 1997: Wedged double Wollaston, a device for single shot polarimetric measurements. *Astron. Astrophys. Suppl.* **123**, 589–592.
- Pancharatnam, S., 1955a: Achromatic combinations of birefringent plates. Part I. An achromatic circular polarizer. *P. Indian Acad. Sci. A* **41**, 130–136.

- Pancharatnam, S., 1955b: Achromatic combinations of birefringent plates. Part II. An achromatic quarter-wave plate. *P. Indian Acad. Sci. A* **41**, 137–144.
- Pernechele, C., E. Giro, and D. Fantinel, 2003: Device for optical linear polarization measurements with a single exposure. *Proc. SPIE* **4843**, 156–163.
- Povel, H. P., 1995: Imaging Stokes polarimetry with piezoelectric modulators and charge-coupled-device image sensors. *Opt. Eng.* **34**, 1870–1878.
- Rodenhuis, M., H. Canovas, S. V. Jeffers, and C. U. Keller, 2008: The Extreme Polarimeter (ExPo): design of a sensitive imaging polarimeter. *Proc. SPIE* **7014**, 70146T.
- Schneider, J., 2010: The extrasolar planets encyclopedia. <http://exoplanet.eu>
- Seager, S., B. A. Whitney, and D. D. Sasselov, 2000: Photometric light curves and polarization of close-in extrasolar giant planets. *Astrophys. J.* **540**, 504–520.
- Serkowski, K., 1974: Polarimeters for optical astronomy. In T. Gehrels, Ed., *Planets, Stars, and Nebulae: Studied with Photopolarimetry* (University of Arizona Press, Tucson, AZ), pp. 135–174
- Sparks, W. B., and D. J. Axon, 1999: Panoramic polarimetry data analysis. *Publ. Astron. Soc. Pac.* **111**, 1298–1315.
- Sparks, W. B., J. Hough, T. A. Germer, et al., 2009: Detection of circular polarization in light scattered from photosynthetic microbes. *Proc. Natl. Acad. Sci. USA* **106**, 7816–7821.
- Tinbergen, J., 1996: *Astronomical Polarimetry* (Cambridge University Press, Cambridge, UK).
- Ulanowski, Z., J. Bailey, P. W. Lucas, et al., 2007: Alignment of atmospheric mineral dust due to electric field. *Atmos. Chem. Phys.* **7**, 6161–6173.
- Ulanowski, Z., O. V. Kalashnikova, P. W. Lucas, and B. Berçot, 2008: Influence of alignment on the scattering properties of atmospheric mineral dust. In J. H. Hough, Ed., *Proceedings of the 11th International Conference on Electromagnetic & Light Scattering* (University of Hertfordshire, Hatfield, UK), pp.145–148.
- Walker, G. A. H., B. Croll, J. M. Matthews, et al., 2008: MOST detects variability on τ Bootis. A possibly induced by its planetary companion. *Astron. Astrophys.* **482**, 691–697.



Jim and Monica Hough enjoying the sight-seeing boat trip along the Dnipro River.

Aerosol retrievals under partly cloudy conditions: challenges and perspectives

Evgueni Kassianov*, Mikhail Ovchinnikov, Larry K. Berg, and Connor Flynn

Pacific Northwest National Laboratory, Richland, WA 99352, USA

Abstract. There are many interesting and intriguing features of aerosols near clouds – many of which can be quite engaging, as well as being useful and climate-related. Exploring aerosols by means of remote sensing, *in situ* observations, and numerical modeling has piqued our curiosity and led to improved insights into the nature of aerosol and clouds and their complex relationship. This chapter conveys the outstanding issues of cloudy-sky aerosol retrievals and outlines fruitful connections between the remote sensing of important climate-related aerosol properties and other research areas such as *in situ* measurements and model simulations. The chapter focuses mostly on treating inverse problems in the context of passive satellite remote sensing and how they can improve our understanding of the cloud-aerosol interactions. The presentation covers basics of the inverse-problem theory, reviews available approaches, and discusses their applications to partly cloudy situations.

Keywords: aerosol retrieval, broken clouds, spectral reflectance, surface albedo, aircraft- and satellite based observations, *in situ* measurements, model simulations

1. Introduction

Splendid blue sky with floating clouds is one of the most beautiful and variable visual displays in nature; throughout the ages it has intrigued and inspired poets, artists, and scholars (e.g., Stephens 2003, and references therein). Clouds vary in color, shape, and size – sometimes resembling brilliant-white castles with a blue background or a red ragged layer covering the sky during the sunset. In addition to being beautiful, the appearance of a cloudy sky has served as a vital visual indicator of fine or destructive weather and its sudden changes. Many seminal experiments in atmospheric science and related profound theoretical studies have been performed to understand the nature of such colorful appearances (e.g., Hey 1983, 1985; Lilienfeld 2004). Yet, the 21st-century explanation with a long his-

* Corresponding author. E-mail: Evgueni.Kassianov@pnl.gov

torical pedigree can be found in the meteorology and related branches of the atmospheric science: water in the form of drops or ice crystals and airborne particles (aerosols) are the most variable and remarkable components of the atmosphere. Their complex interactions with solar light are responsible for the appearance and vitality of the cloudy sky. “The physical and mathematical descriptions of these scattering processes have been pursued for centuries, and this work has contributed greatly to our understanding of both light and the atmosphere” (Hansen and Travis 1974).

Perhaps none of recent theoretical innovations in the atmospheric science can overshadow in importance the development of the classical framework of electromagnetic scattering by macroscopic spherical particles (Mie 1908). An interesting and readable account of its development and evolution can be found in several recent papers (e.g., Mishchenko and Travis 2008; Mishchenko 2009), which are also notable for debunking common “photonic” misinterpretations. Based entirely upon this enormously fruitful framework and its important generalizations (e.g., Gouesbet and Grehan 2000; Gouesbet 2009), great rigor and generality were achieved in the simulation of optical properties of particles with complex morphologies and a wide spectrum of sizes. Moreover, with the aid of this framework, the vector radiative transfer equation (RTE) has been derived with elegance directly from the Maxwell equations (Mishchenko 2008, 2010). The RTE can be considered a language for the description of the particle–radiation interaction and forms the physical basis of remote sensing. The latter uses the scattering of light as an investigative probe (e.g., Stephens 1994; Videen et al. 2004; Kokhanovsky and Chylek 2011), and its countless impressive applications include a variety of objects ranging from tiny nanometer-sized particles (e.g., Khlebtsov et al. 2008; Wax and Backman 2010) to giant solar-system bodies (e.g., Kolokolova et al. 2010; Mishchenko et al. 2010).

Similar to the sailors of the past, who scanned the skies for hints of the upcoming weather, the advent of space-borne instruments has allowed scientists to study weather and climate in ever-increasing detail. Over the past several decades, a wealth of satellite-based information about clouds (e.g., Rossow and Duenas 2004; Stephens and Kummerow 2007) and atmospheric aerosols (Lee et al. 2009; Waquet et al. 2009) as well as their temporal and spatial changes has been obtained. The first recognition of the atmospheric aerosol in cloud formation can be traced back to the 19th century: “Let us quote of the most important facts in the economy of nature, the formation of fogs and of clouds. We know the excellent researches upon this subject by M. Aitken; this physicist has proved by varied and conclusive experiments that aqueous vapor condenses in the air only in the presence of solid particles around which the invisible vapour becomes a liquid; but if the researches of M. Aitken have fully established the mode and the conditions of the formation of fogs and of clouds, they throw little light on the cause of the phenomena” (Mensbrugge 1892). Since this relationship has been in elaborate attire for millennia and has many important climate-related implications (e.g., Lohmann and Feichter 2005; Quaas et al. 2008), there is an increasing need for a better un-

derstanding of the aerosol–cloud interactions (e.g., Small et al. 2009; Stevens and Feingold 2009, and references therein).

The aerosol–cloud relationship is defined by the properties of particles in the environment where clouds form and involves several interactive components such as aerosol chemical composition, radiation, and dynamics (e.g., McComiskey et al. 2009; Leaitch et al. 2010). For example, cloud-related vertical drafts can transport aerosol from low to higher altitudes, altering aerosol stratification in the vicinity of clouds. Because clouds also transport moisture, areas near cloud edges can have elevated relative humidity, making hygroscopic aerosol particles swell. Aerosol growth can also occur via aqueous chemistry in cloud droplets. Typically, aerosol retrievals involve assumptions about the optical properties of hygroscopic aerosols. A fixed relative humidity (RH) is one of these common assumptions, and its application may result in large (up to 40%) errors in clear-sky aerosol optical depth (AOD) (e.g., Wang and Martin 2007), which is a measure of the total aerosol burden. Since RH exhibits large spatial and temporal changes during cloudy conditions, the corresponding cloudy-sky AOD errors associated with aerosol hygroscopicity can exceed significantly their clear-sky counterparts. In addition to the humidification effects, there are so-called cloud adjacency effects (e.g., Kobayashi et al. 2000; Varnai and Marshak 2009): the brightening of clear patches by nearby clouds may result in even larger (up to 140%) errors in the retrieved AOD (e.g., Wen et al. 2006, 2007). Moreover, subpixel cloud contamination effects may come from partly cloudy pixels which can be misclassified as cloud-free (e.g., Zhao et al. 2009). The attempts to improve AOD retrievals under cloudy conditions provided a powerful stimulus to a multitude of integrated studies (e.g., Koren et al. 2010, and references therein).

Interactions and feedbacks among the dynamic components mentioned above represent challenges that both atmospheric science and aerosol remote sensing face today. An integrated approach with a long-term and global-scale perspective is needed to improve understanding of these interactions and feedbacks. The potential afforded by a greater integration across the aerosol remote sensing, *in situ* measurements, and model simulations is yet to be realized. Long-term and global-scale observations of aerosol and cloud properties provide opportunities to analyze their temporal and spatial changes and to propose/validate the corresponding hypotheses about the complex nature of the aerosol–cloud interplay. From this point of view, it is extremely valuable to apply satellite-based global retrievals of aerosols and clouds (e.g., L'Ecuyer and Jiang 2010; Yu et al. 2010).

The main objective of this chapter is to discuss some challenging aspects of the passive satellite aerosol remote sensing between clouds. The following two main sections cover the major topics of the cloudy-sky aerosol remote sensing, including its basis in the inverse problem theory (Section 2), available approaches for reduction of the cloud-adjacency effects and suitability of these approaches to partly cloudy situations (Sections 2 and 3). Before we embark on a summary and outlook (Section 5), the merits of *in situ* aircraft-based aerosol measurements and model simulations of clouds and aerosols will be illustrated (Section 4).

2. Inverse problem: outline

Remote sensing is a very important tool in the study of the dynamics of the atmospheric aerosols and clouds and their interactions. The literature on this topic is extensive and includes several textbooks (e.g., Rodgers 2000; Doicu et al. 2010) and review articles (e.g., Dubovik 2004; Cairns et al. 2009; Doicu et al. 2011) with exhaustive account of the corresponding historical notes and recent developments. A wide range of inverse scattering problems shares a common mathematical foundation. They all require measurement data and a procedure that infers from them information about the model (e.g., actual values of model parameters).

2.1. Clear-sky aerosol inverse problem

In general, the inverse scattering problem can be formulated as

$$\mathbf{y} = \mathbf{F}(\mathbf{x}) + \boldsymbol{\varepsilon}, \quad (1)$$

where \mathbf{x} is the vector of n variables we wish to derive, the so-called state vector, \mathbf{y} is a vector of m measurements we make, the so-called measurement vector, $\boldsymbol{\varepsilon}$ is a vector of random variables that changes with every measurements (also called the error vector), and \mathbf{F} is a forward model that relates the state vector to the measurement vector. For an atmospheric scientist interested in the optical properties of aerosols, observations may consist, for example, of measurements of the polarized visible and near-infrared reflectance \mathbf{R} at different viewing angles and wavelengths.

The adequacy of the forward model defines the quality of the retrieved parameters. The majority of current aerosol retrievals are based on one-dimensional (1D) scalar/vector radiative transfer (RT) calculations. In other words, it is typically assumed that $\mathbf{F}(\mathbf{x}) = \mathbf{F}_{1D}(\mathbf{x})$. Such a model can be appropriate for a scattering medium with relatively small horizontal variations of its optical parameters. Many studies have demonstrated that the aerosol optical properties are characterized by high horizontal homogeneity, and the corresponding autocorrelation scales are typically range from 40 to 400 km (e.g., Anderson et al. 2003; Alexandrov et al. 2004). The observed high homogeneity of aerosol properties explains successful application of the conventional 1D model to clear-sky aerosol retrievals. While the specific form of Eq. (1) depends on the problem in question, its solution typically involves lookup tables (LUTs) wherein the inversion is accomplished via interpolation of observations projected onto the generated LUTs (e.g., Redemann et al. 2000; Stephens and Kummerov 2007). In what follows, we give an example of solving the inverse problem and deriving important clear-sky aerosol properties from polarimetric aircraft-based observations (Waquet et al. 2009) furnished by the NASA Research Scanning Polarimeter (Cairns et al. 1999).

Waquet et al. (2009) apply the LUTs constructed using 1D vector RT calculations to find (i) parameters of a predefined bimodal lognormal size distribution

and (ii) the mode- and wavelength-dependent refractive indices. Such an application provides the best match between measured and calculated polarized reflectance at six wavelengths (412, 443, 555, 672, 865, and 1610 nm) and different viewing geometries with a broad angular range ($\pm 60^\circ$ from nadir). For a given wavelength, the state vector is defined as $\mathbf{x} = [m_r^f, m_i^f, D_g^f, \sigma_g^f, N_a^f; m_r^c, m_i^c, D_g^c, \sigma_g^c, N_a^c; P]$; here superscripts “*f*” and “*c*” denote the fine and coarse modes and *P* is the pressure level corresponding to the altitude at which the top of the aerosol layer is located. The state vector also includes the geometric mean diameter (D_g), the geometric standard deviation (σ_g), the aerosol number concentration (N_a), and the real and imaginary parts of the aerosol refractive index (n_r and n_i , respectively). Thus, this retrieval allows one to convert the polarimetric multi-angle, multi-spectral aircraft-based observations into important aerosol microphysical properties and to identify the chemical composition via the retrieved refractive index and by source type. Similar aerosol properties can be provided by the NASA Glory mission (Mishchenko et al. 2007a). Waquet et al. (2009) utilized the optimal nonlinear inversion framework (Rodgers 2000), which defines the most probable atmospheric state by minimizing a cost function:

$$\Phi = [\mathbf{R} - \mathbf{F}(\mathbf{x})]^T \mathbf{C}_e^{-1} [\mathbf{R} - \mathbf{F}(\mathbf{x})] + [\mathbf{x} - \mathbf{x}_a]^T \mathbf{C}_a^{-1} [\mathbf{x} - \mathbf{x}_a], \quad (2)$$

where \mathbf{R} is the vector of measured polarized reflectance, \mathbf{x}_a is a vector containing *a priori* information about the state vector \mathbf{x} and based on the aerosol climatology (Dubovik et al. 2002), \mathbf{C}_a is the covariance matrix of \mathbf{x}_a , \mathbf{C}_e is the covariance matrix associated with errors, \mathbf{F} is a 1D vector RT model, and *T* stands for “transposed”. Equation (2) can be considered a least-square fit of measurements, weighted by uncertainties (\mathbf{C}_e) and *a priori* values (\mathbf{C}_a). Starting with an initial guess of the atmospheric state $\mathbf{x} = \mathbf{x}_0$, a linear interpolation of the LUTs is applied to give first estimates of reflectances. These estimates are compared to the observed values, and iterations proceed via the Levenberg–Marquadt method. Note that applications of the optimal nonlinear inversion framework are numerous and include the retrievals of aerosol properties (e.g., Goering et al. 2005; Turner 2008; Cairns et al. 2009) and atmospheric sounding (e.g., Marks and Rodgers 1993; Engelen and Stephens 2004).

The majority of current satellite-based aerosol retrievals involve conventional pixel-by-pixel inversions, the so-called Independent Pixel Approximation (IPA). A new multi-pixel framework has been suggested recently for a simultaneous inversion of a large group of clear-sky pixels (Dubovik et al. 2010). The initial application of this framework to one year of PARASOL (Polarization and Anisotropy of Reflectances for Atmospheric Sciences coupled with Observations from a Lidar) observations over two AERONET (Aerosol Robotic Network) sites suggests that it can serve as an efficient tool for gathering information about land surface reflectance and important aerosol properties, such as size, shape, absorption, and composition as well as the aerosol layer elevation. Also, this paper by Dubovik et al. can be considered an important one-source reference to major and most

popular inversion approaches, including the Levenberg–Marquadt method. We emphasize that all aerosol retrievals are currently limited to the cloud-free conditions, and their generalization to cloudy-sky cases remains to be explored.

2.2. Cloudy-sky aerosol inverse problem

In contrast to the atmospheric aerosols, clouds have complex 3D geometries and exhibit strong spatial and temporal variations, with cloud water content ranging from several gm^{-3} to zero over very short distances. For a given viewing angle and wavelength λ , the measured reflectance for clear patches of broken clouds can be expressed as

$$R_{3D}(\lambda) = R_{1D}(\lambda) + \Delta R_{3D}(\lambda), \quad (3)$$

where the subscripts “3D” and “1D” define 3D and 1D components, respectively. The difference between the 3D and 1D reflectances, ΔR_{3D} , is the so-called cloud-induced enhancement characterizing the 3D radiative effects of clouds. Section 2.1 outlined some popular algorithms for the conversion of the R_{1D} into aerosol properties. The 3D component can be estimated by using available numerical solutions of the 3D RTE (e.g., Marshak and Davis 2005; Mayer 2009) and 3D cloud fields obtained from stochastic simulations (Prigarin and Marshak 2009; Venema et al. 2010), observations (e.g., Wen et al. 2006, 2007), and outputs of cloud models (e.g., Barker et al. 2004; Kassianov et al. 2009). Similar to R_{1D} , ΔR_{3D} depends on the surface albedo as well as the optical properties of the molecular atmosphere and aerosols. Unlike R_{1D} , ΔR_{3D} is also a function of cloud optical properties.

Several studies dealing with both polarized (e.g., Cornert et al. 2010) and unpolarized (e.g., Davis and Marshak 2010) RT calculations demonstrated that 3D cloud radiative effects can be large. In particular, it is demonstrated that “polarized reflectances are sensitive to 3D effects in the same way that total reflectances.” (Cornert et al. 2010). Below and in Section 3, we consider approaches based on unpolarized reflectances and outline the corresponding results. These cloud-induced radiative effects can significantly contaminate aerosol retrievals in the vicinity (~ 0.5 km) of clouds (Wen et al. 2007; Yang and Di Girolamo 2008). To address issues associated with cloud adjacency effects and minimize their impact on aerosol retrievals, three different but overlapping approaches have been suggested recently. They are based on conditional sampling, parameterization of 3D effects, and multi-spectral processing, respectively.

Conditional sampling. The first approach includes selection of clear pixels located far away from clouds/shadows, where ΔR_{3D} is relatively small (e.g., Wen et al. 2006, 2007). Statistical analysis of the two-dimensional (2D) horizontal distribution of visible reflectances provides a population of appropriate clear pixels as a function of the nearest cloud distance d , which determines the range of a completely clear area from a clear pixel of interest and is defined as the distance from the clear pixel to the nearest cloudy pixel (Wen et al. 2006). This population de-

creases rapidly with d , and the rate of decrease is a function of both the cloud fraction (CF) and solar zenith angle (SZA). For example, for a cloud field with $CF \sim 0.5$ and $SZA \sim 30^\circ$, the clear pixel population is about 5% and 1% for $d > 2$ km and $d > 3$ km, respectively (Wen et al. 2007). Therefore, the existing 1D operational satellite retrievals of AOD can be successfully applied for a quite limited number of remote (e.g., $d > 2$ km) clear pixels. Note that the aerosol properties of such “remote” pixels may differ substantially from their “near-cloud” cousins (Su et al. 2008; Tackett and Di Girolamo 2009; Twohy et al. 2009).

Parameterization of 3D effects. The second approach involves the parameterization of the 3D radiative effects of clouds on the AOD retrievals. Marshak et al. (2008) proposed such a parameterization based on a simple two-layer model of the atmosphere, with broken clouds in the lower layer and only Rayleigh scatterers in the upper layer, over a dark surface. This parameterization requires several cloud parameters, such as CF, domain-averaged cloud optical depth (COD), and the ratio of cloud thickness to cloud horizontal size, the so-called cloud aspect ratio (CAR). The first two parameters (CF and COD) can be obtained from satellite observations, such as the Moderate Resolution Imaging Spectroradiometer (MODIS) Cloud Product (MOD06). Since atmospheric aerosols are not included, the parameterization assumes that only molecular scattering is responsible for the cloud-induced enhancement of reflectance near clouds. This parameterization allows one to approximate the domain-averaged cloud-induced enhancement as $\Delta\bar{R}_{3D} \approx f(CF, COD, CAR)$. The application of this parameterization is “limited to the case of low-level clouds over dark surfaces with the aerosols below the cloud tops and short wavelengths where molecular scattering dominates” (Marshak et al. 2008).

Multi-spectral processing. The third approach attempts to minimize the 3D radiative effects of clouds by using multi-spectral observations of reflectance. Kassianov and Ovtchinnikov (2008) proposed to exploit reflectance ratios, which are less sensitive to the 3D effects of clouds than the reflectances themselves. As a result, this technique, called the reflectance ratio (RR) method, provides an effective way to reduce substantially the impact of the 3D effects on the retrieved AOD. To evaluate the potential of the RR method, Kassianov et al. (2009) conducted a sensitivity study and demonstrated that it has the ability to detect both “remote” and “nearby” clear pixels appropriate for the RR-based AOD retrievals. Such detection increases the number of appropriate pixels and does not require the statistical analysis of the 2D horizontal distribution of reflectance. Also, the sensitivity study suggested that the RR-based detection of clear pixels and the accuracy of AOD retrievals depend only weakly on the domain-averaged COD. A recently performed case study (Kassianov et al. 2010) demonstrates the ability of the RR method to retrieve AOD from high-resolution aircraft observations. The next section outlines the RR method and discusses its evaluation.

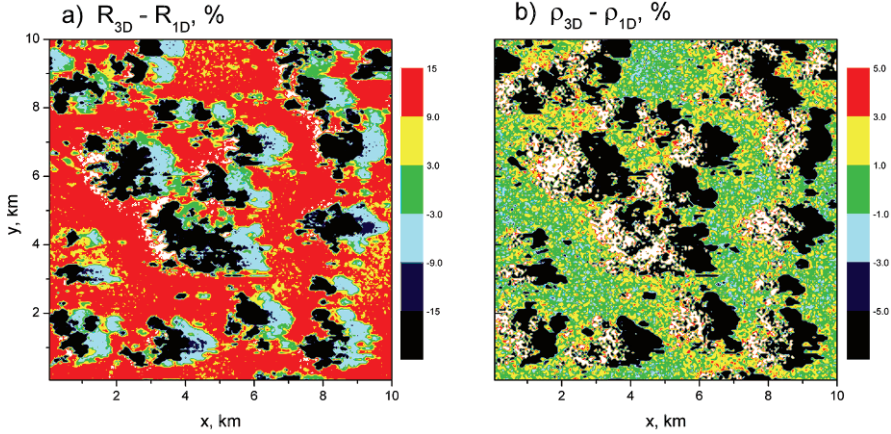


Fig. 1. Relative difference (%) (a) between the reflectances R_{3D} and R_{1D} at a wavelength of 470 nm and (b) between the reflectance ratios $\rho_{3D}(660;470)$ and $\rho_{1D}(660;470)$ (after Kassianov and Ovtchinnikov 2008). On the left: red color represents overestimation of the clear-sky reflectance due to 3D cloud effects, which is in the range from 9% to 15%. Note that a 10% overestimation of reflectance can result in a $\sim 100\%$ overestimation of the retrieved AOD (e.g., Wen et al. 2006, 2007). On the right: green color represents overestimation of the clear-sky reflectance ratio due to 3D cloud effects, which is in the range from -1% to 1% . (Reproduced from Kassianov and Ovtchinnikov 2008 with permission of the American Geophysical Union.)

3. Reflectance ratio method

In traditional aerosol retrievals, AOD is obtained from the observed reflectances using pre-calculated LUTs generated by a 1D RT model for a range of observational conditions and viewing geometries (e.g., Kokhanovsky and de Leeuw 2009). An underlying assumption in this approach is that the 1D reflectances closely approximate the real 3D reflectances. This assumption is violated for partially cloudy scenes with a significant horizontal inhomogeneity.

Similar to traditional aerosol retrievals, the RR method (Kassianov and Ovtchinnikov 2008) is also based on LUTs generated by a 1D radiative transfer model, but it assumes that the 1D reflectance ratio $\rho_{1D}(\lambda_2; \lambda_1)$ approximates the 3D reflectance ratio $\rho_{3D}(\lambda_2; \lambda_1)$, where $\rho(\lambda_2; \lambda_1)$ is the ratio of reflectances $R(\lambda_2)$ and $R(\lambda_1)$ at two wavelengths λ_2 and λ_1 . This assumption is motivated by the well-known fact that clouds reflect about the same amount of sunlight regardless of the wavelength: “Cloud droplets are orders of magnitude larger than aerosol particles. Light scattering by large cloud droplets will be spectrally neutral.” (Kaufman et al. 2005). Consequently, the reflectance ratio is less sensitive to 3D radiative effects of clouds than the reflectances themselves (Fig. 1).

The second assumption in the RR method is that the spectral behavior of AOD τ_a is described by a two-parameter power law: $\tau_a(\lambda) = \beta \lambda^{-\alpha}$. The two param-

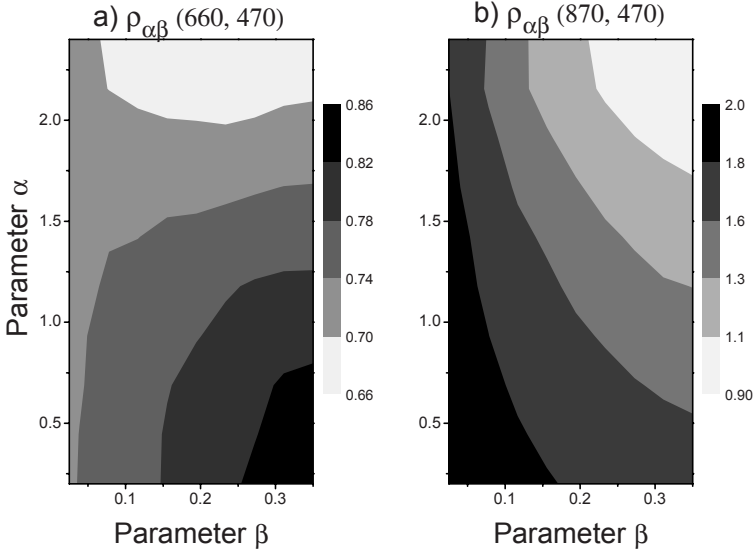


Fig. 2. Example of model reflectance ratios ρ (after Kassianov et al. 2010).

ters, α and β , are known as the Ångström exponent and the turbidity coefficient, respectively. This assumption is motivated by the widespread use of different two-parameter fits of AOD $\tau_a(\lambda)$ in aerosol-related studies and satellite retrievals, such as in the MODIS, Multi-angle Imaging SpectroRadiometer (MISR), and Advanced Very High Resolution Radiometer (AHRR) operational aerosol retrieval algorithms (e.g., Nakajima et al. 2001; Levy et al. 2007; Kahn et al. 2005; Mishchenko et al. 2007b).

In contrast to traditional aerosol retrievals, the RR-based LUTs link two reflectance ratios (two knowns) and the two parameters α and β (two unknowns). For given observational conditions and viewing geometry, the measurement vector \mathbf{y} has two observational components, $\rho(\lambda_2; \lambda_1)$ and $\rho(\lambda_3; \lambda_1)$, and the state vector is defined as $\mathbf{x} = [\alpha, \beta]$. We emphasize that the 1D scalar RT calculations are applied, thus $\mathbf{F}(\mathbf{x}) = \mathbf{F}_{1D}(\mathbf{x})$. The RR method includes two basic steps (Kassianov and Ovtchinnikov 2008) related to the assumptions described above. The first step retrieves the parameters α and β from observed reflectance ratios by applying the first assumption. The second step uses the second assumption and estimates $\tau_a(\lambda)$ by applying the retrieved parameters (α, β) and the assumed power law. The selection of appropriate wavelengths $(\lambda_1, \lambda_2, \lambda_3)$ is governed by the existence and uniqueness of the solution of the inverse problem. Graphically, the solution is the intersection point of two isolines of constant reflectance ratios in the (α, β) domain (Fig. 2), a unique solution for the (α, β) pair can be obtained (Fig. 3).

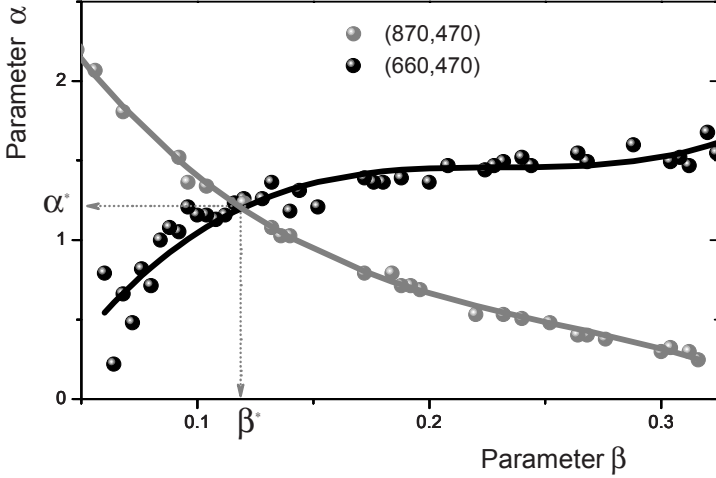


Fig. 3. A family of points where model reflectance ratios match their observed counterparts, the corresponding three-parameter polynomial fits, and the resulting solution (α^* and β^*) (after Kassianov et al. 2010).

For partly cloudy conditions, we examined the performance of the RR method using a sensitivity study (Kassianov et al. 2009). Here we include two important illustrations of the RR performance. The RR method estimates the AOD for clear pixels located both far away from clouds and their shadows and pixels in the vicinity of them (Fig. 4a). The ratio method is able to retrieve AOD in 20 800 pixels, which represents about 70% of the total population of non-cloudy pixels. Note that the application of the conditional sampling (Section 2) based on the nearest cloud distance d for a cloud field with similar CFs (~ 0.5) and similar illumination conditions ($\text{SZA} \sim 30^\circ$) yielded 5% of the clear pixel population at a distance $d > 2$ km (Wen et al. 2007), where the enhancement associated with 3D clouds, ΔR_{3D} , is relatively small. In other words, the traditional retrievals of AOD can be performed quite accurately for 5% of the clear pixel population only, which is far less than 70% enabled by the RR method for similar observational conditions (Fig. 4). Thus, compared to the previously suggested approach based on the nearest cloud distance (Wen et al. 2006, 2007), the RR method increases substantially (several times) the number of acceptable clear pixels for which AOD retrievals can be performed. This demonstrates that the RR method has the ability to (i) detect clear pixels, (ii) increase the “harvest” of such pixels, and (iii) estimate AOD for them. Overall, the RR method provides quite accurate AOD estimations ($\sim 15\%$) for the majority of clear pixels.

Since the 3D cloud-induced enhancement increases with COD (e.g., Wen et al. 2007), the difference between the true and derived values of AOD (or AOD bias) increases with COD as well. The bias is most pronounced for sunlit pixels

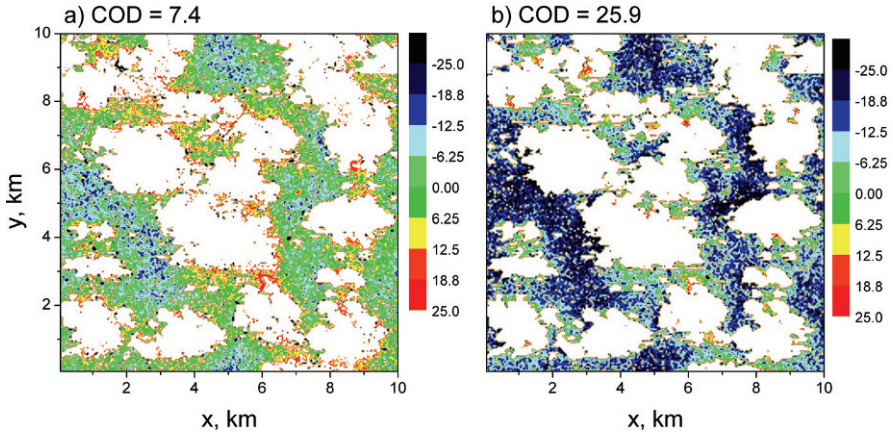


Fig. 4. Horizontal distribution of the AOD bias for two values of the domain-averaged COD (after Kassianov et al. 2009). The AOD values are estimated by the RR method. In simulations, the direct sunlight propagates along the x -axis in the right-hand direction. The largest negative errors (navy and black colors) of the retrieved AOD occur in front of the sunlit sides of the clouds. The white color represents clouds and their shadows. (Reproduced from Kassianov et al. 2009 with permission of Elsevier.)

located in front of large clouds (Fig. 4b). For these sunlit pixels, the cloud contribution to the reflectance is significant, thus the derived AOD can be overestimated substantially and the negative bias of AOD can be as large as 100%. As expected, the histogram of the AOD bias is shifted to negative values (Fig. 5a) with increasing COD. So far we have considered pixel-based differences. Figure 5b shows the domain-averaged values of the AOD bias using both the traditional IPA (Section 2) and the RR method and demonstrates that they overestimate the retrieved AOD (Fig. 2b). However, the IPA-based overestimation is several times greater than the RR-based one (Fig. 5b). Thus, compared to the IPA approach, the RR method is less sensitive to the 3D cloud-induced enhancement. Note that fair-weather cumuli are typically optically thin (averaged COD ~ 10 ; e.g., Chiu et al. 2010). For such clouds, the RR-based overestimation of AOD would be small ($\sim 10\%$).

The initial evaluation of the RR method included extensive airborne and ground-based data sets collected during the Cloud and Land Surface Interaction Campaign (CLASIC) and Cumulus Humilis Aerosol Processing Study (CHAPS), which took place in June 2007 over the U.S. Department of Energy's (DOE) Atmospheric Radiation Measurement (ARM) Southern Great Plains (SGP) site (e.g., Ackerman and Stokes 2003). A detailed case study was performed for a day (June 12, 2007) with single-layer shallow cumuli and a typical aerosol loading. To estimate AOD for clear pixels between clouds, the RR method was applied to reflectance ratios provided by the MODIS Airborne Simulator (MAS) with a high spa-

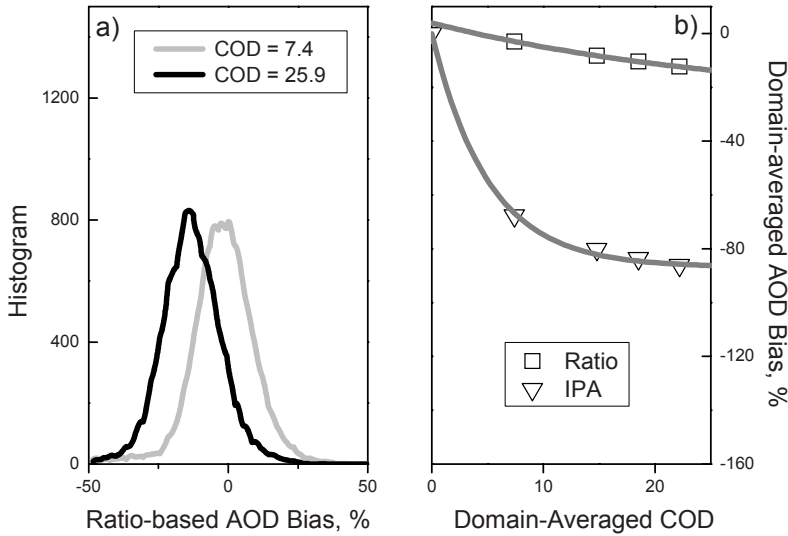


Fig. 5. (a) Histogram and (b) domain-averaged values of the AOD bias as functions of the domain-averaged COD (after Kassianov et al. 2009). The domain-averaged AOD bias is defined by the IPA and ratio methods (b). (Reproduced from Kassianov et al. 2009 with permission of Elsevier.)

tial resolution (0.05 km). Independent measurements of AOD were available from the aircraft-based High Spectral Resolution Lidar (HSRL) and three ground-based Multi-Filter Rotating Shadowband Radiometers (MFRSRs). These HSRL- and MFRSR-derived AODs were used for the development of a dataset to evaluate the RR method. Such development included temporal and spatial interpolations of the MFRSR-derived AODs. Uncertainties associated with this interpolation were reasonably small ($\sim 20\%$). Then, the MFRSR-derived AOD values were interpolated along the MAS flight trajectory, and these interpolated values were considered as observational constraints for the evaluation of the RR method (Fig. 6). Comparisons of the RR- and MFRSR-derived AODs (Fig. 7) revealed that their mean values (along the MAS track) were in a good agreement (within 5%) at the 660-nm wavelength. The RR AODs were greater (by 15–30%) than the MFRSR ones for the 470-nm wavelength. The opposite was true for the 870-nm wavelength.

The focus of this and previous sections is mostly on treating the cloudy-sky aerosol retrieval in the context of a specific setting, namely, passive satellite remote sensing. Current passive observations from satellites with systematic global coverage provide an opportunity for developing long-term records of basic column-integrated aerosol properties, such as AOD and Ångström exponent, mostly for climate-related applications. In addition to these column properties, vertical-resolved aerosol characteristics are needed to improve our understanding of the cloud-aerosol interactions. These properties include the vertical profiles of micro-

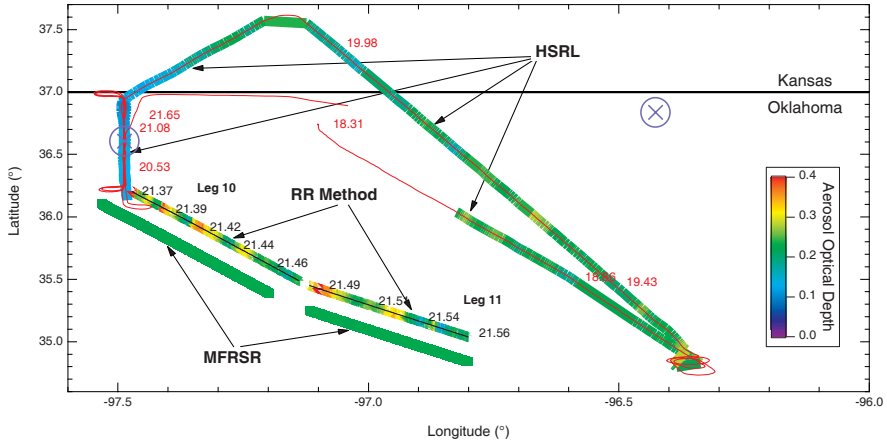


Fig. 6. AOD map obtained from the RR, HSRL, and MFRSR retrievals (colored lines; after Kassianov et al. 2010). Warmer colors indicate larger AODs. The RR retrieval is performed for the MAS-measured reflectances along MAS flight legs 10 and 11. The more Western RR retrieval corresponds to flight leg 10. Numbers along the colored lines denote the corresponding time of day (UTC) and circles define the locations of two MFRSRs. The MFRSR-derived AOD is interpolated spatially and temporally to represent the AOD along the MAS flight trajectory (legs 10 and 11). To ease the comparison of the RR- and MFRSR-derived AODs, the latter are slightly shifted down and to the left (toward the southwest).

physical, chemical, and optical properties. Some of them (e.g., vertical profiles of the aerosol extinction coefficient) can be estimated from active aircraft- and satellite-based observations on both clear- and cloudy-sky days (e.g., Su et al. 2008; Tackett and Di Girolamo 2009; Twohy et al. 2009), while other properties are provided by *in situ* measurements. Typically, these aerosol properties are used for constraining and improving the remote sensing techniques. The next section offers some recent examples of *in situ* aerosol properties near clouds as functions of altitude. The use of model simulations of aerosol changes near clouds in support of the remote sensing studies is discussed as well.

4. Aerosol properties near clouds

Clouds form in regions with enhanced humidity. Clouds also enhance humidity around them by detrainning moist air transported from below by cloud updrafts. Condensed water is certainly one of the main contributors to the AOD even in clear-sky conditions (Hegg et al. 1997). Consequently, knowing and taking into account both the relative humidity profile and aerosol hygroscopicity is of primary importance. Since aerosol hygroscopicity depends on the aerosol chemical composition, the latter is invaluable in studying the humidification impact on the re-

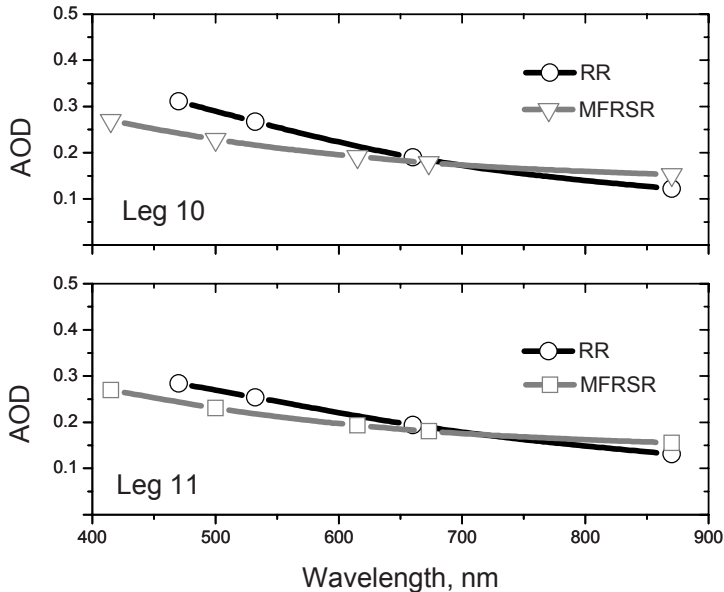


Fig. 7. Spectral dependence of the domain-averaged AOD values obtained from the RR and MFRSR retrievals for legs 10 (top) and 11 (bottom) (after Kassianov et al. 2010).

trieved aerosol optical properties: “...although detailed information of aerosol chemical composition cannot realistically be available for routine retrievals, ...the relative mass ratios of sulfate particles, soot and water-insoluble compounds should be included in the retrieval algorithm” (Wang and Martin 2007).

4.1. *In situ* observations

Long-term *in situ* observations of aerosol optical properties are performed at a number of sites around the world and often include measurements of the scattering and absorption of light by aerosols (e.g., Carrico et al. 1998; Sheridan et al. 2001; Quinn et al. 2002). The primary disadvantage of *in situ* surface measurements is that they do not necessarily reveal features of the aerosol above the sampling inlet. To address this shortcoming, research aircraft have been used to measure aerosol optical and chemical properties at a wide range of altitudes. Some of these studies have made use of relatively small light aircraft that are operated over long periods of time. Research flights supported by the ARM SGP site (Andrews et al. 2004) are an excellent example of such flights. Measurements made during this multi-year study have included light scattering and absorption at three wavelengths. The payload of such aircraft is limited, however, leading to a number of studies that deploy larger aircraft to measure not only the aerosol optical properties, but their chemical composition as well.

For example, during the CHAPS, the U.S. DOE Gulfstream-1 (G-1) was configured to make *in situ* measurements of the chemical and optical properties of aerosols, cloud microphysics, trace gas concentrations, and meteorological variables. The flight pattern used by the G-1 included legs below, within, and above the cloud layer. The instrument payload involved a number of different instruments to measure the aerosol optical properties, including nephelometers to measure the amount of light scattered by aerosol as well as Particle Soot Absorption Photometers (PSAPs) and a photo-acoustic soot spectrometer for measuring the amount of light absorbed by aerosol. In addition to these measurements, an Aerodyne aerosol mass spectrometer (AMS) was deployed. The AMS provides information on non-refractory materials that are chemically and physically stable at high temperatures. A unique feature of this deployment was the use of two aerosol inlets, an isokinetic inlet for sampling particles in clear air and a counterflow virtual impactor (CVI) inlet (Noone et al. 1993) for sampling cloud drops. A full description of the instrument payload used during the CHAPS can be found in Berg et al. (2009).

Measurements of the aerosol chemical composition were made using both the isokinetic and CVI inlets. An example of measurements made upwind- and downwind of Oklahoma City below, within, and above the cloud layer made on June 11, 2007 are shown in Fig. 8. The total aerosol mass generally decreased with height. Sulfate and organics dominated the non-refractory part of the aerosols upwind and downwind of Oklahoma City at all altitudes. The fractional amount of sulfate relative to the other components was smaller, while the fractional amount of organics was larger downwind of Oklahoma City, thereby indicating that a large fraction of the aerosol mass produced near Oklahoma City was organic. There was also an increase in the mass fraction of nitrate within the cloud drops sampled by the CVI.

Additional instruments were deployed in pods mounted on the aircraft nose to measure the particle size distributions over a wide range of sizes. A Droplet Measurement Technology (DMT) Passive Cavity Aerosol Spectrometer Probe (PCASP-100X) was used to measure the number density for particles between 0.1 and 3 μm in diameter. A DMT Cloud, Aerosol, and Precipitation Spectrometer (CAPS) was applied to measure the distribution of cloud droplets with diameters ranging from 0.5 to 50 μm and precipitation particles between 25 and 1550 μm in diameter. Using data collected by the G-1, Berg et al. (2011) found evidence of the so-called first aerosol indirect effect in continental shallow cumuli near Oklahoma City, and demonstrated that pollution had a measurable impact on the cloud optical properties.

Changes of aerosol optical properties can be associated with humidification effects as well. In particular, Perry and Hobbs (1996) observed significant humidity enhancements in the clear air near isolated cumulus clouds over the northeast Pacific Ocean. Using instrumented aircraft, they found detectable humidity halos in about half of all penetrations, with halo width being highly correlated with the wind shear. The halo width was smallest on the upshear side (average width 0.3

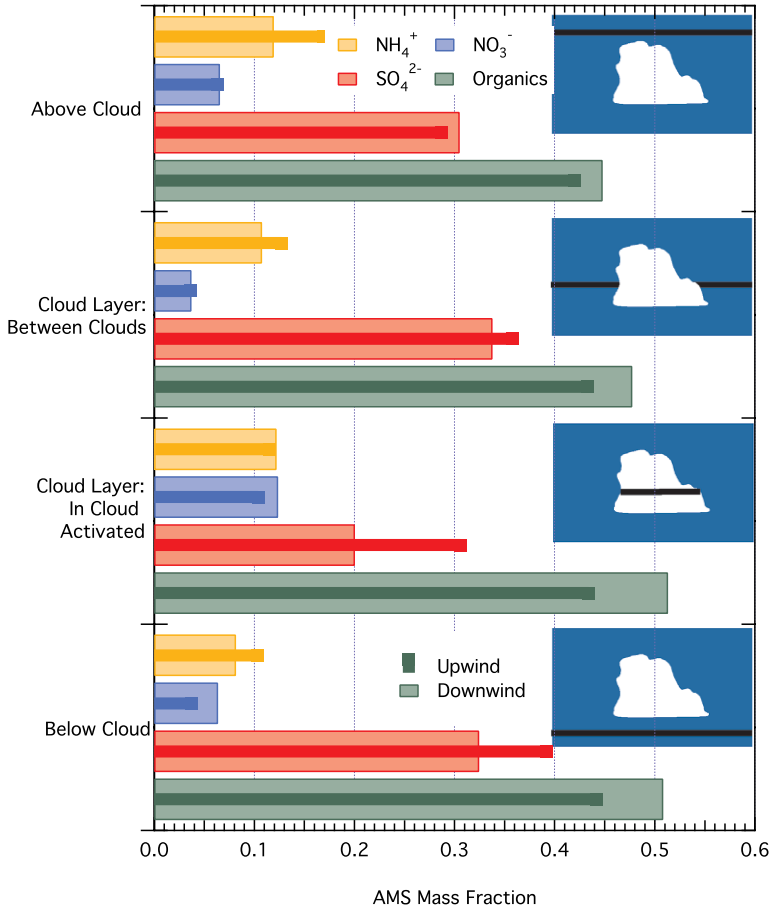


Fig. 8. Mass fraction (mg of component species to total mg measured by the AMS) of organics (green), SO_4^{2-} (red), NO_3^- (blue), and NH_4^+ (yellow) observed on 11 June 2007 upwind (thick bars) and downwind (thin bars) of Oklahoma City below (bottom) cloud. Activated aerosols in clouds and areas (middle) between clouds and (top) above cloud. Black lines on inset pictures illustrate source regions of data. (Reproduced from Berg et al. 2009 with permission of the American Meteorological Society.)

cloud radii) and largest on the downshear side (1.3 cloud radii). It must be noted that because the wind shear tends to tilt the clouds, a cloud-droplet radius determined from an aircraft penetration at any level is likely to be smaller than a vertically projected droplet radius as determined, for example, by a vertically pointing remote sensing instrument, either orbital or ground based. Consequently, a relative halo width relevant to such instruments would be smaller than the above estimates.

Recently, interest in the humidity halos has resurged owing to the recognition of their potential importance to the interpretation of the aerosol radiative effects and aerosol remote sensing (e.g., Koren et al. 2007; Twohy et al. 2009). The next subsection illustrates the temporal/spatial variability of humidity near clouds and the associated changes of aerosol properties by using model simulations.

4.2. Model simulations

Over the past several decades, great progress has been achieved in the development, evaluation, and application of models with an improved description of cloud-aerosol interactions (e.g., Khain 2009). When coupled with 3D radiative transfer calculations, simulations from these models can be compared realistically to remote-sensing observations of the cloudy atmosphere (e.g., Ovtchinnikov and Marchand 2007; Schmidt et al. 2009) and, therefore, can be applied to quantify uncertainties in aerosol retrievals (e.g., Kassianov et al. 2009). Like the measurement-based reports (Wen et al. 2006, 2007), these studies have demonstrated the large impact of the cloud adjacency effects on aerosol retrievals. Section 3 gave some examples of this impact. Also, numerical models are a great tool for the investigation of the complex processes and their interactions such as variations of RH and the corresponding changes of aerosol optical properties.

In an example below, a large-eddy simulation (LES) cloud model is used to characterize the effect of the humidity halos on aerosol extinction coefficient and AOD. The model is described in Fan et al. (2009) and based on the System for Atmospheric Modeling (SAM, Khairoutdinov and Randall 2003) coupled with the spectral bin microphysics for cloud particles and cloud condensation nuclei (Khain et al. 2004). We simulate the same June 12 case that was applied to the evaluation of the RR method in Section 3. The computation domain contains $192 \times 192 \times 128$ grid points at constant 50-meter horizontal and vertical spacing. The simulations span 8 hours with a 2-second time step.

Figure 9 shows the vertical cross-section through the model domain for the liquid water mixing ratio and RH fields. There is a clear correlation between the two fields with clouds located near the middle and at the top of wider regions of elevated RH. It is also clear that the strongest gradient in RH is in the vertical direction. Figure 10a shows that the horizontally averaged RH increases from 60% near the surface to the 86% maximum at the cloud base (Fig. 10b) before decreasing again to below 50% toward the top of the domain.

To evaluate the effect of the elevated RH near clouds on the aerosol optical properties, we compare aerosol extinction coefficients at 500 nm computed in two ways: first, using the actual predicted relative humidity at every grid point and then applying the minimum RH found at each level. By comparing the profiles of the extinction coefficient we can isolate the effects of horizontal variability in the humidity associated with clouds while preserving the altitude dependency due to mean vertical RH stratification (Fig. 10c). At the cloud base level, the humidity halos act to increase extinction by up to 40% on average. Note that this estimate is

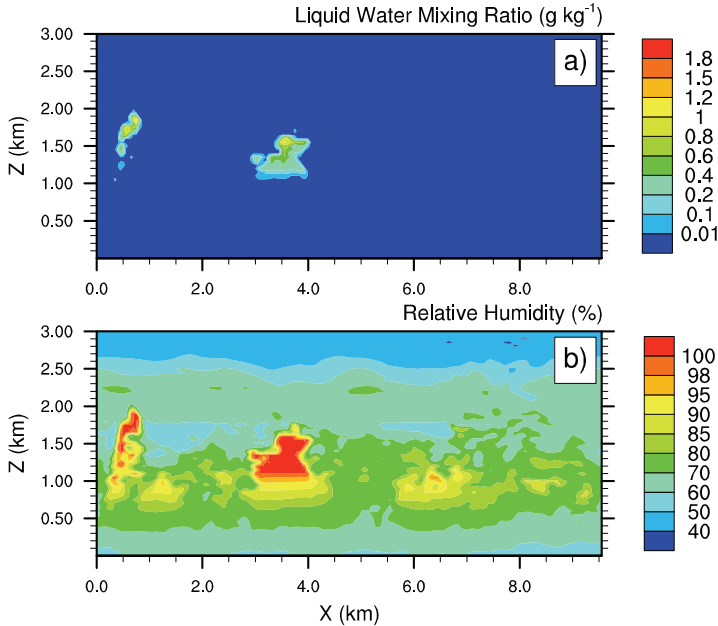


Fig. 9. Vertical cross sections of (a) liquid water mixing ratio and (b) relative humidity. Only the lowest three kilometers of the computational domain are shown to enhance details.

an upper limit for the cloud humidification effect because downdrafts at the sides of the clouds bring drier air to low altitudes and the minimum RH at the lower level tends to be lower in the presence of clouds than it would be otherwise. Enhanced extinction is confined primarily to a layer within 500 m of the cloud base level (~ 1 km). Consequently, the resulting effect on the column-integrated AOD is much smaller: the RH effect enhances the AOD by no more than 10%.

The above estimate is consistent with several previous independent observational studies. Jeong et al. (2007) used aircraft profiles and found that the column aerosol humidification factor, defined as the ratio of the AOD at the ambient RH to that at RH=40% throughout the column, has a mean value of 1.09 over the ARM SGP site. Analyses of HSRL data indicate that the AOD retrieved within about 0.1 km of a cloud is 8%–17% greater than that at distances > 4 km from the cloud (Su et al. 2008).

Uncertainties in aerosol retrievals could arise from inappropriate assumptions and errors associated with the assumed/specified aerosol type and, consequently, with the corresponding aerosol optical properties such as AOD, phase function, and single-scattering albedo (e.g. Kokhanovsky et al. 2010). These properties depend on the aerosol chemical composition and RH. Here we illustrated the impact of RH on the AOD only. Several studies demonstrated the effect of RH on other

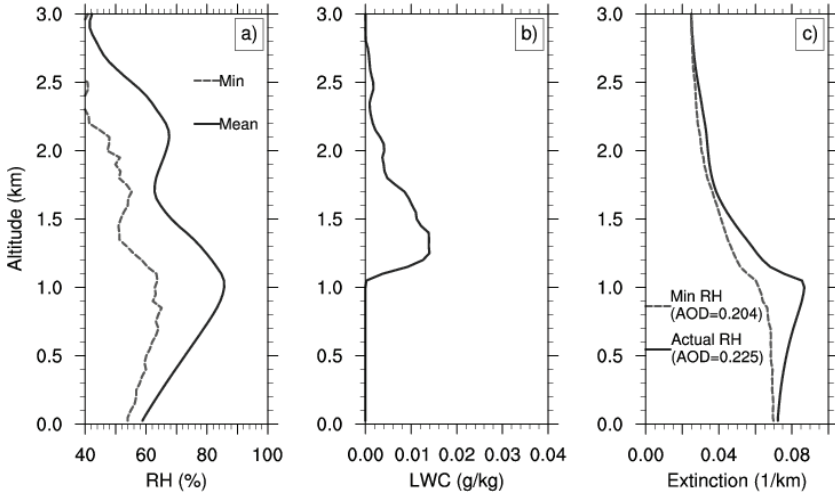


Fig. 10. Profiles of (a) the mean (solid curve) and minimum (dashed curve) relative humidities, (b) mean liquid water mixing ratio, and (c) aerosol extinction coefficients computed using the model-predicted local RH values (solid curve) and the minimum RH for each of the model layers (dashed curve).

aerosol properties as well (e.g., Moosmuller et al. 2009 and references therein). In particular, Wang and Martin (2007) showed that the effects of aerosol hygroscopicity on the aerosol phase function and aerosol single-scattering albedo increase with RH, and these effects are sensitive to the predominant mixing state (interval versus external).

5. Summary and outlook

Satellite-based passive observations of atmospheric aerosols have been performed extensively for several decades around the world under various cloud-free conditions (e.g., Remer et al. 2008; Lee et al. 2009). Typical long-term clear-sky aerosol data products from widely used MODIS and MISR observations include the column AOD as a function of wavelength and sometimes the Ångström exponent, which is a clue to the aerosol size composition. The modern strong demand for cloudy-sky aerosol data products together with improved cloud characteristics (e.g., Kaufman et al. 2005; Koren et al. 2010) brought new incentives for obtaining data from partly cloudy regions, which are particularly important for aerosol–cloud interaction projects. Although the passive clear-sky aerosol remote sensing has facilitated promising progress (e.g., Mishchenko et al. 2007b; van Donkelaar et al. 2010), an extension of existing aerosol retrievals to cloudy-sky conditions represents a great challenge due to the following two main factors.

The first one is defined by the cloud adjacency effects and their impact on passive aerosol retrievals. Solar light scattering by clouds makes a nearby clear patch look brighter, and the cloud-induced brightening is responsible for large (up to 140%) errors of the retrieved AOD (e.g., Wen et al. 2006, 2007). This brightening is associated with the complex 3D radiative effects of clouds (e.g., Davis and Marshak 2010), and these effects can be reduced, at least partially, by using recently suggested approaches based on the conditional sampling (Wen et al. 2006, 2007), parameterization of 3D effects (Marshak et al. 2008), and multi-spectral processing (e.g., Kassianov and Ovtchinnikov 2008). The suggested approaches can be considered as the infancy stage of more sophisticated methods that will be developed in the future for both research and operational mapping of cloudy-sky aerosol inventories (e.g., Davis et al. 2010), and significant efforts related to the cloud adjacency and subpixel cloud contamination effects are anticipated. One of the most potentially important application areas for the refined algorithms is to reuse historical satellite acquisitions of clear- and cloudy-sky AODs and to reexamine the corresponding multi-year trends in the context of the actively discussed global brightening and dimming problem (e.g., Wild et al. 2009). Model simulations of aerosol retrievals are shown to be extremely useful for assessing uncertainty in retrievals, and such modeling capabilities need to be maintained and expanded in the future.

The second outstanding factor is related to large variations of RH under cloudy conditions and simplified assumptions about RH in aerosol retrievals. Since RH changes control the enhancement in light scattering due to the uptake of water vapor by particles, inappropriate specification of RH is responsible for the corresponding uncertainties in aerosol optical properties. Typically, a fixed RH is assumed, often causing large (up to 40%) errors in the retrieved AOD even under cloud-free conditions (e.g., Wang and Martin 2007). To reduce the potentially large uncertainties associated with the aerosol hygroscopicity, additional information about RH, aerosol type, and chemical composition is required. This information can be provided by complementary and near-simultaneous active aerosol remote sensing (e.g., Warneke et al. 2010; Molina et al. 2010), ground-based retrievals (e.g., Schuster et al. 2009), *in situ* observations (e.g., Berg et al. 2009), and detailed model simulations (e.g., Jeong et al. 2007). Thus, the art and science for resolving this outstanding issue should involve combinations of remote sensing observations from various platforms, *in situ* measurements, and numerical modeling.

Compared to reflectance-based MODIS and MISR observations, which can only provide the basic aerosol properties (e.g., AOD and Ångström exponent) with substantial uncertainties (e.g., Li et al. 2009; Mishchenko et al. 2009), multi-spectral and multi-angle polarimetric observations are able to infer critical microphysical characteristics of aerosols and their chemical composition by source type with improved accuracy (e.g., Waquet et al. 2009; Dubovik et al. 2010). Thus, space-based measurements of polarized light scattered by aerosol and cloud particles combined with ground-based aerosol networks and *in situ* and model components

may create great opportunities of a fundamentally new understanding of aerosols and their interactions with clouds. Glory is a future NASA mission (Mishchenko et al. 2007a) that would include the Aerosol Polarimetry Sensor (APS) and offer enhanced aerosol retrievals from accurate multiangle photopolarimetric measurements of the Earth along the satellite ground track over a broad visible and near-infrared spectral range. The addition of the Glory spacecraft to the A-Train constellation (e.g., Stephens et al. 2002; L'Ecuyer and Jiang 2010) is expected to increase the unique capability of the NASA Earth Observing System and serve as a valuable basis for the formulation of future comprehensive satellite missions, such as Aerosol, Clouds and Ecosystem (ACE). The planned ACE mission would combine a next-generation APS, a space-based HSRL, and a multi-band spectrometer with near-ultraviolet, visible, and near-infrared spectral coverage.

Surface (e.g., Dubovik et al. 2002; Michalsky et al. 2010), maritime (e.g., Smirnov et al. 2009), and airborne (e.g., Redemann et al. 2009) observations provide a baseline for constraining and evaluating satellite retrievals and can be beneficial for separating the two main sources of error (3D radiative effects of clouds and RH variability). In particular, the direct-sun retrievals of AOD between the clouds are insensitive to the 3D radiative effects of clouds. An example of recently developed next-generation sensors is the airborne Spectrometer for Sky-Scanning, Sun-Tracking Atmospheric Research (4STAR) developed by the NASA Ames Research Center in collaboration with the Pacific Northwest National Laboratory, which combines the sun-tracking ability of the current 14-Channel NASA Ames Airborne Tracking Sunphotometer (AATS-14, e.g., Redemann et al. 2005) with the sky-scanning ability of the ground-based AERONET sun/sky photometers. The 4STAR measurements can provide improved retrievals of AOD, aerosol size modes, and complex refractive index. When used in conjunction with complementary spectral measurement of upwelling radiance, the airborne spectral measurements also permit the retrieval of cloud optical depth, effective radius, and liquid/ice fraction. Several exciting field campaigns (including the ARM-supported Two-Column Aerosol Project in 2012) are planned with both the observational and modeling perspective. These campaigns will make extensive use of the impressive achievement in the development of next-generation sensors with diverse observational capabilities and expected significant improvements in inverse algorithms.

There is a legendary story that King Arthur's sister Morgan le Fay possessed the ability to create castles in the air. Tiny atmospheric particles share this ability to build clouds and appear to be the critical factor in determining the beautiful appearance of clouds. Moreover, these particles play a critical role in a wide range of optical and radiative phenomena associated with clouds. The understanding of this ability of aerosol had been elusive for centuries. The advent of new and enhanced sensors and instruments together with awe-inspiring improvements in inverse-algorithms and model simulations are expected to lead to the discovery of hidden phenomena which unite aerosol and clouds and tear away the veil of mystery.

Acknowledgments

This work was supported by the National Aeronautics and Space Administration (NASA) through the Radiation Sciences Program and the Office of Biological and Environmental Research (OBER) of the U.S. Department of Energy (DOE) as part of the Atmospheric Radiation Measurement (ARM) Climate Research Facility and Atmospheric System Research (ASR). The Pacific Northwest National Laboratory (PNNL) is operated by Battelle for the DOE under contract DE-AC06-76RLO 1830. This research was performed in part using the Molecular Science Computing Facility (MSCF) on the Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by the U.S. DOE and OBER and located at PNNL.

References

- Ackerman, T. P., and G. M. Stokes, 2003: The Atmospheric Radiation Measurement Program. *Phys. Today* **56**(1), 39–44.
- Alexandrov, M. D., A. Marshak, B. Cairns, et al., 2004: Scaling properties of aerosol optical thickness retrieved from ground-based measurements. *J. Atmos. Sci.* **61**, 1024–1039.
- Anderson, T. L., R. L. Charlson, D. M. Winker, et al., 2003: Mesoscale variations of tropospheric aerosols. *J. Atmos. Sci.* **60**, 119–136.
- Andrews, E., P. J. Sheridan, J. A. Ogren, and R. Ferrare, 2004: *In situ* aerosol profiles over the Southern Great Plains cloud and radiation test bed site. 1. Aerosol optical properties. *J. Geophys. Res.* **109**, D06208.
- Barker, H. W., C. Pavloski, M. Ovtchinnikov, and E. E. Clothiaux., 2004: Assessing a cloud optical depth retrieval algorithm with model-generated data and the frozen turbulence assumption. *J. Atmos. Sci.* **61**, 2951–2956.
- Berg, L. K., C. M. Berkowitz, J. A. Ogren, et al., 2009: Overview of the cumulus humilis aerosol processing study (CHAPS). *Bull. Amer. Meteorol. Soc.* **90**, 1653–1667.
- Berg, L. K., C. M. Berkowitz, J. C. Barnard, et al., 2011: Observations of the first aerosol indirect effect in shallow cumuli. *Geophys. Res. Lett.* **38**, doi:10.1029/2010GL046047, in press.
- Cairns, B., E. E. Russell, and L. D. Travis, 1999: The Research Scanning Polarimeter: calibration and ground-based measurements. *Proc. SPIE* **3754**, 186–196.
- Cairns, B., F. Waquet, K. Knobelspiesse, et al., 2009: Polarimetric remote sensing of aerosols over land surfaces. In A. A. Kokhanovsky and G. de Leeuw, Eds., *Satellite Aerosol Remote Sensing over Land* (Praxis, Chichester, UK), pp. 295–325.
- Carrico, C. M., M. J. Rood, and J. A. Ogren, 1998: Aerosol light scattering properties at Cape Grim, Tasmania, during the First Aerosol Characterization Experiment (ACE 1). *J. Geophys. Res.*, **103**, 16 565–16 574.
- Chiu, J. C., C. H. Huang, A. Marshak, et al., 2010: Cloud optical depth retrievals from the Aerosol Robotic Network (AERONET) cloud mode observations. *J. Geophys. Res.* **115**, D14202.
- Cornert, C., L. C-Labonnote, and F. Szczap, 2010: Three-dimensional polarized Monte Carlo atmospheric radiative transfer model (3DMCPOL): 3D effects on polarized visible reflectances of a cirrus cloud. *J. Quant. Spectrosc. Radiat. Transfer* **111**, 174–186.

- Davis, A. B., and A. Marshak, 2010: Solar radiation transport in the cloudy atmosphere: a 3D perspective on observations and climate impacts. *Rep. Prog. Phys.* **73**, 026801.
- Davis, A. B., P. von Allmen, A. Marshak, and G. Bal, 2010: Toward a new polarization-based approach to the characterization of aerosols very near broken or isolated cumulus-type clouds. *Presentation at the NATO Advanced Study Institute on "Special Detection Technique (Polarimetry) and Remote Sensing"*, Sept. 12–25, Kyiv.
- Doicu, A., T. Trautmann, and F. Schreier, 2010: *Numerical Regularization for Atmospheric Inverse Problems* (Springer, Berlin).
- Doicu, A., T. Trautmann, and F. Schreier, 2011: Regularization of inverse problems in atmospheric remote sensing. In M. I. Mishchenko, Ya. S. Yatskiv, V. K. Rosenbush, and G. Videen, Eds., *Polarimetric Detection, Characterization, and Remote Sensing* (Springer, Berlin) (this volume).
- Dubovik, O., 2004: Optimization of numerical inversion in photopolarimetric remote sensing. In G. Videen, Ya. Yatskiv, and M. Mishchenko, Eds., 2004: *Photopolarimetry in Remote Sensing* (Kluwer, Dordrecht), pp. 65–105.
- Dubovik, O., B. N. Holben, T. F. Eck, et al., 2002: Variability of absorption and optical properties of key aerosol types observed in worldwide locations. *J. Atmos. Sci.* **59**, 590–608.
- Dubovik, O., M. Herman, A. Holdak, et al., 2010: Statistically optimized inversion algorithm for enhanced retrieval of aerosol properties from spectral multi-angle polarimetric satellite observations. *Atmos. Meas. Tech. Discuss.* **3**, 4967–5077.
- Engelen, R. J., and G. L. Stephens, 2004: Information content of infrared satellite sounding measurements with respect to CO₂. *J. Appl. Meteorol.* **43**, 373–378.
- Fan J., M. Ovtchinnikov, J. Comstock, et al., 2009: Ice formation in arctic mixed-phase clouds: insights from a 3D cloud-resolving model with size-resolved aerosol and cloud microphysics. *J. Geophys. Res.* **114**, D04205.
- Goering, C. D., T. S. L'Ecuyer, G. L. Stephens, et al., 2005: Simultaneous retrievals of column ozone and aerosol optical properties from direct and diffuse solar irradiance measurements. *J. Geophys. Res.* **110**, D05204.
- Gouesbet, G., 2009: Generalized Lorenz–Mie theories, the third decade: a perspective. *J. Quant. Spectrosc. Radiat. Transfer* **110**, 1223–1238.
- Gouesbet, G., and G. Grehan, 2000: Generalized Lorenz–Mie theories, from past to future. *Atomization Sprays* **10**, 277–333.
- Hansen, J. E., and L. D. Travis, 1974: Light scattering in planetary atmospheres. *Space Sci. Rev.* **16**, 527–610.
- Hegg, D. A., J. Livingston, P. V. Hobbs, et al., 1997: Chemical apportionment of aerosol column optical depth off the mid-Atlantic coast of the United States. *J. Geophys. Res.* **102**, 25293–25303.
- Hey, J. D., 1983: From Leonardo to the Graser: light scattering in historical perspective. *S. African J. Sci.* **79**, 11–27, 310–324.
- Hey, J. D., 1985: From Leonardo to the Graser: light scattering in historical perspective. *S. African J. Sci.* **81**, 77–91, 601–613.
- Jeong, M.-J., Z. Li, E. Andrews, and S.-C. Tsay, 2007: Effect of aerosol humidification on the column aerosol optical thickness over the Atmospheric Radiation Measurement Southern Great Plains site. *J. Geophys. Res.* **112**, D10202.
- Kahn, R. A., B. J. Gaitley, J. V. Martonchik, et al., 2005: Multiangle Imaging Spectroradiometer (MISR) global aerosol optical depth validation based on 2 years of coincident Aerosol Robotic Network (AERONET) observations. *J. Geophys. Res.* **110**, D10S04.

- Kassianov, E., and M. Ovtchinnikov, 2008: On reflectance ratios and aerosol optical depth retrieval in the presence of cumulus clouds. *Geophys. Res. Lett.* **35**, L06807.
- Kassianov, E., M. Ovtchinnikov, L. K. Berg, et al., 2009: Retrieval of aerosol optical depth in vicinity of broken clouds from reflectance ratios: sensitivity study. *J. Quant. Spectrosc. Radiat. Transfer* **110**, 1677–1689.
- Kassianov, E., M. Ovtchinnikov, L. K. Berg, et al., 2010: Retrieval of aerosol optical depth in vicinity of broken clouds from reflectance ratios: case study. *Atmos. Meas. Tech.* **3**, 1333–1349.
- Kaufman, Y., L. A. Remer, D. Tanré, et al., 2005: A critical examination of the residual cloud contamination and diurnal sampling effects on MODIS estimates of aerosol over ocean. *IEEE Trans. Geosci. Remote Sens.* **43**, 2886–2897.
- Khain, A. P., 2009: Notes on state-of-art investigations of aerosol effects on precipitation: a critical review. *Environ. Res. Lett.* **4**, 015004.
- Khain, A., A. Pokrovsky, M. Pinsky, et al., 2004: Simulation of effects of atmospheric aerosols on deep turbulent convective clouds using a spectral microphysics mixed-phase cumulus cloud model. Part I: Model description and possible applications. *J. Atmos. Sci.* **61**, 2963–2982.
- Khairoutdinov, M. F., and D. A. Randall, 2003: Cloud resolving modeling of the ARM summer 1997 IOP: model formulation, results, uncertainties, and sensitivities. *J. Atmos. Sci.* **60**, 607–625.
- Khlebtsov, B. N., V. A. Khanadeev, and N. G. Khlebtsov, 2008: Determination of the size, concentration, and refractive index of silica nanoparticles from turbidity spectra. *Langmuir* **24**, 8964–8970.
- Kobayashi, T., K. Masuda, M. Sasaki, and J. Mueller, 2000: Monte Carlo simulations of enhanced visible radiance in clear-air satellite fields of view near clouds. *J. Geophys. Res.* **105**, 26 569–26 576.
- Kokhanovsky, A. A. and de G. Leeuw, Eds., 2009: *Satellite Aerosol Remote Sensing over Land* (Praxis, Chichester, UK).
- Kokhanovsky, A. A., and P. Chylek, 2011: *Remote Sensing of the Atmosphere from Space* (Praxis, Chichester, UK).
- Kokhanovsky, J. L. Deuzé, D. J. Diner, et al., 2010: The inter-comparison of major satellite aerosol retrieval algorithms using simulated intensity and polarization characteristics of reflected light. *Atmos. Meas. Tech.* **3**, 909–932.
- Kokolova, L., B. Buratti, and V. Tishkovets, 2010: Impact of coherent backscattering on the spectra of icy satellites of Saturn and the implications of its effects for remote sensing. *Astropys. J.* **711**, L71–L75.
- Koren, I., L. A. Remer, Y. J. Kaufman, et al., 2007: On the twilight zone between clouds and aerosols. *Geophys. Res. Lett.* **34**, L08805.
- Koren, I., G. Feingold, and L. A. Remer, 2010: The invigoration of deep convective clouds over the Atlantic: aerosol effect, meteorology or retrieval artifact? *Atmos. Chem. Phys.* **10**, 8855–8872.
- Leaitch, W. R., U. Lohmann, L. M. Russell, et al., 2010: Cloud albedo increase from carbonaceous aerosol. *Atmos. Chem. Phys.* **10**, 7669–7684.
- L'Ecuyer, T. S., and J. H. Jiang, 2010: Touring the atmosphere aboard the A-Train. *Phys. Today* **63**(7), 36–41.
- Lee, K. H., Z. Li, Y. J. Kim, and A. A. Kokhanovsky, 2009: Atmospheric aerosol monitoring from satellite observations: A history of three decades. In Y. J. Kim et al., Eds., *Atmospheric and Biological Environmental Monitoring* (Springer, Berlin), pp. 13–38.

- Levy, R., L. Remer, S. Mattoo, et al., 2007: Second-generation algorithm for retrieving aerosol properties over land from MODIS spectral reflectance. *J. Geophys. Res.* **112**, D13211.
- Li, Z., X. Zhao, R. Kahn, et al., 2009: Uncertainties in satellite remote sensing of aerosols and impact on monitoring its long-term trend: a review and perspective. *Ann. Geophys.* **27**, 2755–2770.
- Lilienfeld, P., 2004: A blue sky history. *Opt. Photon. News* **15**(6), 32–39.
- Lohmann, U., and J. Feichter, 2005: Global indirect aerosol effects: a review. *Atmos. Chem. Phys.* **5**, 715–737.
- Marks, C., and C. Rodgers, 1993: A retrieval method for atmospheric composition from limb emission measurements. *J. Geophys. Res.* **98**, 14939–14953.
- Marshak, A. and A. B. Davis, Eds., 2005: *3D Radiative Transfer in Cloudy Atmospheres* (Springer, Berlin).
- Marshak, A., G. Wen, J. Coakley, et al., 2008: A simple model for the cloud adjacency effect and the apparent bluing of aerosols near clouds. *J. Geophys. Res.* **113**, D14S17.
- Mayer, B., 2009: Radiative transfer in the cloudy atmosphere. *Eur. Phys. J. Conf.* **1**, 75–99.
- McComiskey, A., G. Feingold, F. A. Shelby, et al. 2009: An assessment of aerosol-cloud interactions in marine stratus clouds based on surface remote sensing. *J. Geophys. Res.* **114**, D09203.
- Mensbrugge, V., 1892: On the formation of fog and of clouds. *Simon's Mon. Meteorol. Mag.* **27**, 40–41.
- Michalsky, J., F. Denn, C. Flynn, et al., 2010: Climatology of aerosol optical depth in north-central Oklahoma: 1992–2008. *J. Geophys. Res.* **115**, D07203.
- Mie, G., 1908: Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen. *Ann. Phys.* **25**, 377–445.
- Mishchenko, M. I., 2008: Multiple scattering, radiative transfer, and weak localization in discrete random media: unified microphysical approach. *Rev. Geophys.* **46**, RG2003.
- Mishchenko, M. I., 2009: Gustav Mie and the fundamental concept of electromagnetic scattering by particles: a perspective. *J. Quant. Spectrosc. Radiat. Transfer* **110**, 1210–1222.
- Mishchenko, M. I., 2010: The Poynting–Stokes tensor and radiative transfer in discrete random media: the microphysical paradigm. *Opt. Express* **18**, 19770–19791.
- Mishchenko, M. I., and L. D. Travis, 2008: Gustav Mie and the evolving discipline of electromagnetic scattering by particles. *Bull. Amer. Meteorol. Soc.* **89**, 1853–1861.
- Mishchenko, M. I., B. Cairns, G. Kopp, et al., 2007a: Accurate monitoring of terrestrial aerosols and total solar irradiance: introducing the Glory mission. *Bull. Amer. Meteorol. Soc.* **88**, 677–691.
- Mishchenko, M. I., I. V. Geogdzhayev, B. Cairns, et al., 2007b: Past, present, and future of global aerosol climatologies derived from satellite observations: a perspective. *J. Quant. Spectrosc. Radiat. Transfer* **106**, 325–347.
- Mishchenko, M. I., I. V. Geogdzhayev, L. Liu, et al., 2009: Toward unified satellite climatology of aerosol properties: what do fully compatible MODIS and MISR aerosol pixels tell us? *J. Quant. Spectrosc. Radiat. Transfer* **110**, 402–408.
- Mishchenko, M. I., V. K. Rosenbush, N. N. Kiselev, et al., 2010: *Polarimetric Remote Sensing of Solar System Objects* (Akademperiodyka, Kyiv) (arXiv:1010.1171).
- Molina, L. T., S. Madronich, J. S. Gaffney, et al., 2010: An overview of the MILAGRO 2006 campaign: Mexico City emissions and their transport and transformation. *Atmos. Chem. Phys.* **10**, 8697–8760.

- Moosmüller, H., R. K. Chakrabarty, and W. P. Arnott, 2009: Aerosol light absorption and its measurement: a review. *J. Quant. Spectrosc. Radiat. Transfer* **110**, 844–878.
- Nakajima, T., A. Higurashi, K. Kawamoto, and J. Penner, 2001: A possible correlation between satellite-derived cloud and aerosol microphysical parameters. *Geophys. Res. Lett.* **28**, 1171–1174.
- Noone, K. B., K. J. Noone, and J. A. Ogren, 1993: *In situ* observations of cloud microphysical properties using the counterflow virtual impactor. *J. Atmos. Oceanic Technol.* **10**, 294–303.
- Ovtchinnikov, M., and R. T. Marchand, 2007: Cloud model evaluation using radiometric measurements from the airborne multiangle imaging spectroradiometer (AirMISR), *Remote Sens. Environ.* **107**, 185–193.
- Perry, K. D., and P. V. Hobbs, 1996: Influences of isolated cumulus clouds on the humidity of their surroundings. *J. Atmos. Sci.* **53**, 159–174.
- Prigarin, S. M., and A. Marshak, 2009: A simple stochastic model for generating broken cloud optical depth and cloud-top height fields. *J. Atmos. Sci.* **66**, 92–104.
- Qaaas, J., O. Boucher, N. Bellouin, and S. Kinne, 2008: Satellite-based estimate of the direct and indirect aerosol climate forcing. *J. Geophys. Res.* **113**, D05204.
- Quinn, P. K., T. L. Miller, T. S. Bates, et al., 2002: A 3-year record of simultaneously measured aerosol chemical and optical properties at Barrow, Alaska. *J. Geophys. Res.* **107**, 4130.
- Redemann, J., R. P. Turco, K. N. Liou, et al., 2000: Retrieving the vertical structure of the effective aerosol complex index of refraction from a combination of aerosol *in situ* and remote sensing measurements during TARFOX. *J. Geophys. Res.* **105**, 9949–9970.
- Redemann, J., B. Schmid, J. A. Eilers, et al., 2005: Suborbital measurements of spectral aerosol optical depth and its variability at sub-satellite grid scales in support of CLAMS 2001. *J. Atmos. Sci.* **62**, 993–1007.
- Redemann, J., Q. Zhang, P. B. Russell, et al., 2009: Case studies of aerosol remote sensing in the vicinity of clouds. *J. Geophys. Res.* **114**, D06209.
- Remer, L. A., R. G. Kleidman, R. C. Levy, et al., 2008: Global aerosol climatology from the MODIS satellite sensors. *J. Geophys. Res.* **113**, D14S07.
- Rodgers, C. D., 2000: *Inverse Methods for Atmospheric Sounding: Theory and Practice* (World Scientific, Singapore).
- Rossow, W. B., and E. N. Duenas, 2004: The International Satellite Cloud Climatology Project (ISCCP) web site: an online resource for research. *Bull. Am. Meteorol. Soc.* **85**, 167–172.
- Schmidt, K. S., G. Feingold, P. Pilewskie, et al., 2009: Irradiance in polluted cumulus fields: measured and modeled cloud-aerosol effects. *Geophys. Res. Lett.* **36**, L07804.
- Schuster, G. L., B. Lin, and O. Dubovik, 2009: Remote sensing of aerosol water uptake. *Geophys. Res. Lett.* **36**, L03814.
- Sheridan, P. J., D. J. Delene, and J. A. Ogren, 2001: Four years of continuous surface aerosol measurements from the Department of Energy's Atmospheric Radiation Measurement Program Southern Great Plains Cloud and Radiation Testbed site. *J. Geophys. Res.* **106**, 20 735–20 747.
- Small, J. D., P. Y. Chuang, G. Feingold, and H. Jiang, 2009: Can aerosol decrease cloud lifetime? *Geophys. Res. Lett.* **36**, L16806.
- Smirnov, A., B. N. Holben, I. Slutsker, et al., 2009: Maritime Aerosol Network as a component of Aerosol Robotic Network. *J. Geophys. Res.* **114**, D06204.
- Stephens, G. L., 1994: *Remote Sensing of the Lower Atmosphere: An Introduction* (Oxford University Press, New York).

- Stephens, G. L., 2003: The useful pursuit of shadows. *Amer. Sci.* **91**, 442–449.
- Stephens, G. L., and C. Kummerow, 2007: The remote sensing of clouds and precipitation from space: a review. *J. Atmos. Sci.* **64**, 3742–3765.
- Stephens, G. L., D. G. Vane, R. J. Boain, et al., 2002: The CloudSat mission and the A-train: a new dimension of space-based observations of clouds and precipitation. *Bull. Am. Meteorol. Soc.* **83**, 1771–1790.
- Stevens, B., and Feingold, G., 2009: Untangling aerosol effects on clouds and precipitation in a buffered system. *Nature* **461**, 607–613.
- Su, W., G. L. Schuster, N. G. Loeb, et al., 2008: Aerosol and cloud interaction observed from high spectral resolution lidar data. *J. Geophys. Res.* **113**, D24202.
- Tackett, J. L., and L. Di Girolamo, 2009: Enhanced aerosol backscatter adjacent to tropical trade wind clouds revealed by satellite-based lidar. *Geophys. Res. Lett.* **36**, L14804.
- Turner, D. D., 2008: Ground-based infrared retrievals of optical depth, effective radius, and composition of airborne mineral dust above the Sahel. *J. Geophys. Res.* **113**, D00E03.
- Twohy, C. H., J. A. Coakley Jr., and W. R. Tahnk, 2009: Effect of changes in relative humidity on aerosol scattering near clouds. *J. Geophys. Res.* **114**, D05205.
- van Donkelaar, A., R. V. Martin, M. Brauer, et al., 2010: Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: development and application. *Environ. Health Persp.* **118**, 847–855.
- Varnai, T., and A. Marshak, 2009: MODIS observations of enhanced clear sky reflectance near clouds. *Geophys. Res. Lett.* **36**, L06807.
- Venema, V., S. G. Garcia, and C. Simmer, 2010: A new algorithm for the downscaling of cloud fields. *Quart. J. R. Meteorol. Soc.* **136**, 91–106.
- Videen, G., Ya. Yatskiv, and M. Mishchenko, Eds., 2004: *Photopolarimetry in Remote Sensing* (Kluwer, Dordrecht).
- Wang, J., and S. T. Martin, 2007: Satellite characterization of urban aerosols: importance of including hygroscopicity and mixing state in the retrieval algorithms. *J. Geophys. Res.* **112**, D17203.
- Waquet, F., B. Cairns, K. Knobelspiesse, et al., 2009: Polarimetric remote sensing of aerosols over land. *J. Geophys. Res.* **114**, D01206.
- Warneke, C., K. D. Froyd, J. Brioude, et al., 2010: An important contribution to springtime Arctic aerosol from biomass burning in Russia. *Geophys. Res. Lett.* **37**, L01801.
- Wax, A., and V. Backman, 2010: *Biomedical Applications of Light Scattering* (McGraw Hill, New York).
- Wen, G., A. Marshak, and R. Cahalan, 2006: Impact of 3-D clouds on clear sky reflectance and aerosol retrieval in a biomass burning region of Brazil. *IEEE Geosci. Remote Sens. Lett.* **3**, 169–172.
- Wen, G., A. Marshak, R. F. Cahalan, et al., 2007: 3D aerosol-cloud radiative interaction observed in collocated MODIS and ASTER images of cumulus cloud fields. *J. Geophys. Res.* **112**, D13204.
- Wild, M., B. Trüssel, A. Ohmura, et al., 2009: Global dimming and brightening: an update beyond 2000. *J. Geophys. Res.* **114**, D00D13.
- Yang, Y., and L. Di Girolamo, 2008: Impacts of 3-D radiative effects on satellite cloud detection and their consequences on cloud fraction and aerosol optical depth retrievals. *J. Geophys. Res.* **113**, D04213.
- Yu, H., M. Chin, D. M. Winker, et al., 2010: Global view of aerosol vertical distributions from CALIPSO lidar measurements and GOCART simulations: regional and seasonal variations. *J. Geophys. Res.* **115**, D00H30.

Zhao, G., L. Di Girolamo, S. Dey, et al., 2009: Examination of direct cumulus contamination on MISR-retrieved aerosol optical depth and Ångström coefficient over ocean. *Geophys. Res. Lett.* **36**, L13811.



The Kassianov family: Ivan, Rita, Marina, and Zhenya (from left to right).



Left: Ludmilla Kolokolova and Jim Hough discuss astrophysical polarimetry at the first poster session. Right: Irina Kulyk and Kristan Gurton enjoy the field trip to the Main Astronomical Observatory.

Astrophysical polarimetry in Ukraine

Nikolai N. Kiselev^{1*} and Michael I. Mishchenko²

¹Main Astronomical Observatory of the National Academy of Sciences of Ukraine, 27 Akademika Zabolotnoho St., 03680 Kyiv, Ukraine

²NASA Goddard Institute for Space Studies, 2880 Broadway, New York, NY 10025, USA

Abstract. The history of astrophysical polarimetry in Ukraine since the mid 1920s is broadly outlined. It is demonstrated that Ukrainian scientists have made fundamental contributions to the development and implementation of polarimetric methods in astrophysical research, including polarimetric measurement techniques, registration and data reduction methodologies, and a rigorous theory of electromagnetic scattering by particulate media. They have also carried out extensive polarimetric observations of numerous celestial objects and contributed to their physically-based interpretation.

Keywords: polarimeters, remote sensing, polarimetric techniques and methods, theory of electromagnetic scattering, observations and interpretations

1. Introduction

The definitive discovery of the electromagnetic wave nature of light by James Clerk Maxwell (see Maxwell 1891) has enabled the study of not only the simplest scalar attribute of light (the intensity or brightness), but also its “vectorial” parameters, i.e., the degree and orientation of linear polarization and the degree and helicity of circular polarization (Stokes 1852). Different types of scattering particles (e.g., aerosols, haze particles, ice crystals, and dust with different compositions and morphologies) exhibit characteristic polarization properties in the scattered, transmitted, and emitted light (Hansen and Travis 1974; Mishchenko et al. 2002, 2010). As a consequence, polarimetric observations can provide valuable information on the nature of scattering particles and ambient physical processes that cannot be retrieved from other types of observation. However, only in the first half of the 19th century polarimetry ceased to be an exclusively laboratory method of research. Polarization was discovered for the light scattered by celestial bodies such as the Moon (1811) and comets (1819) (Arago 1854–1857). The work by Lyot (1929) contributed profoundly to polarimetry of planets, planetary satellites, and Saturn’s rings. For the first time, he demonstrated the presence of the negative

* Corresponding author. E-mail: kiselevnn@yandex.ru

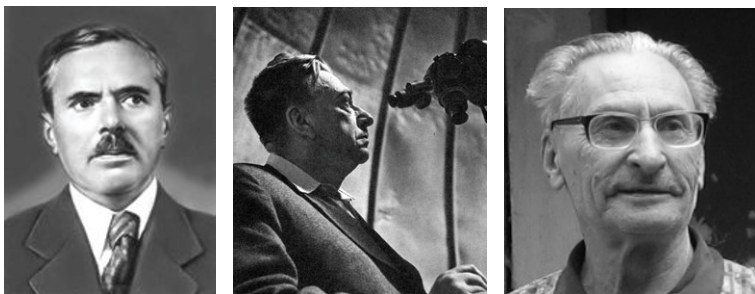


Fig. 1. From left to right: Grigory A. Shain (1892–1956), Nikolai P. Barabashov (1894–1971), and Nikolay M. Shakhovskoy (1931–2011).

polarization branch (NPB) at small phase angles for the Moon. In the late 1940s, Hiltner (1949), Hall (1949), and Dombrovsky (1949) discovered independently that light from many stars is partly linearly polarized.

Nevertheless, until the mid 1950s only a handful of astrophysicists worked in the field of polarimetry. The instrumentation was simple, and the accuracy on the order of $\pm 0.2\%$ was typical. It was even worse than the accuracy of Lyot's visual polarimetry, which was on the order of $\pm 0.1\%$. Despite offering the obvious benefit of quadrupling the number of observable parameters, polarimetry had not been widely applied for two reasons. The accuracy of polarization measurements was insufficient, and the interpretation of measurement data had suffered from the lack of a proper theoretical understanding of the relevant scattering, absorption, and emission phenomena. Efficient practical applications of high-precision polarimetry became possible only after the advent of the photoelectric technique and the development of adequate physically-based theoretical tools. At present, astrophysical polarimetry—along with polarimetric remote sensing and polarimetric laboratory characterization—is a thriving area of research, as this volume vividly demonstrates.

Ukrainian astrophysicists have contributed profoundly to polarimetry as an independent and uniquely informative method of remote-sensing and laboratory research, as will be outlined in the following sections. Our objective is not to present a comprehensive survey of numerous, multi-decadal studies but rather to summarize the most important developments according to our personal and, thus, somewhat subjective viewpoint.

2. Development of techniques and methodologies of polarimetric measurements and observations

2.1. Polarimeters for telescopic observations

Active applications of polarimetry in astrophysics began in the mid 20th century. The first director of the Crimean Astrophysical Observatory (CrAO)

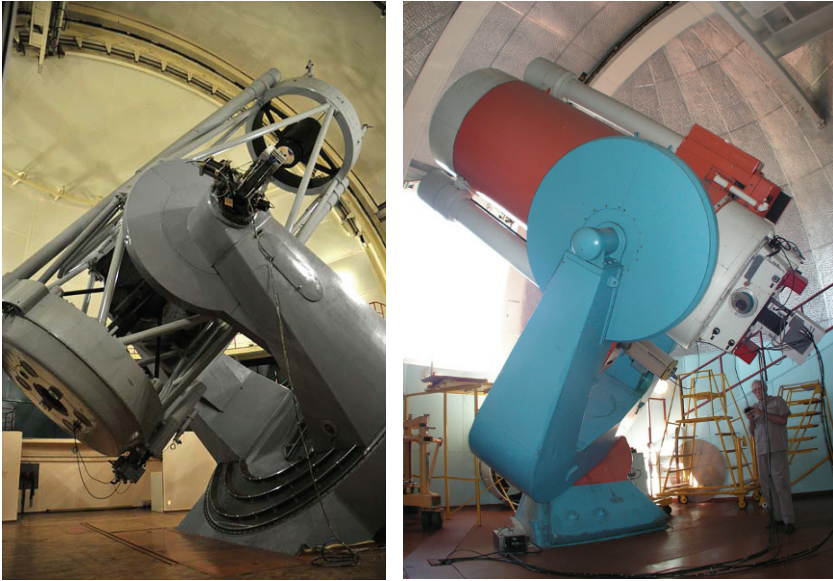


Fig. 2. 2.6-m (left) and 1.25-m CrAO telescopes used for polarimetric observations.

G. A. Shain (Fig. 1) identified a correlation between the orientations of filaments in diffuse nebulae and the plane of polarization of neighboring stars and suggested that it could be caused by a global magnetic field of our Galaxy (Shain 1955). Shain determined the general structure of the magnetic field and obtained one of the first estimates of its strength; the latter agrees well with more recent assessments. Shain et al. (1955) used photographic techniques and the 0.4-m CrAO astrograph to study the Crab nebula and showed that its light is polarized, the polarization being different for different parts of the nebula. This result provided an independent demonstration of the synchrotron nature of the emitted radiation and the complex morphology of the magnetic field of the nebula.

Systematic polarization observations in Ukraine began in the early 1960s. Measurements of circular polarization were identified as an effective method of study of the magnetic fields of the Sun and bright stars based on the Zeeman effect. CrAO astrophysicists Nikulin et al. (1958) and Severny (1970) developed the technique of solar magnetography. Specifically, they placed a combination of an ADP crystal and a polaroid behind the entrance slit of the spectrograph mounted at a solar telescope. Measurements of the difference between the left- and right-hand circularly polarized intensity components of the profiles of magnetically sensitive spectral lines proved to be efficient in systematic recordings (since 1967) of the general magnetic field of the Sun. This method was successfully implemented on the Coude spectrograph of the 2.6-m CrAO telescope (Fig. 2) for the measurement of magnetic fields (30–300 Gauss) of certain bright stars (Severny 1970; Severny et al. 1974).

The development of a dual-beam polarimeter with an acousto-optical modulator became the next stage in the measurement of stellar magnetic fields (Weiss et al. 1990). This method is based on observations of circular polarization in the wings of the H_{β} line and has been shown to enable reliable measurements of magnetic fields of about 500–1000 Gauss for stars as faint as 7–8 mag. Currently, the measurements of stellar magnetic fields are performed by using a polarization analyser (Stokes-meter) (Plachinda 1997). The instrument consists of an entrance retarding achromatic 90° -step quarter-wave plate, an Iceland spar plate, and an exit achromatic quarter-wave plate mounted in front of the entrance slit of the spectrograph in the Coude focus of the 2.6-m telescope. During two successive exposures using two orientations (0° and 90°) of the entrance quarter-wave plate, the contour of a spectral line detected in right- and left-hand circular polarizations is sequentially projected onto the same part of the CCD receiver.

In 1960s, measurements of linear polarization for different objects were carried out with the CrAO integrating polarimeter by Shakhovskoy and Dimov (1962) and with an automatic polarimeter of the Main Astronomical Observatory (MAO) of the National Academy of Sciences of Ukraine by O. I. Bugaenko using an analog registration technique and a slowly rotating polaroid. Since 1968, the polarimeters of these observatories have been substantially modernised for operation in the photon-counting mode using the principle of synchronous detection (Bugaenko et al. 1968; Shakhovskoy and Efimov 1972, 1976).

An important stage in the advancement of astrophysical polarimetry was the development of instruments capable of measuring circular polarization. However, the truly widespread use of polarimeters allowing one to measure both linear and circular polarization of light became possible in 1991–92 owing to the emergence of achromatic phase plates designed by Kucherov (1985). Kucherov developed a general theory of multicomponent superachromatic retarders that can be used to design optical systems with any degree of achromatism by including a sufficiently large number of combinations of optical components.

Crimean astrophysics, in collaboration with astrophysics from the Institute of Astronomy of the Odessa National University (hereinafter OIA), modified the single-channel polarimeter of the 2.6-m CrAO telescope (Shakhovskoy et al. 2001). The polaroid was replaced with a rotating achromatic quarter-wave phase plate followed by a fixed polaroid and, in the most recent modification, the Glan prism.

The installation of the 1.25-m telescope (Fig. 2) and the availability of the five-channel photopolarimeter of the Helsinki University Observatory (Piirola 1988) have dramatically expanded the range of polarimetric observations in CrAO. Polarization measurements can be performed simultaneously in five spectral channels U, B, V, R, and I. The software package used in the registration of data was further improved by Berdyugin and Shakhovskoy (1993).

An original spectropolarimeter called “Planetary Patrol” includes a rotating plate causing a 127° phase shift (Kucherov 1986), thereby allowing one to perform measurements of linear and circular polarization with the same efficiency. This instrument was designed and built in MAO (Bugaenko and Gural’chuk 1985)

and was used to collect a large volume of polarimetric and photometric data for comet 1P/Halley in the framework of the International Halley Watch Program at the 0.6-m telescope of the Soviet–Bolivian Astronomical Observatory in Tarija, Bolivia. Recently, this spectropolarimeter has been significantly modernized (Vid'machenko et al. 2004). The new photomultiplier allows one to collect data in the extended spectral range 330–1060 nm. The software has also been improved, allowing one to control the instrument from a personal computer. Most recently a new project was initiated with the purpose of designing and building a spectropolarimeter for ground-based observations accompanying the space experiment “Planetary Monitoring” (Vid'machenko et al. 2007).

Polarimetric instruments and methods of polarimetric observations have also been developed in the Institute of Astronomy of the Kharkiv V. N. Karazin National University (hereinafter KhIA). Since 1983, polarimetry of asteroids has been carried out in collaboration with scientists from CrAO and the Institute of Astrophysics of the Academy of Sciences of Tadjikistan using the 1.25-m reflector and a 1.0-m RCC telescope, respectively. Since 1995, regular observations of asteroids, comets, and planetary satellites have been carried out with a single-channel photopolarimeter mounted at a 0.7-m reflecting telescope. The polarimeter used and the polarimetric observation techniques are similar to those developed in CrAO (Kiselev and Velichko 1997). Korokhin et al. (1993, 2000) have developed methods, instrumentation, and original software packages for polarimetric observations of the Sun, the Moon, and Jupiter using CCD matrices.

2.2. Polarimeters for laboratory measurements

Besides polarimeters intended for telescopic observations, MAO scientists have developed instrumentation for laboratory polarization measurements. These laboratory setups have been used extensively for the study of polarization characteristics of different terrestrial and man-made particulate surfaces (Degtjarev and Kolokolova 1990; Degtjarev et al. 1992).

Laboratory simulations of light scattering by regolith surfaces have become one of the main foci of research at KhIA. It is well known that the phase-angle dependence of the Stokes parameters of the reflected light carries important information on macro- and microphysical properties of particulate surfaces. To exploit this information content, the Kharkiv astrophysicists have developed several generations of laboratory polarimetric goniometers, including a unique instrument working at extremely small phase angles, down to 0.008° (Fig. 3; Shkuratov and Ovcharenko 2002a; Shkuratov et al. 2008a,c).

2.3. Space-borne polarimeters

In 1974, CrAO scientists in collaboration with their Swedish colleagues developed a spectrometer–polarimeter intended to record polarization in the 120–150 nm region of the solar spectrum. It was launched onboard the satellite Inter-



Fig. 3. KhIA photometer–polarimeter operating at extremely small phase angles.

cosmos-16 on 27 July 1976. The experiment demonstrated that Lyman- α polarization measurements of the solar limb can be a useful diagnostic tool for the study of chromospheric and coronal magnetic fields of the Sun (Stenflo et al. 1980).

MAO and CrAO scientists proposed an innovative design of a low-spectral-resolution ultraviolet spectropolarimeter (UVSPEPOL) without a slit intended for use on space missions (Kucherov et al. 1997; Yefimov 2004). The instrument design involves original optical elements. Unlike the conventional classical scheme, the rotating superachromatic quarter-wave phase plate is followed by a composite Wollaston prism with two off-centered elements and deformed surfaces. This prism functions as the dispersive element as well as the polarization analyzer. This allows one to obtain two spectra with orthogonal polarization directions on one CCD image. UVSPEPOL was designed for the orbital telescope T-170. A ground-based prototype of this spectropolarimeter has also been built.

In recent years, MAO scientists have proposed several projects involving ultraviolet spectropolarimeters intended for planetary monitoring onboard the International Space Station (Ivanov et al. 2002, 2004) as well as for the study of terrestrial stratospheric aerosols from a microsatellite “UMS-1” (Nevodovskij et al. 2004; Nevodovsky and Morozhenko 2009).

2.4. Experimental methodologies, analysis techniques, and observational programs

The development of polarimetric instrumentation in Ukraine has been accompanied by the development of basic principles of polarimetric measurements, detection and registration methodologies, and data processing algorithms as well as

their practical implementation. Polarimetric measurements have several distinctive features, the most important one being that the difference between two intensity components can be very small compared to the intensities themselves. Furthermore, one needs to correct very carefully for instrumental and atmospheric effects distorting the measurement results. There are also significant differences in the measurement processes and data reduction. Great attention has, therefore, been paid to the analysis of random and systematic errors of polarimetric measurements. Efimov (1970) and Shakhovskoy (1971) performed a detailed study of common sources of random and systematic errors and their effects on the resulting polarimetric accuracy. The reduction of polarimetric data and practical aspects of measurements of linear and circular polarization of light scattered by celestial bodies have been discussed by Shakhovskoy and Efimov (1972, 1976), Bugaenko and Gural'chuk (1985), Berdyugin and Shakhovskoy (1993), and Shakhovskoy et al. (2001). Practical implementations of the theory of multicomponent superachromatic retarders, methods and specific aspects of polarimetric observations of satellites and comets as well as stellar occultations by comets have been summarized in Mishchenko et al. (2010).

As a result of these scientific and engineering developments, several centers of polarimetric research (CrAO, MAO, KhIA, and OIA) have been very active in Ukraine since the 1960s. Polarimetric studies in CrAO have included the following:

- development of polarimetric instrumentation and measurement methodologies;
- development of theory and data reduction techniques;
- observations of different types of variable stars, nebulae, and galaxies; and
- observations of planetary satellites, asteroids, and comets (in collaboration with KhIA and MAO).

Research programs in MAO have primarily been focused on:

- development of polarimetric instrumentation and methods of polarization measurements;
- development of fundamental theory of electromagnetic scattering and data interpretation; and
- polarimetric studies of various solar-system bodies (planets, planetary satellites, asteroids, and comets).

The foci of polarimetric research at KhIA have been the following:

- development of polarimetric instrumentation;
- theoretical studies, computer modeling, and laboratory simulations of light scattering by regolith surfaces; and
- studies of polarimetric properties of planets and comets as well as of atmosphereless solar-system bodies (ASSBs) such as the Moon, asteroids, and trans-Neptunian objects (TNOs).

Finally, polarimetric investigations at OIA have focused on the study of cataclysmic variable stars, in collaboration with the CrAO colleagues.

Less systematic polarization studies, mostly of comets, have been performed at the Astronomical Observatory of the T. G. Shevchenko Kyiv National University (e.g., Churyumov et al. 1986).

3. Multi-decadal observational programs and physically-based data analyses

3.1. The Moon and other planetary satellites

Apparently the first polarimetric observations in Ukraine were carried out by N. P. Barabashov (Fig. 1) at KhIA in 1923 (Barabaschoff 1926). He measured polarization for different areas of the Moon with a Cornu polarimeter mounted at a 3-inch reflector and compared his results with laboratory data for different terrestrial powders. Thus Barabashov, in essence, pioneered the technique of comparisons of the observed scattering properties of ASSBs with those of appropriately selected laboratory analogues. In more recent studies (e.g., Akimov et al. 1985; Shkuratov and Ovcharenko 2002b), measurements of the scattering phase function and the degree of linear polarization have been carried out for a wide range of samples: from terrestrial and meteoritic ones to lunar soils.

The most important area of lunar research in KhIA over the past several decades has been the synthesis of images or maps of the Moon for different photometric and polarization parameters and analyses of their correlation with properties of the lunar regolith. For example, Shkuratov et al. (1980) identified a relationship between structural characteristics of the lunar surface (such as porosity and the median particle size) and deviations from the line of regression of the cross-correlation dependence “albedo (A) versus degree of linear polarization (P)”. A photopolarimetric catalogue of particular lunar details was compiled by Opanasenko and Shkuratov (1994). Opanasenko et al. (1994) composed the first image of the lunar surface in terms of the distribution of the minimum-polarization parameter (P_{\min}). Maps (at the 88° phase angle) of the lunar brightness, degree of polarization, and a parameter representing the median particle size of the lunar regolith were compiled by Shkuratov et al. (2008b). The first two maps strongly anti-correlate, thereby exhibiting the famous so-called Umov effect. The latter map demonstrates that the surfaces of lunar continents are, on average, more fine-grained, while bright young craters are covered with medium-sized regolith grains indicative of relatively immature soils.

The search for common polarimetric features typical of high-albedo planetary satellites observed near opposition has become one of the principal foci of astrophysical research in Ukraine, the main stimulus coming from physically-based theoretical studies of the effect of coherent backscattering (CB; Mishchenko 1992a, b, 1993). Observations during the 2000 and 2008 oppositions (Rosenbush and Kiselev 2005; Kiselev et al. 2009) have fully corroborated the earlier conclusion by Rosenbush et al. (1997a) that the Jovian satellites Io, Europa, and Ganymede exhibit a pronounced and narrow secondary minimum of polarization at phase angles smaller than 2° superposed on a regular NPB typical of the majority

of ASSBs. This result constitutes the observational discovery of the polarization opposition effect (POE) predicted theoretically by Mishchenko (1993). The simultaneous presence of the pronounced and narrow brightness opposition effect (BOE) and POE for the high-albedo satellites suggests that a significant fraction of the satellite surfaces is covered with fine-grained material causing strong CB (Mishchenko et al. 2006a).

Rosenbush (2002) developed a method for the separation of phase-angle and longitudinal dependences of polarization for another Jovian satellite, Callisto. This method enables one to study the distribution of albedo over satellite surfaces. The application of this technique to Callisto revealed no POE.

3.2. Asteroids, trans-Neptunian objects, and comets

Regular and systematic polarimetric observations of asteroids began in 1983 and have been carried out at KhIA in collaboration with CrAO, MAO, the Sanglok Observatory (Tadjikistan; features a 1-m telescope), the Peak Terskol Observatory (Russia; 2-m telescope), the Complejo Astronomico El Leoncito (Argentina; 2.15-m telescope), the European Southern Observatory (ESO, Chile; 8.2-m telescope), and other astronomical organizations. Over the past 25 years, polarimetric data for ~ 200 asteroids of different sizes and types, 10 of which are near-Earth asteroids (NEAs), have been accumulated. These numbers represent $\sim 70\%$ of all asteroids and $\sim 70\%$ of all NEAs studied by means of polarimetry. The main foci of research have been the measurement of polarization phase curves, the study of optical and geometrical characteristics of asteroids, the asteroid type classification, analyses of the spectral behavior of polarization, observations of NEAs over a wide range of illumination–observation geometries, the identification and analysis of asteroids with anomalous characteristics, etc.

These results, combined with data from the Asteroid Polarimetric Database (Lupishko and Vasil'ev 2008), allow one to derive and analyze the average parameters of the phase dependence of polarization — P_{\min} , α_{\min} (the phase angle of minimal polarization), α_{inv} (the inversion phase angle), and the polarimetric slope h at the inversion phase angle — for basic F-, C-, M-, S-, A-, and E-type asteroids (Lupishko 1997; Belskaya 2007). In general, asteroids from the same compositional type show very similar values of the respective polarimetric parameters. However, $\sim 10\%$ of asteroids studied so far have been found to exhibit significant deviations of their polarization phase curves from the average polarization curves of their respective compositional types (Mishchenko et al. 2010, and references therein). It has been found that the inversion angles α_{inv} of asteroids cover a wide range of phase angles from uniquely small, $\sim 14^\circ$ for (419) Aurelia (F-type), to uniquely large, $\sim 28^\circ$ for (234) Barbara (S-type) (Belskaya et al. 2005; Cellino et al. 2006). Furthermore, it has been shown that the minimal polarization values P_{\min} for these asteroids violate significantly the well-known correlation “ P_{\min} – albedo”. The anomalous polarimetric properties of these asteroids could be related to their specific surface compositions.

It has been shown that the wavelength dependence of P_{\min} is primarily defined by the asteroid type, although the variability within a specific asteroid type can be significant. Belskaya et al. (1989) measured $P_{\min}(\lambda)$ in the laboratory for samples of 15 different meteorites as well as for 4 samples of terrestrial silicates (olivine, bronzite, hypersthene, and hedenbergite). The measurement results have shown that asteroids of the basic types and their meteoritic analogues have similar spectral dependences of P_{\min} . This result is important both for establishing a genetic relationship between asteroids and meteorites and for a more reliable classification of asteroids into types. While the absolute value of P_{\min} for S-type asteroids increases with wavelength, the positive polarization becomes substantially weaker. On the other hand, the spectral dependence of polarization for C-type asteroids (2100) Ra-Shalom (Kiselev et al. 1999) and (1580) Betulia (Tedesco et al. 1978) reveals no inversion: both the negative and positive branches of polarization become slightly weaker with wavelength.

Belskaya and Lupishko with colleagues carried out polarization studies of asteroids (1) Ceres, (4) Vesta, (21) Lutetia, and (2867) Steins, which are the targets of the NASA and the ESA space missions “Dawn” and “Rosetta”, respectively. Belskaya et al. (1987) identified a unique feature of Ceres: a monotonous increase of the orientation angle of the polarization plane with wavelength. Lupishko et al. (1992) hypothesized that this effect can be related to a thin gaseous (e.g., water loss?) or gas–dust shell around Ceres. However, it seems more likely that the observed polarization effect is related to the existence of H₂O hoarfrost on the Ceres’ surface near one of its poles. Simultaneous photometric and polarimetric observations of Vesta, performed by Kiselev and Morozhenko in Bolivia, revealed an inverse correlation of the degree of negative polarization and the asteroid brightness (Lupishko et al. 1988). It is recognized now that Vesta’s lightcurve is controlled primarily by the spatial distribution of the surface albedo. Therefore, the inverse polarization–brightness correlation observed for Vesta extends the validity of the well-known Umov law to the case of negative polarization. Polarimetric observations of Steins were carried out by Fornasier et al. (2006) with the 8.2-m VLT (ESO, Chile) in the phase-angle range from 10.3° to 28.3°. By using the empirical relation “polarimetric slope–albedo”, the geometric albedo in the V band was estimated to be $p_V = 0.45 \pm 0.10$.

The main objective of observations of potentially hazardous asteroids has been to estimate their polarization characteristics, albedos, and sizes. The Apollo-group object (23817) 2000 PN₉ approached the Earth to within 0.02 AU in March of 2006. Polarimetric observations by Belskaya et al. (2009) revealed the maximal polarization value $P = 7.7\%$ at phase angles in the range 90°–115°, which agrees very well with observations of two other S-type NEAs (1685) Toro and (4179) Toutatis. In December of 2001, asteroid (33342) 1998 WT₂₄ from the Aten group passed at a distance of 0.0125 AU of the Earth. The resulting brightness and polarization phase curves allowed a reliable identification of this object as a high-albedo E-type asteroid as well as the determination of its basic characteristics ($P_{\max} = 1.6\% - 1.8\%$, $\alpha_{\max} = 72^\circ - 80^\circ$, sizes 420×330 m, and albedo $p_V = 0.43$;

Kiselev et al. 2002). The albedo and size were also estimated from polarimetry of asteroid (144898) 2004 VD₁₇ (De Luise et al. 2007). The relatively small polarization $P = 2.35\%$ measured at a large phase angle of $\alpha = 76.8^\circ$ is indicative of a high-albedo E-type asteroid. These results agree very well with data available for NEA (33342) 1998 WT₂₄.

The totality of NEA observations allows one to conclude that P_{\max} in the V filter can reach values $8.5\% - 10.5\%$ and $\approx 2\%$ for S- and E-type asteroids, respectively. The angle of maximal positive polarization for the E-type asteroids is unexpectedly small: $\alpha_{\max} \approx 76^\circ$, compared to 110° for S-type asteroids. The application of the Geake and Dollfus (1986) calibration to the measured values of h and P_{\max} yielded an estimate of the average size of surface regolith particles of $\sim 25 \mu\text{m}$ (Kiselev et al. 1990, 2002).

The theoretical prediction of the POE by Mishchenko (1993) stimulated systematic observations of high-albedo asteroids intended to verify the presence of a narrow secondary minimum of polarization near the exact backscattering direction accompanying an equally narrow BOE peak. POE has been found for asteroids (64) Angelina and (44) Nysa (Rosenbush et al. 2005, 2009). These objects were intentionally selected for the search of POE since they are high-albedo bodies for which a spike-like BOE had been previously found by Harris et al. (1989).

The polarimetric method for the determination of asteroid albedos is based on empirical linear relations between the polarimetric slope h or the depth of the NPB P_{\min} on one hand and the asteroid geometric albedo p_V on the other hand: $\log p_V = A \log h + B$ and $\log p_V = C \log |P_{\min}| + D$. The values of the constants were determined using laboratory measurements of meteorite samples (Zellner and Gradie 1976; Zellner et al. 1977). New empirical relations were obtained by Lupishko and Mohamed (1996) using data on IRAS-based albedos, occultation albedos, ground-based radiometric albedos, and their own observations of many asteroids. These relations allow one to determine the values of the constants A , B , C , and D more accurately.

I. N. Belskaya and her collaborators on the international team using telescopes from the 8–10-m class have carried out polarimetric studies of Centaurs and TNOs. Nine objects were selected: dwarf planets (136199) Eris and (136108) Haumea, classical objects (20000) Varuna and (50000) Quaoar, a resonant object (38628) Huya, a scattered-disk object (26375) 1999 DE₉, and Centaurs (2060) Chiron, (5145) Pholus, and (1019) Chariklo. Two different types of behavior of the polarization phase dependence for TNOs have been found: a slowly changing negative polarization for the largest objects (Pluto, Eris, Quaoar) and a rapidly changing negative polarization (a $\sim 1\%$ /deg slope in the phase-angle range $0.1^\circ - 1^\circ$) for smaller TNOs (Ixion, Varuna, Huya, and 1999 DE₉) (Belskaya et al. 2008; Bagnulo et al. 2008; Mishchenko et al. 2010, and references therein). Centaurs exhibit a great diversity of polarization properties (Belskaya et al. 2010). All TNOs with diameters smaller than 750 km possess similar polarization properties despite the fact that they have different albedos and belong to different dynamic groups. The larger objects have a higher albedo than the smaller ones, and have the capa-

bility of retaining volatiles such as CO, N₂, and CH₄. Both facts can be linked to their different polarimetric behavior compared to smaller objects (Bagnulo et al. 2008).

Extensive multi-decadal measurements of the degrees of linear and circular polarization of comets as functions of the phase angle and wavelength in maximally wide ranges of observation geometries have been carried out at KhIA, MAO, and CrAO (Kiselev 2003; Rosenbush 2006; and references therein). Two typical linear-polarization phase curves have been identified, specifically, those of dust-rich comets with a high positive-polarization maximum and gas-rich comets with a low maximum. It has been demonstrated that the *apparent* segregation of comets into two groups—dust-rich and gas-rich—based on polarimetric data at large phase angles is an artifact caused by the low spatial and spectral resolution of the instrumentation traditionally used in cometary observations (Kiselev et al. 2004; Jockes et al. 2005). This conclusion has led to the following binary classification of comets (Kolokolova et al. 2007). Type I comets exhibit strong polarization of scattered light coming from the circumnuclear area, weak or no silicate emissions, and a concentration of compact aggregate dust particles in the vicinity of the nucleus. This class is populated by short-period comets which have experienced a relatively long exposure to intense solar insolation. Type II comets exhibit strong silicate emissions and strong polarization throughout their extended dusty atmospheres composed of fluffy aggregate particles.

The degree of linear polarization measured for comet C/1995 O1 (Hale–Bopp) at phase angles 34°–49° is ~4% higher than that for any other comet observed so far (Kiselev and Velichko 1997; Rosenbush et al. 1997b). Peculiarities in the wavelength dependence of polarization for several comets have been detected (Rosenbush et al. 2008b). These results may justify the separation of such comets into a peculiar group with inherently atypical dust properties (Rosenbush et al. 2008a,b; Kolokolova et al. 2008).

A drastic increase of linear polarization was observed during a complete disruption of the nucleus of comet D/1999 S4 (LINEAR) (Kiselev et al. 2001). The parameters of the polarization maximum of dust-rich comets ($P_{\max} \approx 25\%–28\%$, $\alpha_{\max} = 94^\circ$) were determined by Kiselev and Rosenbush (2004).

A nontrivial amount of circular polarization (up to 0.3%) has been reliably detected for several comets (Rosenbush et al. 1997b, 2007a,b). Several models of its origin have been proposed and discussed (Rosenbush et al. 2007a). It has been found that the observed circular polarization is predominantly left-handed. The predominance of left-handed polarization observed in comets and recent modeling results coupled with the spectroscopic detection of organically enriched chemical composition of comets allow one to put forth a reasonable hypothesis that the observed circular polarization implies the presence of prebiotic homochiral organics in comets (Rosenbush et al. 2007a, 2008a).

Unique opportunities to observe the occultation of a star by a cometary coma were presented by comets C/1990 K1 (Levy) and Hale–Bopp. As a result, Rosenbush et al. (1994, 1997b) could measure, for the first time, the parameters of linear

polarization of a star during its visible movement across the coma. These observations helped estimate the upper limit on the optical thickness of the Hale–Bopp dust coma along the line of sight at 0.3 ± 0.1 .

With the goal of archiving all existing results of linear- and circular-polarization observations of comets, the Database of Comet Polarimetry was created as part of the NASA Planetary Data System (Kiselev et al. 2006).

3.3. Planets

Morozhenko (1973, 1975) carried out polarimetric observations of Mars during the 1971 and 1973 oppositions. He determined the dependence of the degree of linear polarization on the longitude of the Martian central meridian for the wavelength range $0.316 - 0.798 \mu\text{m}$. The decrease of the degree of polarization and inversion angle with decreasing wavelength was caused by the increasing influence of the atmosphere. The dust storm of 1973 resulted in a decrease in the degree of polarization. A theoretical analysis of these results yielded the aerosol refractive index, average size, and optical thickness as well as the atmospheric pressure at the surface (Morozhenko and Yanovitskii 1973).

During the 2003 Martian opposition, an international team of American and Ukrainian astrophysicists carried out joint observations with the Hubble Space Telescope (Shkuratov et al. 2005). The most interesting phenomena revealed by these observations were transient high-polarization (exceeding 2%) features. The observed polarization of light scattered by these optically thin clouds indicates that they consisted of strongly polarizing particles. Theoretical modeling suggests that irregular crystals of $\sim 1 \mu\text{m}$ size could cause such strong polarization (Zubko et al. 2006). Perhaps, highly polarizing clouds can be formed at the very beginning of the nucleation of H_2O ice crystals on submicrometer dust grains. Kaydash et al. (2006) showed that the polarizing clouds were located at 30–40 km above the Martian surface.

In the 1960s and 1970s, extensive ground-based polarimetric observations of the center of the Jovian disk were carried out at MAO by Bugaenko et al. (1974). The results of measurements of both the entire Jovian disk and its central area were interpreted using two model atmospheres and an approximate theoretical technique to simulate the observed degree of linear polarization (Morozhenko and Yanovitskii 1973). The cloud-particle refractive index and mean geometrical radius were estimated to be 1.36 ± 0.01 and $0.2 \mu\text{m}$, respectively. This refractive index value is consistent with ammonia ice. Mishchenko (1990a) reanalyzed these observations using numerically exact computer solutions of the vector radiative transfer equation and obtained similar results. More recently, Dlugach and Mishchenko (2008) used the same dataset to analyze implications of the likely nonsphericity of cloud particles in the Jovian atmosphere.

Since 1981, KhIA astronomers carried out a 28-year program of polarimetric observations of Jupiter, thereby capturing two Jovian years. Seasonal variations in the North–South asymmetry of polarized light and its correlation with the Jovian

magnetic field were detected (Starodubtseva et al. 2002). An explanation of the origin of substantial linear polarization in the polar regions of Jupiter at small phase angles has been proposed in terms of a two-layer atmospheric model. According to this explanation, the light coming from invisible underlying clouds is scattered by a layer of stratospheric haze and contributes substantially to the resulting polarization (Shalygina et al. 2009).

Bugaenko and Morozhenko (1981) summarized polarimetric observations of Saturn carried at MAO in the 1970s. Based on the observed spectral behavior of the direction of the polarization plane for the equatorial region of the planet, they identified the presence of an optically thin layer composed of submicrometer-sized, preferentially oriented nonspherical particles.

3.4. Magnetic fields of the Sun and bright stars

On the basis of high-spatial-resolution observations, Severny (1966a,b) discovered a heterogeneity of the solar magnetic field and a fine structure in spots as well as in active areas between spots. It was shown that the majority of small sunspots have magnetic fields ~ 1500 Gauss. Weak longitudinal fields not exceeding 200 Gauss were found for the supergiant γ Cyg. Sporadically appearing weak fields of the order of 50 Gauss were found for α CMa, α Lyr, β Gem, α CMi, α Tau, and other stars (Severny et al. 1974). The presence of weak general magnetic fields for many stars with vigorous convection of spectral types F9–M3 and luminosity classes I–V has been detected. The variation of the general magnetic field as a function of stellar rotation has been identified for several stars of the solar type (Plachinda and Tarasova 1999; Plachinda 2003).

3.5. Stars

N. M. Shakhovskoy (Fig. 1) and Yu. S. Efimov carried out extensive multi-decadal observations of polarization for different types of variable stars, including double systems, Mira stars, flare stars, polars (stars of the AM Her type with magnetic fields reaching millions of Gauss), symbiotic stars, novae and supernovae, nebulae, and galaxies with active nuclei. Shakhovskoi (1962) discovered variable polarization for the early-type eclipsing binary β Lyr. This allowed him, for the first time, to estimate the mass of clusters in its gaseous envelope at 10^{-9} solar masses. His studies led to the refinement of models of gas shells and estimates of their masses and mass loss rates, in particular, in application to other close binary systems such as V448 Cyg and V444 Cyg (Shakhovskoi 1964).

The nature of the flare activity of UV Cet stars was studied using measurements of polarization during their outbursts. From observations of ten flashes, no change in polarization with brightness of stars was found. This helps one to refine considerably the models of red-dwarf flares and their nature, as summarized in the monograph by Gershberg (2002).

In contrast, studies of the spectral dependence of polarization for red giants and supergiants have revealed its complex nature and temporal variability (Shakhovskoy 1969; Efimov 1990a). The complex nature of the spectral dependence of polarization has been detected for T Tau. It turns out that polarization changes occur while the brightness remains almost constant. These studies formed the basis of a “magneto-hydrodynamic” model of stars of this type (Grinin et al. 1980).

Polarimetric observations of the star R CrB during deep minima of brightness performed by Yu. S. Efimov have revealed a relationship between the brightness and polarization variations. The decrease in brightness of R CrB-type stars is caused by the ejection of a cloud of gas and dust along the line of sight. The expansion of the cloud is accompanied by the growth of dust grains, thereby causing changes in both photometric and polarimetric characteristics of a star (Efimov 1980, 1990b).

Variable intrinsic polarization was revealed for a number of novae (DQ Her, a new symbiotic PU Vul, N Cyg 1975). The reason is the light scattering on inhomogeneities of the extended dust shell which were ejected during the nova outburst or formed later (Belyakina et al. 1989; Efimov et al. 1977).

A high degree of polarization was first discovered for young Ae/Be Herbig stars with circumstellar dust disks (so-called UXORs) during a cooperative program of patrol observations (Voshchinnikov et al. 1988; Grinin et al. 1991). The observations have fully confirmed the model by Grinin (1988), which implies that during a brightness minimum there should be a significant increase in polarization, a weak blue emission, and other effects caused by the scattering of light by dust grains forming the circumstellar envelope. Such observations imply the presence of gas–dust protoplanetary disks and opaque dust clouds around these stars. The properties of the dust disks as well as the opaque dust clouds were investigated. Moreover, it was found that the intrinsic polarization of stars measured in the deep minima of brightness allows one to determine the orientation of the circumstellar disks relative to the direction of the local interstellar magnetic field. It was found that the axes of the circumstellar disks are parallel to the Galactic magnetic field, which is indicative of the important role of the Galactic magnetic field in the formation of circumstellar envelopes (Grinin et al. 1988).

The measurement of linear polarization for the microquasar SS 433 exhibited changes in its proper polarization following the precession of the accretion disk and relativistic jets (Efimov et al. 1984). Variable linear polarization in the visible spectral range was found for several X-ray sources such as Her-X1, Cyg-X1, Cyg-X2, Sco-X1 (Shakhovskoy and Efimov 1976), and Vela X-1 (Beskrovnyaya et al. 1992). The occasional appearance of significant and variable circular polarization was detected for Sco X-1, Cyg X-1, and 3C 273 by Severny and Kuvshinov (1975).

Measurements of circular polarization have been particularly useful in studies of “polars”, i.e., interacting close binary systems in which the white dwarf’s magnetic field can reach tens of megaGauss. A series of photometric and polarimetric studies of polars and related objects (Shakhovskoy and Efimov 1977; Efimov and

Shakhovskoy 1982; Shakhovskoy et al. 1996; Efimov and Shakhovskoy 1996) helped create a database of observations of cataclysmic systems with different degrees of the magnetic-field impact on accretion. It has been shown that changes in brightness and polarization can be interrelated; they can also be related to changes of rate and nature of accretion onto a compact component as well as to changes in the geometric characteristics of systems and the topology of their magnetic fields. Changes in the orientation of accretion columns, both in latitude and longitude, of white dwarfs belonging to the system, over a period of a few years have been detected for the polars AM Her and QQ Vul. This finding confirmed a previously proposed “swinging dipole” model (Andronov 1987). Thus, the presence of “active” and “inactive” states of the system has been demonstrated, when the circular polarization almost vanishes.

Substantial linear polarization has been observed for selected white dwarfs of various spectral classes. The results of multicolor polarimetric observations for 14 white dwarfs and hot subdwarfs were summarized by Efimov and Shakhovskoy (1974). Linear-polarization measurements for more than 80 bright white dwarfs and hot subdwarfs were reviewed by Efimov (1981).

A new approximation of the wavelength dependence of interstellar polarization in the spectral range from 0.2 to 4 micrometers was derived on the basis of satellite and ground-based observations of stars located near the Galactic equator (Efimov 2009).

3.6. Extragalactic objects

The first polarimetric observations of several galaxies with active nuclei, such as NGC 1068, 1275, 3516, 4151, and 7469, were carried out in 1965–66 using the 1.25-m telescope of the Southern Station of the Astronomical Institute of the Moscow State University in Nauchny, Crimea. Subsequently, observations were performed using the 2.6- and 1.25-m CrAO telescopes; the program was extended to include a polarimetric study of active galaxies of the BL Lac type (so-called blazars), quasars, and supernovae. From 1994 to 2005, CrAO astrophysicists participated in the large-scale international program OJ-94 focused on the monitoring of a number of blazars. The main target of the program was blazar OJ 287 with a binary black hole in the center and a weak variability with a period of 11.6 years. Among the results of this program was the detection of rapid changes of the spectral dependence of polarization in the quasar 3C 273 and several blazars. The nature of these changes is consistent with the predictions of the model according to which a variable polarized emission is associated with shock waves propagating along the relativistic jets.

A significant result of CrAO astrophysicists was the first observational confirmation of the spiral structure of the magnetic field in the inner parts of jets as well as the estimate of the magnetic field and its topology in blazars OJ287 and PKS 0735+178 (Efimov 1999). Analyses of the observed relationship between changes in the degree of linear polarization and the spectral index of blazars

OJ287 and 3C 66A allows one to trace temporal changes in the degree of inhomogeneity of the blazar magnetic fields (Efimov et al. 2002; Efimov and Primak 2006). A detailed overview of the main objectives and results of polarimetric observations in studies of the structure and magnetism of blazars was published by Efimov (1999).

Five-color photometric and polarization observations of I- and II-type supernovae in several galaxies were carried out and analyzed by Shakhovskoy (1976), while Gnedin et al. (1992) studied the unique supernova SN 1987A in the Large Magellanic Cloud. It was shown that different stages of a supernova exhibit different mechanisms of polarization, such as the scattering of light in asymmetrically ejected shells and the effect of “light echo” in an inhomogeneous dusty medium surrounding the supernova.

4. Development of theory and modeling techniques

A number of important results have been obtained by Ukrainian scientists in the framework of various theoretical and modeling research programs focused on astrophysical polarimetry. Grinin (1988) developed a model of photometric activity and blue emission of young Herbig Ae/Be stars with circumstellar dust disks (so-called UXORs) and predicted a significant increase in polarization during brightness minima for these stars.

Shkuratov (1988, 1989) pioneered the application of the theory of CB to the interpretation of the BOE and NPB ubiquitously observed for ASSBs (see also Shkuratov et al. 1994). Mishchenko (1993) (see also Mishchenko et al. 2006b, 2010) demonstrated that CB results in POE rather than NPB. He also used the fundamental principle of electromagnetic reciprocity to derive a rigorous relationship between the diffuse and coherent reflection matrices in the exact backscattering direction and used it to explain polarimetric radar observations of high-albedo planetary surfaces (Mishchenko 1992a,b).

Mishchenko (1990b) pioneered the analytical theory of radiative transfer in anisotropic scattering media composed of preferentially oriented nonspherical particles. He also developed an extremely efficient analytical technique for the averaging of light-scattering observables over orientations of nonspherical particles in the framework of the T -matrix method (Mishchenko 1990c, 1991a,b). The resulting software package has been used worldwide in hundreds of research publications. An interesting extension of the analytical averaging approach was proposed by Petrov et al. (2008, 2009).

Extensive research efforts resulted in the development of the microphysical vector theory of radiative transfer and CB (see the summary in Mishchenko et al. 2010), including near-field corrections (Tishkovets and Litvinov 1999; Tishkovets 2002, 2007, 2008, 2009; Tishkovets and Mishchenko 2004; Litvinov et al. 2006; Petrova et al. 2007). In particular, the structure of the scattered wavelet in the close vicinity of a scatterer and the peculiarities of the scattering of this wavelet by a neighboring particle were analyzed. In the far-field zone of the particle the scat-

tered field becomes an outgoing spherical wave. However, in the immediate vicinity of the particle the field is strongly inhomogeneous, which leads to a number of significant photometric and polarization effects.

More recent research has resulted in the development of a new branch of statistical optics, wherein electromagnetic scattering by discrete random media is modeled using direct computer solvers of the Maxwell equations (Mishchenko et al. 2009a,b, 2011; Dlugach et al. 2011). Numerically exact modeling has reproduced all observable manifestations of CB, including POE, despite large values of the particle packing density (Mishchenko et al. 2009a,b; Muinonen and Zubko 2010).

Acknowledgments

N. Kiselev appreciates a grant from the NATO ASI. This study was partly supported by the NASA Radiation Sciences Program managed by Hal Maring.

References

- Akimov, L. A., V. P. Tishkovets, and Yu. G. Shkuratov, 1985: Experimental modeling of polarization properties of atmosphereless cosmic bodies at small phase angles. In A. V. Morozhenko, Ed., *Photometric and Polarimetric Investigations of Celestial Bodies* (Naukova Dumka, Kyiv), pp. 42–47 (in Russian).
- Andronov, I. L., 1987: ‘Swinging dipoles’ in magnetic close binary stars. *Astrophys. Space Sci.* **131**, 557–570.
- Arago, F. (1854–1872). *Astronomie Populaire, Vols. 1–4* (Paris, Gide et Baudry éditeurs).
- Bagnulo, S., I. Belskaya, K. Muinonen, et al., 2008: Discovery of two distinct polarimetric behaviours of trans-Neptunian objects. *Astron. Astrophys.* **491**, L33–L36.
- Barabaschoff, N., 1926: Polarimetrische beobachtungen an der mondoberfläche und an gesteinen. *Astron. Nachr.* **226**, 7–14.
- Belskaya, I. N., 2007: Surface optical properties of asteroids, Centaurs, and Kuiper Belt objects. Habilitation dissertation (V. N. Karazin Kharkiv National University, Kharkiv, Ukraine).
- Belskaya, I. N., D. F. Lupishko, and N. M. Shakhovskoy, 1987: Spectral dependence of the negative polarization of certain asteroids. *Pis'ma Astron. Zhurn.* **13**, 530–534 (in Russian).
- Belskaya, I. N., D. F. Lupishko, Yu. G. Shkuratov, et al., 1989: Spectral dependence of negative polarization of meteorites and terrestrial silicates. *Meteoritika* **48**, 116–120 (in Russian).
- Belskaya, I. N., Yu. G. Shkuratov, Yu. S. Efimov, et al., 2005: The F-type asteroids with small inversion angles of polarization. *Icarus* **178**, 213–221.
- Belskaya, I. N., A.-C. Levasseur-Regourd, Yu. G. Shkuratov, and K. Muinonen, 2008: Surface properties of Kuiper-Belt objects and Centaurs from photometry and polarimetry. In M. A. Barucci, H. Boehnhardt, D. P. Cruikshank, and A. Morbidelli, Eds., *The Solar system Beyond Neptune* (University of Arizona Press, Tucson, AZ), pp. 115–127.
- Belskaya, I. N., S. Fornasier, and Y. N. Krugly, 2009: Polarimetry and BVRI photometry of the potentially hazardous near-Earth asteroid (23187) 2000 PN9. *Icarus* **201**, 167–171.

- Belskaya, I. N., S. Bagnulo, M. A. Barucci, et al., 2010: Polarimetry of Centaurs (2060) Chiron, (5145) Pholus and (10199) Chariklo. *Icarus* **210**, 472–479.
- Belyakina, T. S., N. I. Bondar, D. Chochol, et al., 1989: The Kuwano–Honda’s peculiar object (PU Vulpeculae) in 1983–1986. *Astron. Astrophys.* **223**, 119–135.
- Berdyyugin, A. V., and N. M. Shakhovskoy, 1993: The method of the simultaneous measurements of four Stokes parameters in the UVRI bands. *Izv. Krym. Astrofiz. Obs.* **87**, 122–134 (in Russian).
- Beskrovnaya, N. G., Y. N. Gnedin, N. N. Kiselev, et al., 1992: Polarimetric study of early type southern objects with circumstellar envelopes – the X-ray binary VELA-X-1. *Sov. Astron. Lett.* **18**, 203–207.
- Bugaenko, O. I., and A. V. Morozhenko, 1981: Oriented particles in the upper layers of the Saturnian atmosphere. In A. V. Morozhenko, Ed., *Physics of Planetary Atmospheres* (Naukova Dumka, Kyiv), pp. 108–112 (in Russian).
- Bugaenko, L. A., O. I. Bugaenko, V. D. Krugov, and V. G. Parusimov, 1968: Electro-photometer of faint fluxes. *Astrometr. Astrophiz.* **1**, 193–210 (in Russian).
- Bugaenko, O. I., and A. L. Gural’chuk, 1985: Astronomical spectropolarimeter. I. Basic principles of operation. In A. V. Morozhenko, Ed., *Photometric and Polarimetric Investigations of Celestial Bodies* (Naukova Dumka, Kyiv), pp. 160–164 (in Russian).
- Bugaenko, O. I., A. V. Morozhenko, and E. G. Yanovitskii, 1974: Polarization investigations of the planets carried out at the Main Astronomical Observatory of the Ukrainian Academy of Sciences. In T. Gehrels, Ed., *Planets, Stars, and Nebulae: Studied with Photopolarimetry* (University of Arizona Press, Tucson, AZ), pp. 599–606.
- Cellino, A., I. N. Belskaya, Ph. Bendjoya, et al., 2006: The strange polarimetric behavior of asteroid (234) Barbara. *Icarus* **180**, 565–567.
- Churyumov, K. I., F. R. Rspaev, and D. I. Gorodetskij, 1986: Electrophotometry and polarimetry of comet Halley. *Komet. Tsyrk.* **356**, 2–3 (in Russian).
- De Luise, F., D. Perna, E. Dotto, et al., 2007: Physical investigation of the potentially hazardous asteroid (144898) 2004 VD₁₇. *Icarus* **191**, 628–635.
- Degtjarev, V. S., and L. O. Kolokolova, 1992: Possible application of circular polarization for remote sensing of cosmic bodies. *Earth Moon Planets* **57**, 213–223.
- Degtjarev, V. S., L. O. Kolokolova, and A. V. Morozhenko, 1990: Rotation of polarization plane near inversion point for light scattered on rough surfaces. *Astron. Tsyrk.* **1545**, 35–36 (in Russian).
- Dlugach, J. M., and M. I. Mishchenko, 2008: Photopolarimetry of planetary atmospheres: what observational data are essential for a unique retrieval of aerosol microphysics? *Mon. Not. R. Astron. Soc.* **384**, 64–70.
- Dlugach, J. M., M. I. Mishchenko, L. Liu, and D. W. Mackowski, 2011: Numerically-exact computer simulations of light scattering by densely packed, random particulate media. *J. Quant. Spectrosc. Radiat. Transfer* **112**, in press.
- Dombrovskij, V. A., 1949: On the polarization of stars of early spectral classes. *Proc. Natl Acad. Sci. Arm. SSR* **10**, 199–203.
- Efimov, Yu. S., 1970: On the polarimetric measurements of UV Ceti-type star flash radiation. *Izv. Krym. Astrofiz. Obs.* **41–42**, 357–366 (in Russian).
- Efimov, Yu. S., 1980: Polarization of radiation from R CrB. *Izv. Krym. Astrofiz. Obs.* **61**, 110–119 (in Russian).
- Efimov, Yu. S., 1981: Polarimetric observations of white dwarf stars. *Izv. Krym. Astrofiz. Obs.* **61**, 118–139 (in Russian).

- Efimov, Yu. S., 1990a: Polarimetry of Mira-type stars. *Veröff. Sternw. Sonneberg* **10**, 428–435.
- Efimov, Yu. S., 1990b: R-Coronae in the brightness minimum of 1988/1989. *Sov. Astron.* **34**, 247–254.
- Efimov, Yu., 1999: Blazars optical polarimetry. In C. M. Raiteri, M. Villata, and L. O. Takalo, Eds., *Blazar Monitoring towards the Third Millennium* (Observatorio Astronomico di Torino, Pino Torinese, Italy), pp. 55–56.
- Efimov, Yu. S., 2009: Interstellar polarization: new approximation. *Izv. Krym. Astrofiz. Obs.* **105**, 82–114 (in Russian).
- Efimov, Yu. S., and N. M. Shakhovskoy, 1974: Observations of linear polarization of white dwarf radiation. *Izv. Krym. Astrofiz. Obs.* **50**, 68–83 (in Russian).
- Efimov, Yu. S., and N. M. Shakhovskoy, 1982: Polarization and photometric observations of the polar AM Her. *Izv. Krym. Astrofiz. Obs.* **65**, 143–155 (in Russian).
- Efimov, Yu. S. and N. M. Shakhovskoy, 1996: Photopolarimetry of OJ 287 at the Crimean Astrophysical Observatory in 1994–1995. In L. O. Takalo, Ed., *Workshop on Two Years of Intensive Monitoring of OJ 287 and 3C66A* (University of Turku, Piikkiö, Finland), p. 32.
- Efimov, Yu. S., and N. V. Primak, 2006: Photopolarimetry of 3C 66A in 2003. *Astrophysics* **49**, 36–53.
- Efimov, Yu. S., N. V. Narizhnaia, and N. M. Shakhovskoi, 1977: Studies of Nova Cygni 1975 at the Crimean Astrophysical Observatory. V. Linear polarization. *Sov. Astron.* **21**, 278–287.
- Efimov, Yu. S., V. Piirola, and N. M. Shakhovskoy, 1984: The linear polarization of SS433=V1343 Aql. *Astron. Astrophys.* **138**, 62–70.
- Efimov, Yu. S., N. M. Shakhovskoy, L. O. Takalo, and A. Sillanpää, 2002: Photopolarimetric monitoring of OJ 287 in 1994–1997. *Astron. Astrophys.* **381**, 408–419.
- Fornasier, S., I. Belskaya, M. Fulchignoni, et al., 2006: First albedo determination of 2867 Steins, target of the Rosetta mission. *Astron. Astrophys.* **449**, L9–L12.
- Geake, J. E., and A. Dollfus, 1986: Planetary surface texture and albedo from parameter plots of optical polarization data. *Mon. Not. R. Astron. Soc.* **218**, 75–91.
- Gershberg, R. E., 2003: Activity of solar-type main sequence stars (Astroprint, Odessa, Ukraine) (in Russian).
- Gnedin, Yu. N., K. L. Maslennikov, A. E. Rozenbush, et al., 1992: Polarimetric investigation of southern early spectral type objects with circumstellar envelopes – the supernova SN 1987A in 1988–1989. *Sov. Astron. Lett.* **18**, 163–215.
- Grinin, V. P., 1988: The origin of the blue emission observed in the deep minima of young irregular variable stars. *Pis'ma Astron. Zhurn.* **14**, 65–69 (in Russian).
- Grinin, V. P., Yu. S. Efimov, V. I. Krasnobabtsevet, et al., 1980: Results of the joint program observations of the T Tauri-type star DI Cephei. *Perem. Zvezdy, Byull.* **21**, 247–271 (in Russian).
- Grinin, V. P., N. N. Kiselev, N. Kh. Minikulov, and G. P. Chernova, 1988: Observations of linear polarization in the deep minima of WW Vul. *Pis'ma Astron Zhurn.* **14**, 514–525 (in Russian).
- Grinin, V. P., N. N. Kiselev, G. P. Chernova, et al., 1991: The investigations of 'zodiacal light' of isolated AE-Herbig stars with nonperiodic algol-type minima. *Astrophys. Space Sci.* **186**, 283–298.
- Hall, J. S., 1949: Observations of the polarized light from stars. *Science* **109**, 166–167.
- Hansen, J. E., and L. D. Travis, 1974: Light scattering in planetary atmospheres. *Space Sci. Rev.* **16**, 527–610.

- Harris, A. W., J. W. Young, L. Contreiras, et al., 1989: Phase relations of high albedo asteroids: the unusual opposition brightening of 44 Nysa and 64 Angelina. *Icarus* **81**, 365–374.
- Hilltner, W. A., 1949: Polarization of light from distant stars by interstellar medium. *Science* **109**, 165.
- Ivanov, Yu. S., A. V. Morozhenko, A. P. Vidmachenko, and M. G. Sosonkin, 2002: Ultra-violet spectropolarimeter for planetary monitoring on board of the International Space Station. *Kosm. Nauka Tekhnol.* **8**, 45–50 (in Russian).
- Ivanov, Yu. S., I. I. Sinyavsky, and M. G. Sosonkin, 2004: On-board UV-spectropolarimeter for monitoring of planetary atmospheres. *Kosm. Nauka Tekhnol.* **10**, 32–35 (in Russian).
- Jockers, K., N. Kiselev, T. Bonev, et al., 2005: CCD imaging and aperture polarimetry of comet 2P/Encke: are there two polarimetric classes of comets? *Astron. Astrophys.* **441**, 773–782.
- Kaydash, V. G., M. A. Kreslavsky, Yu. G. Shkuratov, et al., 2006: Measurements of winds on Mars with Hubble Space Telescope images in 2003 opposition. *Icarus* **185**, 97–101.
- Kiselev, N. N., 2003: Light scattering by dust particles of comets, asteroids, and circumstellar shells: observations and interpretation. Habilitation dissertation (V. N. Karazin Kharkiv National University, Kharkiv, Ukraine).
- Kiselev, N. N., and F. P. Velichko, 1997: Aperture polarimetry and photometry of comet Hale–Bopp. *Earth Moon Planets* **78**, 347–352.
- Kiselev, N., and V. Rosenbush, 2004: Polarimetry of comets: progress and problems. In G. Videen, Ya. Yatskiv, and M. Mishchenko, Eds., *Photopolarimetry in Remote Sensing* (Kluwer Academic Publishers, Dordrecht, The Netherlands), pp. 411–429.
- Kiselev, N. N., D. F. Lupishko, G. P. Chernova, and Yu. G. Shkuratov, 1990: Polarimetry of the asteroid 1685 Toro. *Kinemat. Fiz. Nebesn. Tel* **6**, 77–82 (in Russian).
- Kiselev, N. N., V. K. Rosenbush, and K. Jockers, 1999: Polarimetry of asteroid 2100 Ra-Shalon at large phase angle. *Icarus* **140**, 464–466.
- Kiselev, N. N., F. P. Velichko, and S. F. Velichko, 2001: Polarimetry of comet C/1999 S4 (LINEAR) before and after breake-up. In K. I. Churyumov, Ed., *Fourth Vsekhsvyatsky Readings. Modern Problems of Physics and Dynamics of the Solar System* (Kyiv University, Kyiv), pp. 127–131.
- Kiselev, N. N., V. K. Rosenbush, K. Jockers, et al., 2002: Polarimetry of near-Earth asteroid 33342 (1998 WT24). Synthetic phase angle dependence of polarization for the E-type asteroids. In B. Warmbein, Ed., *Asteroids, Comets, Meteors (ACM 2002)*. ESA SP–500, pp. 887–890.
- Kiselev, N. N., K. Jockers, and T. Bonev, 2004: CCD imaging polarimetry of comet 2P/Encke. *Icarus* **168**, 385–391.
- Kiselev, N., S. Velichko, K. Jockers, V. Rosenbush, and S. Kikuchi, Eds., 2006: Database of Comet Polarimetry. EAR-C-COMPIL-5-COMET-POLARIMETRY-V1.0. NASA Planetary Data System.
- Kiselev, N., V., Rosenbush, F. Velichko, and S. Zaitsev, 2009: Polarimetry of the Galilean satellites and Jupiter near opposition. *J. Quant. Spectrosc. Radiat. Transfer* **110**, 1713–1718.
- Kolokolova, L., H. Kimura, N. Kiselev, and V. Rosenbush, 2007: Two different evolutionary types of comets proved by polarimetric and infrared properties of their dust. *Astron. Astrophys.* **463**, 1189–1196.

- Kokolova, L., N. Kiselev, V. Rosenbush, and H. Kimura, 2008: Spectral dependence of polarization in comets: recent observations and interpretation. *LPI Contribution No. 1405*, paper id. 8240.
- Korokhin, V. V., L. A. Akimov, G. P. Marchenko, et al., 1993: Observation of Jupiter, Moon and Sun with linear CCD detector. *Astron. Vestnik* **27**, 56–64 (in Russian).
- Korokhin, V. V., S. A. Beletzkij, and Yu. I. Velikodzkij, 2000: Experience of using CCD matrix in the Astronomical Observatory of Kharkiv National University. *Kinemat. Fiz. Nebesn. Tel.* **16**, 80–86 (in Russian).
- Kucherov, V. A., 1985: Multicomponent symmetric achromatic phase plates. I. Extension of the Pancharatnam system to the case of an arbitrary number of components. In A. V. Morozhenko, Ed., *Photometric and Polarimetric Investigations of Celestial Bodies* (Naukova Dumka, Kyiv), pp. 152–159 (in Russian).
- Kucherov, V. A., 1986: Multicomponent symmetric achromatic phase plates. III. Theoretical modeling of the optical properties of a 10-element quartz phase-shifting system. *Kinemat. Fiz. Nebesn. Tel.* **2**, 59–66 (in Russian).
- Kucherov, V. A., Yu. S. Ivanov, Yu. S. Efimov, et al., 1997: Ultraviolet low-resolution spectropolarimeter for the space mission SPECTRUM-UV (UVSPEPOL project). *Kosm. Nauka Tekhnol.* **3**, 3–27 (in Russian).
- Litvinov, P., V. Tishkovets, and K. Ziegler, 2006: Coherent backscattering effects for discrete random media. *J. Quant. Spectrosc. Radiat. Transfer* **103**, 131–145.
- Lupishko, D. F., 1997: Photometry and polarimetry of asteroids: results of observations and data analysis. Habilitation dissertation (V. N. Karazin Kharkiv National University, Kharkiv, Ukraine).
- Lupishko, D. F., and R. A. Mohamed, 1996: A new calibration of the polarimetric albedo scale of asteroids. *Icarus* **119**, 209–213.
- Lupishko, D. F., and S. V. Vasil'ev, Eds., 2008: Asteroid Polarimetric Database V6.0. EAR-A-3-RDR-APD-POLARIMETRY-V6.0. NASA Planetary Data System.
- Lupishko, D. F., I. N. Bel'skaia, O. I. Kvaratskheliia, et al., 1988: The polarimetry of Vesta during the 1986 opposition. *Astron. Vestnik* **22**, 142–146 (in Russian).
- Lupishko, D. F., Yu. S. Efimov, and N. M. Shakhovskoj, 1992: Ceres' peculiar polarization. *Bull. Am. Astron. Soc.* **24**, 874.
- Lyot, B., 1929: Recherches sur la polarisation de la lumière des planetes et de quelques substances terrestres. *Ann. Observ. Paris, Sect. Meudon*. Vol. 8, No. 1. English translation: Research on the polarization of light from planets and from some terrestrial substances (National Aeronautics and Space Administration, Washington, DC), NASA Tech. Transl. NASA TT F-187, 1964.
- Maxwell, J. C., 1891: *A Treatise on Electricity and Magnetism* (Clarendon Press, Oxford). (Also: Dover Publications, New York, 1954).
- Mishchenko, M. I., 1990a: Physical properties of the upper tropospheric aerosols in the Equatorial Region of Jupiter. *Icarus* **84**, 296–304.
- Mishchenko, M. I., 1990b: Multiple scattering of polarized light in anisotropic plane-parallel media. *Transp. Theory Statist. Phys.* **19**, 293–316.
- Mishchenko, M. I., 1990c: Extinction of light by randomly-oriented non-spherical grains. *Astrophys. Space Sci.* **164**, 1–13.
- Mishchenko, M. I., 1991a: Extinction and polarization of transmitted light by partially aligned nonspherical grains. *Astrophys. J.* **367**, 561–574.
- Mishchenko, M. I., 1991b: Light scattering by randomly oriented axially symmetric particles. *J. Opt. Soc. Am. A* **8**, 871–882.

- Mishchenko, M. I., 1992a: Enhanced backscattering of polarized light from discrete random media: calculations in exactly the backscattering direction. *J. Opt. Soc. Am. A* **9**, 978–982.
- Mishchenko, M. I., 1992b: Polarization characteristics of the coherent backscatter opposition effect. *Earth Moon Planets* **58**, 127–144.
- Mishchenko, M. I., 1993: On the nature of the polarization opposition effect exhibited by Saturn's rings. *Astrophys. J.* **411**, 351–361.
- Mishchenko, M. I., L. D. Travis, and A. A. Lacis, 2002: *Scattering, Absorption, and Emission of Light by Small Particles* (Cambridge University Press, Cambridge, UK).
- Mishchenko, M. I., V. K. Rosenbush, and N. N. Kiselev, 2006a: Weak localization of electromagnetic waves and opposition phenomena exhibited by high-albedo atmosphereless solar system objects. *Appl. Opt.* **45**, 4459–4463.
- Mishchenko, M. I., L. D. Travis, and A. A. Lacis, 2006b: *Multiple Scattering of Light by Particles: Radiative Transfer and Coherent Backscattering* (Cambridge University Press, Cambridge, UK).
- Mishchenko, M. I., J. M. Dlugach, and L. Liu, 2009a: Azimuthal asymmetry of the coherent backscattering cone: theoretical results. *Phys. Rev. A* **80**, 053824.
- Mishchenko, M. I., J. M. Dlugach, L. Liu, et al., 2009b: Direct solutions of the Maxwell equations explain opposition phenomena observed for high-albedo solar system objects. *Astrophys. J.* **705**, L118–L122.
- Mishchenko, M. I., V. K. Rosenbush, N. N. Kiselev, et al., 2010: *Polarimetric Remote Sensing of Solar System Objects* (Akademperiodyka, Kyiv) (arXiv:1010.1171).
- Mishchenko, M. I., V. P. Tishkovets, L. D. Travis, et al., 2011: Electromagnetic scattering by a morphologically complex object: fundamental concepts and common misconceptions. *J. Quant. Spectrosc. Radiat. Transfer* **112**, 671–692.
- Morozhenko, A. V., 1973: Longitude effect in the polarization of Mars. *Sov. Astron.* **17**, 668–671 (in Russian).
- Morozhenko, A. V., 1975: Results of polarimetric observations of Mars in 1971 and 1973. *Astrometr. Astrofiz.* **26**, 97–107 (in Russian).
- Morozhenko, A. V., and E. G. Yanovitskii, 1973: The optical properties of Venus and the Jovian planets. I. The atmosphere of Jupiter according to polarimetric observations. *Icarus* **18**, 583–592.
- Muinonen, K., and E. Zubko, 2010: Coherent backscattering by a finite medium of particles. In K. Muinonen, A. Penttilä, H. Lindqvist, et al., Eds., *Electromagnetic and Light Scattering XII* (University of Helsinki, Helsinki), pp. 194–197.
- Nevodovsky, E. P., and A. V. Morozhenko, 2009: Studies of stratospheric ozone layer from near-Earth orbit utilizing ultraviolet polarimeter. *Acta Astron.* **64**, 54–58.
- Nevodovsky, E. P., A. P. Vid'machenko, A. V. Morozhenko, and I. I. Syniavskiy, 2004: Ukrainian youth sputnik: research of aerosol layer of atmosphere of the Earth by means of a UV-polarimeter. *Kosm. Nauka Tekhnol.* **10**, 27–32 (in Russian).
- Nikulin, N. S., A. B. Severny, and V. E. Stepanov, 1958: The solar magnetograph of the Crimean Astrophysical Observatory. *Izv. Krym. Astrofiz. Obs.* **19**, 3–19 (in Russian).
- Opanasenko, N. V., and Yu. G. Shkuratov, 1994: Results of simultaneous polarimetric and photometric observations of the Moon. *Astron. Vestnik* **28**, 133–154 (in Russian).
- Opanasenko, N. V., A. A. Dolukhanjan, Yu. G. Shkuratov, et al., 1994: First image of the Moon in parameter P_{\min} . *Astron. Vestnik* **28**, 27–36 (in Russian).
- Petrov, D., Yu. Shkuratov, and G. Videen, 2008: Analytic light-scattering solution of two merging spheres using *Sh*-matrices. *Opt. Commun.* **281**, 2411–2423.

- Petrov, D., Yu. Shkuratov, and G. Videen, 2009: The *Sh*-matrices method applied to light scattering by small lenses. *J. Quant. Spectrosc. Radiat. Transfer* **110**, 1448–1459.
- Petrova, E., V. Tishkovets, and K. Jockers, 2007: Modeling of opposition effects with ensembles of clusters: interplay of various scattering mechanisms. *Icarus* **188**, 233–245.
- Pirola, V., 1988: Simultaneous five-colour (UBVRI) photopolarimeter. In G. V. Coyne et al., Eds., *Polarized Radiation of Circumstellar Origin* (Vatican Observatory/University Arizona Press, Vatican City/Tucson), pp. 735–746.
- Plachinda, S. I., 1997: Achromatic Stokesmeter for precision measurements of stellar magnetic fields using CCD-detector. In *Abstracts of the Joint European and National Astronomical Meeting, JENAM-97*; 6th European and 3rd Hellenic Astronomical Conference, Thessaloniki, Greece, p. 240.
- Plachinda, S. I., 2003: General magnetic field on convective stars. In Yu. Glagolevskij, D. Kudryavtsev, I. Romanyuk, Eds., *Magnetic Stars* (Special Astrophysical Observatory, Nizhnij Arkhyz, Russia), pp. 241–246.
- Plachinda, S. I., and T. N. Tarasova, 1999: Precise spectropolarimetric measurements of magnetic fields of some solar-like stars. *Astrophys. J.* **514**, 402–410.
- Rosenbush, V. K., 2002: The phase-angle and longitude dependence of polarization for Callisto. *Icarus* **159**, 145–155.
- Rosenbush, V. K., 2006: Light-scattering properties of small Solar System bodies. Habilitation dissertation (Main Astronomical Observatory, Kyiv).
- Rosenbush, V. K., and N. N. Kiselev, 2005: Polarization opposition effect for the Galilean satellites of Jupiter. *Icarus* **179**, 490–496.
- Rosenbush, V. K., A. E. Rosenbush, and M. S. Dement'ev, 1994: Comets Okazaki–Levy–Rudenko (1989 XIX) and Levy (1990 XX): polarimetry and stellar occultations. *Icarus* **108**, 81–91.
- Rosenbush, V. K., V. V. Avramchuk, A. E. Rosenbush, and M. I. Mishchenko, 1997a: Polarization properties of the Galilean satellites of Jupiter: observations and preliminary analysis. *Astrophys. J.* **487**, 402–414.
- Rosenbush, V. K., N. M. Shakhovskoj, and A. E. Rosenbush, 1997b: Polarimetry of comet Hale–Bopp: linear and circular polarization, stellar occultation. *Earth Moon Planets* **78**, 381–386.
- Rosenbush, V. K., N. N. Kiselev, V. G. Shevchenko, et al., 2005: Polarization and brightness opposition effects for the E-type asteroid 64 Angelina. *Icarus* **178**, 222–234.
- Rosenbush, V., L. Kolokolova, A. Lazarian, et al., 2007a: Circular polarization in comets: observations of comet C/1999 S4 (LINEAR) and tentative interpretation. *Icarus* **186**, 317–330.
- Rosenbush, V., N. Kiselev, N. Shakhovskoy, et al., 2007b: Circular and linear polarization of comet C/2001 Q4 (NEAT). Why circular polarization in comets is predominantly left-handed? In G. Videen, M. Mishchenko, M. P. Mengüç, and N. Zakharova, Eds., *Peer-Reviewed Abstracts of the Tenth Conference on Electromagnetic & Light Scattering*, pp. 181–184.
- Rosenbush, V., N. Kiselev, and L. Kolokolova, 2008a: Predominantly left-handed circular polarization in comets: does it indicate L-enantiometric excess in cometary organics? In S. Kwok and S. Sandford, Eds., *Proc. IAU Symposium 251 on Organic Matter in Space* (Hong Kong), pp. 311–312.
- Rosenbush, V., N. Kiselev, L. Kolokolova, et al., 2008b: Atypical polarization in some recent comets. In *Proceedings of the 11th Electromagnetic and Light Scattering Conference* (University of Hertfordshire, UK), pp. 193–196.

- Rosenbush, V., V. Shevchenko, N. Kiselev, et al., 2009: Polarization and brightness opposition effects for E-type asteroid 44 Nysa. *Icarus* **201**, 655–665.
- Severnyi, A. B., 1966a: Magnetic fields of the Sun and stars. *Sov. Phys. Usp.* **9**, 1–30.
- Severnyi, A. B., 1966b: The nature of solar magnetic fields (the fine structure of the field). *Sov. Astron.* **9**, 171–182.
- Severnyi, A. B., 1970: The weak magnetic fields of some bright stars. *Astrophys. J.* **159**, L73–L76.
- Severnyi, A. B., and V. M. Kuvshinov, 1975: Circular polarization of the X-ray sources Scorpius X-1, Cygnus X-1, and 3C 273. *Astrophys. J.* **200**, L13–L17.
- Severnyi, A. B., V. M. Kuvshinov, and N. S. Nikulin, 1974: Photoelectric studies of weak stellar magnetic fields. *Izv. Krym. Astrofiz. Obs.* **50**, 3–51 (in Russian).
- Shain, G. A., 1955: Diffuse nebulae and interstellar magnetic field. *Astron. Zhurn.* **32**, 110–117 (in Russian).
- Shain, G. A., S. B. Pikel'ner, and R. N. Ikhanov, 1955: The measurement of the polarization of the Crab nebula. *Astron. Zhurn.* **32**, 395–400 (in Russian).
- Shakhovskoi, N. M., 1962: Polarization observations of β Lyrae. *Sov. Astron.* **6**, 587–589.
- Shakhovskoi, N. M., 1964: Polarization in variable stars. II. Eclipsing binaries. *Sov. Astron.* **8**, 833–842.
- Shakhovskoy, N. M., 1969: On the intrinsic polarization of red giants and supergiants. *Izv. Krym. Astrofiz. Obs.* **39**, 11–33 (in Russian).
- Shakhovskoy, N. M., 1971: Methods of investigation of the polarization of variable stars' radiation. In B. V. Nikonov, Ed., *Stationary Stars and Methods of their Investigation. Methods of Investigation of Variable Stars* (Nauka, Moscow), pp. 199–224 (in Russian).
- Shakhovskoy, N. M., 1976: Observations of linear polarization of optical radiation from a Type I supernova in galaxy NGC 7723. *Pis'ma Astron. Zhurn.* **2**, 275–279 (in Russian).
- Shakhovskoy, N. M., and N. A. Dimov, 1962: Integrating stellar electropolarimeter. *Izv. Krym. Astrofiz. Obs.* **27**, 291–308 (in Russian).
- Shakhovskoy, N. M., and Yu. S. Efimov, 1972: Polarization observations of unstable stars and extragalactic objects. I: Equipment, method of observation and reduction. *Izv. Krym. Astrofiz. Obs.* **45**, 90–110 (in Russian).
- Shakhovskoy, N. M., and Yu. S. Efimov, 1976: Observations of linear polarization of optical emission from X-ray sources. *Izv. Krym. Astrofiz. Obs.* **54**, 99–119 (in Russian).
- Shakhovskoy, N. M., and Yu. S. Efimov, 1977: Study of linear polarization of optical emission from BL Lac objects. *Izv. Krym. Astrofiz. Obs.* **56**, 39–51 (in Russian).
- Shakhovskoy, N. M., I. L. Andronov, and S. V. Kolesnikov, 1996: Different types of photometric and polarimetric behavior of AM HER in 1989–1994. In A. Evans, and J. H. Wood, Eds., *Cataclysmic Variables and Related Objects* (Kluwer Academic Publishers, Dordrecht, The Netherlands), p. 209.
- Shakhovskoy, N. M., I. L. Andronov, S. V. Kolesnikov, and A. V. Khalevin, 2001: Methods of measurements of circular polarization by using single-channel photopolarimeter. *Izv. Krym. Astrofiz. Obs.* **97**, 91–100 (in Russian).
- Shalygina, O., V. Korokhin, L. Starukhina, et al., 2009: Properties of polar aerosol haze in Jupiter's stratosphere. *EPSC Abstracts*, Vol. 4, EPSC2009-337-1, 337–338.
- Shkuratov, Yu. G., 1988: A diffraction mechanism for the formation of the opposition effect of the brightness of surfaces having a complex structure. *Kinem. Fiz. Nebes. Tel* **4**, 33–39 (in Russian).

- Shkuratov, Yu. G., 1989: A new mechanism for the negative polarization of light scattered by the solid surfaces of cosmic bodies. *Astron. Vestnik* **23**, 176–180 (in Russian).
- Shkuratov, Yu. G., K. Muinonen, E. Bowell, et al., 1994: A critical review of theoretical models of negatively polarized light scattered by atmosphereless solar system bodies. *Earth Moon Planets* **65**, 201–246.
- Shkuratov, Yu. G., and A. A. Ovcharenko, 2002a: Polarization light scattered by surfaces with complicated microstructure at small phase angles 0.1–3.5°. *Solar Syst. Res.* **36**, 62–67.
- Shkuratov, Yu. G., and A. A. Ovcharenko, 2002b: Experimental modeling of opposition effect and negative polarization of regolith-like surfaces. In G. Videen and M. Kocifaj, Eds., *Optics of Cosmic Dust* (Kluwer Academic Publishers, Dordrecht, The Netherlands), pp. 225–238.
- Shkuratov, Yu. G., S. P. Redkin, N. V. Bitanova, and N. B. Il'inskii, 1980: The relationship of albedo and polarization properties of the Moon. 1. A new optical parameter (preliminary study). *Astron. Tsyrk.* **1112**, 3–6 (in Russian).
- Shkuratov, Yu., M. Kreslavsky, V. Kaydash, et al., 2005: Hubble Space Telescope imaging polarimetry of Mars during the 2003 opposition. *Icarus* **176**, 1–11.
- Shkuratov, Yu., V. A. Psarev, and A. A. Ovcharenko, 2008a: Laboratory modeling of scattered light by regolith surfaces. In Yu. Shkuratov, Ed., *200 Years of Astronomy in Kharkiv University* (V. N. Karazin Kharkiv National University, Kharkiv), pp. 282–287 (in Russian).
- Shkuratov, Yu. G., N. V. Opanasenko, and V. G. Kaydash, 2008b: Investigation of Moon. In Yu. Shkuratov, Ed., *200 Years of Astronomy in Kharkiv University* (V. N. Karazin Kharkiv National University, Kharkiv), pp. 282–287 (in Russian).
- Shkuratov, Yu. G., A. A. Ovcharenko, V. A. Psarev, and S. Y. Bondarenko, 2008c: Laboratory measurements of reflected light intensity and polarization for selected particulate surfaces. *Light Scattering Rev.* **3**, 383–402.
- Starodubtseva, O. M., L. A. Akimov, and V. V. Korokhin, 2002: Seasonal variations in the north-south asymmetry of polarized light of Jupiter. *Icarus* **157**, 419–425.
- Stenflo, J. O., D. Dravins, N. Wihlborg, et al., 1980: Search for spectral line polarization in the solar vacuum ultraviolet. *Solar Phys.* **66**, 13–20.
- Stokes, G. G., 1852: On the composition and resolution of streams of polarized light from different sources. *Trans. Cambridge Philos. Soc.* **9**, 399–416.
- Tedesco, E., J. Drummond, M. Gandy, et al., 1978: 1580 Betulia: an unusual asteroid with an extraordinary lightcurve. *Icarus* **35**, 344–359.
- Tishkovets, V. P., 2002: Multiple scattering of light by a layer of discrete random medium: backscattering. *J. Quant. Spectrosc. Radiat. Transfer* **72**, 123–137.
- Tishkovets, V. P., 2007: Incoherent and coherent backscattering of light by a layer of densely packed random medium. *J. Quant. Spectrosc. Radiat. Transfer* **108**, 454–463.
- Tishkovets, V. P., 2008: Light scattering by closely packed clusters: shielding of particles by each other in the near field. *J. Quant. Spectrosc. Radiat. Transfer* **109**, 2665–2672.
- Tishkovets, V. P., 2009: Multiple scattering of electromagnetic waves by discrete random media. Habilitation dissertation (V. N. Karazin Kharkiv National University, Kharkiv, Ukraine).
- Tishkovets, V. P., and P. V. Litvinov, 1999: Opposition effects in light scattering by regolith-type media. *Solar Syst. Res.* **33**, 162–167.
- Tishkovets, V. P., and M. I. Mishchenko, 2004: Coherent backscattering of light by a layer of discrete random medium. *J. Quant. Spectrosc. Radiat. Transfer* **86**, 161–180.

- Vid'machenko, A. P., A. V. Morozhenko, and E. G. Yanovitskii, 2004: Department of the Solar System Bodies. In Ya. S. Yatskiv, Ed., *The 60th Anniversary of the Main Astronomical Observatory of the NAS of Ukraine* (Academpriodyka, Kyiv), pp. 224–281 (in Ukrainian).
- Vid'machenko, A. P., Yu. S. Ivanov, A. V. Morozhenko, et al., 2007: Spectropolarimeter for ground-based support of the space experiment “Planetary Monitoring”. *Kosm. Nauka Tekhnol.* **13**, 63–70 (in Russian).
- Voshchinnikov, N. V., V. P. Grinin, N. N. Kiselev, and N. K. Minikulov, 1988: Dust around young stars – observations of the polarization of Ux-Orionis in deep minima. *Astrophysics* **28**, 182–193.
- Weiss, W., V. P. Malanushenko, and N. M. Shakhovskoy, 1990: A method for measuring stellar magnetic fields according to hydrogen lines using a double-beam polarimeter with an acousto-optic modulator. *Izv. Krym. Astrofiz. Obs.* **82**, 69–86 (in Russian).
- Yefimov, Yu., 2004: Invitation to spectropolarimetry. In G. Videen, Ya. Yatskiv, and M. Mishchenko, Eds., *Photopolarimetry in Remote Sensing* (Kluwer Academic Publishers, Dordrecht, The Netherlands), pp. 455–478.
- Zellner, B., and J. Gradie, 1976: Minor planets and related objects. XX. Polarimetric evidence for the albedos and compositions of 94 asteroids. *Astron. J.* **81**, 262–280.
- Zellner, B., M. Leake, T. Lebertre, et al., 1977: The asteroid albedo scale. I – Laboratory polarimetry of meteorites. In *Lunar Science Conference* (Pergamon Press, New York), pp. 1091–1110.
- Zubko, E., Yu. Shkuratov, N. Kiselev, and G. Videen, 2006: DDA simulations of light scattering by small irregular particles with various structures. *J. Quant. Spectrosc. Radiat. Transfer* **101**, 416–434.



At Kyiv Pechersk Lavra.



From left to right: Nadiya Kostohryz, Jim Hough, Anny-Chantal Levasseur-Regourd, and Nikolai Kiselev in *Koncha-Zaspa*.



Directors of the NATO ASI Yaroslav Yatskiv (right) and Michael Mishchenko open the field tour of the Main Astronomical Observatory.

Photopolarimetric remote sensing of aggregates in cosmic dust

Ludmilla Kolokolova*

University of Maryland, College Park, MD 20742, USA

Abstract. Current theories of the origin and evolution of cosmic dust agree that in many cases dust particles are clusters (aggregates) of small grains. This was confirmed by observations of a variety of types of cosmic dust including protostellar, circumstellar, interplanetary and cometary dust, and many types of planetary aerosols. One of the most important remote-sensing methods to reveal the aggregated nature of cosmic dust and study properties of aggregates has been polarimetry. This paper reviews recent achievements in laboratory and theoretical modeling of light scattering by aggregates and considers how polarimetric observations allow one to identify aggregated dust particles in space by studying their angular and wavelength dependence of polarization. It is also shown how polarimetry helps to find out the structure and porosity of aggregates, and the size and composition of their constituent particles. This can be done successfully only when angular and spectral polarimetric data are considered together, and when they are combined with photometric and thermal infrared data. Consistency of the model with the ideas of origin and evolution of a given type of cosmic dust and with the results of space missions should also be checked to prove the reliability of the results.

Keywords: polarimetry, photometry, dust, aggregate, modeling

1. Introduction

For a long time, cosmic dust particles were considered as compact solid particles similar to sand or clay grains that can be rather realistically simulated by homogeneous spherical particles, i.e., using the Mie theory. Probably Fred Whipple [53] was the first who noticed that interplanetary dust is dominated by fragile porous particles; he concluded this from studying meteor light curves and meteor radar studies. Soon after that interplanetary dust particles (IDPs) were collected in the upper atmosphere and many of them appeared to be aggregates of small grains [17]. This was later confirmed by numerous studies of IDPs by Brownlee et al. (e.g., [2]). [Figure 1](#) shows a typical IDP that is clearly an aggregate of submicron particles.

Thus, the idea of the aggregated nature of cosmic dust particles came from di-

* Corresponding author. E-mail: ludmilla@astro.umd.edu

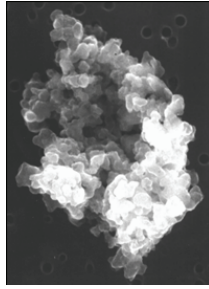


Fig. 1. 11- μm interplanetary dust particle. From the collection of the Astromaterials Acquisition and Curation Office of the Johnson Space Center, NASA.

rect studies rather than from remote sensing. Only 15 years later, Giese [9,10], using laboratory light-scattering simulations, showed that the angular dependence of brightness and polarization of the light scattered by interplanetary dust does not resemble any curve produced by spherical particles but can be reproduced by irregular fluffy particles. Later Giese [11] showed that the same is true for cometary dust. In 1980 Greenberg [13] described a possible formation of interplanetary and cometary particles via agglomeration of submicron interstellar grains during their low-speed collisions that made the aggregate model of cosmic dust even more popular. A real triumph of the aggregate model of cometary dust was direct evidence of the aggregate nature of the majority of particles returned by the Stardust mission after its rendezvous with comet 81P/Wild 2 [7,22,46]. Although the Stardust mission confirmed the results of in situ study of comet 1P/Halley [8,20] that cometary dust also contains some solid irregular silicate particles, the important role of aggregates in cosmic dust is now out of the question.

Based on modern, powerful and complex instrumentation and sophisticated light-scattering modeling, remote sensing of cosmic dust is now focusing on studying detailed properties of cosmic aggregates, such as their size, structure, and composition so that through these characteristics the story of their origin and evolution can be discovered.

In this chapter we consider the existing observational data, focusing on photopolarimetry of different types of cosmic dust: interplanetary dust, cometary dust, dust on the surface of asteroids, and dust in debris disks. We describe laboratory and computer simulations of the dust properties as well as remote-sensing capabilities of those simulations.

This chapter mainly considers recent results that were published after the reviews [28,29] of observations and simulations of photopolarimetric properties of cosmic dust.

2. Observations

Brightness is considered to be the most powerful observational characteristic of cosmic objects. However, in the case of ensembles of dust particles, the bright-

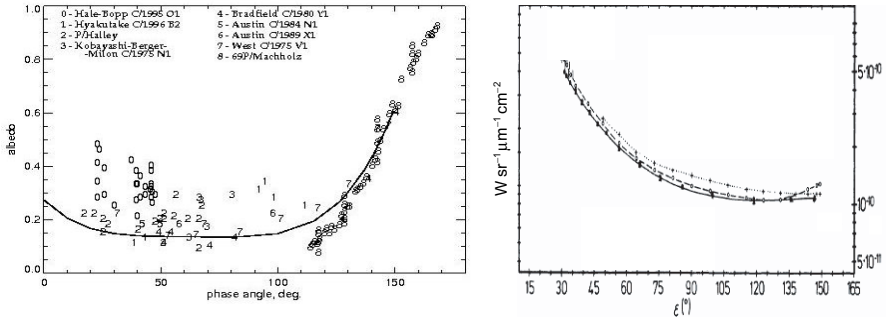


Fig. 2. Photometric phase curves for comets (left, adapted from [28]) and interplanetary dust (right, adapted from [37]). For interplanetary dust that is a highly extended object, the general phase curve cannot be determined and is usually substituted by the dependence of the brightness on elongation ϵ (that represents $180^\circ - \alpha$) and is slightly different for different heliocentric distances and latitudes β .

ness is affected by the concentration (number density) of the dust particles as well as their physical properties (e.g., absorption). This makes brightness an ambiguous characteristic if we would like to learn physical properties of the dust particles. Astronomical characteristics that are ratios of brightness quantities, e.g., color or polarization, allow us to eliminate effects of number density and are determined only by physical characteristics of the dust particles, such as their size, structure, and composition. These will be the main focus of this paper.

Henceforth, talking about properties of aggregates we will use the following terminology: a single constituent particle of the aggregate will be called a monomer; fluffiness of the aggregate will be defined through the packing factor p , i.e. the ratio of the volume of the monomers to the total volume of the aggregate, or its porosity, i.e., $1 - p$. Finally, in discussing the composition we will base our judgments on the complex refractive index of the particle material $m = n + i\kappa$, where $i = (-1)^{1/2}$. To characterize the size of a monomer or aggregate we will use a size parameter $x = 2\pi r/\lambda$, where r is the radius of the particle and λ is the wavelength. Dependence of the observational characteristics on the geometry of observations will be described by the phase angle α , the angle sun–dust–Earth (the scattering angle is equal to $180^\circ - \alpha$). The term “phase curve” will mean the dependence of some observational characteristic on the phase angle.

2.1. Photometry

As any ensemble of rather small particles, cosmic dust has a photometric phase curve characterized by an increase of brightness in the backward and especially the forward scattering directions (small and large phase angles), as shown in Fig. 2 for cometary and interplanetary dust. The same shape of the photometric phase curve is typical for debris disks [12]. The observed photometric curves are always

smooth, which is good evidence that cosmic dust is a polydisperse and, probably, polyshaped ensemble of particles. Some conclusions about the properties of particles can be drawn from the amplitude and width of the backward and forward scattering peaks. A significant asymmetry of the photometric phase curve immediately tells us that the cosmic dust particles are larger than Rayleigh particles, i.e., characterized by $x > 1$. A great advantage of using the forward scattering peak is that its characteristics do not depend on the refractive index and shape of a particle and can be used to characterize the particle size [1, chapter 13.5]. However, this is not straightforward in the case of aggregates as, depending on the porosity of the aggregate, the characteristics of the forward scattering peak can be dominated by the properties of the overall aggregate (compact aggregate) or properties of its monomers (very porous aggregate) [14]. The situation when the backscattering is stronger than the forward scattering is characteristic for small metal particles or very large particles, [18, sections 10.3 and 10.6]. It is clear from Fig. 2 that both these cases are not realistic for any type of cosmic dust. Thus, the photometric curve can give us some main ideas about the particles; however, in the case of aggregates it is not sensitive to details of their structure and composition.

As we mentioned above, a more informative characteristic is spectral dependence of brightness, usually defined as color, i.e., the logarithm of the ratio of intensity in longer (usually in R filter, i.e., around 690 nm) and shorter (usually in B filter, i.e., around 440 nm) wavelengths. The color is positive if the light is scattered more efficiently at longer wavelengths, and it is called “red”. In the other case, it is negative and is called “blue”. As a ratio of intensities, color is not sensitive to the number of dust particles but depends on the properties of the individual dust particles. In the case of Rayleigh particles, the color is blue independently of the composition of particles, since particles of larger size parameter (shorter wavelength) scatter light more efficiently than particles of smaller size parameter (longer wavelength). For larger particles, the color results from the interplay of spectral changes in the refractive index, particularly the change of the absorption with wavelength, and the change in scattering cross section mainly defined by the particle size parameter. In the case when the composition and size of particles are both unknown, it is hard to say what defines the color. Considering such a complex dependence of color on the particle composition and size, it is rather surprising that the majority of observations of cosmic dust show red color that indicates some similarity of their compositional and structural properties.

2.2. Linear polarization

As was mentioned above, to study properties of complex particles, including aggregates, it is beneficial to consider a ratio of two intensities to eliminate the influence of the number density of particles in the ensemble. One such ratio is linear polarization. In planetary physics it is usually defined as $P = (I_{\perp} - I_{\parallel}) / (I_{\perp} + I_{\parallel})$, i.e., the ratio of the difference of the intensity in the plane perpendicular to the scattering plane and intensity in the plane parallel to it, to the sum of these inten-

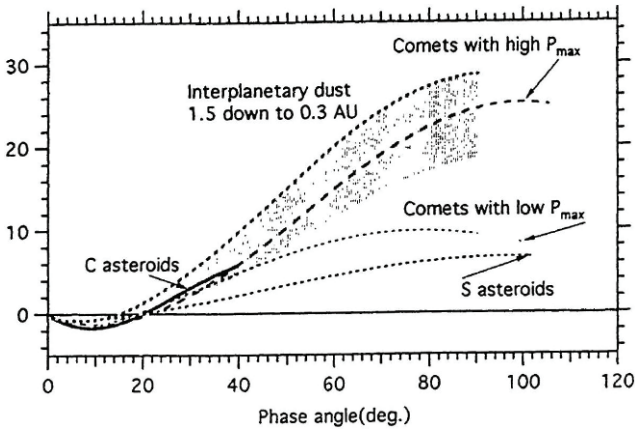


Fig. 3. Polarimetric phase curves for interplanetary dust, comets, and asteroids (adapted from [39]). Note the existence of two classes of comets with high and low polarization.

sities. In terms of Stokes parameters it corresponds to the ratio Q/I . Such a definition results from the fact that the vast majority of observations of dust in the solar system demonstrate the Stokes parameter U close to zero. Amazingly, different kinds of cosmic dust have polarimetric phase curves of a very similar shape (Fig. 3). More so, similar polarimetric curves are typical for the dust-covered surfaces of asteroids (Fig. 3) and debris disks [12]. They are characterized by a small bowl-like branch of negative polarization (i.e., when $I_{\parallel} > I_{\perp}$) with the minimum around 10° and the inversion angle (at which the polarization changes sign) about $20^{\circ} - 25^{\circ}$; after that the polarization becomes positive and has a bell-like shape with the maximum around $90^{\circ} - 110^{\circ}$. The polarimetric phase curves differ only in the value of polarization minimum and maximum, their position, and have a slight difference in the position of the inversion angle. Despite the fact that those differences are small, they indicate a difference in the composition and structure of the dust particles. To reveal these differences and to find out the properties of the dust particles, one needs to find an analogical behavior of the particles modeled using computer or laboratory simulations.

In the same way that color in photometry provides us with information about the properties of the scattering particles, spectral dependence of polarization is a good supplement to its angular dependence. The spectral gradient of polarization, also called “polarimetric color”, is defined as a difference of polarization in longer and shorter wavelengths. It is usually considered in the case of positive polarization mainly because data for negative polarization are scarce. As for photometric color, polarimetric color is called “blue” (or negative) if polarization is larger at shorter wavelengths and “red” (positive) if it is larger at longer wavelengths. An interesting fact is that despite the similarity of polarimetric phase curves for comets, asteroids, and interplanetary dust, their polarimetric color is different: it is

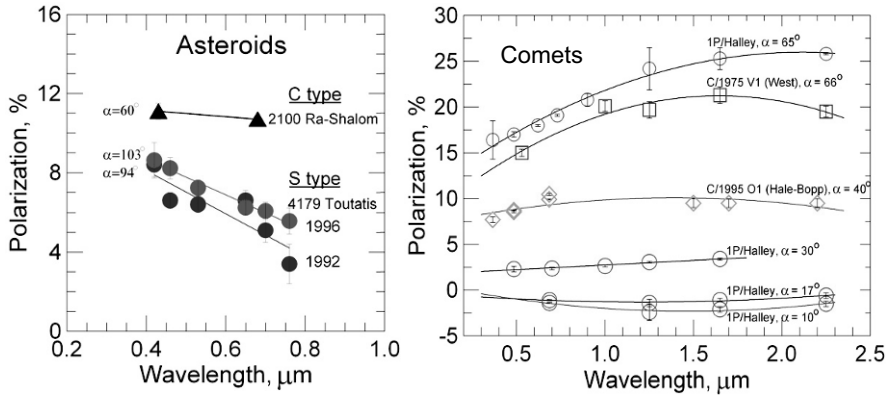


Fig. 4. Typical change of polarization with wavelength for asteroids (left) and comets (right) (courtesy V. Rosenbush and N. Kiselev). Notice the data for comet Halley at small phase angles where polarization is negative; its absolute value changes the same way as the positive polarization changes: it increases in the visible and decreases in the near infrared.

blue for asteroids (Fig. 4) and interplanetary dust [38]. Typical cometary polarimetric color in the visible is red [28, 29], although recently it has been found that some comets demonstrate blue polarimetric color [27]. Moreover, the positive polarimetric color observed in the visible changes to the negative in observations at longer, near infrared wavelengths as shown in Fig. 4. In the case of debris disks only data for Beta Pictoris have been published [49]; they show different signs of polarimetric colors in different parts of the disk. It is clear that the sign of the polarimetric color contains important information about properties of the dust particles. However, without modeling we cannot say how to extract this information from the observational data.

3. Modeling of light scattering by aggregates

As was mentioned in Section 1, this paper is focused on the results obtained after the reviews [28,29]. Interpretation of observational data for cosmic dust has been based on the experimental and theoretical modeling of the light scattering by aggregates.

The experimental study of light scattering by aggregates largely has been carried out by two groups: a Dutch–Spanish group [51] and a French group (e.g., [15,16]). The former group created a database (<http://www.iaa.es/scattering>) of the angular and spectral characteristics of the scattering matrix for a variety of terrestrial samples. Their samples were natural particles: powdered rocks, aerosols, hydrosols, and, among them, cosmic dust analogs, i.e., aggregates. The aggregates were produced using the Condensation Flow Apparatus at the NASA Goddard Space Flight Center, simulating the conditions of aggregate growth in the circumstellar environment. For these particles, the group published the elements F_{11} , F_{12} ,

and F_{22} of the scattering matrix measured at $\alpha = 6^\circ - 175^\circ$ and $\lambda = 633$ nm [52]. The shape of the measured brightness and polarization curves was found to be similar to that observed for cosmic dust (shown in Fig. 3), although the maximum polarization was too high and minimum polarization was too low. This most likely resulted from the small size of the monomers (mainly around 50 nm) and aggregates themselves that were not larger than 650 nm.

The French group accomplished a systematic study of light-scattering properties of manmade aggregates of submicron monomers [15,16]. The aggregates had controlled size, shape and composition of the monomers, although the size and structure of the aggregates was not controlled. The group published photometric and polarimetric phase curves for $\alpha = 5^\circ - 170^\circ$ using red (633 nm) and green (544 nm) filters. The results represent a large database of the light-scattering characteristics of aggregates of a variety of sizes and composition. The advantage of this study is that they measured not only angular but also spectral photopolarimetric characteristics of the aggregates. Even though this study provided interesting findings (e.g., they showed that cometary data fits best to the mixture of silicates and carbon and that red polarimetric color is typical for aggregates whose material contains carbon thus indirectly confirming presence of carbon in comets), lack of control of the measured samples and their not very realistic composition (silica SiO_2 , magnesiosilica MgSiO , ferrosilica FeSiO , carbon C, and their mixtures) limit application of these results to the remote sensing of cosmic dust.

Fundamental results in interpretation of the observations of cosmic dust have been achieved based on theoretical simulations. This became possible due to increasing computer power that improved the performance of earlier developed computer codes and allowed modeling of aggregates whose characteristics approach characteristics of the large inhomogeneous aggregates realistic for cosmic dust. The most popular computer codes used for aggregate modeling are the superposition T -matrix code for clusters of spheres by Mackowski and Mishchenko [41] (<ftp://ftp.eng.auburn.edu/pub/dmckwski/scatcodes/index.htm> and http://www.giss.nasa.gov/staff/mmishchenko/t_matrix.html) and various versions of the discrete dipole approximation (DDA) mainly originated from the DDA code by Draine and Flatau [6] that regularly gets upgraded to improve its efficiency and to cover more research tasks (<http://www.astro.princeton.edu/~draine/DDSCAT.html>).

Besides the fact that opportunities for computer simulations dramatically improved since 2004, the main progress in solving the inverse problem for cosmic dust was achieved due to a new, integrated approach to the modeling of cosmic dust. Earlier studies, reviewed in [28,29], were mainly concentrated on obtaining a reasonable fit to some light-scattering characteristics, most often the polarization phase curve, ignoring the other data. However, it is well known that the polarization phase curve can be fit even with polydisperse spherical particles of rather reasonable refractive index $m \sim 1.4 + i0.03$ [44]. Thus, fitting just one or two light scattering characteristics without checking the consistency with other observational facts not only does not allow finding a unique solution for the inversion

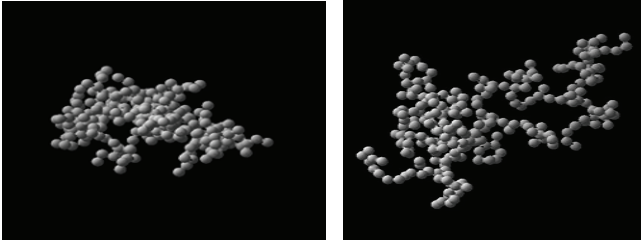


Fig. 5. Samples of aggregates used in [24,25]. Left: more compact ballistic particle–cluster aggregate (BPCA). Right: porous ballistic cluster–cluster aggregate (BCCA). The aggregates were built using the ballistic approach by Meakin [43].

problem but can even result in wrong conclusions.

Unfortunately, the approach based on limited photopolarimetric data continues to be popular in modeling light scattering by cosmic dust, with the only difference that the majority of modelers base their simulations on an aggregated model of dust particles. The ease of obtaining negative polarization using transparent (silicate) monomers can confuse researchers who may become satisfied with the results that fit the polarimetric data without checking if their photometric results also fit the observational data. For example, a great theoretical survey of light scattering properties of aggregates by Shen et al. [47] concludes that cometary dust particles are aggregates of silicate monomers. This conclusion is based on a good fit to the shape of the brightness and polarization phase curves and ignores the facts that these aggregates show a high albedo and wrong (blue) color. Even more limited are results by Das et al. [4,5] who modeled only the negative branch of polarization and found a good fit to the data for comets Halley and Hale–Bopp. The best fit was achieved at the imaginary part of the refractive index about 0.07–0.08. This characterizes the comet material as transparent and would definitely result in high albedo and blue color as well as probably wrong values of the maximum polarization and polarimetric color which have not been checked in [4,5].

The first paper that was based on the integrated approach to the observational data and found the model of the comet dust qualitatively consistent with the photometric and polarimetric angular and spectral data was that of Kimura et al. [24]. The composition of the dust considered in this paper was also consistent with the composition of comet dust obtained *in situ* for comet 1P/Halley. This approach was extended in [25] where a theoretical survey of the light-scattering properties of ballistic aggregates (Fig. 5) was accomplished. It was shown that the best fit to the cometary data is provided by large aggregates of monomers of radius $\sim 0.1 \mu\text{m}$ made of dark (mixture of silicate, carbon, and organics) material.

Combining the results by Kimura et al. [25] with the modeling of the thermal emission by aggregate particles, Kolokolova et al. [31] could explain the existence of two classes of comets, characterized by low and high maximum polarization (see Fig. 3). As was shown in [31], based on [40,48], these two polarimetric

classes have their counterparts in the thermal infrared: high-polarization comets usually show a strong 10-micron silicate feature; whereas, low-polarization comets show weak or no silicate feature. The modeling in [31] showed that a strong silicate feature is typical for highly porous aggregates; whereas, a weak silicate feature results from thermal emission of large compact aggregates. This is consistent with the fact discussed in [21,26] that comets have low polarization when their dust is mainly concentrated near the nucleus (large, slow particles); whereas, high polarization comets have evenly distributed dust throughout the coma (light porous particles easily dragged by gas). This result is also consistent with the dynamical classes of comets: compact particles appeared to be typical for old comets with processed materials; whereas, highly porous particles were found typical for “fresh” dust in new comets. One more result of [31] is that cometary aggregates, especially compact ones, should be really large aggregates, containing thousands of monomers as small compact aggregates still demonstrate a significant silicate feature. Notice that large aggregates are required for all types of porosity, since for small aggregates, negative polarization cannot reach the observed values. This is confirmed by the results of Zubko et al. [54] who considered a different type of aggregated particles, so called agglomerated debris (built of rather large irregular constituents), and showed that in the case of low-albedo (absorptive) materials negative polarization appears if the particles are of size parameter larger than 20.

To make the model of comet dust consistent with the results of the Stardust and Giotto–Vega space missions, Kolokolova and Kimura [33] modeled a mixture of solid silicate particles and aggregates made of organics and Halley-like material (mixture of organics, carbon, silicates, etc., see [24]). Solid particles were simulated as a polydispersed (power-law) polyshaped ensemble of spheroids. Aggregates were those that provided the best qualitative fit to the cometary observational data in [25]. This allowed them to achieve a quantitative fit to the polarimetric curve keeping all photometric and spectral characteristics correct (Fig. 6). It also provided the mass ratio of silicates to organics equal to unity, corresponding to the value obtained based on the elemental abundance of comet Halley dust [19]. The best-fit result appears to be also in accordance with the abundance of rock, organic, and mixed particles found in *in situ* measurements of the dust of comet 1P/Halley. It was achieved when the ratio of rock particles to the organic ones and to the Halley-like aggregates was equal to 0.30/0.44/0.26 that was consistent with the results by Fomenkova [8] who showed that the rock particles represented 1/3 of the Halley dust. The model also agrees with the fact that the rock particles in the Halley dust were compact; whereas, other particles were fluffy [20]. Using a mixture of aggregates and solid particles, Lasue et al. [35,36] also achieved a good fit to observational data for cometary [35] and interplanetary [36] dust. However, those papers provided only a fit to the polarization curve without checking consistency with photometric, cosmogonic, and *in situ* characteristics of the dust. Interestingly, a similar result, that interpretation of polarimetric observations requires a mixture of aggregates and solid (although Rayleigh ones) particles, was achieved for aerosols in Titan’s atmosphere [50]. Thus, it is

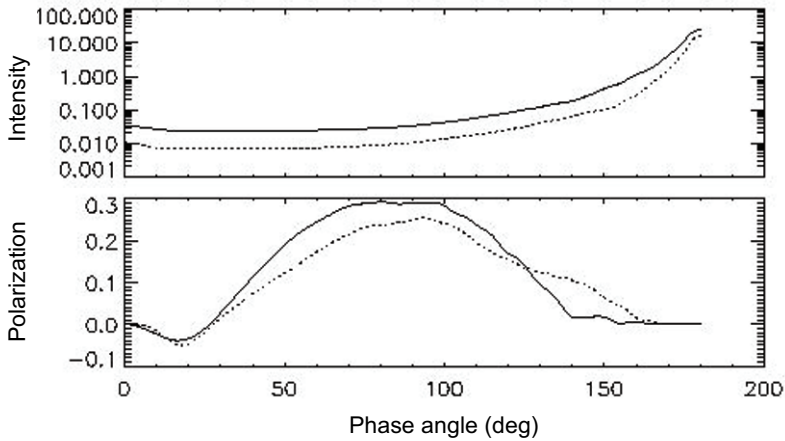


Fig. 6. Photometric and polarimetric phase curves obtained by modeling cometary dust as a mixture of aggregates and solid silicate particles [33]. The solid curves are for the 0.648 μm , the dotted curves are for 0.444 μm .

very likely that the mixture of solid and aggregate particles is the best model for dust in a variety of cosmic environments.

With the goal to explain the observations of spectral dependence of polarization, Kolokolova and Kimura [32] studied the strength of electromagnetic interaction between the monomers in aggregates. They showed that electromagnetic interaction depends on how many monomers are covered by a single wavelength. That is consistent with earlier findings by Chen et al. [3], Kimura and Mann [23], Kolokolova et al. [30] and recent studies by Shen et al. [47]. In the case when more monomers are covered (stronger electromagnetic interaction) the light gets more depolarized (similar to the effect of multiple scattering). As described above, comets usually show red polarimetric color (positive spectral gradient) in the visible that is typical for aggregates of submicron particles made of low-albedo material. However, their spectral gradient changes to negative in the near infrared. Besides, it appeared that some comets have blue polarimetric color even in the visible similar to interplanetary dust or asteroids. Since longer wavelengths cover more monomers, it is not surprising that the polarization decreases as the observations move to the near-infrared. Electromagnetic interaction can also explain why in some cases aggregate particles show blue polarimetric color in the visible. This happens when aggregates are so compact that a slightly longer wavelength already covers more monomers than a shorter one (Fig. 7), and the depolarizing effect of the stronger interaction dominates the polarizing effect of decreasing the size parameter of the monomer. This appears to be consistent with the fact that blue polarimetric color is usually observed for old comets, whose dust is highly processed and for asteroids whose regolith contains more compact aggregates than comets [45]. Thus, studying polarimetric color over a broad range of wavelengths, espe-

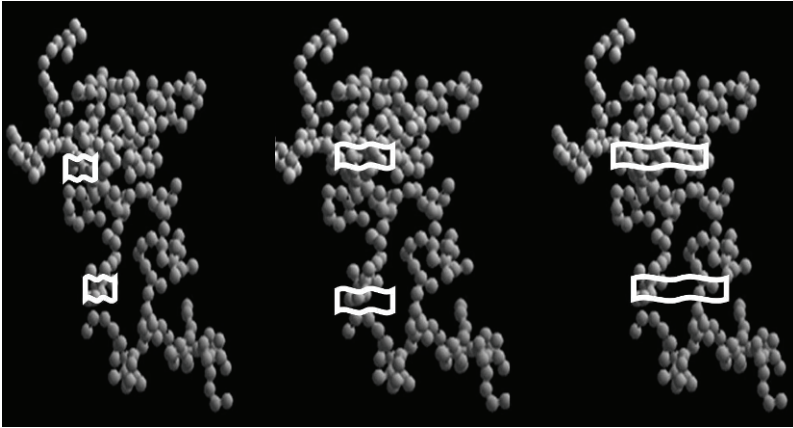


Fig. 7. Illustration of the effect of increasing wavelength for an aggregate. In a compact aggregate (top part of the aggregate) the longer the wavelength the more monomers it covers, the stronger is the interaction between the particles, and the more depolarized becomes the light. This results in a decrease of polarization with wavelength, i.e. blue polarimetric color. For a porous aggregate (low part of the aggregate), the number of monomers covered by a single wavelength does not change much as the wavelength increases, i.e. the change in the interaction between the monomers cannot overpower the change in the monomer size parameter; the polarimetric color stays red. However, as the wavelength reaches some critical value, the number of covered monomers changes significantly and interaction becomes the main factor that defines the polarimetric color which then becomes blue.

cially focusing on those wavelengths where the polarimetric color changes its sign, is a promising way to determine quantitatively the porosity of cosmic dust aggregates.

4. Conclusions

Theoretical means to study light scattering by aggregates are now as efficient as experimental ones and can be used for interpretation of remote-sensing data on cosmic dust. Moreover, theoretical modeling is more suitable for studying regularities in light-scattering by dust particles as it uses highly controlled models of particles. However, the results of the modeling cannot be recognized as realistic if they describe only one light scattering characteristic, e.g. angular dependence of polarization. To get reliable results and approach the uniqueness of the interpretation of the observational data, one needs to consider polarization together with other observational data. It is a must to consider a combination of photometric and polarimetric data – their angular and spectral characteristics as well as angular change in their colors. These should include albedo and, for completeness and to assure the uniqueness of the model, circular polarization (see [34,42]). A very useful constraint on the properties of particles can be obtained from a combination of

polarimetric and thermal emission data such as shape and strength of the silicate features and temperature of the dust. A realistic model should also be consistent with (i) cosmic abundances; (ii) the object evolution and dynamics; (iii) space mission data.

Combining the power of modern computer codes with such an integrated approach is a promising way to achieve uniqueness in interpretation of the observational data for cosmic dust despite the fact that it consists of such complex particles as aggregates.

Acknowledgments

This work would not be possible without the financial support from the Japanese Society for the Promotion of Science, (JSPS), the NASA Astrobiology Program, and the NATO ASI. The author is also grateful to T. Yamamoto, H. Kimura, V. Rosenbush, and N. Kiselev for collaborations in studying comet dust.

References

1. Bohren, C., and D. Huffman, 1983: *Absorption and Scattering of Light by Small Particles* (Wiley, New York).
2. Brownlee, D. E., D. A. Tomandl, and E. Olszewski, 1977: Interplanetary dust – a new source of extraterrestrial material for laboratory studies. In *8th Lunar Science Conference* (Pergamon, New York), Vol. 1, pp. 149–160.
3. Chen, Zh., P. Sheng, D. A. Weitz, et al., 1988: Optical properties of aggregate clusters. *Phys. Rev. B* **37**, 5232–5235.
4. Das, H. S., S. R. Das, T. Paul, et al., 2007: Aggregate model of cometary dust: an application to comet Levy 1990XX. *Mon. Not. R. Astron. Soc.* **389**, 787–791.
5. Das, H. S., S. R. Das, and A. K. Sen, 2008: Aggregate dust model to describe polarization properties of Comet Hale–Bopp. *Mon. Not. R. Astron. Soc.* **390**, 1195–1199.
6. Draine, B. T., and P. J. Flatau, 1994: Discrete-dipole approximation for scattering calculations. *J. Opt. Soc. Amer. A* **11**, 1491–1499.
7. Flynn, G. J., 2008: Physical, chemical, and mineralogical properties of Comet 81P/Wild 2: particles collected by Stardust. *Earth Moon Planets* **102**, 447–459.
8. Fomenkova, M., 1999: On the organic refractory component of cometary dust. *Space Sci. Rev.* **90**, 109–114.
9. Giese, R. H., 1977: Interpretation of the optical properties of interplanetary dust. *Z. Geophys.* **42**, 705–716.
10. Giese, R. H., K. Weiss, R. H. Zerull, and T. Ono, 1978: Large fluffy particles – a possible explanation of the optical properties of interplanetary dust. *Astron. Astrophys.* **65**, 265–272.
11. Giese, R. H., 1980: Optical investigation of dust in the solar system. In *Solid Particles in the Solar System* (Reidel, Dordrecht), pp. 1–13.
12. Graham, J. R., P. G. Kalas, and B. C. Matthews, 2007: The signature of primordial grain growth in the polarized light of the AU Microscopii Debris Disk. *Astrophys. J.* **654**, 595–605.
13. Greenberg, J. M., 1980: From interstellar dust to comets to the zodiacal light. In *Solid Particles in the Solar System* (Reidel, Dordrecht), pp. 343–349.

14. Gustafson, B. Å. S., L. Kolokolova, J. Thomas-Osip, et al., 1999: Scattering by complex systems II: results from microwave measurements. In J. M. Greenberg and A. Li, Eds., *Formation and Evolution of Solids in Space* (Kluwer, Dordrecht), pp. 549–564.
15. Hadamcik, E., J. B. Renard, A. C. Levasseur-Regourd, and J. Lasue, 2006: Light scattering by fluffy particles with the PROGRA² experiment: mixtures of materials. *J. Quant. Spectrosc. Radiat. Transfer* **100**, 143–156.
16. Hadamcik, E., J.-B. Renard, F. J. M. Rietmeijer, et al., 2007: Light scattering by fluffy Mg Fe SiO and C mixtures as cometary analogs (PROGRA² experiment). *Icarus* **190**, 660–671.
17. Hemenway, C. L., and R. K. Soberman, 1962: Small meteoric particles in the Earth's neighborhood: studies of micrometeorites obtained from a recoverable sounding rocket. *Astron. J.* **67**, 256–266.
18. van de Hulst, H. C., 1981: *Light Scattering by Small Particles* (Dover, New York).
19. Jessberger, E. K., A. Christoforidis, and J. Kissel, 1988: Aspects of the major element composition of Halley's dust, *Nature* **332**, 691–695.
20. Jessberger, E. K., and J. Kissel, 1991: Chemical properties of cometary dust and a note on carbon isotopes. In R. L. Newburn, M. Neugebauer, and J. H. Rahe, Eds., *Comets in the Post-Halley Era* (Kluwer, Dordrecht), Vol. 2, pp. 1075–1092.
21. Jockers, K., N. Kiselev, T. Bonev, et al. 2005: CCD imaging and aperture polarimetry of comet 2P/Encke: are there two polarimetric classes of comets? *Astron. Astrophys.* **441**, 773–782.
22. Kearsley, A. T., J. Borg, G. A. Graham, et al., 2008: Dust from comet Wild 2: interpreting particle size, shape, structure, and composition from impact features on the Stardust aluminum foils. *Meteor. Planet. Sci.* **43**, 41–73.
23. Kimura, H., and I. Mann, 2004: Light scattering by large clusters of dipoles as an analog for cometary dust aggregates. *J. Quant. Spectrosc. Radiat. Transfer* **89**, 155–164.
24. Kimura, H., L. Kolokolova, and I. Mann, 2003: Optical properties of cometary dust: constraints from numerical studies on light scattering by aggregate particles. *Astron. Astrophys.* **407**, L5–L8.
25. Kimura, H., L. Kolokolova, and I. Mann, 2006: Light scattering by cometary dust numerically simulated with aggregate particles consisting of identical spheres. *Astron. Astrophys.* **449**, 1243–1254.
26. Kiselev, N., and V. Rosenbush, 2004: Polarimetry of comets: progress and problems. In G. Videen, Ya. Yatskiv, and M. Mishchenko, Eds., *Photopolarimetry in Remote Sensing* (Kluwer, Dordrecht), pp. 411–430.
27. Kiselev, N., V. Rosenbush, L. Kolokolova, and K. Antonyuk, 2008: The anomalous spectral dependence of polarization in comets. *J. Quant. Spectrosc. Radiat. Transfer* **109**, 1384–1391.
28. Kolokolova, L., H. Kimura, and I. Mann, 2004: Characterization of dust particles using photopolarimetric data: Example of cometary dust. In G. Videen, Ya. Yatskiv, and M. Mishchenko, Eds., *Photopolarimetry in Remote Sensing* (Kluwer, Dordrecht), pp. 431–454.
29. Kolokolova, L., M. Hanner, A.-Ch. Levasseur-Regourd, and B. Å. S. Gustafson, 2004: Physical properties of cometary dust from light scattering emission. In M. C. Festou, H. U. Keller, and H. A. Weaver, Eds., *Comets II* (University of Arizona Press, Tucson, AZ), pp. 577–614.
30. Kolokolova, L., H. Kimura, K. Ziegler, and I. Mann, 2006: Light-scattering properties of random-oriented aggregates: do they represent the properties of an ensemble of aggregates? *J. Quant. Spectrosc. Radiat. Transfer* **100**, 199–206.

31. Kolokolova, L., H. Kimura, N. Kiselev, and V. Rosenbush, 2007: Two different evolutionary types of comets proved by polarimetric and infrared properties of their dust. *Astron. Astrophys.* **463**, 1189–1196.
32. Kolokolova, L., and H. Kimura, 2010: Effects of electromagnetic interaction in the polarization of light scattered by cometary and other types of cosmic dust. *Astron. Astrophys.* **513**, A40.
33. Kolokolova, L., and H. Kimura, 2010: Comet dust as a mixture of aggregates and solid particles: model consistent with ground-based and space-mission results. *Earth Planets Space* **62**, 17–21.
34. Kolokolova, L., W. Sparks, and D. Mackowski, 2011: Astrobiological remote sensing with circular polarization. In M. I. Mishchenko, Ya. S. Yatskiv, V. K. Rosenbush, and G. Videen, Eds., *Polarimetric Detection, Characterization, and Remote Sensing* (Springer, Berlin), pp. 277–294.
35. Lasue, J., and A. C. Levasseur-Regourd, 2006: Porous irregular aggregates of sub-micron sized grains to reproduce cometary dust light scattering observations. *J. Quant. Spectrosc. Radiat. Transfer* **100**, 220–236.
36. Lasue, J., A. C. Levasseur-Regourd, N. Fray, and H. Cottin, 2007: Inferring the interplanetary dust properties from remote observations and simulations. *Astron. Astrophys.* **473**, 641–649.
37. Leinert, C., I. Richter, E. Pitz, and B. Planck, 1981: The zodiacal light from 1.0 to 0.3 A.U. as observed by the HELIOS space probes. *Astron. Astrophys.* **103**, 177–188.
38. Leinert, Ch., S. Bowyer, L. K. Haikala, et al., 1998: The 1997 reference of diffuse night sky brightness. *Astron. Astrophys. Suppl.* **127**, 1–99.
39. Levasseur-Regourd, A. Ch., 1996: Optical and thermal properties of zodiacal dust. In B. Å. S. Gustafson and M. Hanner, Eds., *Physics, Chemistry, and Dynamics of Interplanetary Dust* (Astronomical Society of Pacific, San Francisco), pp. 301–309.
40. Lisse, C. M., M. F. A’Hearn, Y. R. Fernández, and S. B. Peschke, 2002: A search for trends in cometary dust emission. In S. F. Green, I. P. Williams, J. A. M. McDonnell, and N. McBride, Eds., *Dust in the Solar System and Other Planetary Systems* (Pergamon, Oxford), pp.259–269.
41. Mackowski, D. W., and M. I. Mishchenko, 1996: Calculation of the T matrix and the scattering matrix for ensembles of spheres. *J. Opt. Soc. Am. A* **13**, 2266–2278.
42. Mackowski, D., L. Kolokolova, and W. Sparks, 2011: T -matrix approach to calculating circular polarization of aggregates made of optically active (chiral) materials and its applications to cometary dust observations. *J. Quant. Spectrosc. Radiat. Transfer* **112**, in press.
43. Meakin, P., 1984: Effects of cluster trajectories on cluster-cluster aggregation: a comparison of linear and Brownian trajectories in two- and three-dimensional simulations. *Phys. Rev. A* **29**, 997–999.
44. Mukai, T., S. Mukai, and S. Kikuchi, 1987: Complex refractive index of grain material deduced from the visible polarimetry of comet P/Halley. *Astron. Astrophys.* **187**, 650–652.
45. Petrova, E. V., V. P. Tishkovets, and K. Jockers, 2004: Polarization of light scattered by Solar System bodies and the aggregate model of dust particles. *Solar Syst. Res.* **38**, 309–324.
46. Rietmeijer, F. A., 2009: Cometary aggregate interplanetary dust particle as an analog for comet Wild 2 grain chemistry preserved in silica-rich Stardust glass. *Meteor. Planetary Sci.* **44**, 1589–1609.

47. Shen, Y., B. T. Draine, and E. T. Johnson, 2009: Modeling porous dust grains with ballistic aggregates. II. Light scattering properties. *Astrophys. J.* **696**, 2126–2137.
48. Sitko, M., D. Lynch, R. Russell, and M. Hanner, 2004: 3–14 micron spectroscopy of comets C/2002 O4 (Hönig), C/2002 V1 (NEAT), C/2002 X5 (Kudo–Fujikawa), C/2002 Y1 (Juels–Holvorcem), and 69P/Taylor and the relationships among grain temperature, silicate band strength, and structure among comet families. *Astrophys. J.* **612**, 576–587.
49. Tamura, M., M. Fukagawa, H. Kimura, et al., 2006: First two-micron imaging polarimetry of β Pictoris. *Astrophys. J.* **641**, 1172–1177.
50. Thomas–Osip, J. E., B. Å. S. Gustafson, L. Kolokolova, and Y.-I. Xu, 2005: An investigation of Titan’s aerosols using microwave analog measurements and radiative transfer modeling. *Icarus* **179**, 511–522.
51. Volten, H., O. Muñoz, J. W. Hovenier, et al., 2005: WWW scattering matrix database for small mineral particles at 441.6 nm and 632.8 nm. *J. Quant. Spectrosc. Radiat. Transfer* **90**, 191–206.
52. Volten, H., O. Muñoz, J. W. Hovenier, et al., 2007: Experimental light scattering by fluffy aggregates of magnesiosilica, ferrosilica, and alumina cosmic dust analogs. *Astron. Astrophys.* **470**, 377–386.
53. Whipple, F. L., 1957: Some problems of meteor astronomy. In H. C. van de Hulst, Ed., *Radio Astronomy* (Cambridge University Press, Cambridge, UK), pp. 375–389.
54. Zubko, E., H. Kimura, Yu. Shkuratov, et al., 2009: Effect of absorption on light scattering by agglomerated debris particles. *J. Quant. Spectrosc. Radiat. Transfer* **110**, 1741–1749.



From left to right: Theodor Kostiuk, Michael Mishchenko, and Vera Rosenbush enjoy the tour of the Dnipro River.



Touring the Dnipro River (top) and the Saint Sophia Cathedral (bottom).

Astrobiological remote sensing with circular polarization

Ludmilla Kolokolova^{1*}, William Sparks², and Daniel Mackowski³

¹ University of Maryland, College Park, MD 20742, USA

² Space Telescope Science Institute, Baltimore, MD 21218, USA

³ Department of Mechanical Engineering, Auburn University, AL 36849, USA

Abstract. One of the main drivers of astrobiology is the search for life in the Universe. Important evidence relevant to extraterrestrial life is the existence in space of organic molecules of prebiological or biological significance. Such molecules are often characterized by a special type of asymmetry called “homochirality” (domination of molecules of a specific handedness). This results in optical activity of materials containing such molecules. Due to optical activity, the light scattered by such materials is characterized by non-zero circular polarization. We review existing observations of circular polarization in space, including observations of molecular clouds, comets, and the planet Mars. We also review laboratory measurements of light scattered by biological (e.g., bacteria and leaves) and non-biological (minerals) samples. These reveal distinctive features in the circular polarization spectra in absorption bands for the biological samples. We also consider theoretical simulations of light scattering by homochiral materials. Significant progress in this direction has been achieved after the development of the superposition *T*-matrix code for clusters of optically-active spheres. This allows us to simulate light scattering by biological objects, e.g., colonies of bacteria, and by materials of prebiological value, e.g., cometary dust. We explore how circular polarization depends on the porosity and size of aggregates as well as on the degree of their homochirality.

Keywords: circular polarization, optical activity, homochirality, comets, aggregates, laboratory measurements, *T*-matrix method

1. Introduction

The search for life in the Universe starts by defining what we believe is a biosignature, that is, ideally, an observational consequence that can appear only as a result of biological activity. Among such biosignatures are the presence of particular constituents (e.g., methane and oxygen) or combinations of apparently in-

* Corresponding author. E-mail: ludmilla@astro.umd.edu

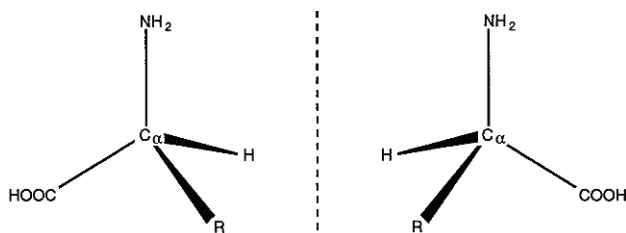


Fig. 1. Two chiral aminoacids: L-aminoacid (left) and D-aminoacid (right).

compatible constituents in the atmospheres of Earth-like planets; observations of the so-called “vegetation red edge” (a rapid change in reflectance for chlorophyll in the near infrared); and identification of homochiral organic molecules. The last of these biosignatures is of special interest in this review as it can be identified remotely, in principle, using polarimetry.

Homochirality is a special feature of organics of biological origin and manifests itself as a violation of mirror symmetry in the material containing biological organic molecules. Complex organic molecules exist in two forms that are identical compositionally and structurally except that they are mirror images of each other (Fig. 1). In laboratory chemical experiments, usually equal numbers of left- and right-handed molecules are produced. However, biological organics are characterized by a major excess of left-handed amino acids and right-handed sugars compared to their oppositely handed counterparts. It is this phenomenon that is referred to as homochirality and is apparently unique to biology. Hence, finding in space a preponderance of one form of these organic molecules over the other may well signal the presence of processes related to biological or prebiological organic chemistry. The presence of strong homochirality would serve as a powerful biosignature.

How can homochirality be identified polarimetrically? It is well known that organic molecules with a predominant chirality (left-handed, called L-molecules, or right handed, called D-molecules) are typically optically active and thus can produce circular polarization (hereafter CP) upon the scattering of natural unpolarized light. Using optical terminology, they possess circular birefringence (the real part of their refractive index is different for light with different handedness of circular polarization) and circular dichroism (the absorption, i.e., the imaginary part of the refractive index, is different for light with different handedness of circular polarization). As a result, the light scattered by such molecules can become either negatively (left-handed) circularly polarized or positively (right-handed) circularly polarized (note that the handedness of the polarization is not necessarily the same as the structural handedness of the molecules). If we have a scattering medium with an equal number of L and D molecules then the circular polarization remains zero. However, as soon as one type of molecules dominates, the scattered light can become circularly polarized. Thus, we should observe CP of the light scattered by any object whose composition is characterized by homochirality.

2. Astronomical observations of circular polarization

In the majority of cases, observations of CP for cosmic objects are not associated with any biological activity. Here, we are not talking about such sources of CP as, for example, light scattering by charged particles. Even for light scattered by neutral dust, circular polarization can be produced by multiple scattering in asymmetric media [47] or by the scattering of light by nonspherical particles aligned by a gas flow, magnetic or electric field, or radiative torque [24,25]. In these situations CP is not caused by homochirality; however, they have an important astrobiological value as they may be the reason for chemical homochirality. For example, an active discussion of the origin of homochirality was ignited by observations of strong CP (up to 20%) for star forming regions (see, e.g., the review [15]). It has been suggested [1,2,6] that CP, mainly caused by the interaction of dust with the magnetic field, can affect synthesis of organic molecules due to so-called asymmetric photolysis involving the preferential destruction of molecules of a specific chirality by circularly polarized UV light [7]. It is plausible that the subsequent evolution of the molecular cloud, in which the organics were affected by asymmetric photolysis, into a protoplanetary disk and then into a planetary system can result in the organic material in this system being characterized by a built-in chiral bias for one handedness over the other. This hypothesis is now considered as one of the possible origins of homochirality in terrestrial bioorganics.

If this hypothesis is correct, homochirality, or at least a chiral excess, should be found in the organics of other Solar-system objects. An important confirmation of this hypothesis is the discovery of an excess of L amino acids in meteorites [10,11]. As an evidence of wide-spread homochiral organics in the Solar system, we may also consider observations of CP in comets.

The presence of non-zero CP in the light scattered by cometary dust has been confirmed for several comets. The first comet for which CP was detected is 1P/Halley; left-handed (negative) CP was detected by Dollfus and Suchail [12]. The first systematic and high-precision measurements of CP were obtained for comet C/1995 O1 (Hale–Bopp) in 1997 [37]. Left-handed CP with a maximum value of $-0.26\% \pm 0.02\%$ was detected for all measured areas across the coma (Fig. 2a). These results are consistent with those obtained by Manset and Bastien [28].

The spatial distribution of CP along cuts through the coma was investigated for comet C/1999 S4 (LINEAR) (hereafter S4) during its splitting in the summer of 2000 [38]. The maximum CP value reached 1%. Left-handed as well as right-handed polarization was observed over the coma, although immediately after the complete disintegration of the nucleus the polarization was mainly left-handed, on average $-0.41 \pm 0.07\%$ (Fig. 2b). CP variations coincided in time with outbursts caused by the fragmentation of the nucleus. Simultaneous measurements of circular and linear polarization were performed for comet C/2001 Q4 (NEAT) (hereafter Q4) [39]. These measurements were made along cuts which passed over the

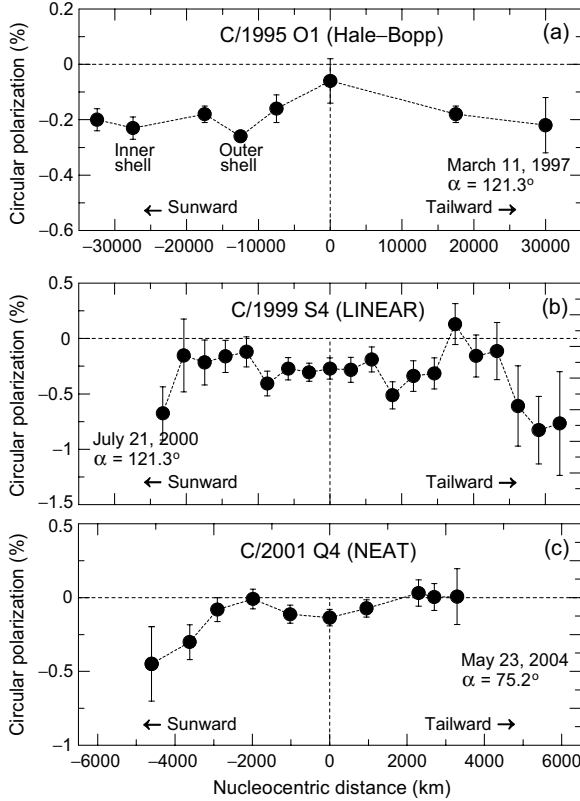


Fig. 2. Variations of CP with the nucleocentric distance for comets (a) C/1995 O1 (Hale-Bopp) [37], (b) C/1999 S4 (LINEAR) [38], and (c) C/2001 Q4 (NEAT) [39].

coma and the nucleus as well as along the dust jet (Fig. 2c). The first CCD maps of cometary CP were obtained at the European Southern Observatory (ESO) for comets 9P/Tempel 1 and 73P/Schwassmann–Wachmann 3 (hereafter SW3). For SW3, the values of CP did not exceed 0.2%, however they could be identified as left-handed [46]. Data for 9P/Tempel 1 obtained four days after the Deep Impact event [5] showed zero CP. The absence of CP for Tempel 1 may indicate that only materials that are insufficiently processed (as those in comets Hale-Bopp and Q4) or freshly released from the nucleus interior (as in comets S4 and SW3) can produce non-zero CP.

The traditional explanation of CP as a result of multiple scattering or particle alignment does not work in the case of cometary dust; this was discussed in detail in [38]. Multiple scattering is negligible in comets. Moreover, the only comet with a noticeable optical thickness for which multiple scattering could play a role [13], i.e., comet Hale-Bopp, showed rather small values of CP, whereas other, more “transparent” comets like Q4 or S4 showed much higher CP values. Comets do

not possess a magnetic field. The solar magnetic field can penetrate into the coma. However, due to interaction with ions of the cometary coma, it does not penetrate deeply enough (e.g., for comet Halley it did not penetrate deeper than ~ 5000 km from the nucleus [9, 16]), whereas CP in comets is observed as close as ~ 100 km to the nucleus. The gas drag alignment of the dust particles can work in the vicinity of the nucleus. However, cometary dust gets separated from the gas very close to the nucleus (~ 10 nucleus radii [8]) and later this alignment gets destroyed by radiative torque. In turn, the alignment due to radiative torque cannot be responsible for the CP observed for comets as it aligns particles perpendicularly to the direction of illumination and produces zero CP after averaging over particle orientations in the plane perpendicular to the radiation. Currently, the scattering on homochiral molecules appears to be the least contradictory way to explain cometary CP, especially since, unlike all other mechanisms, it provides a logical explanation for the fact that the CP is predominantly left-handed in all observed comets [39], just as meteorites consistently reveal the dominance of L-amino acids.

There is also observational evidence of a circularly polarized component in the light scattered by planetary surfaces [18, 19]. CP at the level of 1%–2% has been measured. In all cases it was shown that non-zero CP could be explained by multiple scattering, e.g., on a rough (regolith) surface that was “asymmetric”, meaning that it was in an oblique configuration relative to the observer. Attempts to search for homochiral organics in specific locations on the Martian surface [41] by observing their polarimetric characteristics did not reveal any CP. Although this negative result appears to be rather discouraging at first sight, it is consistent with the finding from the Viking mission and later analyses that UV radiation destroys all complex organics on the Martian surface (see, e.g., [31]). However, only a very small fraction of the surface and the spectrum has been characterized, while this work has also the additional encouraging outcome of indicating the absence of “false positives”, i.e., non-biological CP.

3. Theoretical modeling of circular polarization produced by particles containing homochiral organics

Additional evidence of the origin of CP in comets as a result of light scattering on homochiral organic molecules would be obtained if theoretical modeling could reproduce the observed values of CP and its variation with phase angle. The first attempt [38] to simulate CP in the light scattered by cometary dust was undertaken using the solution for an optically active sphere by Bohren [3]. A code based on this solution is available through the Library of Light-Scattering Codes maintained by Thomas Wriedt at <http://www.t-matrix.de>. To make the calculations more realistic, particles having the power-law size distribution measured *in situ* for the dust in comet 1P/Halley [30] were considered. Values of the refractive indices and specific optical rotation were taken typical of the amino-acids discovered in the Murchison meteorite [33–35]. According to Mason [29], the corresponding rotation angle is $\sim 100^\circ$, which causes a difference in the refractive indices for the left-

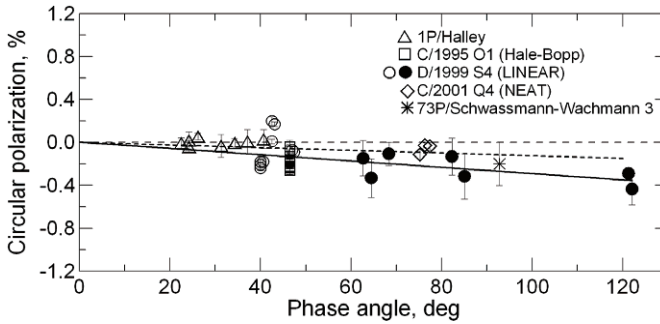


Fig. 3. Composite dependence of CP on the phase angle (i.e., 180° minus scattering angle) for comets 1P/Halley, C/1995 O1 (Hale-Bopp), S4, Q4, and SW3. The solid line is the linear fit to the observed data; the thick dashed line depicts the results of calculations for an ensemble of polydisperse optically active spheres [38].

and right-hand circular polarizations of 3×10^{-6} (in these calculations, 10% of the material was assumed to be optically active, i.e., chiral). Under these conditions, the values of CP greater than 0.005% could not be achieved. However, after taking into consideration the circular dichroism of organic molecules (i.e., different absorptions for left- and right-hand circularly polarized light), the calculated polarization increases significantly. Even though circular dichroism is usually very small, causing a difference in the imaginary part of the refractive index on the order of 1×10^{-8} [40], this appeared to be sufficient to produce CP that had the observed phase-angle trend and reached 0.15% at the phase angle equal to 120° (the thick dashed curve in Fig. 3).

Even though the results of calculations for an optically active sphere looked promising, the calculated values of CP were smaller than the observed ones. However, the shape and morphology of the grains can be a crucial factor in the genesis of the observed CP. A more realistic model of cometary dust as aggregates of submicron particles can significantly increase the value of CP owing to optical interactions between the monomers in the aggregate that provide an effect similar to the multiple scattering effects discussed above, or even stronger due to the close proximity of the interacting particles. Note that each single aggregate can provide some CP in the scattered light due to structural features of the aggregate (mirror asymmetry of particle arrangement). However, a natural ensemble of aggregates should contain equal numbers of aggregates whose structure is responsible for producing right-handed and left-handed polarization, thus, the average CP of such an ensemble should be zero and only optical activity of the material can be responsible for non-zero CP.

The notion of cometary particles being aggregates is not only a good approach to interpret their CP but also the most recognized model of the cometary dust. It has been shown [20,21] that the aggregate model is the only one that can simulate the full scope of observational data. This model allows one to fit cometary pho-

tometry (including the observed low albedo, phase function of brightness, color of comet dust and its change with phase angle), polarimetry (including the shape of the phase function and spectral trend of cometary polarization), and thermal infrared data. This model can explain why comets form two polarimetrically- and infrared-distinct classes. It also is consistent with the findings of the Stardust mission and with the evolutionary models of comet formation (for more details, see [23]).

3.1. Modeling of light scattering by clusters of optically active spheres

The superposition T -matrix code described in [26] (available online at <ftp://ftp.eng.auburn.edu/pub/dmckwski/scatcodes/index.html>) is based on the assumption that light is scattered by a cluster (aggregate) consisting of optically isotropic spheres. However, there is no limitation in the code that prevents it from being extended to clusters (aggregates) of optically active spheres. The exact solution for a single optically active sphere by Bohren [3] is formulated in the vector-harmonic basis, which makes it amenable to the multiple-sphere problem. We only need to replace the Lorenz–Mie coefficients in the code with those corresponding to optically active spheres. The formulation for interactive scattering among a cluster of optically active spheres was obtained by a merging of the formulations for optically active single spheres and multiple, non-active spheres; see for detail [27]. Briefly, it was done the following way. As in the case for isotropic spheres, the electric field incident on and scattered by an optically active sphere was represented by expansions of regular and outgoing vector wave harmonics, respectively, which appear as

$$\mathbf{E}_{\text{inc}}(\mathbf{r}) = \sum_{n=1}^{\infty} \sum_{m=-n}^n \sum_{p=1}^2 f_{mnp} \mathbf{N}^{(1)}(k\mathbf{r}), \quad (1)$$

$$\mathbf{E}_{\text{sca}}(\mathbf{r}) = \sum_{n=1}^{\infty} \sum_{m=-n}^n \sum_{p=1}^2 a_{mnp} \mathbf{N}^{(3)}(k\mathbf{r}), \quad (2)$$

in which f and a denote the incident and scattered expansion coefficients, while $(m; n; p)$ denote the degree, order, and mode (TM or TE) of the harmonic. Unlike the isotropic case, however, the optically active sphere case results in a coupling of TE and TM modes between the incident and scattered field coefficients for a given harmonic order [3]. That is, the Lorenz–Mie relation for the active sphere appears as

$$a_{mnp} = \sum_{q=1}^2 \tilde{a}_{n;pq} f_{mnq}, \quad (3)$$

in which the coefficients a are functions of the sphere size parameter and the left- and right-handed refractive indices.

The modified Lorenz–Mie relation for the optically active sphere was incorporated directly into the formulation for multiple spheres, and resulted in the fol-

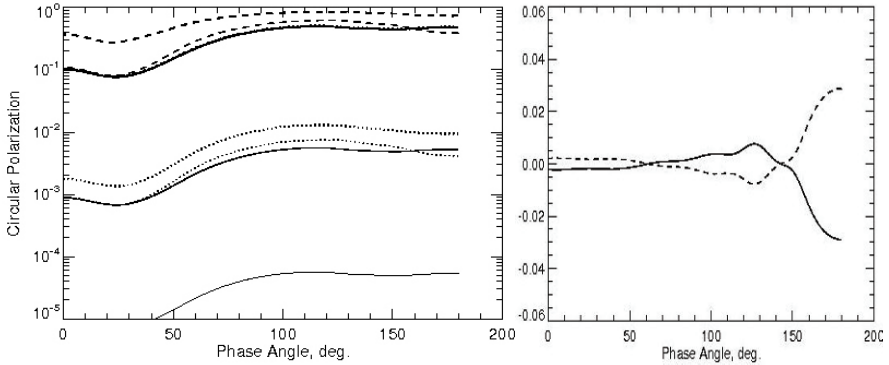


Fig. 4. Dependence of CP on the optical activity of the material. We depict the results for a pair of mirror-symmetric fractal aggregates, one of which is shown in the inset. The number of monomers is 50, their radius is $0.05 \mu\text{m}$, and the wavelength is $0.65 \mu\text{m}$. The right-hand panel shows the results for $\beta = 0.1 + i0.1$ (solid curve) and β of the opposite sign (dashed curve). As expected, the results differ only in the sign of CP. The left-hand panel shows the results for a variety of β values. Solid curves correspond to $\beta_i = -0.8692 \times 10^{-8}$, dotted curves are for $\beta_i = -0.86922 \times 10^{-4}$, and dashed curves are for $\beta_i = -0.8692 \times 10^{-2}$. The thinnest curves are for $\beta_r = 7.0342 \times 10^{-6}$, the moderately thick curves are for $\beta_r = 7.034 \times 10^{-4}$, and the thick curves are for $\beta_r = 7.0342 \times 10^{-2}$. The thin and moderately thick dashed curves and the thick solid and dotted curves are hardly distinguishable as the data are very close.

lowing interaction equations for the sphere scattering coefficients:

$$a_{mnp}^i = \sum_{p'=1}^2 \bar{a}_{n,pp'} \sum_{\substack{j=1 \\ j \neq i}}^{N_p} \sum_{l=1}^{L_j} \sum_{k=-l}^l \sum_{q=1}^2 H_{mnp'klq}^{i-j} a_{klq}^j = \sum_{p'=1}^2 \bar{a}_{n,pp'}^i f_{mnp'}^i. \quad (4)$$

In the above, H^{i-j} is an outgoing harmonic translation matrix depending solely on the distance and direction between origins i and j . Following the procedures developed in [26], a T matrix for the cluster of active spheres was obtained from the solution of the interaction equations. The orientation-averaged scattering matrix elements can be analytically determined from operations on the T matrix.

The Stokes vector of the scattered light can be expressed in terms of the parallel and perpendicular components of the electric field vector \mathbf{E} [4]. Once the electric field vector \mathbf{E} of the scattered light is known, we can calculate all four components of the Stokes vector, I , Q , U , and V and find the CP defined as V/I .

After thorough testing described in [27], we performed some calculations to check remote-sensing capabilities of CP. As is well known (see, e.g., [14, 22]) individual aggregates often demonstrate circular polarization due to the mirror asymmetry of the arrangement of particles in the aggregate. However, this effect disappears in an ensemble of natural aggregates characterized by equal numbers of left- and right-handed arrangements of monomers. To see the effect related only to

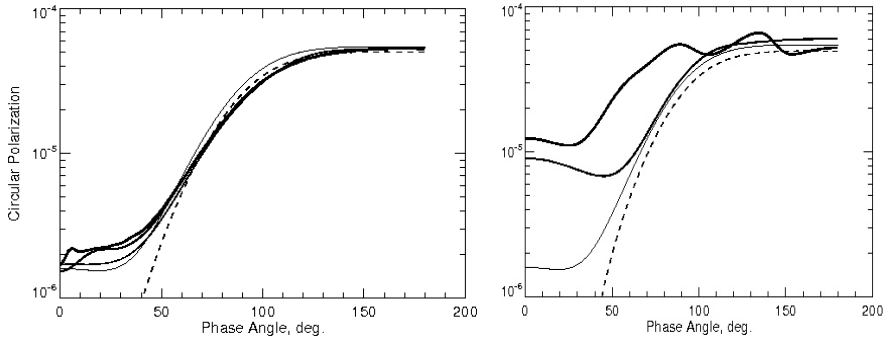


Fig. 5. Dependence of absolute values of CP on the size of a 3D-cross aggregate (left) and size of its constituent monomers (right). An example of the 3D-cross aggregate is shown in inset. The radius of the monomers for the left-hand panel is $0.05 \mu\text{m}$. The dashed curve is for a single monomer; the solid curves are for the aggregates of 13, 31, 127, and 343 monomers (the thickness of the curves increases with the number of monomers). The number of monomers for the right-hand panel is 13. The radius of the smallest monomer is $0.025 \mu\text{m}$ (dashed curve); the solid curve is for monomers of radii 0.05 , 0.1 , and $0.15 \mu\text{m}$ (the thickness of the curves increases with the monomer size). The wavelength is $0.65 \mu\text{m}$.

the optical activity of the material, we considered either mirror-symmetric pairs of aggregates or completely symmetric aggregates like 3D-crosses or cubes. Some results are shown in Figs. 4–6. In all cases we obtained negative values of circular polarization which we reversed in order to exploit the logarithmic scale. Thus, all plots (except the middle panel of Fig. 4) show absolute values of CP.

Figure 4 demonstrates how the CP depends on the optical activity of the aggregates. The refractive index of the material was taken from [38] where it was used for modeling cometary CP using optically active spheres. As in [38], we assumed that only 10% of the material was optically active. We described optical activity as a complex parameter $\beta = \beta_r + i\beta_i$ characterizing the difference between the left- and right-handed refractive indices, where $i = (-1)^{1/2}$. Based on the assumptions described above, we got the refractive index $m = 1.55002 + i0.0006002$, to which should be added $\beta_r = 7.034 \times 10^{-6}$ to parameterize circular birefringence and $\beta_i = -0.8692 \times 10^{-8}$ to represent circular dichroism. Calculations were performed for the wavelength $0.65 \mu\text{m}$. It is evident (and not surprising) that CP increases rapidly if the values of either birefringence or dichroism increase.

Figures 5 and 6 show how CP depends on the structure and size of the aggregates and the size of their monomers. Figure 5 represents an extreme case of a porous symmetric aggregate in the form of a 3D-cross. Figure 6 corresponds to an extreme case of a very compact symmetric aggregate in the form of a cube made of spherical monomers. One can see that the CP produced by very porous aggregates is very much the same for all sizes of the aggregate. At small phase angles, it is larger than the CP for an individual monomer; at larger phase angles, it follows

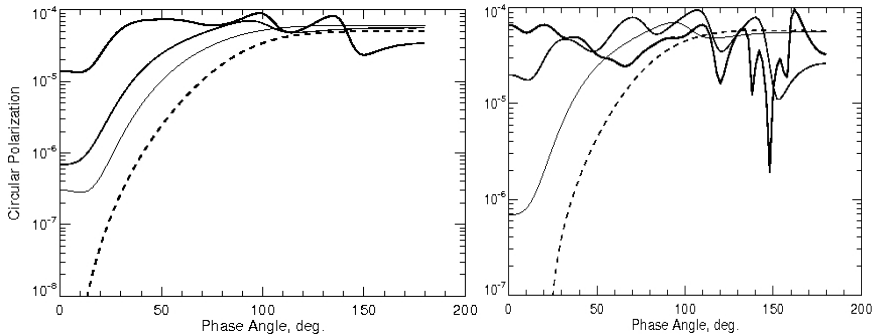


Fig. 6. Dependence of absolute values of CP on the size of a cubic aggregate (left) and size of its constituent monomers (right). An example of the cubic aggregate is shown in the inset. The radius of the monomers for the left-hand panel is $0.05 \mu\text{m}$. The dashed curve is for a single monomer; the solid curves are for the aggregates of 9, 125, and 343 monomers (the thickness of the curves increases with the number of monomers). The number of monomers for the right-hand panel is equal 125 to keep the overall size of the cubic aggregate close to the size of the 3D-cross used to obtain the curves in the right-hand panel of Fig. 5. The radius of the smallest monomer is $0.025 \mu\text{m}$ (dashed curve); the solid curve is for monomers of radii 0.05 , 0.1 , and $0.15 \mu\text{m}$ (the thickness of the curves increases with the monomer size). The wavelength is $0.65 \mu\text{m}$.

the behavior of the individual sphere. For compact aggregates, the CP increases with the size of aggregates, especially at small and medium phase angles where it becomes orders of magnitude larger as the size of the aggregate increases. When studying the effect of the aggregate size (the left-hand panels in Figs. 5 and 6), we tried to keep the same number of monomers in both the 3D-cross and the cube. Studying the effect of the monomer size (the right-hand panels in Figs. 5 and 6) was facilitated by considering aggregates of almost the same geometric cross-section; in the case of the 3D-cross this was an aggregate of 13 monomers, while the corresponding cube consisted of 125 monomers. One can see that increase of the monomer size also increases CP and, again, the effect is more pronounced in the case of compact aggregates.

Our computations presented in Figs. 4–6 demonstrate the general trend in the phase curve of CP similar to the one observed for comets: CP is negative and increases in absolute value with phase angle. However, the computed values of CP do not reach the values observed for comets. Perhaps we need to consider larger and, probably, rather compact aggregates to further increase the theoretical values of CP. This does not contradict our knowledge of properties of cometary dust which, most likely, consists of aggregates built of thousands of monomers [20,23].

It is clear that observations and analyses of CP can provide information not only on optical activity but also on packing density of the aggregates, their size, and size of their monomers. These three factors produce competing effects, and an

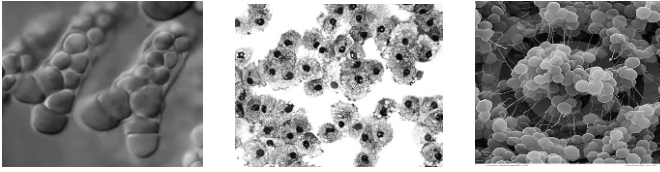


Fig. 7. *Hyella stella*, a cyanobacterium that lives in Marine Limestone (left, adapted from <http://www.biologie.uni-hamburg.de/b-online>); white blood cells (middle, adapted from <http://www.fi.edu/learn/heart/blood/images>); staphylococcus epidermidis bacteria (right, adapted from www.scharfphoto.com).

extended computational survey is necessary to find a way to distinguish between the compositional and structural factors when using CP in remote sensing.

3.2. Circular polarization of light scattered by biological objects: laboratory measurements

The previous section demonstrated that to study chirality of molecules in cometary dust through the use of CP, one needs to consider the dust particles as aggregates (clusters) of small monomers. Aggregates are a very typical form of natural particles. They constitute interstellar and interplanetary dust, dust in protoplanetary nebulae, and many planetary aerosols, including those in the Titan atmosphere [44,45]; aggregates are a very likely kind of particle to be found in the atmospheres of exoplanets.

Many biological particles (bacteria, spores, blood cells) also tend to form aggregates (Fig. 7). Modeling their optical properties also requires a computer code that works for clusters of small monomers. The approach of modeling biological particles as clusters of spheres has been widely used (see, e.g., [32,48,49]). However, none of the previous computer modeling studies has involved optical activity resulting from the homochirality of the constituent molecules and thus the resulting CP. However, CP is a key characteristic that allows us to distinguish biotic objects from abiotic ones. This was certainly proved by our laboratory measurements described below. Here we only briefly summarize our measurements and refer the reader to [42,43] for more details.

Using a dedicated precision polarimeter located at the National Institute of Standards (NIST), we measured CP produced by the reflection and transmission of light by microorganisms. It is reasonable to suppose that photosynthetic life is concentrated at the surface, uses windows of atmospheric transparency, and exploits regions of the spectrum where the host star shines brightly, which makes such life forms maximally observable. The strong electronic absorption bands characteristic of photosynthesis are known to exhibit circular dichroism, so we anticipated a consequent polarization signature in the scattered light, although this was not entirely obvious beforehand owing to the complexities of the scattering process.

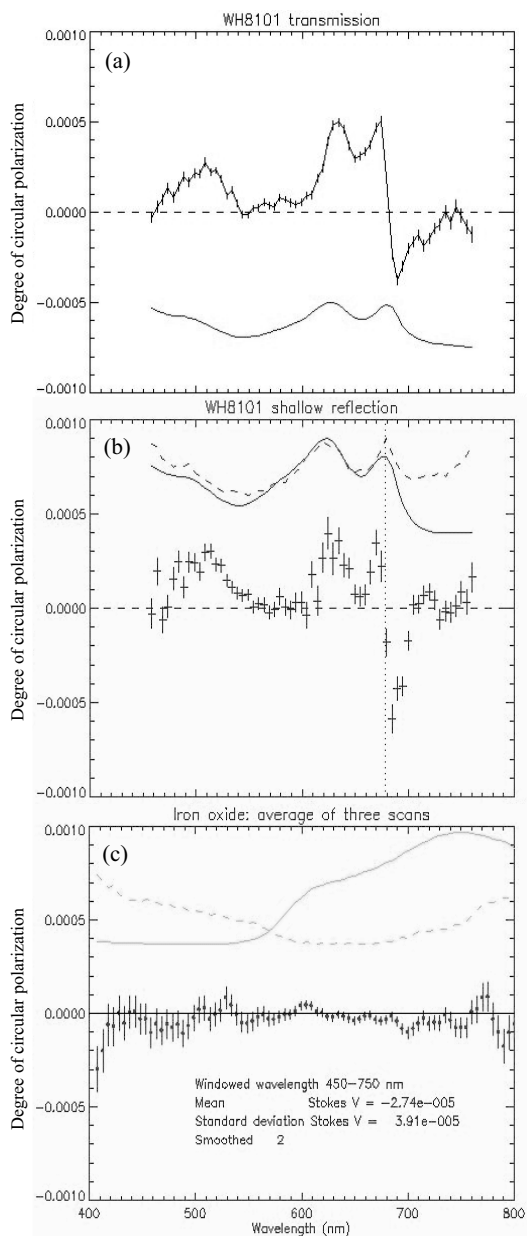


Fig. 8. CP spectra of *Synechococcus* WH7805 in (a) transmission (the darker curve with error bars) and (b) reflection (error-bar crosses). Note the high degree of similarity between the two. The solid curve in (b) shows an arbitrarily scaled intensity spectrum; the dashed curve is a scaled plot of linear polarization. (c) The dotted curve with one-sigma error bars shows the CP reflection spectrum of iron oxide; the lighter solid and dashed curves show the intensity and linear polarization, respectively.

To relate the experimental results to known biophysical properties, we compared the scattered polarization spectrum to the circular dichroism spectrum. For this, the experiment was configured to acquire both transmission and scattered polarization spectra of our target microbial cultures. As expected, the cyanobacteria (Figs. 7a,b) show chlorophyll absorption in the red ($\approx 0.68 \mu\text{m}$) and the antenna pigments phycocyanin ($\approx 0.62 \mu\text{m}$) and phycoerythrin ($\approx 0.56 \mu\text{m}$) absorption bands. The transmission CP spectra (Fig. 8a) are analogous to a classical circular dichroism experiment used in protein structure and conformation analysis [17]. Hence we expected to see Cotton effect circular dichroism signatures of the strong electronic absorption features, and this is exactly what was revealed. For chlorophyll, the sign of CP reverses precisely at the wavelength of the absorption maximum. This well-known effect is due to the presence of exciton-coupled chlorophyll molecule dimers in which chlorophyll molecules in close proximity function in pairs, effectively acting as a macromolecule. The essential characteristics of the CP transmission spectrum are fully reproduced in the polarization spectrum of reflected light (Fig. 8b). Figure 8c shows the polarization of a red iron oxide powder, chosen because it has a spectral edge not unlike chlorophyll and might present a false positive in a chlorophyll red-edge detection experiment. The spectropolarization signature of the iron oxide is very close to the noise limit of the instrument; there is a lack of any pronounced spectral features in CP, and there is no correlation with the absorption spectrum. Pospergelis [36] showed circular spectropolarimetry of a variety of other minerals with similar results. Hence the CP properties of light scattered by chiral biological material differ markedly from those of abiotic substances.

4. Conclusions

An important biosignature is the presence of homochiral organics, i.e., organics dominated by left- or right-handed molecules. Such molecules are optically active, i.e., possess circular birefringence and circular dichroism. As a result, the light they scatter can become circularly polarized. Thus, circular polarization observed for an object that is not associated with a strong magnetic field or particle alignment may be a manifestation of the presence of homochiral organics of biological or prebiological origin.

One such example is cometary dust. Simulations of light scattering by aggregates made of optically active materials are required to characterize the polarization spectrum arising from light scattered by cometary dust. A powerful computer code to simulate light scattering by optically active aggregates was developed based on the *T*-matrix approach [27]. Preliminary computations using this code showed that aggregates demonstrate CP that can be several orders of magnitude stronger than the CP from the individual spheres constituting the aggregate. We expect that simulations for typical cometary aggregates composed of thousands of monomers will yield polarization at the level observed for comets, even with very low values of optical activity based on the homochirality typical of meteorites.

This may depend on the porosity of the aggregates as CP increases much faster in the case of compact aggregates. An increase in the circular birefringence or dichroism of the monomers also strengthens the CP dramatically. Hence it is possible that the observed levels of CP can result from just a small admixture of homochiral organics in cosmic dust or planetary aerosols. Conversely, this may make CP a good marker of prebiological organics provided that other sources of polarization can be excluded.

Even better characterization of CP can be provided for biological objects such as bacteria or vegetation. Our laboratory measurements show that CP is much larger for biological objects than for non-biological ones and reaches especially high values inside absorption bands. Moreover, it can experience a reversal of sign at the center of an absorption band. Interestingly, this happens only for biological objects and the absorption bands of “biological” origin, e.g., chlorophyll or carotene. Such distinctive behavior indicates that CP can be an exceptionally good biomarker in the search for extraterrestrial life.

The recent studies discussed in this chapter show CP is an important astrobiological tool. The next steps in its application to the search for life in space will be the simulation of light scattering by cometary-like aggregates, colonies of bacteria, leaves, and other objects using the newly developed *T*-matrix code. Such simulations will allow us to identify the best angular and spectral regimes to study CP of the light scattered by biological objects and prebiological chiral molecules. Ultimately, it will be very interesting to obtain CP spectra of exoplanets and interplanetary/interstellar media in the context of these results and to assess their character in the framework of the search for life in the Universe.

Acknowledgments

This work was supported by the NASA Astrobiology Program. L. Kolokolova acknowledges a NATO ASI travel grant.

References

1. Bailey, J., 2000: Circular polarization and the origin of biomolecular homochirality. In G. Lemarchand and K. Meech, Eds., *Bioastronomy 99: A New Era in the Search for Life* (Astronomical Society of the Pacific, San Francisco), pp. 349–354.
2. Bailey, J., 2004: Extraterrestrial chirality. In R. P. Norris and F. H. Stootman, Eds., *Bioastronomy 2002: Life among the Stars* (Astronomical Society of the Pacific, San Francisco), pp. 139–144.
3. Bohren, C. F., 1975: Scattering of electromagnetic waves by an optically active spherical shell. *J. Chem. Phys.* **62**, 1566–1571.
4. Bohren, C. F., and D. Huffman, 1983: *Absorption and Scattering of Light by Small Particles* (Wiley, New York).
5. Boehnhardt, H., and ESO DI team, 2005: The Deep Impact campaign at ESO: dust and nucleus characterization. In *IAU symposium ACM-2005*, Vol. 229 (Buzios, Brazil).
6. Bonner, W. A., and B. D. Bean, 2000: Asymmetric photolysis with elliptically polarized light. *Orig. Life Evol. Biosph.* **30**, 513–517.

7. Cerf, C., and A. Jorissen, 2000: Is amino-acid homochirality due to asymmetric photolysis in space? *Space Sci. Rev.* **92**, 603–612.
8. Combi, M. R., K. Kabin, D. L. Dezeew, et al., 1997: Dust-gas interrelations in comets: observations and theory. *Earth Moon Planets* **79**, 275–306.
9. Cravens, T. E., and T. I. Gombosi, 2004: Cometary magnetospheres: a tutorial. *Adv. Space Res.* **33**, 1968–1976.
10. Cronin, J. R., and S. Pizzarello, 1997: Enantiomeric excesses in meteoritic amino acids. *Science* **275**, 951–955.
11. Cronin, J. R., and S. Pizzarello, 1999: Amino acid enantiomeric excesses in meteorites: origin and significance. *Adv. Space Res.* **23**, 293–299.
12. Dollfus, A., and J. -L. Suchail, 1987: Polarimetry of grains in the coma of P/Halley. I. Observations. *Astron. Astrophys.* **187**, 669–688.
13. Fernandez, Y. R., D. D. Wellnitz, M. W. Buie, et al., 1999: The inner coma and nucleus of Comet Hale–Bopp: results from a stellar occultation. *Icarus* **140**, 205–220.
14. Guirado, D., J. W. Hovenier, and F. Moreno, 2007: Circular polarization of light scattered by asymmetrical particles. *J. Quant. Spectrosc. Radiat. Transfer* **106**, 63–73.
15. Hough, J. H., J. Bailey, A. Chrysostomou, et al., 2001: Circular polarization in star-forming regions: possible implications for homochirality. *Adv. Space Res.* **27**, 313–322.
16. Israelevich, P., A. Ershkovich, T. Gombosi, et al., 2003: Fine structure of the diamagnetic cavity boundary in comet Halley. *J. Geophys. Res.* **108**, 12–19.
17. Kelly, S. M., and N. C. Price, 2000: The use of circular dichroism in the investigation of protein structure and function. *Current Protein Peptide Sci.* **1**, 349–384.
18. Kemp, J. C., R. D. Wolstencroft, and J. B. Swedlund, 1971: Circular polarization: Jupiter and other planets. *Nature* **232**, 165–168.
19. Kemp, J. C., 1974: Circular polarization of planets. In T. Gehrels, Ed., *Planets, Stars, and Nebulae: Studied with Photopolarimetry* (University of Arizona Press, Tucson, AZ), pp. 607–616.
20. Kolokolova, L., H. Kimura, and I. Mann, 2004: Characterization of dust particles using photopolarimetric data: example of cometary dust. In G. Videen, Ya. Yatskiv, and M. Mishchenko, Eds., *Photopolarimetry in Remote Sensing* (Kluwer, Dordrecht), pp. 431–454.
21. Kolokolova, L., M. Hanner, A.-Ch. Levasseur-Regourd, and B. Å. S. Gustafson, 2004: Physical properties of cometary dust from light scattering and emission. In M. C. Festou, H. U. Keller, and H. A. Weaver, Eds., *Comets II* (University of Arizona Press, Tucson, AZ), pp. 577–605.
22. Kolokolova, L., H. Kimura, K. Ziegler, and I. Mann, 2006: Light-scattering properties of random-oriented aggregates: do they represent the properties of an ensemble of aggregates? *J. Quant. Spectrosc. Radiat. Transfer* **100**, 199–206.
23. Kolokolova, L., 2011: Photopolarimetric remote sensing of aggregates in cosmic dust. In M. I. Mishchenko, Ya. S. Yatskiv, V. K. Rosenbush, and G. Videen, Eds., *Polarimetric Detection, Characterization, and Remote Sensing* (Springer, Berlin) (this volume).
24. Lazarian, A., 2003: Magnetic fields via polarimetry: progress of grain alignment theory. *J. Quant. Spectrosc. Radiat. Transfer* **79–80**, 881–902.
25. Lazarian, A., 2009: Quantitative theory of grain alignment: probing grain environment and grain composition. In Th. Henning, E. Grün, and J. Steinacker, Eds., *Cosmic Dust – Near and Far* (Astronomical Society of the Pacific, San Francisco), pp. 482–497.

26. Mackowski, D. W., and M. I. Mishchenko, 1996: Calculation of the T matrix and the scattering matrix for ensembles of spheres. *J. Opt. Soc. Am. A* **13**, 2266–2278.
27. Mackowski, D. W., L. Kolokolova, and W. Sparks, 2011: T -matrix approach to calculating circular polarization of aggregates made of optically active materials. *J. Quant. Spectrosc. Radiat. Transfer* **112**, in press.
28. Manset, N., and P. Bastien, 2000: Polarimetric observations of comets C/1995 O1 Hale–Bopp and C/1996 B2 Hyakutake. *Icarus* **145**, 203–219.
29. Mason, S. F., 1982: *Molecular Optical Activity and the Chiral Discriminations* (Cambridge University Press, New York).
30. McDonnell, J. A. M., G. C. Evans, S. T. Evans, et al., 1987: The dust distribution within the inner coma of comet P/Halley 1982i – encounter by Giotto’s impact detectors. *Astron. Astrophys.* **187**, 719–741.
31. Pang, K. D., S. F. S. Chun, J. M. Ajello, et al., 1982: Organic and inorganic interpretations of the martian UV–IR reflectance spectrum. *Nature* **295**, 43–46.
32. Petrov, D., G. Videen, Yu. Shkuratov, and V. Kaydash, 2007: Analytic T -matrix solution of light scattering from capsule and bi-sphere particles: applications to spore detection. *J. Quant. Spectrosc. Radiat. Transfer* **108**, 81–105.
33. Pizzarello, S., and G. R. Cronin, 2000: Non-racemic amino acids in the Murray and Murchison meteorites. *Geochim. Cosmochim. Acta* **64**, 329–338.
34. Pizzarello, S., and G. Cooper, 2001: Molecular and chiral analyses of some protein amino acid derivatives in the Murchison and Murray meteorite. *Meteorit. Planet. Sci.* **36**, 897–909.
35. Pizzarello, S., 2004: Chemical evolution and meteorites: an update. *Orig. Life Evol. Biosph.* **34**, 25–34.
36. Pospergelis, M. M., 1969: Spectroscopic measurements of the four Stokes parameters for light scattered by natural objects. *Soviet Astron.* **12**, 973–977.
37. Rosenbush, V., N. Shakhovskoy, and A. Rosenbush, 1999: Polarimetry of Comet Hale–Bopp: linear and circular polarization, stellar occultation. *Earth Moon Planets* **78**, 373–379.
38. Rosenbush, V., L. Kolokolova, A. Lazarian, et al., 2007: Circular polarization in comets: observations of Comet C/1999 S4 (LINEAR) and tentative interpretation. *Icarus* **186**, 317–330.
39. Rosenbush, V., N. Kiselev, N. Shakhovskoy, et al., 2007: Circular and linear polarization of comet C/2001 Q4 (NEAT). Why circular polarization in comets is predominantly left-handed? In G. Videen, M. Mishchenko, M. P. Mengüç, and N. Zakharova, Eds., *Peer-Reviewed Abstracts of the Tenth Conference on Electromagnetic & Light Scattering* (Bodrum, Turkey), pp.181–184.
40. Schreier, P., A. Bernreuther, and M. Huffer, 1995: *Analysis of Chiral Organic Molecules: Methodology and Applications* (Walter de Gruyter Publ., Berlin).
41. Sparks, W., J. Hough, and L. Bergeron, 2005: A search for chiral signatures on Mars. *Astrobiology* **5**, 737–748.
42. Sparks, W., J. H. Hough, L. Kolokolova, et al., 2009: Circular polarization in scattered light as a possible biomarker. *J. Quant. Spectrosc. Radiat. Transfer* **110**, 1771–1779.
43. Sparks, W., J. H. Hough, T. A. Germer, et al., 2010: Detection of circular polarization in light scattered from photosynthetic microbes. *Proc. Natl. Acad. Sci. USA* **106**, 7816–7821.
44. Tomasko, M. G., B. Archinal, T. Becker, et al., 2005: Rain, winds and haze during the Huygens probe’s descent to Titan’s surface. *Nature* **438**, 765–778.

45. Tomasko, M. G., L. Doose, S. Engel, et al., 2008: A model of Titan's aerosols based on measurements made inside the atmosphere. *Planet. Space Sci.* **56**, 669–707.
46. Tozzi, G. P., S. Bagnulo, H. Boehnhardt, et al., 2006: Observations of comet 73P/SW3 at its closest approach to the Earth. In *European Planetary Science Congress 2006* (Berlin, Germany), p. 725.
47. Whitney, B. A., and M. J. Wolff, 2002: Scattering and absorption by aligned grains in circumstellar environments. *Astrophys. J.* **574**, 205–231.
48. Zubko, E., Yu. Shkuratov, M. Hart, et al., 2003: Backscattering and negative polarization of agglomerate particles. *Opt. Lett.* **28**, 1504–1506.
49. Zubko, E., A. Ovcharenko, S. Bondarenko, et al., 2004: Backscattering of agglomerate particles. *Proc. SPIE* **5617**, 407–415.



At the Museum of the Main Astronomical Observatory.



Waiting for lunch while sailing the Dnipro River.



On the Bohdan Khmelnytsky Square.

Inferring properties of dust within small bodies of the solar system through observations and simulations of the linear polarization of scattered solar light

Anny-Chantal Levasseur-Regourd*

UPMC Univ. Paris 06, UMR 8190 (LATMOS–CNRS), BC 102, 4 place Jussieu,
75005 Paris, France

Abstract. Measuring properties of solar light scattered by dust on the surfaces of asteroids and within dust clouds (cometary comae, zodiacal cloud) is often a technical challenge, since the signals need to be isolated from any contaminating source on the line of sight. However, measurements of linear polarization are of major interest, since they enable comparisons between data obtained at different distances from the Sun and the observer on a given object or on various objects. They provide some classifications and efficiently complement unique *in situ* and sample return missions. Some clear observational trends can be pointed out, from the results obtained, mostly on relatively bright objects, by various teams all over the world. It is worth mentioning that all polarimetric phase curves present a shallow negative branch in the backscattering region, with some significant dispersion of the inversion angle for asteroids, and that the wavelength dependence at a fixed phase angle is, for a given object, mostly linear in the visible domain; besides, changes in the properties of the dust within cometary comae are identified through imaging polarimetry. Promising interpretations of the observed variations of the polarization are inferred from experimental and numerical simulations with irregular compact grains and aggregates of dust particles. They already provide information on the structure, composition, and size distribution of dust within the zodiacal cloud and within comae that are monitored over a wide range of phase angles, as well as comparisons between the properties of asteroidal regoliths and those of some meteorites.

Keywords: dust, solar system, comets, asteroids, regolith, albedo, light scattering, linear polarization, optical probe, imaging polarimetry, polarimetric halo

* Corresponding author. E-mail: aclr@aerov.jussieu.fr

1. Introduction

1.1. Dust and small solar system bodies

Dust particles are all over the solar system: in cometary comae, tails, and trails; in planetary atmospheres; in the interplanetary dust cloud and meteoroids; as well as on surfaces of planets, satellites, asteroids and cometary nuclei. Learning about the properties of such media is of major importance for the inference of their formation and evolution processes.

Cometary nuclei, the structure and dust-to-ices ratio of which are still unknown (e.g., Kofman et al. 1998; Levasseur-Regourd et al. 2009), are likely to be built of ices and dust particles. *In situ* studies provide clear evidence for the presence of dust on nuclei surfaces, and within sub-surfaces, as established by the Deep Impact mission (A'Hearn et al. 2005). Dust particles ejected with gases from sublimating ices are found within cometary comae and tails, as already suggested by Arago (1858) through polarimetric observations of comets, and nowadays analyzed through light scattering and spectroscopic observations (e.g., Kolokolova et al. 2004; Bockelée-Morvan et al. 2004; Mishchenko et al. 2010). Finally larger dust particles are present along cometary trails, as monitored by near-infrared observations (e.g., Kelley et al. 2008).

Cometary dust contributes significantly to the replenishment of the zodiacal cloud (e.g., Nesvorný et al. 2010), at least below 1.5 AU. This lenticular circum-solar cloud of interplanetary dust scatters solar light, giving rise to the zodiacal light, visible from the Earth after sunset or before sunrise in the absence of any light pollution. Amongst the interplanetary dust particles collected in the Earth's stratosphere, so-called IDPs, aggregates (in the size range of a few tens of micrometers) of smaller grains are assumed to be of cometary origin (e.g., Levasseur-Regourd et al. 2001; Jessberger et al. 2001).

Dust particles resulting from successive impacts of meteoroids and micrometeoroids are also found on asteroidal surfaces and on surfaces of other atmosphereless bodies, such as planetary moons and trans-Neptunian objects. They build up a regolith which corresponds to layers of dust and rocks formed over billion years (e.g., Muinonen et al. 2002; Clark et al. 2002). Regoliths may be more or less thick, and have a coarse or fine-grained size distribution.

1.2. Properties revealed by *in situ* missions

Since the mid 1980s, six comets (1P/Halley, 26P/Grigg-Skjellerup, 19P/Borrelly, 81P/Wild 2, 9P/Tempel 1, and 103P/Hartley 2, all being periodic comets) and about ten asteroids have been explored by space probes. The most conspicuous result is possibly the vast diversity of these objects which are not at all spherical, as opposed to the legendary asteroids imagined in 1943 by Saint-Exupéry for Le Petit Prince, or even to the dirty cometary snowballs cleverly predicted in 1950 by Whipple. More specifically, amongst a wealth of scientific results, space missions to comets have helped to discover that cometary dust parti-

cles are rich in refractory organics, so-called CHON from their constitutive elements (Kissel et al. 1986, 2004), that their density is very low (Fulle et al. 2000), and that the dust properties may be quite heterogeneous within the coma (Clark et al. 2004). Meanwhile, space missions to asteroids have revealed that regolith layers cover their cratered surfaces. The regolith, as on (433) Eros, may present a fine-grain size distribution and behave like granular matter to form flat surfaces perpendicular to the local gravity field (Thomas et al. 2002) or, as on (2867) Steins, may be rock-dominated and rather coarse (Gulkis et al. 2010).

Such studies provide unique information on cometary and asteroidal dust particles. They can even provide a ground truth, with Stardust samples collected in 81P/Wild 2 coma (Brownlee et al. 2006) – although they may have suffered while impacting into aerogel cells, and, hopefully, with Hayabusa samples collected on the surface of (25143) Itokawa. Space missions are nevertheless limited to a few objects. Clues to the bulk properties of the dust thus come significantly from remote spectroscopic observations that provide information on the chemical composition as well as from remote light-scattering observations, including polarization measurements, that provide information on the bulk properties.

1.3. Motivation for polarization measurements

Solar light scattered by low-density particulate media is predominantly linearly polarized. The degree of linear polarization, thereafter called P , is the ratio of the difference to the sum of the intensity components, respectively, perpendicular and parallel to the scattering plane (defined by the Sun, the scattering medium, and the observer):

$$P = (I_{\perp} - I_{\parallel})/I = (I_{\perp} - I_{\parallel})/(I_{\perp} + I_{\parallel}).$$

It thus remains within the $[-1, +1]$ interval, with negative values corresponding to $I_{\perp} < I_{\parallel}$; it varies with the phase angle α (or the scattering angle $\theta = \pi - \alpha$) and the wavelength of observations λ , as well as with the properties of the scattering medium. Comparisons between data obtained for different distances to the Sun and to the observer and on different comets or asteroids are then possible.

Tentative interpretations in terms of physical properties (e.g., size distribution, morphology, albedo) of the observed variations of some polarization parameters, defined through the dependence of the polarization upon the phase angle and the wavelength, stem from experimental and numerical simulations with various dust particles. The derived results may be indicative of the physical processes that allowed the formation and evolution of the dust particles (e.g., accretion, agglomeration, ejection, fragmentation, sublimation, and collisions).

2. Back to the golden age of polarimetry of small bodies

Pioneering observations in the field of asteroidal properties were initiated by Lyot, who discovered the negative polarization of (1) Ceres and (4) Vesta back in 1934 (Dollfus et al. 1989). Measurements of the linear polarization of solar light

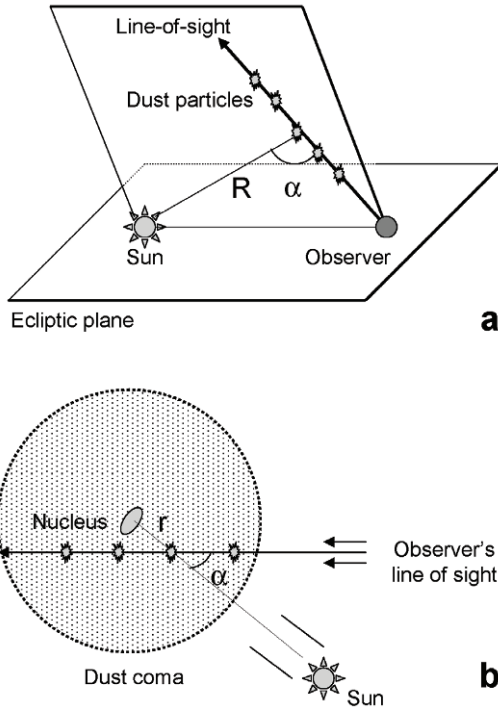


Fig. 1. Geometry of observations of scattered solar light (a) within the interplanetary dust cloud and (b) within a cometary coma. The data gathered along the line of sight correspond to different distances R to the Sun or r to the cometary nucleus, and thus possibly to different properties of the dust. For interplanetary dust observations, the phase angle α varies along the line of sight.

scattered on the surfaces of various asteroids at different phase angles and wavelengths (B, G, also U, V, R), published in the 1960s–1970s, have been used to draw curves providing the dependence of P upon α (e.g., Veverka and Liller 1969; Zellner et al. 1974; Zellner and Gradie 1976). Such curves, limited in phase angle coverage, except for near-Earth asteroids, are reminiscent of those previously obtained for the Moon (Lyot 1929).

Obtaining significant polarimetric measurements within the zodiacal cloud or a cometary coma is somehow more tricky. In both cases, the observed signals are integrated over the line of sight along which the intrinsic properties of the dust are likely to change with the distance to the Sun in the first case (Fig. 1a) or to the nucleus in the second case (Fig. 1b). Besides, for interplanetary dust observations, the phase angle varies drastically along the line-of-sight. However, assuming a given dust cloud to be steady within two measurements from a space probe, the difference between these polarized intensities is, for a line-of-sight parallel to the

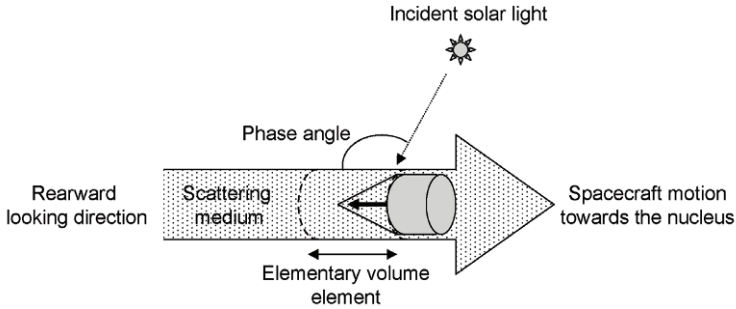


Fig. 2. The optical probe concept. The figure illustrates its application to the HOPE polarimetric instrument onboard the Giotto spacecraft within comet 1P/Halley coma.

trajectory of the moving probe, proportional to the light scattered by dust in a small elementary volume at the probe. This so-called optical probe concept was first applied for the moving Earth, in order to derive properties of the interplanetary dust at 1 AU from the Sun in the ecliptic (Dumont 1973; Schuerman 1979). A typical result is that one cubic centimeter of interplanetary space near the Earth orbit scatters sunlight at 90° phase angle with an intensity that is 4×10^{-34} times smaller than that of the Sun and a local polarization of about 0.3 (Levasseur-Regourd et al. 2001). Just to illustrate the fact that *in situ* solar system exploration is extremely demanding and sometimes disappointing, it may be added that out-of-ecliptic zodiacal measurements were anticipated to take place along the trajectory of the International Solar Polar mission, which was unfortunately cancelled in 1981.

The return of the famous 1P/Halley comet in 1985–1986 had triggered the development of cometary polarimetric observations, not only from various ground-based telescopes, but also *in situ*. The Halley Optical Probe Experiment (HOPE) that we had proposed was selected for measurements along the trajectory of the Giotto spacecraft through Halley's cometary coma (Fig. 2). Solar system exploration, as already mentioned, is always a major challenge, full of both programmatic and technical disappointments and unexpected discoveries. The positioning of HOPE on Giotto rearward-facing platform implied significant high stray light owing to reflections from the high-gain antenna and its tripod. However, this positioning meant that the instrument would survive dust impacts during the Halley encounter, allowing the gathering of new data during Giotto encounter with comet 26P/Grigg–Skjellerup in 1992. Although the geometry of this second flyby was not appropriate for a direct inversion of the data, the development of an updated science data pipeline allowed us to point out an error in the data reduction code used for the Halley encounter, which had resulted in an overestimation of the signal-to-noise ratio by a factor of 16. It was then estimated, from a comparison of light scattering and dust flux data (Levasseur-Regourd et al. 1999) that the dust within the comet Halley coma had a low albedo of about 0.04, a size distribution

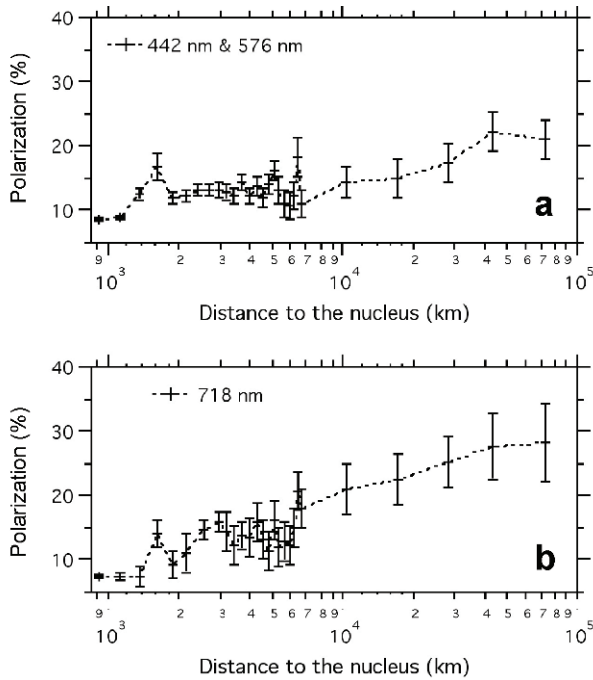


Fig. 3. Linear polarization of dust as a function of the distance to the Halley comet nucleus. *In situ* data were obtained from Giotto Optical Probe observations at 73° phase angle, at (a) 442 nm and 578 nm and (b) 718 nm. The error bars are smaller in the inner coma than in the outer coma and the polarization in red is higher than in blue–green, except in the innermost coma (from Levasseur-Regourd et al. 2005) The decrease in polarization in the innermost region corresponds to the polarimetric halo, while the increase in a narrow region might result from the crossing of a dust jet.

approximately following a power law with a -2.6 index, and an extremely low density, of about 100 kg m^{-3} (Fulle et al. 2000). Finally, the return of Halley, together with the development of CCD detectors, had triggered the development of imaging polarimetry, which has confirmed the presence of dust particles with different properties within various cometary comae (Eaton et al. 1988; Renard et al. 1992; Hadamcik et al. 2003a).

3. Main observational trends

Certain constraints apply to remote polarimetric observations of small solar system bodies, which suffer not only from random errors but also from systematic errors that may be caused by instrumental polarization and (de)polarization of the sky background. As already mentioned, measurements within the interplanetary dust cloud need to be inverted to derive results representative of a given phase angle and a given solar distance (Fig. 1a). Cometary measurements require narrow-

band filters that prevent depolarization from contaminating gaseous emissions. Besides, measurements on wide cometary comae cover different regions with different scattering properties (Fig. 1b). However, once spatial changes are taken into account, the main typical observational trends can be identified, from all the polarimetric results obtained by various teams, mostly on relatively bright objects, in the visible and near-infrared spectral domains.

3.1. Spatial changes within comae

Spatial changes in the polarimetric properties of a comet observed at a fixed phase angle and solar distance, which are clues to significant changes in the intrinsic properties of the scattering dust particles, were first noticed within the 1P/Halley coma, both from *in situ* measurements (Levasseur-Regourd et al. 1999) and from a synthesis of remote measurements through given apertures by several independent groups (Dollfus et al. 1988). Giotto measurements (Fig. 3) show that, in March 1986, the polarization varied along the trajectory, from 10^5 to 10^3 km nucleus distance, being lower in the innermost coma and larger in some narrow regions; also, it was higher at 718 nm than at 442 nm and 578 nm, except within the innermost coma, where it was higher in the blue-green than in the red (Levasseur-Regourd et al. 1999, 2005). Remote observations (Dollfus et al. 1988) indicate that, in 1985–1986, the polarization was, as compared to the outer coma, lower in the innermost coma and larger in the inner coma.

Since then, polarization maps of quite a few comets by CCD imaging polarimetry have pointed out three main regions in comae, as illustrated with comet C/1995 O1 (Hale–Bopp) in Fig. 4: (i) background coma; (ii) so-called polarimetric halo near the photometric center; and (iii) jet-like or arc-like features (e.g., Tozzi et al. 1997; Jones and Gehrz 2000; Hadamcik and Levasseur-Regourd 2003b; Hadamcik and Levasseur-Regourd 2009). The halo, which may be partially or totally hidden by jets, corresponds to a lower polarization whenever the polarization is positive, while the jets or arcs of fresh dust correspond to a higher polarization. The linear polarization reaches an asymptote with increasing aperture; it is thus possible, from polarimetric imaging with a synthetic aperture, to define the whole-coma polarization, which takes into account the actual size of the dust coma though polarimetric images.

3.2. Phase angle dependence

The whole-coma linear polarization dependence upon the phase angle, $P(\alpha)$, is monitored through the changing geometry for an Earth-based observer. Interpolations, typically by polynomial or trigonometric fits, whenever a sufficient amount of well-distributed data points is available, provide polarization phase curves. Such curves are quite smooth and typical of scattering by irregular particles with sizes greater than the wavelengths, i.e., a few microns. However, as usual in the absence of any robust theory, any extrapolation of the curves towards lower or larger phase angles may lead to inappropriate conclusions. These curves

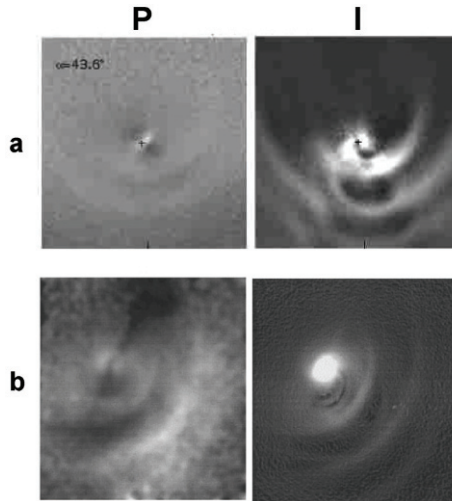


Fig. 4. Polarization and intensity (with highlighted gradients) images of comet C/1995 O1 (Hale–Bopp), as measured (a) at $0.68\ \mu\text{m}$ on 9 April 1997 by Hadamcik and Levasseur-Regourd (2003b) and (b) at $2.2\ \mu\text{m}$ on 16 April 1997 by Jones and Gehrz (2000). The fields of view are of about 150 000 km and 67 000 km, respectively. The dark areas on polarization images correspond to about 8% and the bright ones to at least 12% (the dark diagonal streak in the infrared image being an artifact). Polarization points out intrinsic differences between the properties of the dust particles inside the curved jets and those in the background coma.

exhibit a shallow negative branch near the backscattering region and, after the inversion region, a wide positive branch with a maximum in the 90° – 100° range. Quite interesting features leading to classifications may be noticed, once the data points relative to a given object are separated in different wavelengths bins.

For cometary dust, the minimum in polarization is about -0.02 , the inversion angle is mostly in the 21° – 22° phase-angle range, and a significant dispersion is noticed for α greater than $\sim 30^\circ$. Data retrieved for tens of comets within a given wavelength range actually suggest the existence of at least two classes of comets (Levasseur-Regourd et al. 1996) corresponding to different properties of the ejected dust particles: comets with a low polarization maximum (in the 0.10 to 0.15 range, depending on the wavelength), comets with a high polarization maximum (in the 0.25 to 0.30 range), and comet C/1999 O1 (Hale–Bopp), whose polarization is the highest ever observed (Levasseur-Regourd 1999; Hadamcik and Levasseur-Regourd 2003b). The maximum in polarization is high whenever a silicate emission feature is detected in the $11\text{-}\mu\text{m}$ region (Hanner 2002). It may increase after the release of fresh dust particles from inside the nucleus. Such a behavior has typically been noticed after the fragmentation of C/1999 S4 (LINEAR) and within the dust plume released from the sub-surface of 9P/Tempel 1 after Deep Impact event (e.g., Furusho et al. 2007; Hadamcik et al. 2007a).

For asteroidal surfaces, different trends, which somehow correspond to different taxonomic classes, have also been noticed in the polarization phase curves (Goidet-Devel et al. 1995; Pentillä et al. 2005), with higher slopes at inversion corresponding to asteroids with lower geometric albedos. The slope at inversion, as well as the absolute value of the minimum in polarization, increases when the geometric albedo decreases. Asteroids belonging to the C-type actually exhibit a deeper negative branch, about -0.015 , than those belonging to the S- and M-types, for which the minimum is rather about -0.01 (Muinonen et al. 2002). Relations linking the slope at inversion or the minimum in polarization to the albedo have been derived from laboratory measurements on various samples and thermal radiometry (Zellner and Gradie 1976; Dollfus and Zellner 1979; Lupishko and Mohamed 1996; Cellino et al. 1999, 2005a). Although these relations may be called empirical laws, they need to be considered carefully, with certainly the same cautiousness as the above-mentioned extrapolations, since they do not rely on any physical theory and since the coefficients have mostly been estimated for observations through V-filters of moderately bright S-type asteroids. A noticeable example is provided by unique polarimetric observations of the bare nucleus of comet 2P/Encke (Boehnhardt et al. 2008), which could be misinterpreted as a clue to an unrealistic albedo of about 0.15.

Other peculiar features have been noticed for asteroids. A polarimetric opposition effect is clearly found for some E-type objects, e.g., (44) Nysa and (64) Angelina (Mishchenko et al. 2010). Besides, the inversion angles present a large dispersion. While they remain in the 20° – 23° range for S-, C-, and M-type asteroids, they are about 14° for F-type asteroid (419) Aurelia and 18° for (2867) Steins and other E-type asteroids (Belskaya et al. 2005; Fornasier et al. 2006); on the other hand, they reach 25° for (21) Lutetia and even about 30° for objects such as (234) Barbara (387) and (387) Aquitania, appropriately and humorously called Barbarians (Cellino et al. 2006; Masiero and Cellino 2009).

For the zodiacal light, the local polarization in the vicinity of the ecliptic plane and at 1.5 AU from the Sun has been derived using mathematical inversion techniques, at least up to 90° phase angle (Levasseur-Regourd et al. 1990). The polarimetric phase curve is smooth, with a shallow negative branch, an inversion at $15^{\circ} \pm 5^{\circ}$, and a slope at inversion comparable to that of cometary dust and C-type asteroids. Although mathematical inversion techniques may imply large error bars and data are mostly obtained at 550 nm, the polarization at 90° (again in the vicinity of the ecliptic) can also be derived in the 0.1 to 1.5 AU range (Levasseur-Regourd et al. 2001). It decreases with decreasing solar distance, i.e., with increasing temperature and time as the dust particles spiral towards the Sun under Poynting–Robertson effect.

3.3. Wavelength dependence

The dependence of the degree of linear polarization upon the wavelength, $P_{\alpha}(\lambda)$, may also be monitored for a given phase angle, as soon as measurements

are available for a few wavelengths. An interesting feature is that the wavelength dependence seems to be quite linear, at least in the visible domain.

For comets, the polarization on the positive branch usually linearly increases with the wavelength (Levasseur-Regourd and Hadamcik 2003), up to a maximum near $1.5 \mu\text{m}$ for some extensively observed comets (Hadamcik and Levasseur-Regourd 2003b). Some exceptions may nevertheless be noticed in the innermost coma of, e.g., 1P/Halley (see Fig. 3), as well as during disruption events of, e.g., C/1999 S4 (LINEAR) and 73P/Schwassmann–Wachmann 3 (Kiselev et al. 2008). They reveal drastic changes between the properties of freshly ejected dust and of dust that may have already experienced some evaporation and/or fragmentation within the inner coma.

As far as asteroids are concerned, two specific behaviors are worth noting: (i) the polarization at a given phase angle depends fairly linearly upon the wavelength (Belskaya et al. 2009); and (ii) the polarization on the positive branch seems to either decrease or increase with wavelength, depending on the taxonomic type. The polarization of the positive branch, well documented for near-Earth asteroids, decreases linearly with increasing wavelength for S-type objects (Levasseur-Regourd and Hadamcik 2003), as derived from observations of, e.g., (4179) Toutatis and (25143) Itokawa (Mukai et al. 1997; Cellino et al. 2005b). On the contrary, the polarization of the positive branch seems to increase with increasing wavelength for C-type asteroids, as well as for (21) Lutetia (Hadamcik et al. 2011).

4. Interpretation of polarimetric data

Tentative interpretation in terms of physical properties of the above-mentioned variations of the polarization parameters (e.g., the values of the minimum in polarization, of the inversion angle, of the slope at inversion, of the maximum in polarization), as well as the quasi-linear dependence upon the wavelength of the polarization at a given phase angle, should stem from both experimental and numerical simulations with real and virtual dust particles, the morphology, size and complex refractive index of which somehow agrees with our present understanding of the corresponding scattering medium. It would indeed be of major importance to unequivocally interpret the origin of cometary polarimetric halos (as attempted by Zubko et al. 2010), to disentangle from higher polarization observed in some cometary jets the effect of a possible alignment of elongated particles, and to systematically compare the polarimetric properties of asteroidal surfaces to those of meteorites.

4.1. Experimental simulations

Laboratory measurements have been initiated in the 1970s to infer properties of some regolith layers. Since then, various teams have developed laboratory measurements of the light scattered by particulate surfaces or by dust (e.g., Dollfus and Zellner 1979; Shkuratov et al. 2002), including measurements in jet streams

and steady-state gas flows, which provide the whole Mueller matrix for particles below 1 μm (Hovenier and Muñoz 2009).

We have, since the mid-1990s, developed the PROGRA² series of experiments to infer, through measurements made in the laboratory and/or under microgravity conditions, the properties of solar system dust clouds (cometary comae, interplanetary dust, planetary atmospheres) and dusty surfaces of atmosphereless bodies (Levasseur-Regourd et al. 1997; Renard et al. 2002; Hadamcik et al. 2009). For a wide variety of samples (<http://www.icare.uni-lille.fr/progra2>), the intensity and linear polarization are monitored over a large range of phase angles for at least two wavelengths (543.5 and 632.8 nm), the microgravity conditions during parabolic flight campaigns achieving conditions close to those prevailing in dust clouds and on surfaces of small bodies, while avoiding sedimentation and mass or morphology segregation.

As far as comets are concerned, excellent matches have typically been obtained with porous aggregates of sub-micrometer ($\text{MgSiO} + \text{FeSiO} + \text{C}$) grains and compact Mg-silicates (Hadamcik et al. 2007b). As far as asteroidal surfaces are concerned, remarkable agreements have been obtained, for the phase angle and the wavelength dependences, between polarimetric observations and measurements, on the one hand for (2867) Steins (or other E-type asteroids) and powdered samples of an aubrite meteorite, on the other hand for (21) Lutetia and powdered samples of Allende meteorite (Levasseur-Regourd et al. 2011; Hadamcik et al. 2011).

4.2. Numerical simulations

Numerical simulations have first been made with the use of the Mie theory, in order to derive complex refractive indices and size distributions (e.g., Mukai et al. 1987). Although computations for homogeneous spherical particles hardly represent the scattering behavior of irregular particles with a size greater than the wavelength of observations, they have nevertheless allowed certain comparisons between different scattering objects. More realistic numerical simulations have subsequently been performed, thanks to the progress in computational codes (DDA, *T*-matrix, ray-tracing), for inhomogeneous non-spherical particles, and especially for aggregates of grains that might build up cometary and interplanetary dust particles (e.g., Petrova et al. 2004; Kimura et al. 2006; Lasue et al. 2006).

We have developed a versatile model, with a distribution of particles, consisting of a mixture of spheroidal grains and aggregates of small spheroids (or of spheres with a size distribution), with bimodal complex refractive indices, typical of astronomical silicates and more absorbing organics (Lasue et al. 2006). From data corresponding to a comet extensively observed, Hale–Bopp as an example, a fit in two colours indicates that both compact and fluffy particles are required, with a size of the grains in the 0.1–20 μm range and a power law of their size distribution of about -3 , and that there are 40% to 65% of silicates in mass and 60% to 35% of organics. Besides, computations with the same values in other colors fit very well the observational data (Levasseur-Regourd et al. 2007; Lasue et al.

2009). The relevance of the results of such simulations, which agree with those of experimental simulations, is demonstrated by Stardust ground truth, with dust impacts on aluminium foils and dust tracks in aerogel cells demonstrating the existence of both compact grains and loosely-bound aggregates of tiny grains in a cometary coma (Hörz et al. 2006). Finally, the same approach used for interplanetary dust particles indicates that their equivalent radius is in the 0.1–100 μm range and there are 50% to 75% of silicates in mass and 50% to 25% in organics; it also strongly suggests that the decrease in polarization with decreasing solar distance originates in a thermal decay of organic compounds (Lasue et al. 2007) and that the contribution of comets to the interplanetary dust cloud is most significant, at least below 1.5 AU.

5. Polarimetric properties of some objects of special interest

Progress in our understanding of small bodies, including through polarimetry, has been triggered by special events, such as cometary outbursts or fragmentations, and by the preparation and analysis of space missions. Detailed below are a couple of points of present interest.

5.1. 17P/Holmes outburst

The unexpected outburst of comet 17P/Holmes, which increased its brightness by a factor of about 700 000 in October 2007, was monitored from many instruments, including polarimeters. Observations from a couple of teams, at phase angles in the 11° to 16° range, suggest that the near-minimum polarization is less negative than usually observed for other comets and that it strongly depends upon the color, being significantly less negative with increasing wavelength (Rosenbush et al. 2009; Zubko et al. 2011). Such a behavior, although far from being well understood, is compatible with highly adsorbing debris particles, and might be a clue to the existence of a big dust crust on the surface of the nucleus.

5.2. Preparation of Rosetta rendezvous with 67P/Churyumov–Gerasimenko

Rosetta spacecraft mission has already flown by two main belt asteroids, quite unique as compared to those that had been previously explored, (2867) Steins and (21) Lutetia. As mentioned in Section 4.1, remote polarimetric observations have been compared with laboratory measurements on meteorites, the parent bodies of which could be fragments of these two asteroids or of other objects of the same type.

The main objective of the Rosetta mission is to rendezvous comet 67P/Churyumov–Gerasimenko in 2014–2015 and tentatively release a module on its nucleus (Glassmeier et al. 2007). In 2008–2009, that is to say for the last return of the comet before the rendezvous, polarimetric observations have been made for solar distances in the 1.7 AU to 2 AU range and phase angles in the 29°

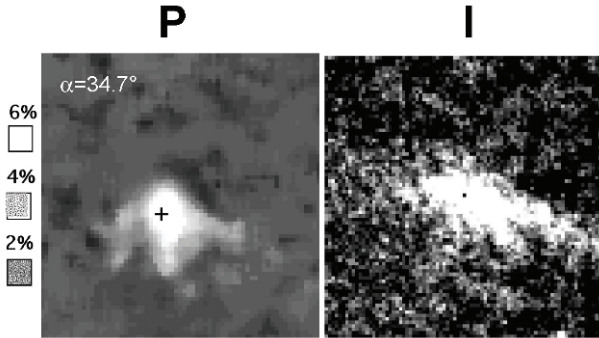


Fig. 5. Polarization and intensity (with highlighted gradients) images of comet 67P/Churyumov–Gerasimenko, the target of the ESA Rosetta rendezvous mission. Field-of-view is of about 80000 km. Measurements were obtained in the red domain in March 2009, 18 days after perihelion, by Hadamcik et al. (2010b). The polarization image gives clue to the presence of the emission of particles with different properties in the coma, possibly fluffy aggregates of micron-sized grains.

to 36° range, as illustrated in Fig. 5. The purpose is to contribute to the mission safety and the choice of optimal sequences for the on-board dust experiments, MIDAS and GIADA. Polarimetric results confirm that the comet exhibits a brightness asymmetry and is more active after its perihelion; they suggest that, before perihelion, the dust is dominated by large slowly moving particles and that, after perihelion, a new material with different properties, possibly aggregates of tiny grains, is released into the jets (Hadamcik et al. 2010). It is nevertheless mandatory to stress the fact that the properties of the dust particles in the very innermost coma, below 1000 km nucleus distance, are quite impossible to detect from remote observations. Only the rendezvous, allowing to trace the activity of the comet at solar distances from about 3 AU to 1.25 AU, will reveal such properties, and perhaps provide images even more amazing than those of the “cosmic snow storm” revealed in the innermost coma of comet 103P/Hartley 2 in November 2010 by the successful EPOXI mission (<http://epoxi.umd.edu>).

6. Summary and future work

Observations of the linear polarisation of solar light scattered by small bodies present significant constraints, typical of polarimetric techniques, as well as specific of the objects that are studied. However, such observations, over a wide range of phase angles and tentatively at a few different wavelengths, provide interesting comparisons, giving evidence for changing dust properties within an object and for classes of objects with similar dust properties. It is better, while trying to interpret such observations, to avoid empirical laws and disputable extrapolations.

Sophisticated programs of numerical simulations have been developed; various advanced programs of laboratory simulations (including under microgravity) have also been developed for large series of relevant samples. They tentatively rely on hypotheses that are as close as possible to what is already known about the properties of the scattering medium. They typically indicate that cometary dust particles are built of both very fluffy aggregates and of more compact grains, with significant proportions of rather transparent silicates and of absorbing materials, and that a significant proportion of interplanetary dust is of cometary origin. They are also used to suggest some links between asteroidal surfaces and specific meteorites.

As usual whenever observations of complex media are concerned, more polarimetric observations (again over a large range of phase angles and at several wavelengths) are needed, with emphasis on (i) the monitoring of new Oort cloud comets, asteroids of relatively rare or unknown types, and the zodiacal light; (ii) the preparation of future space missions to small bodies; and hopefully (iii) the revival of polarimetric instruments on board spacecraft devoted to solar system exploration. While the development of numerical and experimental simulations is likely to continue, future breakthroughs might be expected from successful numerical simulations on layers of dust particles and from experimental simulations of the aggregation of tiny dust particles, with continuous monitoring of the polarization (or, even better, of their Stokes parameters), as anticipated to take place on board the Columbus module of the International Space Station, not to mention the resolution of some open problems, e.g., the precise dependence(s) of the linear polarization upon the wavelength or upon the albedo.

Acknowledgements

It is a real pleasure to thank colleagues who have most directly and efficiently contributed to the development of these studies over the past years, e.g., E. Hadamcik (UPMC-LATMOS), J. Lasue (LANM), and J. B. Renard (LPCEE). Numerous fruitful discussions with P. Bastien, I. Belskaya, M. Cabane, A. Cellino, A. Dollfus, R. Dumont, M. Fulle, B. Goidet-Devel, M. Hanner, L. McFadden, K. Muinonen, T. Mukai, A. Penttillä, F. Rietmeijer, A. K. Sen, J. C. Worms, and many other colleagues, especially during the NATO ASI in September 2010, are acknowledged. This research was partly supported by Centre national d'études spatiales (CNES), Programme national de planétologie (PNP), and European Space Agency (ESA).

References

- A'Hearn, M. F., M. J. S. Belton, W. A. Delamere, et al., 2005: Deep Impact: excavating Comet Tempel 1. *Science* **310**, 258–264.
- Arago, F., 1858: *Astronomie populaire*, Livre XVII (Gide, Paris).
- Belskaya, I. N., Yu. G. Shkuratov, Yu. S. Efimov, et al., 2005: The F-type asteroids with small inversion angles of polarization. *Icarus* **178**, 213–221.

- Belskaya, I. N., A.-C. Lvasseur-Regourd, A. Cellino, et al., 2009: Polarimetry of main belt asteroids: wavelength dependence. *Icarus* **199**, 97–105.
- Bockelée-Morvan, D., J. Crovisier, M. J. Mumma, and H. A. Weaver, 2004: The composition of cometary volatiles. In M. Festou, H. U. Keller, and H. A. Weaver, Eds., *Comets II* (University of Arizona Press, Tucson, AZ), pp. 391–423.
- Boehnhardt, H., G. P. Tozzi, S. Bagnulo, et al., 2008: Photometry and polarimetry of the nucleus of comet 2P/Encke. *Astron. Astrophys.* **489**, 1337–1343.
- Brownlee, D., P. Tsou, J. Aléon, et al., 2006: Comet 81P/Wild 2 under a microscope. *Science* **314**, 1711–1716.
- Cellino, A., R. Gil-Hutton, E. F. Tedesco, et al., 1999: Polarimetric observations of small asteroids: preliminary results. *Icarus* **138**, 129–140.
- Cellino, A., R. G. Hutton, M. Di Martino, et al., 2005a: Asteroid polarimetric observations using the Torino UBVRi photopolarimeter. *Icarus* **179**, 304–324.
- Cellino, A., F. Yoshida, E. Anderlucci, et al., 2005b: A polarimetric study of asteroid 25143 Itokawa. *Icarus* **179**, 297–303.
- Cellino, A., I. N. Belskaya, P. Bendjoya, et al., 2006: The strange polarimetric behavior of asteroid (234) Barbara. *Icarus* **180**, 565–567.
- Clark, B. E., B. Hapke, C. Pieters, and D. Britt, 2002: Asteroid space weathering and regolith evolution. In W. Botke, A. Cellino, P. Paolocchi, and R. P. Binzel, Eds., *Asteroids III* (University of Arizona Press, Tucson, AZ), pp. 585–599.
- Clark, B. C., S. F. Green, T. E. Economou, et al., 2004: Release and fragmentation of aggregates to produce heterogeneous lumpy coma streams. *J. Geophys. Res.* **109**, E12S03.
- Dollfus, A., and B. Zellner, 1979: Optical polarimetry of asteroids and laboratory samples. In T. Gehrels, Ed., *Asteroids* (University Arizona Press, Tucson, AZ), pp. 170–183.
- Dollfus, A., P. Bastien, J. L. Le Borgne, et al., 1988: Optical polarimetry of P/Halley: synthesis of the measurements in the continuum. *Astron. Astrophys.* **206**, 348–356.
- Dollfus, A., M. Wolff, J. E. Geake, et al., 1989: Photopolarimetry of asteroids. In P. Binzel and T. Gehrels, Eds., *Asteroids II* (University of Arizona Press, Tucson, AZ), 594–616.
- Dumont, R., 1973: Phase function and polarization of interplanetary scatters from zodiacal light photopolarimetry. *Planet. Space Sci.* **21**, 2149–2155.
- Eaton, N., S. M. Scarrott, and R. F. Warren-Smith, 1988: Polarization images of the inner coma of comet Halley. *Icarus* **76**, 270–278.
- Fornasier, S., I. Belskaya, M. Fulchignoni, et al., 2006: First albedo determination of 2867 Steins, target of the Rosetta mission. *Astron. Astrophys.* **449**, L9–L12.
- Fulle, M., A. C. Lvasseur-Regourd, N. McBride, and E. Hadamcik, 2000: *In situ* dust measurements from within the coma of 1P/Halley: first order approximation with a dust dynamical model. *Astron. J.* **119**, 1968–1977.
- Furusho, R., Y. Ikeda, and D. Kinoshita, et al., 2007: Imaging polarimetry of comet 9P/Tempel 1 before and after the Deep Impact. *Icarus* **190**, 454–458.
- Glassmeier, K. H., H. Boehnhardt, D. Koschny, et al., 2007: The Rosetta mission: flying towards the origin of the solar system. *Space Sci. Rev.* **128**, 1–21.
- Goidet-Devel, B., J.-B. Renard, and A. C. Lvasseur-Regourd, 1995: Polarization of asteroids: synthetic curves and characteristic parameters. *Planet. Space Sci.* **43**, 779–786.
- Gulkis, S., S. Keihm, L. Kamp, et al., 2010: Millimeter and submillimeter measurements of asteroid (2867) Steins during the Rosetta fly-by. *Planet. Space Sci.* **58**, 1077–1087.
- Hadamcik, E., and A. C. Lvasseur-Regourd, 2003a: Imaging polarimetry of cometary dust: different comets and phase angles. *J. Quant. Spectrosc. Radiat. Transfer* **79**, 661–678.

- Hadamcik, E., and A. C. Lvasseur-Regourd, 2003b: Dust evolution of comet C/1995 O1 (Hale–Bopp) by imaging polarimetric observations. *Astron. Astrophys.* **403**, 757–768.
- Hadamcik, E., and A. C. Lvasseur-Regourd, 2009: Optical properties of dust from Jupiter Family comets. *Planet. Space Sci.* **57**, 1118–1132.
- Hadamcik, E., A. C. Lvasseur-Regourd, V. Leroi, and D. Bardin, 2007a: Imaging polarimetry of the dust coma of comet Tempel 1 before and after Deep Impact. *Icarus* **190**, 459–468.
- Hadamcik, E., J.-B. Renard, F. J. M. Rietmeijer, et al., 2007b: Light scattering by fluffy Mg-Fe-SiO and C mixtures as cometary analogs (PROGRA² experiment). *Icarus* **190**, 660–671.
- Hadamcik, E., J.-B. Renard, A. C. Lvasseur-Regourd, and J. C. Worms, 2009: Laboratory measurements of the light scattered by clouds of solid particles by an imaging technique. *Light Scattering Rev.* **4**, 31–70.
- Hadamcik, E, A. K. Sen, A. C. Lvasseur-Regourd, et al., 2010: Polarimetric observations of comet 67P/Churyumov–Gerasimenko during its 2008–2009 apparition. *Astron. Astrophys.* **517**, A86.
- Hadamcik, E., A.C. Lvasseur-Regourd, J.-B. Renard, et al., 2011: Polarimetric observations and laboratory simulations of asteroidal surfaces: the case of 21 Lutetia. *J. Quant. Spectrosc. Radiat. Transfer* **112**, in press.
- Hanner, M. S., 2002: Comet dust: the view after Hale–Bopp. In S. F. Green, I. P. Williams, J. A. M. McDonnell, and N. McBride, Eds., *Dust in the Solar System and other Planetary Systems* (Elsevier, Amsterdam), pp. 239–254.
- Hörz, F., R. Bastien, J. Borg, et al., 2006: Impact features on Stardust: implications for comet 81P/Wild 2 dust. *Science* **314**, 1716–1719.
- Hovenier, J. W., and O. Muñoz, 2009: Light scattering in the solar system: an introductory review. *J. Quant. Spectrosc. Radiat. Transfer* **110**, 1280–1292.
- Jessberger, E. K., T. Stephan, D. Rost, et al., 2001: Properties of interplanetary dust: information from collected samples. In E. Grün, B. Å. S. Gustafson, S. F. Dermott, and H. Fechtig, Eds., *Interplanetary Dust* (Springer, Berlin), pp. 253–294.
- Jones, T. J., and R. D. Gehrz, 2000: Imaging polarimetry of comet C/1995 O1 (Hale–Bopp). *Icarus* **143**, 338–346.
- Kelley, M. S., W. T. Reach, and D. J. Lien, 2008: The dust trail of comet 67P/Churyumov–Gerasimenko. *Icarus* **193**, 572–587.
- Kimura, H., L. Kolokolova, and I. Mann, 2006: Light scattering by cometary dust numerically simulated with aggregate particles consisting of identical spheres. *Astron. Astrophys.* **449**, 1243–1254.
- Kiselev, N., V. Rosenbush, L. Kolokolova, and K. Antonyuk, 2008: The anomalous spectral dependence of polarization in comets. *J. Quant. Spectrosc. Radiat. Transfer* **109**, 1384–1391.
- Kissel, J., R. Z. Sagdeev, J. L. Bertaux, et al., 1986: Composition of comet Halley dust particles from Vega observations. *Nature* **321**, 280–282.
- Kissel, J., F. R. Krueger, J. Silén, and B. C. Clark, 2004: The cometary and interstellar dust analyzer at comet 81P/Wild 2. *Science* **304**, 1774–1776.
- Kofman, W., Y. Barbin, J. Klinger, et al., 1998: Comet nucleus sounding experiment by radiowave transmission. *Adv. Space Res.* **21**, 1589–1598.
- Kolokolova L., M. S. Hanner, A. C. Lvasseur-Regourd, and B. Å. S. Gustafson, 2004: Physical properties of cometary dust obtained using their light scattering and emission. In M. Festou, H. U. Keller, and H. A. Weaver, Eds., *Comets II* (University of Arizona Press, Tucson, AZ), pp. 577–604.

- Lasue, J., and A. C. Lvasseur-Regourd, 2006: Porous aggregates of irregular sub-micron sized grains to reproduce cometary dust light scattering observations. *J. Quant. Spectrosc. Radiat. Transfer* **100**, 220–236.
- Lasue, J., A. C. Lvasseur-Regourd, N. Fray, and H. Cottin, 2007: Inferring the interplanetary dust properties from remote observations and simulations. *Astron. Astrophys.* **473**, 641–649.
- Lasue, J., A. C. Lvasseur-Regourd, E. Hadamcik, and G. Alcouffe, 2009: Cometary dust properties retrieved from polarization observations: application to C/1995 O1 Hale–Bopp and 1P/Halley. *Icarus* **199**, 129–144.
- Lvasseur-Regourd, A.-C., 1999: Polarization of light scattered by cometary dust particles: observations and tentative interpretations. *Space Sci. Rev.* **90**, 163–168.
- Lvasseur-Regourd, A. C., and E. Hadamcik, 2003: Light scattering by irregular dust particles in the solar system: observations and interpretation by laboratory measurements. *J. Quant. Spectrosc. Radiat. Transfer* **79**, 903–910.
- Lvasseur-Regourd, A. C., R. Dumont, and J.-B. Renard, 1990: A comparison between polarimetric properties of cometary dust and interplanetary dust particles. *Icarus* **86**, 264–272.
- Lvasseur-Regourd, A. C., E. Hadamcik, and J.-B. Renard, 1996: Evidence for two classes of comets from their polarimetric properties at large phase angles. *Astron. Astrophys.* **313**, 327–333.
- Lvasseur-Regourd, A. C., M. Cabane, J. C. Worms, and V. Haudebourg, 1997: Physical properties of dust in the solar system: relevance of the computational approach and of measurements under microgravity conditions. *Adv. Space Res.* **20**, 1585–1594.
- Lvasseur-Regourd A. C., N. McBride, E. Hadamcik, and M. Fulle, 1999: Similarities between *in situ* measurements of local dust scattering and dust flux impact data within the coma of 1P/Halley. *Astron. Astrophys.* **348**, 636–641.
- Lvasseur-Regourd, A. C., I. Mann, R. Dumont, and M. S. Hanner, 2001: Optical and thermal properties of interplanetary dust. In E. Grün, B. Å. S. Gustafson, S. F. Dermott, and H. Fechtig, Eds., *Interplanetary Dust* (Springer, Berlin), pp. 57–94.
- Lvasseur-Regourd, A. C., E. Hadamcik, and J. Lasue, 2005: Light scattering as a clue to cometary dust structure. *Highlights Astron.* **13**, 498–500.
- Lvasseur-Regourd, A. C., T. Mukai, J. Lasue, and Y. Okada, 2007: Physical properties of cometary and interplanetary dust. *Planet. Space Sci.* **55**, 1010–1020.
- Lvasseur-Regourd, A. C., E. Hadamcik, E. Desvoivres, and J. Lasue, 2009: Probing the internal structure of the nuclei of comets. *Planet. Space Sci.* **57**, 221–228.
- Lvasseur-Regourd, A. C., E. Hadamcik, J.-B. Renard, and L. A. McFadden, 2011: Clues to links between aubrites and (2867) Steins, from comparisons between polarimetric laboratory measurements on meteoritic samples and observations. *Planet. Space Sci.*, in press.
- Lupishko, D. F., and R. A. Mohamed, 1996: A new calibration of the polarimetric albedo scale of asteroids. *Icarus* **119**, 209–213.
- Lytot, B., 1929: Recherches sur la polarisation de la lumière des planètes et de quelques substances terrestres. *Ann. Obs. Meudon* **8**, 1–161.
- Masiero, J., and A. Cellino, 2009: Polarization of asteroid (387) Aquitania: the newest member of a class of large inversion angle asteroids. *Icarus* **199**, 333–337.
- Mishchenko, M. I., V. K. Rosenbush, N. N. Kiselev, et al., 2010: *Polarimetric Remote Sensing of Solar System Objects* (Akadempriodyka, Kyiv) (arXiv:1010.1171).

- Muironen, K., J. Piironen, Yu. G. Shkuratov, et al., 2002: Asteroid photometric and polarimetric phase effects. In W. Botke, A. Cellino, P. Paolocci, and R. P. Binzel, Eds., *Asteroids III* (University of Arizona Press, Tucson, AZ), pp. 123–136.
- Mukai, T., S. Mukai, and S. Kikuchi, 1987: Complex refractive index of grain material deduced from the visible polarimetry of comet P/Halley. *Astron. Astrophys.* **187**, 650–652.
- Mukai, T., T. Iwata, S. Kikuchi, et al., 1997: Polarimetric observations of 4179 Toutatis in 1992/1993. *Icarus* **127**, 452–460.
- Nesvorny, D., P. Jenniskens, H. F. Levison, et al., 2010: Cometary origin of the zodiacal cloud and carbonaceous meteorites, implication for hot debris disk. *Astrophys. J.* **713**, 816–836.
- Penttilä, A., K. Lumme, E. Hadamcik, and A. C. Lvasseur-Regourd, 2005: Statistical analysis of asteroidal and cometary phase curves. *Astron. Astrophys.* **432**, 1081–1090.
- Petrova, E. V., V. P. Tishkovets, and K. Jockers, 2004: Polarization of light scattered by solar system bodies and the aggregate model of dust particles. *Solar Syst. Res.* **38**, 309–324.
- Renard, J.-B., A. C. Lvasseur-Regourd, and A. Dollfus, 1992: Polarimetric CCD imaging of comet Levy. *Ann. Geophys.* **10**, 288–292.
- Renard, J.-B., J. C. Worms, T. Lemaire, et al., 2002: Light scattering by dust particles in microgravity: polarization and brightness imaging with the new version of the PROGRA² instrument. *Appl. Opt.* **41**, 609–618.
- Rosenbush, V., N. Kiselev, L. Kolokolova, et al., 2009: Polarization properties of odd comet 17P/Holmes. *J. Quant. Spectrosc. Radiat. Transfer* **110**, 1719–1725.
- Saint-Exupéry, A. de, 1943: *Le Petit Prince* (Reynal and Hitchcock, New York).
- Schuerman, D. W., 1979: Inverting the zodiacal brightness integral. *Planet. Space Sci.* **27**, 551–556.
- Shkuratov, Yu., A. Ovcharenko, and E. Zubko, 2002: The opposition effect and negative polarization of structural analogs for planetary regoliths. *Icarus* **159**, 396–416.
- Thomas, P. C., J. Joseph, B. Carcich, et al., 2002: Eros: shape, topography, and slope processes. *Icarus* **155**, 18–37.
- Tozzi, G. P., A. Cimatti, S. di Serego Alighieri, and A. Cellino, 1997: Imaging polarimetry of comet C/1996 B2 (Hyakutake) at the perigee. *Planet. Space Sci.* **45**, 535–540.
- Veverka, J., and W. Liller, 1969: Observations of Icarus: 1968. *Icarus* **10**, 441–444.
- Whipple, F. L., 1950: A comet model. The acceleration of comet Encke. *Astrophys. J.* **111**, 375–394.
- Zellner, B., and J. Gradie, 1976: Minor planets and related objects. XX. Polarimetric evidence for the albedos and composition of 94 asteroids. *Astron. J.* **81**, 262–280.
- Zellner, B., T. Gehrels, and J. Gradie, 1974: Minor planets and related objects. XVI. Polarimetric diameters. *Astron. J.* **79**, 1100–1110.
- Zubko, E., R. Furusho, K. S. Kawabata, et al., 2010: Interpretation of spectro-polarimetry of comet 17P/Holmes during outburst in 2007. In K. Muironen, A. Penttilä, H. Lindqvist, et al., Eds., *Electromagnetic and Light Scattering XII: Conference Proceedings* (University of Helsinki, Helsinki), pp. 337–341.
- Zubko, E., G. Videen, Yu. Shkuratov, et al., 2011: The Umov effect for single irregularly shaped particles with sizes comparable with wavelength. *Icarus* **212**, 403–415.

Semi-empirical BRDF and BPDF models applied to the problem of aerosol retrievals over land: testing on airborne data and implications for modeling of top-of-atmosphere measurements

Pavel Litvinov^{1*}, Otto Hasekamp¹, Brian Cairns², and Michael Mishchenko²

¹SRON Netherlands Institute for Space Research, Sorbonnelaan 2, 3584 CA Utrecht, The Netherlands

²NASA Goddard Institute for Space Studies, 2880 Broadway, New York, NY 10025, USA

Abstract. For the retrieval of atmospheric aerosol properties from satellite measurements, the atmospheric signal should be correctly separated from the surface signal. This represents one of the most important challenges in the development of algorithms for the retrieval of aerosol properties over land surfaces. Intrinsic reflectance properties of surfaces can be described by the Bidirectional Reflectance and Polarization Distribution Functions (BRDF and BPDF). In this chapter, we investigate the performance of different semi-empirical BRDF and BPDF models as they relate to the requirements for aerosol retrievals over land. First, we test BRDF and BPDF for bare soil and vegetation surfaces using multi-angle, multi-spectral photopolarimetric airborne measurements of the Research Scanning Polarimeter. Then, we investigate the capability of the different models to represent top-of-atmosphere measurements.

Keywords: surface reflectance, bidirectional reflection matrix, bidirectional reflectance and polarization distribution functions, land surfaces, aerosol retrieval, Research Scanning Polarimeter

1. Introduction

The land surface of the Earth is an important component of the climate system. Its interactions with incoming solar radiation and the atmosphere have a substantial impact on the Earth's energy budget. To account for such interactions, the radiative properties of both land surfaces and the atmosphere must be robustly understood on a global scale.

The intrinsic reflectance properties of surfaces can be described by the bidirectional reflectance distribution function (BRDF) and the bidirectional polariza-

* Corresponding author. E-mail: PVLitvinov@mail.ru

tion distribution function (BPDF). They provide a relation between the Stokes parameters of scattered and incident radiation fields. Both BRDF and BPDF are a potential source of information about physical and optical properties of land surfaces. Moreover, accurate models of BRDF and BPDF at visible and infrared wavelength are required for retrieval of atmospheric aerosols properties over land surfaces.

Anthropogenic aerosols are believed to cause the second most important anthropogenic forcing of climate change after greenhouse gases. In contrast to the climate effect of greenhouse gases, which is understood relatively well, the negative forcing (cooling effect) caused by aerosols represents the largest uncertainty in climate change research (Hansen et al. 2005). To reduce this uncertainty, multiple-viewing-angle and multi-spectral photopolarimetric satellite measurements at a global scale are necessary (Mishchenko et al. 2007a; Hasekamp and Landgraf 2007).

An essential part of algorithms for the retrieval of aerosol properties is to accurately account for reflection of the surface of the Earth. Over the ocean, the surface contribution is relatively small and can for most scenes be modeled with sufficient accuracy (Mishchenko and Travis 1997; Tanré et al. 1997; Chowdhary et al. 2005; Hasekamp and Landgraf 2005b). Over land, the surface reflection contribution is in general much larger, shows significant spatial variability, and represents one of the most important problems for aerosol retrieval algorithms.

Different algorithms have been proposed for aerosol properties retrievals over land from space (e.g., Kokhanovsky and de Leeuw 2009). MODIS retrievals (Remer et al. 2005) use an empirical relationship between the albedo retrieved at the 2100 nm band (where the aerosol contribution is small) and the albedo at other wavelengths. Here, uncertainties in the surface reflectance model represent one of the largest error sources on the retrieved aerosol properties. Retrievals from multiple-viewing-angle measurements can take advantage of the different angular reflectance signatures of the surface and the atmosphere to accomplish the retrieval of aerosol optical thickness over land surfaces (Martonchik et al. 1998; Diner et al. 2005). Also, retrieval methods have been proposed using only measurements of polarized reflectance, which have a relatively small and spectrally flat contribution from surface reflection (Deuzé et al. 2001; Waquet et al. 2009a).

To make full use of the information contained in multi-angle photopolarimetric measurements, it is necessary to perform a simultaneous retrieval of aerosol and surface properties using both radiance and polarization measurements. For this purpose the atmospheric signal must be accurately separated from the surface signal. Hence, the following aspects are important for BDRF and BPDF models: (i) constraints on and invariances of the spectral and angular dependences of the BDRF and BPDF; (ii) BRDF and PBDF models should be able to correctly describe the surface signal with the geometry of the measurements (flexibility), as well as (iii) for all other geometries (predictability).

For surface BRDF characterization on the basis of airborne and satellite data, semi-empirical models are often used (Hapke 1981; Roujean et al. 1992; Rahman

et al. 1993; Wanner et al. 1995; Spurr 2004; Maignan et al. 2004), whereas surface BPDF is usually considered as spectrally independent in the visible and infrared regions and described by models based on the assumption of single Fresnel reflection from the surface facets (see, for example, Tsang et al. 1985; Rondeaux and Herman 1991; Bréon et al. 1995; Nadal and Bréon 1999; Maignan et al. 2009; Waquet et al. 2009b). Below we present results of testing existing BRDF and BPDF models with respect to the three criteria listed above (Litvinov et al. 2010, 2011).

Aspects (i), (ii), and (iii) of BRDF and BPDF models are evaluated using measurements performed with the Research Scanning Polarimeter (RSP). The RSP provides multi-spectral, multi-angle photopolarimetric measurements. It is a prototype for the Aerosol Polarimetry Sensor (APS) instrument of the NASA Glory Mission (Cairns et al. 1999; Mishchenko et al. 2007b). RSP measurements provide a valuable source of observations for testing different models of the BRDF and BPDF of Earth surfaces.

In assessing aerosol retrievals over land surfaces, it is the manifestation of errors in the top-of-atmosphere signals caused by BRDF and BPDF model errors that are of interest. The different semi-empirical BRDF and BPDF models were, therefore, used in radiative transfer calculations for the coupled atmosphere-surface system, and their capability to represent top-of-atmosphere measurements is analyzed in this chapter.

2. Research Scanning Polarimeter

The RSP measures intensity and polarization characteristics of the radiation at viewing zenith angles in the range $-60^\circ \leq \vartheta_v \leq 60^\circ$ from the nadir direction in 9 spectral bands in the range 410–2250 nm. We used RSP data obtained during the Aerosol Lidar Validation Experiment (ALIVE) measurement campaign performed in Oklahoma (Southern Great Plains, USA) in September of 2005. There were several flights in the ALIVE campaign that had measurements at low altitudes over land (about 200–600 m), with low aerosol optical thicknesses ($\tau_{aer} \approx 0.04$ in the “red” channel, $\lambda = 670$ nm, and $\tau_{aer} \approx 0.0075$ in the “short-wave infrared” channel, $\lambda = 1589$ nm), and in clear sky conditions. The values of τ_{aer} for different wavelengths were taken from an AERONET station in Oklahoma (the U.S. Southern Great Plains Cloud and Radiation Testbed (CART) Site). These RSP measurements provide a good opportunity for testing different models of the BRDF and BPDF of land surfaces, since surface signals can be separated from atmospheric signals very accurately.

Table 1 describes the flights that were used in this study. The flights were carried out over the same area at different times during the same day and with similar weather conditions. Thus the data for these flights were obtained for different illumination and scattering geometries and pertain, generally, to the same types of soil and vegetation surfaces.

Table 1. Summary of RSP flights used in this analysis

	Flight 1	Flight 2	Flight 3	Flight 4
Day, month, and year	16.09.2005	16.09.2005	16.09.2005	16.09.2005
Time (UTC)	16:32:25	22:01:49	16:25:24	22:09:32
Average altitude over sea level	510 m	942 m	640 m	476 m
Average solar zenith angle	42.68°	60.8°	43.67°	62°
Average solar azimuth angle	43.3°	290.9°	45.5°	289.7°
Average observation azimuth angle	89.25°	312.7°	268.8°	134°
Average ARVI, soil	0.033	0.024	0.031	–
Average ARVI, vegetation	0.63	0.65	–	0.577
Aerosol optical thickness at 670 nm	0.039	0.046	0.039	0.038
Aerosol optical thickness at 1589 nm	0.0075	0.01256	0.0075	0.0083
Aerosol optical thickness at 2264 nm	0.0037	0.0075	0.0037	0.0046

Soil and other non-vegetated surfaces have much smaller spectral contrast between the “red” and “near-infrared” bands. Following Knobelspiesse et al. (2008) and Kaufman and Tanré (1992), we used the Atmospherically Resistant Vegetation Index (ARVI) to distinguish soil and vegetation types of surfaces (Litvinov et al. 2010). The data with $-0.25 < \text{ARVI} < 0.075$ and $0.375 < \text{ARVI} < 0.775$ were classified as “soil” and “vegetation” respectively. A detailed description of the approach to splitting the data into “soil” and “vegetation” classes for flights 1 and 4 from Table 1 is presented by Knobelspiesse et al. (2008). Using this classification for each flight from Table 1, we averaged the intensity and polarization measurements over different realizations (scans) separately for soil and vegetated surfaces. Then a small atmospheric correction to the RSP data was performed (for details, see Litvinov et al. 2010).

3. Semi-empirical BRDF and BPDF models

The intrinsic reflectance properties of surfaces are described by the 4×4 bidirectional reflection matrix (BRM) \mathbf{R} . It provides a relation between the Stokes parameters of scattered and incident radiation fields (see, e.g., Mishchenko and Travis 1997):

$$\mathbf{I} = \frac{1}{\pi} \mathbf{R}(\lambda, \vartheta_v, \vartheta_0, \phi) \mathbf{I}_0(\lambda) \cos \vartheta_0. \quad (1)$$

Here, $\mathbf{I} = (I, Q, U, V)^T$ is the intensity column vector describing the radiance and polarization state of scattered radiation (T stands for “transposed”); $\mathbf{I}_0 = (I_0, Q_0, U_0, V_0)^T$ is the Stokes vector, describing total and polarized incident irradiances; λ is the wavelength of the incident and scattered radiation; ϕ is the azimuth angle difference $\varphi_v - \varphi_0$, with φ_0 and φ_v being the solar and viewing

azimuth angles, respectively; and ϑ_0 and ϑ_v are the solar and viewing zenith angles, respectively ($\vartheta_0 = \pi - \vartheta_{inc}$, ϑ_{inc} is the incident zenith angle). Below we use positive and negative values of ϑ_v to denote the cases when ϕ is changed by π : $\vartheta_v < 0$ when $0 \leq |\phi| \leq \pi/2$ and $3\pi/2 \leq |\phi| \leq 2\pi$ (these ranges include the backscattering direction when $|\vartheta_v| = \vartheta_0$ and $\phi = 0$ or $|\phi| = 2\pi$), $\vartheta_v > 0$ when $\pi/2 \leq |\phi| \leq 3\pi/2$ (this range includes the specular reflection direction when $\vartheta_v = \vartheta_0$ and $\phi = \pi$).

When the incident radiation is unpolarized, the element R_{11} of the matrix \mathbf{R} is the surface total reflectance (denoted hereinafter as R_I), and the elements R_{21} and R_{31} define surface polarized reflectances (denoted hereinafter as R_p):

$$I = \frac{1}{\pi} R_I(\lambda, \vartheta_v, \vartheta_0, \phi) F_0(\lambda) \cos \vartheta_0, \quad (2)$$

$$Q = \frac{1}{\pi} R_{21}(\lambda, \vartheta_v, \vartheta_0, \phi) F_0(\lambda) \cos \vartheta_0, \quad (3)$$

$$U = \frac{1}{\pi} R_{31}(\lambda, \vartheta_v, \vartheta_0, \phi) F_0(\lambda) \cos \vartheta_0, \quad (4)$$

$$R_p(\lambda, \vartheta_v, \vartheta_0, \phi) = \sqrt{R_{21}^2 + R_{31}^2}. \quad (5)$$

Here, F_0 is the incident energy flux per unit area perpendicular to the incident beam. This definition of the total and polarized reflectances has been used by different authors (Roujean et al. 1992; Nadal and Bréon 1999; Maignan et al. 2009). The definition of the surface total reflectance used here is equivalent to the definition of the bidirectional reflectance factor (BRF) (see, e.g., Schaepman-Strub et al. 2006). In other words, we suppose that directional surface reflection properties vary weakly within the instrument instantaneous field of view (IFOV), and the conical reflectance quantities are equivalent to the directional ones. That is the case in the directions far away from the specular or the exact backscattering directions, and for small IFOVs. The RSP data we are working with satisfy these conditions (see [Table 1](#) for the RSP geometry description; the RSP IFOV is 14 mrad or 0.8° ; Mishchenko et al. 2007a).

It must be noted that radiative transfer calculations for a coupled atmosphere–surface system require all elements of the BRM for surfaces rather than only the surface total and polarized reflectances (R_I and R_p). However it is usually assumed that surface reflections are strongly depolarizing (i.e., $|R_{22}| \ll R_{11}$ and $|R_{33}| \ll R_{11}$) over a wide range of scattering angles. This is the case, for example, for complex media causing considerable diffuse scattering (Woolley 1971; Savchenkov et al. 2003; Tishkovets et al. 2004; Muñoz et al. 2007). Under these conditions the elements R_{11} , R_{21} , and R_{31} of the BRM for surfaces give the main contribution to the top-of-atmosphere total and polarized reflectances. In general, it is not possible to characterize the elements other than R_{11} , R_{21} , and R_{31} (for example, R_{22} , R_{33} , etc.) if the incident radiation is not polarized and semi-empirical models for the surface total and polarized reflectances R_I and R_p are used. If single scat-

tering by randomly oriented elementary surface (or volume) scattering elements gives the main contribution to the polarization of the scattered signal, then R_{21} and R_{31} are related to R_p via the following simple relations (Hovenier et al. 2004):

$$R_{21} = -R_p \cos 2\eta_v, \quad (6)$$

$$R_{31} = R_p \sin 2\eta_v, \quad (7)$$

where the dihedral angle η_v is the angle between the scattering plane (the plane containing the solar and viewing directions) and the meridional plane containing the zenith and viewing directions. It can be found, for example, from the equations

$$\cos \eta_v = -\frac{\cos \vartheta_0 + \cos \vartheta_v \cos \gamma}{\sin |\vartheta_v| \sin \gamma} \quad \text{and} \quad \sin \eta_v = \frac{\sin \vartheta_0 \sin \phi}{\sin \gamma}, \quad (8)$$

where γ is the scattering angle defined in the scattering plane, that

$$\cos \gamma = -\cos \vartheta_v \cos \vartheta_0 - \sin |\vartheta_v| \sin \vartheta_0 \cos \phi. \quad (9)$$

As shown by Litvinov et al. (2010), the relations (6) and (7) hold for soil and vegetated surfaces measured with the RSP instrument.

For surface reflectance description on the basis of satellite data, the bidirectional reflection distribution function (BRDF) and bidirectional polarization distribution function (BPDF) are used. When the definition of surface total and polarized reflectances (R_I and R_p) is given by Eqs. (1)–(5), the BRDF and BPDF differ from R_I and R_p by the following normalization (Schaeppman-Strub et al. 2006):

$$\text{BRDF} = \frac{R_I}{\pi}, \quad \text{BPDF} = \frac{R_p}{\pi}. \quad (10)$$

Throughout the rest of this chapter, when we refer to BRDF and BPDF we mean the relations given by Eq. (10).

For surface reflection characterization from the Multi-angle Imaging Spectro-Radiometer (MISR), the MODerate resolution Imaging Spectroradiometer (MODIS), and the Polarization and Directionality of Earth's Reflectances (POLDER) instrument, the Rahman–Pinty–Verstraete (RPV) model and kernel-driven models (Ross–Li and Ross–Roujean models) for surface BRDF are used.

For characterization of atmospheric aerosol over land surfaces using POLDER data, the Nadal–Bréon model for polarized reflectance is used (Nadal and Bréon 1999). Recently, for POLDER surface polarized reflectance characterization a new linear BPDF model with only one free parameter was introduced by Maignan et al. (2009). The RSP airborne instrument and the APS instrument onboard of the Glory satellite (will) provide very accurate surface polarized reflectance measurements (Cairns et al. 1999; Mishchenko et al. 2007b). For accurate description of such measurements, a three-parameter semi-empirical model was proposed by Litvinov et al. (2011).

3.1. Semi-empirical BRDF models

3.1.1. Rahman–Pinty–Verstraete model for surface reflectance

The semi-empirical Rahman–Pinty–Verstraete (RPV) BRDF model is based on a modification of the Minnaert empirical model of bidirectional reflectance for the Moon (Minnaert 1941). In the RPV model, the surface total reflectance can be presented in the following form (Rahman et al. 1993):

$$R_I(\lambda, \vartheta_v, \vartheta_0, \phi) = \frac{(\cos \vartheta_0 \cos \vartheta_v)^{k-1}}{(\cos \vartheta_0 + \cos \vartheta_v)^{1-k}} \rho_0(\lambda) F(\gamma) [1 + R(G)], \quad (11)$$

$$F(\gamma) = \frac{1 - g^2}{(1 + g^2 - 2g \cos \gamma)^{3/2}}, \quad (12)$$

$$1 + R(G) = 1 + \frac{1 - \rho_0}{1 + G}, \quad (13)$$

$$G = \sqrt{\tan^2 \vartheta_0 + \tan^2 \vartheta_v - 2 \tan \vartheta_0 \tan \vartheta_v \cos \phi}. \quad (14)$$

Here, ρ_0 , g , and k are free parameters of the model; $F(\gamma)$ is the Henyey–Greenstein phase function; and the function $1 + R(G)$ is used to approximate the shadowing hot spot effect (Rahman et al. 1993).

There are no analytical relations between the parameters of the RPV model and actual physical parameters of the scattering surface. In general, all three parameters of the RPV model may be wavelength dependent. It should be mentioned that the RPV model is reciprocal: it remains invariant with respect to switching the variables ϑ_0 and ϑ_v .

3.1.2. Kernel driven Ross–Roujean and Ross–Li models for surface reflectance

The most common kernel-driven model uses a linear combination of three kernels f_{iso} , f_{vol} , and f_{geom} representing isotropic, volumetric, and geometric-optics surface scattering, respectively (Roujean et al. 1992; Wanner et al. 1995):

$$R_I(\lambda, \vartheta_v, \vartheta_0, \phi) = f_{iso}(\lambda) + k_1(\lambda) f_{geom}(\vartheta_v, \vartheta_0, \phi) + k_2(\lambda) f_{vol}(\vartheta_v, \vartheta_0, \phi). \quad (15)$$

For surface BRDF characterization on the basis of airborne and satellite data, the Ross-thick kernel is often used as the volumetric scattering kernel f_{vol} (Ross 1981; Roujean et al. 1992; Wanner et al. 1995). It is derived in the single-scattering approximation from the radiative transfer theory for a layer of randomly oriented and randomly positioned facets with fixed reflectance and transmittance (Roujean et al. 1992). Unaccounted multiple scattering between different facets is usually considered to be isotropic and is described by the isotropic kernel $f_{iso}(\lambda)$ in the kernel-driven models. The $f_{vol}(\vartheta_v, \vartheta_0, \phi)$ has the form

$$f_{vol}(\vartheta_v, \vartheta_0, \phi) = \frac{(\pi/2 - \gamma) \cos \gamma + \sin \gamma}{\cos \vartheta_v + \cos \vartheta_0} - \frac{\pi}{4}. \quad (16)$$

The Ross–Roujean and Ross–Li models differ by the description of the geometric-optical scattering kernels $f_{geom}(\vartheta_v, \vartheta_0, \phi)$. The $f_{geom}(\vartheta_v, \vartheta_0, \phi)$ takes into account the geometrical structure of reflectors as well as shadowing by reflectors (Wanner et al. 1995).

The Roujean geometric scattering kernel does not have any additional model parameters. It depends on the illumination and viewing geometry as follows (Roujean et al. 1992):

$$f_{geom}^{Rouj}(\vartheta_v, \vartheta_0, \phi) = \frac{1}{2\pi} [(\pi - \phi') + \sin \phi'] \tan \vartheta_0 \tan |\vartheta_v| - \frac{1}{\pi} (\tan \vartheta_0 + \tan |\vartheta_v| + G). \quad (17)$$

Here, $\phi' = |\varphi_v - \varphi_0|$ is chosen in the range $0 \leq \phi' \leq \pi$, and this kernel therefore has the symmetry $f_{geom}(\vartheta_v, \vartheta_0, \phi) = f_{geom}(\vartheta_v, \vartheta_0, 2\pi - \phi)$, with G being defined by Eq. (14). Let us note that both the Ross-thick volumetric and the Roujean geometric kernels are reciprocal, resulting in a reciprocal Ross–Roujean model.

For the description of $f_{geom}(\vartheta_v, \vartheta_0, \phi)$, the Li-dense (f_{geom}^{LiDen}) and Li-sparse (f_{geom}^{LiSp}) kernels are also often used (Wanner et al. 1995). In the reciprocal form, they can be presented as follows:

$$f_{geom}^{LiSp}(\vartheta_v, \vartheta_0, \phi) = O(\vartheta_v, \vartheta_0, \phi) - \sec \vartheta'_0 - \sec \vartheta'_v + \frac{1}{2} (1 - \cos \vartheta') \sec \vartheta'_v \sec \vartheta'_0, \quad (18)$$

$$f_{geom}^{LiDen}(\vartheta_v, \vartheta_0, \phi) = \frac{(1 - \cos \vartheta') \sec \vartheta'_0 \sec \vartheta'_v}{\sec \vartheta'_v + \sec \vartheta'_0 - O(\vartheta_v, \vartheta_0, \phi)}, \quad (19)$$

$$O = \frac{1}{\pi} (t - \sin t) (\sec \vartheta'_v + \sec \vartheta'_0), \quad (20)$$

$$\cos t = \frac{h}{b} \frac{\sqrt{D^2 + (\tan \vartheta'_0 \tan \vartheta'_v \sin \phi)^2}}{\sec \vartheta'_v + \sec \vartheta'_0}, \quad (21)$$

$$D = \sqrt{\tan^2 \vartheta'_0 + \tan^2 \vartheta'_v - 2 \tan \vartheta'_0 \tan \vartheta'_v \cos \phi}, \quad (22)$$

$$\cos \vartheta' = -\cos \vartheta'_v \cos \vartheta'_0 - \sin \vartheta'_v \sin \vartheta'_0 \cos \phi, \quad (23)$$

$$\vartheta'_v = \tan^{-1} \left(\frac{b}{r} \tan |\vartheta_v| \right), \quad \vartheta'_0 = \tan^{-1} \left(\frac{b}{r} \tan |\vartheta_0| \right). \quad (24)$$

The term $\cos t$ in Eq. (21) is equal to 1 if $|\cos t| > 1$. Both sparse and dense kernels contain two parameters h/b and b/r (Wanner et al. 1995), which are fixed in linear kernel-driven models. For example, the MODIS BRDF retrieval algorithm employs the Li-sparse kernel in the reciprocal form, Eq. (18), with $h/b = 2$ and $b/r = 1$ (Strahler et al. 1999).

3.2. Semi-empirical BPDF models

3.2.1. Nadal–Bréon model for surface polarized reflectance

This BPDF model was introduced by Nadal and Bréon (1999) as a parameterization of the polarized reflectance for different type of surfaces. It can be written as follows:

$$R_p(\vartheta_v, \vartheta_0, \phi) = \alpha \left[1 - \exp \left(-\beta \frac{F_p(m, \gamma)}{\cos \vartheta_0 + \cos \vartheta_v} \right) \right], \quad (25)$$

$$F_p(m, \gamma) = \frac{1}{2} \left[\left(\frac{m\mu_t - \mu_r}{m\mu_t + \mu_r} \right)^2 - \left(\frac{m\mu_r - \mu_t}{m\mu_r + \mu_t} \right)^2 \right], \quad (26)$$

$$\mu_r = \cos \theta_r, \quad \mu_t = \cos \theta_t, \quad (27)$$

$$\sin \theta_r = m \sin \theta_t, \quad \theta_r = (\pi - \gamma)/2. \quad (28)$$

Here, $-F_p(m, \gamma)$ is the element F_{21} of the Fresnel scattering matrix; m is the refractive index; θ_r and θ_t are angles of specular reflection and refraction, respectively; α and β are parameters of the model. It must be noted that in most cases the calculation of R_p for land surfaces is based on the refractive index m fixed at 1.5 (Nadal and Bréon 1999).

3.2.2. Linear one-parameter model for surface polarized reflectance

Maignan et al. (2009) introduced a linear one-parameter BPDF model as a simplification of two-parameter models based on the assumption of Fresnel reflection from soil and vegetated surfaces. It can be applied both to soil and vegetated surfaces and provides a similar goodness of fit to measurements as the Nadal–Bréon model. This model is written as follows (Maignan et al. 2009):

$$R_p(\vartheta_v, \vartheta_0, \phi) = \alpha \frac{\exp(-\tan \theta_r) \exp(-\nu) F_p(m, \gamma)}{4(\cos \vartheta_0 + \cos \vartheta_v)}. \quad (29)$$

Here, the parameter ν is supposed to be related to the Normalized Difference Vegetation Index (NDVI). In the calculations below we take ν equal to Atmospherically Resistant Vegetation Index (ARVI) (Kaufman and Tanré 1992). The parameter α is the only free linear parameter of the model. The $-F_p(m, \gamma)$, as before, is the element F_{21} of the Fresnel scattering matrix (see Eq. (26)). The angle θ_r is the angle of specular reflection (see Eq. (28)).

3.2.3. Modified Fresnel models for surface polarized reflectance

The uncertainty in the degree of linear polarization of RSP and APS measurements is less than 0.002 (Cairns et al. 1999; Mishchenko et al. 2007b). To provide a precise description of the RSP polarimetric data, a new model for surface

polarized reflectance was introduced (Litvinov et al. 2011). It is based on a model of Fresnel reflection from a Gaussian random rough surface in the Kirchhoff approximation (Tsang et al. 1985; Mishchenko and Travis 1997). To suppress the value of polarized reflectance in the forward reflection region, a shadowing function that has its maximum in the backscattering direction ($\gamma = 180^\circ$) was introduced (Litvinov et al. 2010) instead of the shadowing function for Gaussian surfaces, which has a maximum at $\vartheta_v = 0^\circ$ (Tsang et al. 1985; Mishchenko and Travis 1997). Also, to use this model both for soil and vegetated surfaces, we assume that scattering facets are distributed in space rather than on a surface. The modified model for polarized reflectance can be written as follows (Litvinov et al. 2011):

$$R_p(\vartheta_v, \vartheta_0, \phi) = \frac{\alpha \pi F_p(m, \gamma)}{4\mu_n(\cos \vartheta_0 + \cos \vartheta_v)} f(\mathbf{n}_v, \mathbf{n}_0) f_{sh}(\gamma), \quad (30)$$

$$f(\mathbf{n}_v, \mathbf{n}_0) = \frac{1}{\pi \mu_n^3 2\sigma^2} \exp\left(-\frac{1 - \mu_n^2}{\mu_n^2 2\sigma^2}\right), \quad (31)$$

$$f_{sh}(\gamma) = \left(\frac{1 + \cos k_\gamma(\pi - \gamma)}{2}\right)^3, \quad (32)$$

$$\mu_n = \frac{n_v^z + n_0^z}{|\mathbf{n}_v + \mathbf{n}_0|}, \quad (33)$$

$$\mathbf{n}_0 = (\sin \vartheta_0 \cos \varphi_0; \sin \vartheta_0 \sin \varphi_0; \cos \vartheta_0), \quad (34)$$

$$\mathbf{n}_v = (\sin |\vartheta_v| \cos \varphi_v; \sin |\vartheta_v| \sin \varphi_v; \cos \vartheta_v). \quad (35)$$

Here, the function $f(\mathbf{n}_v, \mathbf{n}_0)$ describes the distribution of facets over orientation (in our calculations we assume a Gaussian distribution of facet slopes, Eq. (31), where σ^2 is the mean square facet slope); n_0^z and n_v^z are the z -components of the unit vectors \mathbf{n}_0 and \mathbf{n}_v pointing in the solar and viewing directions, respectively (see Eqs. (34) and (35)); $f_{sh}(\gamma)$ is a shadowing function which is modeled by Eq. (32) with a free parameter k_γ that controls the width of the shadowing region ($0 < k_\gamma < 1$).

The modified BPDF model presented here is similar to the two-parameter analytical BPDF model for vegetation surfaces presented in (Maignan et al. 2009) when an additional, third, linear parameter is introduced. We apply the modified model both for soil and vegetation surfaces. It has three parameters: α , σ , and k_γ (for land surfaces m is taken equal to 1.5 but it should be noted that changing α is equivalent to changing the refractive index). These are empirical free model parameters rather than physical ones, and can be obtained from remote-sensing data in the short-wave infrared (SWIR) where, in general, the atmospheric contribution is small. Since the surface polarized reflectance depends weakly on the wavelength, these parameters can also be used for other wavelengths. This method for retrieving the surface polarized reflectance can be applied to observations from

the RSP and APS instruments, which perform measurements in the SWIR region (Waquet et al. 2009b). The number of model parameters can be reduced to two, assuming random orientation of surface facets (in this case $f(\mathbf{n}_v, \mathbf{n}_0) = 1/2\pi$ in Eqs. (30) and (31)). This decreases the accuracy of the model, but it may still be suitable for a number of applications.

4. Testing BRDF and BPDF models

As mentioned in the Introduction, for the application to aerosol retrievals over land, three aspects of BRDF and BPDF models are important. These aspects will be discussed in this section

4.1. Constraints on spectral and angular dependence

The *ad hoc*, or semi-empirical, nature of existing BRDF and BPDF models used for the interpretation of satellite data means that their controlling parameters are only tenuously, or not at all, related to the actual physical parameters of the scattering surfaces. Moreover, because of this lack of physical basis, the parameters of BRDF models are not necessarily consistent with the parameters of BPDF models. There are, therefore, uncertainties regarding the physical constraints, which can be imposed on the spectral and angular dependences of BRDF and BPDF models.

Instead of physical constraints, empirical constraints can be used. To determine what these constraints are, multi-spectral and multi-angle measurements are required. To investigate the empirical constraints on BRDF and BPDF model parameters, we used RSP measurements at low altitudes over land performed in clear-sky conditions and a low aerosol optical thickness. This type of data allows one to accurately separate surface and atmospheric contributions to the observed signal.

Below, we consider the ratios K_I and K_P of the total reflectances ($R_I(\lambda_1; \vartheta_v, \vartheta_0, \phi)$ and $R_I(\lambda_2; \vartheta_v, \vartheta_0, \phi)$) and of the polarized reflectances ($R_P(\lambda_1; \vartheta_v, \vartheta_0, \phi)$ and $R_P(\lambda_2; \vartheta_v, \vartheta_0, \phi)$), evaluated for two different wavelengths λ_1 and λ_2 :

$$K_I(\lambda_1, \lambda_2; \vartheta_v, \vartheta_0, \phi) = \frac{R_I(\lambda_1; \vartheta_v, \vartheta_0, \phi)}{R_I(\lambda_2; \vartheta_v, \vartheta_0, \phi)}, \quad (36)$$

$$K_P(\lambda_1, \lambda_2; \vartheta_v, \vartheta_0, \phi) = \frac{R_P(\lambda_1; \vartheta_v, \vartheta_0, \phi)}{R_P(\lambda_2; \vartheta_v, \vartheta_0, \phi)}. \quad (37)$$

Figures 1 and 2 show the angular dependences of the ratios obtained from RSP measurements averaged over different scans after atmospheric correction. Different curves (solid, dashed, and dotted) correspond to different geometries of illumination and measurements (see Fig. 1 and 2 captions). Both figures demonstrate good agreement between the angular dependences of the ratios K_I for the considered flights. In particular, for the same type of soil or vegetated surfaces the ratio

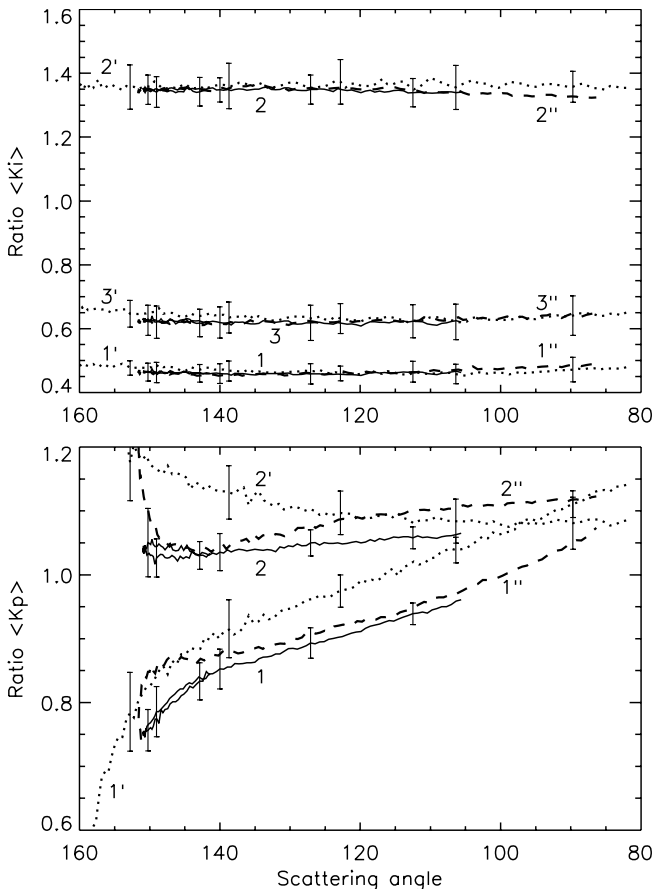


Fig. 1. Angular dependences of the ratios $K_I(\lambda_1, \lambda_2; \vartheta_v, \vartheta_0, \phi)$ and $K_P(\lambda_1, \lambda_2; \vartheta_v, \vartheta_0, \phi)$ obtained from RSP data for soil surfaces. The solid, dotted, and dashed curves show the ratios for flights 1, 2, and 3 respectively (see Table 1). The curves 1, 1', and 1'' correspond to the ratios with $\lambda_1 = 670$ nm and $\lambda_2 = 1589$ nm. The curves 2, 2', and 2'' correspond to the ratios with $\lambda_1 = 1589$ nm and $\lambda_2 = 2264$ nm. The curves 3, 3', and 3'' correspond to the ratios with $\lambda_1 = 670$ nm and $\lambda_2 = 2264$ nm. The error bars show one standard deviation from the average values.

of total reflectances K_I , taken at two different wavelengths, is almost independent of the scattering angle and is the same for different illumination and scattering geometries (for different flights). For example, flights 1 and 2, 2 and 3, 1 and 4 had different illumination and scattering geometries (see Table 1), but the respective intensity ratios are almost the same.

To explain qualitatively this behavior of the ratios, let us note that, in general, the reflection matrix for a surface can be decomposed into the sum of matrices describing the single- and multiple-scattering contributions:

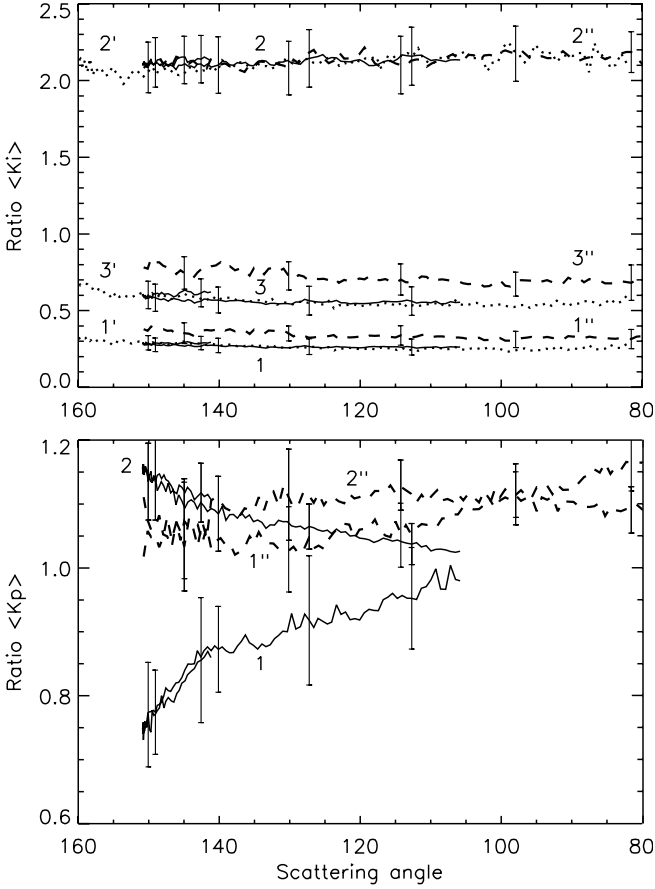


Fig. 2. As in Fig. 1, but for vegetated surfaces. The curves 1', 2'', and 3'' correspond to flight 4 from Table 1.

$$\mathbf{R} = \mathbf{R}_{single} + \mathbf{R}_{mult}. \quad (38)$$

The single-scattering contribution can be further decomposed into the sum of a Fresnel surface reflection matrix \mathbf{R}_{single}^{Fr} and a diffuse scattering matrix $\mathbf{R}_{single}^{dif}$:

$$\mathbf{R}_{single} = \mathbf{R}_{single}^{Fr} + \mathbf{R}_{single}^{dif}. \quad (39)$$

For vegetated surfaces, an illustrative example of this representation would be single scattering by leaves of vegetation canopies and multiple scattering between different leaves. Then, the single-scattering contributions from the surface reflection off waxy cuticles (Fresnel reflection) and scattering off structures inside the leaf (diffuse, or volume scattering) can be represented separately. For soil surfaces, the single and multiple scattering can be considered as scattering by a sur-

face element and that between different elements, respectively. Fresnel reflection for soil surfaces can be produced by reflection from facets of large soil particles.

According to the representation of the reflection matrix given by Eqs. (38) and (39), the ratio K_I is independent of the scattering angle and the geometry of illumination and measurements if one of the following conditions is fulfilled (Litvinov et al. 2010):

1. one of the matrices gives the main contribution to the total reflection matrix and can be approximately expressed as a product of wavelength-dependent and geometry-dependent terms;
2. the angular dependences of the matrices are similar and can be expressed as a product of wavelength-dependent and geometry-dependent terms.

The results presented in Figs. 1 and 2 are from flights performed far away from the principle plane (the plane containing the incident and viewing directions as well as the zenith direction). Because of this, and also taking into account the strong spectral dependence of the surface total reflectance which cannot be produced by the Fresnel reflection, one can assume that for all the spectral channels and geometries considered $R_{11mult} + R_{11single} \gg R_{11single}^{Fr}$. Then $K_I(\lambda_1, \lambda_2; \mathcal{G}_v, \mathcal{G}_0, \phi) \approx \text{const}(\lambda_1, \lambda_2)$ when

1. $R_{11single} \gg R_{11mult}$ and $R_{11single}$ has the functional form

$$R_{11single}(\lambda; \mathcal{G}_v, \mathcal{G}_0, \phi) \approx a_{single}(\lambda) f_I(\mathcal{G}_v, \mathcal{G}_0, \phi); \quad (40)$$

2. $R_{11single} \sim R_{11mult}$, $R_{11single}$ has the form given by Eq. (40), and R_{11mult} is similar to $R_{11single}$ (Litvinov et al. 2010), viz.,

$$R_{11mult}(\lambda; \mathcal{G}_v, \mathcal{G}_0, \phi) \approx a_{mult}(\lambda) f_I(\mathcal{G}_v, \mathcal{G}_0, \phi). \quad (41)$$

Here, the function $f_I(\mathcal{G}_v, \mathcal{G}_0, \phi)$ depends on the illumination and scattering geometries but not on the wavelength, and $a_{single}(\lambda)$ and $a_{mult}(\lambda)$ are the wavelength dependent terms. According to Eqs. (40) and (41), for the surface total reflectance one can write:

$$R_I(\lambda; \mathcal{G}_v, \mathcal{G}_0, \phi) \approx a(\lambda) f_I(\mathcal{G}_v, \mathcal{G}_0, \phi). \quad (42)$$

The relation (42) shows that the surface total reflectance depends almost linearly on the wavelength-dependent model parameter $a(\lambda)$. The parameters describing the angular profiles of R_I are wavelength independent. It suggests that it is possible to retrieve the wavelength-independent part of the surface reflectance model from measurements in the short-wave infrared range where, in general, the effect of aerosols on the measurements is weak. A similar method has been used by Waquet et al. (2009a) to retrieve surface polarized reflectance.

It follows from the relation (42) that the ratio of total reflectances measured at two different viewing angles is nearly spectrally invariant. The spectral invariance of the geometry-dependent term has already been exploited in the aerosol retrieval algorithm over land developed for the Along Track Scanning Radiometer-2

(AATSR-2) instrument (Flowerdew and Haigh 1996; Veefkind et al. 1998) and discussed for use in the MISR algorithm (Diner et al. 2005; Kokhanovsky and de Leeuw 2009).

The functional form of Eq. (42) provides constraints on the spectral and angular dependences of BRDF models, but it may not hold for all possible geometries and wavelengths. For example, in the principle plane and in directions close to specular the condition $R_{11mult} + R_{11single} \gg R_{11single}^{Fr}$ may be not fulfilled for surfaces which have significant Fresnel reflection since $R_{11single}^{Fr}$ will, in general, depend differently on ϑ_v, ϑ_0 , and ϕ than $R_{11single}$ and R_{11mult} . Moreover, some media at certain wavelengths (for example, vegetated surfaces in near-infrared bands) can cause strong multiple scattering with an angular dependence different from that of single scattering (in which case the relation (41) is not fulfilled). Under these conditions the relation (42) for the reflection matrix R_I is not fulfilled, and this may cause uncertainties in the retrieval of wavelength-dependent aerosol parameters.

Let us consider now the angular profiles of K_p for the different illumination and scattering geometries of the flights from Table 1. As can be seen in Figs. 1 and 2, there is a small spectral dependence of the surface polarized reflectance. Thus if one assumes that the surface polarized reflectance is spectrally independent then a small error δ_p is introduced:

$$\delta_p = \frac{R_p(\lambda_2; \vartheta_v, \vartheta_0, \phi) - R_p(\lambda_1; \vartheta_v, \vartheta_0, \phi)}{R_I(\lambda_1; \vartheta_v, \vartheta_0, \phi)} \times 100\%, \quad (43)$$

where $R_p(\lambda_2; \vartheta_v, \vartheta_0, \phi)$ and $R_p(\lambda_1; \vartheta_v, \vartheta_0, \phi)$ are the polarized reflectances in two different channels with central wavelengths λ_2 and λ_1 , while $R_I(\lambda_1; \vartheta_v, \vartheta_0, \phi)$ is the total reflectance at λ_1 . For example, for the data presented in Fig. 1, this error between the “red” and “short-wave infrared” channels is of the order of 0.1%–0.8% in the range $95^\circ \leq \gamma \leq 160^\circ$. Estimation of this error at the top of the atmosphere leads to values of the order of 0.1%–0.6% for the aerosol optical thickness 0.04 and of the order of 0.1%–0.3% for the aerosol optical thickness 0.35 (for both cases, $95^\circ \leq \gamma \leq 160^\circ$). If we compare these values with the 0.2% polarimetric accuracy of the Glory APS then it follows that the wavelength dependence of the surface polarization may be important to take into account when the aerosol optical thickness is small ($\tau_{aer} < 0.1$). For larger values ($\tau_{aer} > 0.1$), one can neglect the spectral dependence of the surface polarized reflectance. The spectral independence of the surface polarized reflectance has already been used in aerosol-retrieval algorithms over land on the basis of other polarized reflectance measurements (Deuzé et al. 2001; Waquet et al. 2009a).

4.2. Description of observations for given geometry

An extensive comparison of different BRDF and BPDF models with POLDER satellite data has been performed by Maignan et al. (2004, 2009). For the BRDF, Maignan et al. (2004) found that the RPV model (Rahman et al. 1993) and the Ross–Li model (Ross 1981; Li and Strahler 1992; Wanner et al. 1995) are

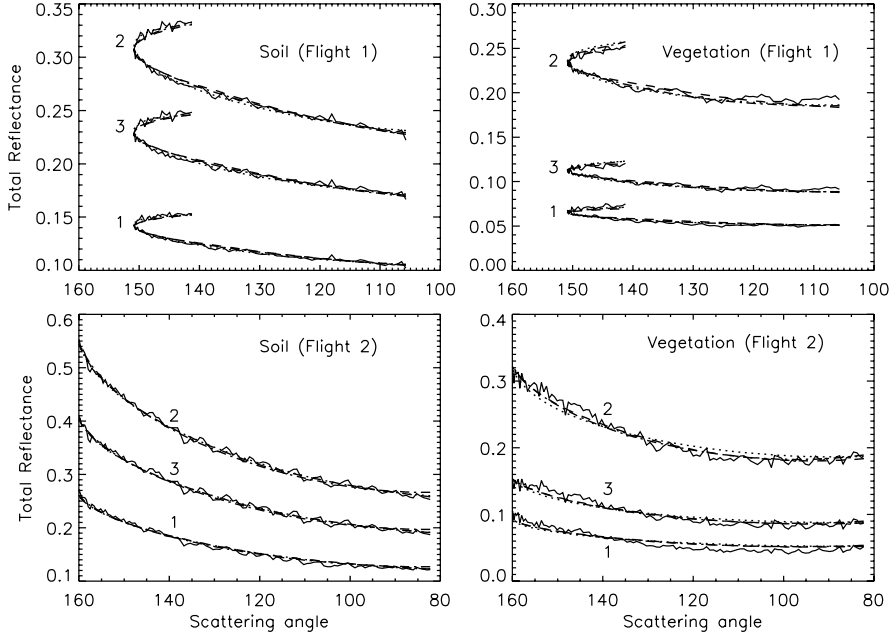


Fig. 3. Angular dependences of the average total reflectance for soil and vegetated surfaces and for different flights (see Table 1). The solid curves 1, 2, and 3 correspond to RSP data obtained in channels 4 ($\lambda = 670$ nm), 7 ($\lambda = 1589$ nm), and 9 ($\lambda = 2264$ nm), respectively. The dotted, dashed, and dash-dotted curves show the angular dependences of the BRDF according to the RPV, Ross–Roujean and Ross–Li models, respectively.

both capable of reproducing the POLDER measurements, except for the so-called hot spot region (the region near exact backscattering). To take into account the hot spot effect (also known as the opposition effect), BRDF models must be modified (see, for example, Rahman et al. 1993; Maignan et al. 2004). For the BPDF, it was found that the one-parameter model allows a similar fit to the POLDER data as the previously developed Nadal–Bréon model (Maignan et al. 2009).

Now, we analyze how well the BRDF and BPDF models can reproduce RSP observations for a given geometry (solar zenith angle, relative azimuth angle). When fitting the BRDF and BPDF models to the RSP measurements, we use the fact that BRDF models for the geometries and surfaces we use can be presented in the form given by Eq. (42) (see Figs. 1 and 2). Thus, in the RPV models the parameters g and k are nearly independent of the wavelength, and the kernel-driven models (see Eq. (15)) can be presented as:

$$R_I(\lambda; \mathcal{G}_v, \mathcal{G}_0, \phi) = k(\lambda)[1 + k_1 f_{geom}(\mathcal{G}_v, \mathcal{G}_0, \phi) + k_2 f_{vol}(\mathcal{G}_v, \mathcal{G}_0, \phi)], \quad (44)$$

where k_1 and k_2 are wavelength-independent linear model parameters and $k(\lambda)$ is a wavelength-dependent model parameter. Below, we test the kernel-driven

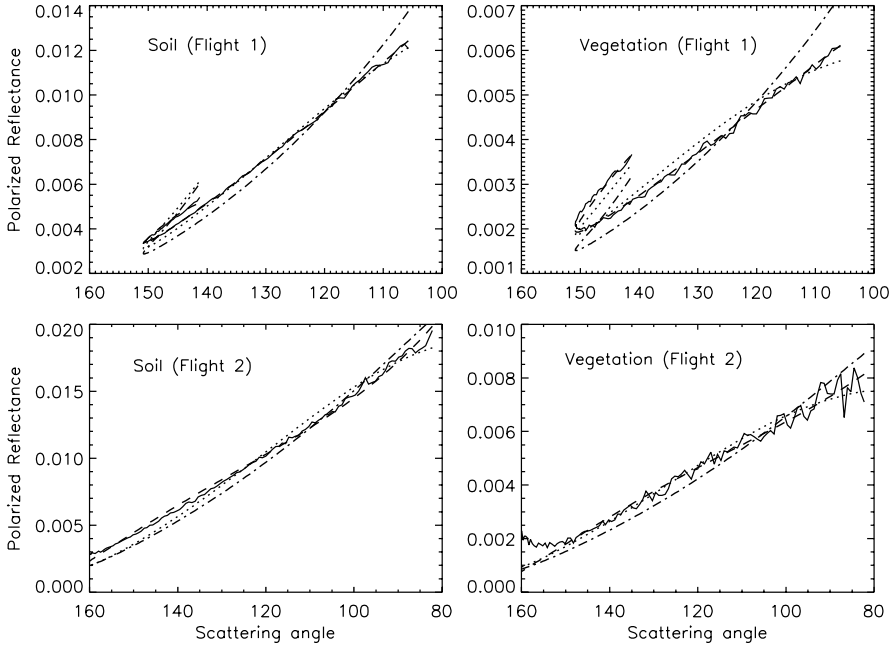


Fig. 4. Angular dependences of the average surface polarized reflectance for soil and vegetated surfaces and for different flights (see Table 1). The solid curve corresponds to RSP data obtained in channel 7 ($\lambda = 1589$ nm). The dotted, dashed, and dash-dotted curves show the angular dependences of the polarized reflectance according to the Nadal–Bréon, modified Fresnel, and linear one-parameter model, respectively.

models in the form given by Eq. (42) and assume that the RPV model parameters g and k are independent of the wavelength.

Figure 3 shows the results of fitting the RPV, Ross–Roujean, and Ross–Li BRDF models to RSP data. It can be seen that the three BRDF models describe the RSP measurements well, with no substantial difference in performance between the models. On the basis of this limited comparison no conclusions can be drawn regarding which model is the most accurate.

Figure 4 shows the results of fitting the BPDF models to the RSP data. It can be seen that the Nadal–Bréon model and the one-parameter linear model (Maignan et al. 2009) show differences with the RSP data, whereas the modified model, Eq. (30), describes the data significantly better for all data and geometries considered.

Overall, we conclude that the semi-empirical BRDF and BPDF models satisfy the requisite flexibility aspect (ii) of such models. When this is not the case, an appropriate simple modification of such models appears to be sufficient (Maignan et al. 2004, 2009; Litvinov et al. 2011).

Table 2. Best-fit parameters of the surface total reflectance models

λ , nm	RPV model			Ross–Roujean model			Ross–Li model			h/b	b/r
	$\rho_0(\lambda)$	g	k	$k(\lambda)$	k_1	k_2	$k(\lambda)$	k_1	k_2		
Flight 1 (soil)											
670	0.071	-0.097	0.746	0.145	0.201	0.640	0.139	0.158	0.547	1	1
1589	0.159	-0.097	0.746	0.315	0.201	0.640	0.301	0.158	0.547	1	1
2264	0.116	-0.097	0.746	0.234	0.201	0.640	0.224	0.158	0.547	1	1
Flight 2 (soil)											
670	0.090	-0.133	0.756	0.223	0.276	0.351	0.183	0.158	0.547	1	1
1589	0.195	-0.133	0.756	0.469	0.276	0.351	0.385	0.158	0.547	1	1
2264	0.142	-0.133	0.756	0.347	0.276	0.351	0.285	0.158	0.547	1	1
Flight 1 (vegetation)											
670	0.034	-0.071	0.725	0.065	0.155	0.600	0.064	0.087	0.688	2	1
1589	0.128	-0.071	0.725	0.235	0.155	0.600	0.232	0.087	0.688	2	1
2264	0.060	-0.071	0.725	0.113	0.155	0.600	0.111	0.087	0.688	2	1
Flight 2 (vegetation)											
670	0.035	-0.071	0.725	0.069	0.155	0.600	0.066	0.087	0.688	2	1
1589	0.128	-0.071	0.725	0.244	0.155	0.600	0.231	0.087	0.688	2	1
2264	0.060	-0.071	0.725	0.116	0.155	0.600	0.110	0.087	0.688	2	1

4.3. BRDF and BPDF angular dependence uncertainties: implications for modeling top-of-atmosphere measurements

For the retrieval of aerosol properties over land surfaces, the radiative transfer equation for coupled atmosphere–surface system must be used. It requires the surface BRDF and BPDF for all possible illumination and viewing geometries (viewing and solar zenith angles, the azimuth angles of incident and viewing directions; see Hovenier et al. 2004; Hasekamp and Landgraf 2005a; Mishchenko et al. 2006). Thus BRDF and BPDF models must describe correctly the angular profiles of the surface total and polarized reflectances at all possible geometries of illumination and viewing (requirement (iii) for BRDF and BPDF models). In this section we consider how the semi-empirical BRDF and BPDF models meet this third requisite aspect of such models, which can be regarded as the ability to predict the variations of surface reflectance and polarized reflectance away from the observational geometry, and estimate the uncertainties in angular dependences of the total and polarized reflectances at the top of the atmosphere caused by the models.

Tables 2 and 3 show the parameters of the BRDF and BPDF models that yield the best fit to RSP measurements. It can be seen that the best-fit model parameters of both BRDF and BPDF models, obtained for a particular geometry (\mathcal{G}_0, ϕ),

Table 3. Best-fit parameters of the surface polarized reflectance models

Flight	Nadal–Breon model		Modified Fresnel model			One-parameter linear model	
	α	β	α	σ^2	k_γ	α	ν
Flight 1 (soil)	0.0141	111.410	4.260	0.347	0.788	6.9	0.03
Flight 2 (soil)	0.0193	71.536	4.440	0.643	0.543	6.5	0.03
Flight 1 (vegetation)	0.0061	160.924	2.707	0.421	0.830	6.57	0.62
Flight 1 (vegetation)	0.0072	76.298	1.850	0.506	0.628	5.17	0.62

$m = 1.5$, $\lambda = 1589$ nm

may be different for other geometries, especially for soil surfaces. This demonstrates the empirical nature of the BRDF and BPDF models. Moreover, it indicates that, strictly speaking, the fitted surface model parameters only describe the surface reflection accurately for the illumination geometry of the measurement. This is demonstrated in Figs. 5 and 6, which compare the different

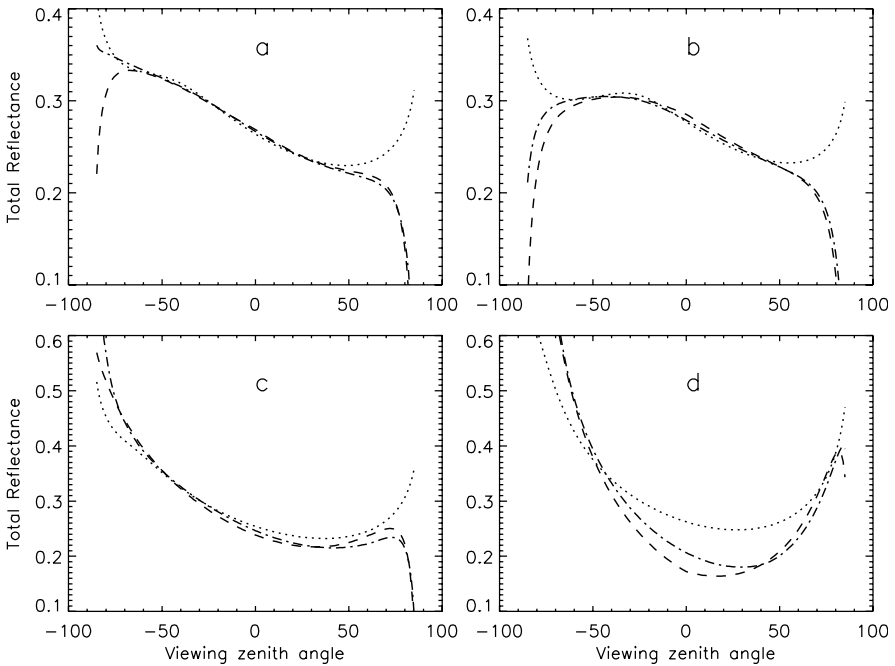


Fig. 5. Angular profiles of the total reflectance for soil surfaces according to the RPV, Ross–Roujean, and Ross–Li models (dotted, dashed, and dash-dotted curves, respectively) for different solar zenith angles: (a) $\mathcal{S}_0 = 42.68^\circ$, (b) $\mathcal{S}_0 = 30^\circ$, (c) $\mathcal{S}_0 = 60^\circ$, and (d) $\mathcal{S}_0 = 75^\circ$. For all cases, $\phi = 45.95^\circ$ for $\mathcal{S}_v < 0$ and $\phi = 225.95^\circ$ for $\mathcal{S}_v > 0$.

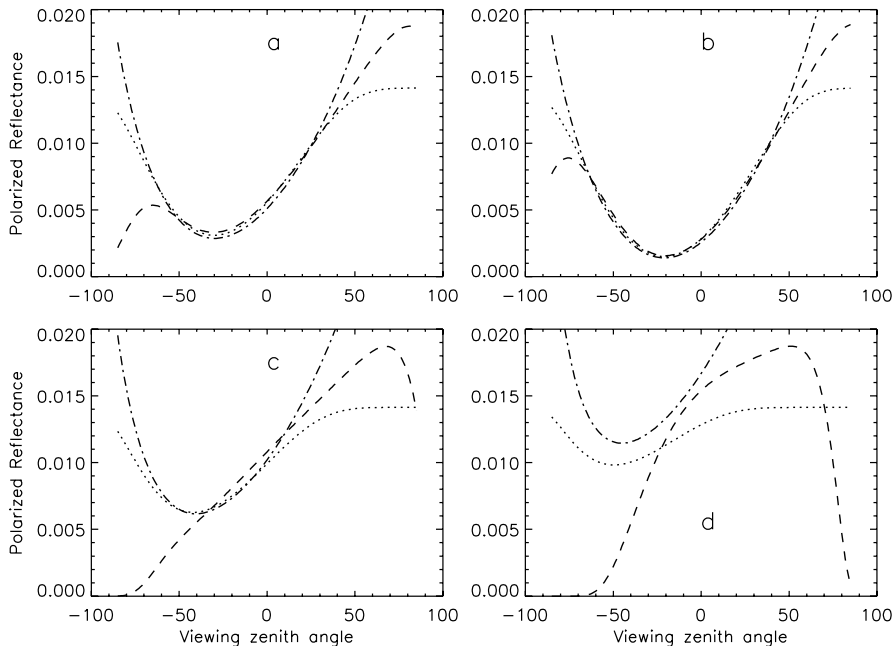


Fig. 6. Angular profiles of the polarized reflectance according to the Nadal–Bréon, modified Fresnel, and linear one-parameter model (dotted, dashed, and dash-dotted curves, respectively) for different solar zenith angles: (a) $\vartheta_0 = 42.68^\circ$, (b) $\vartheta_0 = 30^\circ$, (c) $\vartheta_0 = 60^\circ$, and (d) $\vartheta_0 = 75^\circ$. For all cases $\phi = 45.95^\circ$ for $\vartheta_v < 0$ and $\phi = 225.95^\circ$ for $\vartheta_v > 0$.

BRDF and BPDF models for the specific RSP geometry of flight 1 (corresponding to the best-fit parameters) in the upper left panels, and for different angles of incidence in the other panels. Based on these comparisons for different geometries, it can be seen that the different BRDF and BPDF models diverge outside the viewing zenith angle range of the RSP measurements. Also, the comparison for other angles of incidence shows different angular profiles of the BRDF and BPDF models over the whole range of viewing zenith angles. The differences between the different BRDF as well as BPDF models increase with increasing solar zenith angle (see Figs. 5 and 6).

The differences between the different BRDF and BPDF models outside the angular range of the RSP observations, the dependence of the fitted surface parameters on geometry, and the difference between the different models for geometries other than the one for which the parameters are obtained will result in differences in the top-of-atmosphere total and polarized reflectances, which may affect the retrieval of aerosol properties over land.

In order to investigate the manifestation of these BRDF and BPDF model angular dependence uncertainties on the top-of-atmosphere total and polarized re-

flectance, radiative transfer calculations for a coupled atmosphere–surface system were performed. The top-of-atmosphere total and polarized reflectances are calculated for the same model atmosphere but for the different BRDF and BPDF models tested in the preceding section, with the best-fit parameters listed in [Tables 2](#) and [3](#). We use a radiative transfer model for the coupled atmosphere–surface system (Hasekamp and Landgraf 2002; Hasekamp and Landgraf 2005a) that requires as input the aerosol optical thickness, single-scattering albedo, and scattering matrix. The aerosol parameters are taken from an aerosol model representative of a US background scenario taken from the ECHAM5-HAM model (Stier et al. 2005).

Here, we will quantify two types of errors at the top of the atmosphere:

1. Differences in the top-of-atmosphere total reflectance $\delta_{R_I}^{top}$ caused by differences in the BRDF models $\delta_{R_I}^{sur}$, i.e.,

$$\delta_{R_I}^{top} = \frac{R_I^{top1} - R_I^{top2}}{R_I^{top1}}, \quad \delta_{R_I}^{sur} = \frac{R_I^{sur1} - R_I^{sur2}}{R_I^{sur1}}, \quad (45)$$

where R_I^{top1} and R_I^{top2} are the top-of-atmosphere total reflectances calculated for the same model of the atmosphere but for different BRDF models R_I^{sur1}/π and R_I^{sur2}/π , respectively.

2. Differences in the top-of-atmosphere degree of linear polarization $\Delta_{R_P}^{top}$ caused by differences in BPDF models $\Delta_{R_P}^{sur}$, i.e.,

$$\Delta_{R_P}^{top} = \frac{R_P^{top1} - R_P^{top2}}{R_I^{top}}, \quad \Delta_{R_P}^{sur} = \frac{R_P^{sur1} - R_P^{sur2}}{R_I^{sur}}, \quad (46)$$

where R_P^{top1} and R_P^{top2} are the top-of-atmosphere polarized reflectances calculated for different BPDF models R_P^{sur1}/π and R_P^{sur2}/π , respectively, and the same BRDF model $R_I^{sur}/\pi =$ either R_I^{sur1}/π or R_I^{sur2}/π . The top-of-atmosphere total reflectance R_I^{top} (either R_I^{top1} or R_I^{top2}) depends weakly on the BPDF model.

In general, $\delta_{R_I}^{top}$ and $\Delta_{R_P}^{top}$ depend on the uncertainties of the BRDF and BPDF models (i.e., on $\delta_{R_I}^{sur}$ and $\Delta_{R_P}^{sur}$) at all possible illumination and viewing geometries. As was shown in [Figs. 5](#) and [6](#), the model uncertainties increase with the solar zenith angle. Also, both $\delta_{R_I}^{top}$ and $\Delta_{R_P}^{top}$ depend on the role of the surface–atmosphere interaction in the formation of the top-of-atmosphere signal. Thus, they depend on the atmosphere optical thickness τ_0 and the surface directional albedo $a(\lambda, \vartheta_0)$. Below, we present the results of an analysis of these dependences at two different wavelengths in the visible ($\lambda = 670$ nm) and short-wave infrared ($\lambda = 1589$ nm) regions providing variation of both τ_0 and $a(\lambda, \vartheta_0)$.

[Figure 7](#) shows $\delta_{R_I}^{top}$ and $\delta_{R_I}^{sur}$ when the RPV and Ross–Li BRDF models are used to describe the surface total reflectance. We consider the differences $\delta_{R_I}^{top}$ and $\delta_{R_I}^{sur}$ in the range $-60^\circ < \vartheta_v < 40^\circ$, i.e., the range for which RSP data were available for fitting the BRDF parameters. For soil surfaces, as one can see from

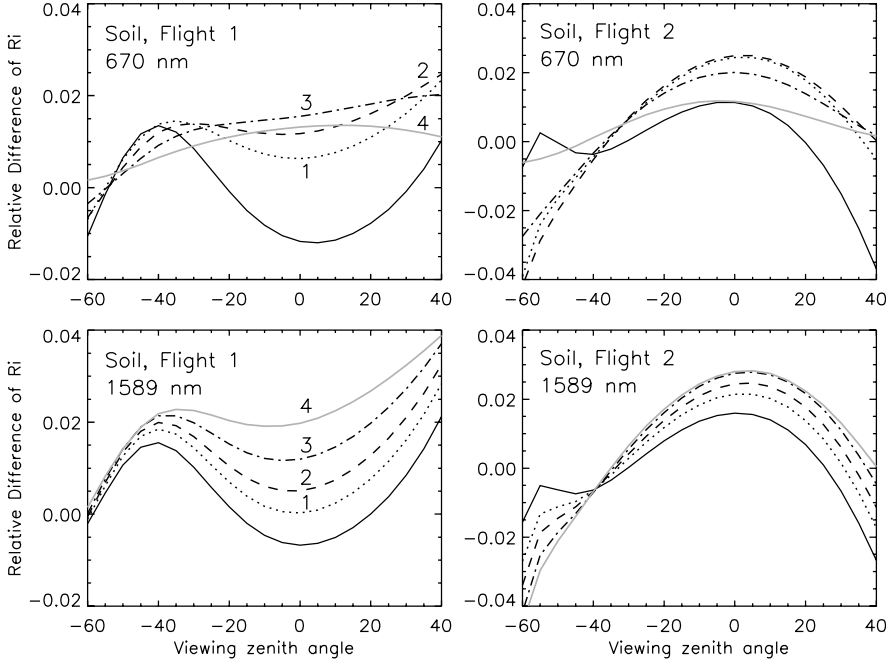


Fig. 7. Angular profiles of the relative differences $\delta_{R_I}^{sur}$ (solid curves) and $\delta_{R_I}^{top}$ (other curves) (see Eq. (45)) for soil surfaces according to the RPV and Ross–Li BRDF models. The results are presented for two wavelengths as well as for different atmosphere optical thicknesses τ_0 and aerosol optical thicknesses τ_{aer} , $\lambda = 670$ nm: (1) dotted curves, $\tau_0 = 0.239$, $\tau_{aer} = 0.179$; (2) dashed curves, $\tau_0 = 0.417$, $\tau_{aer} = 0.357$; (3) dash-dotted curves, $\tau_0 = 0.774$, $\tau_{aer} = 0.714$; and (4) gray curves, $\tau_0 = 1.489$, $\tau_{aer} = 1.429$. $\lambda = 1589$ nm: (1) dotted curves, $\tau_0 = 0.0461$, $\tau_{aer} = 0.0346$; (2) dashed curves, $\tau_0 = 0.081$, $\tau_{aer} = 0.069$; (3) dash-dotted curves, $\tau_0 = 0.15$, $\tau_{aer} = 0.1385$; and (4) gray curves, $\tau_0 = 0.2885$, $\tau_0 = 0.277$.

Fig. 7. $|\delta_{R_I}^{sur}| < 2\%$ almost in the entire range $-60^\circ < \vartheta_v < 40^\circ$ and at ϑ_0 and ϕ corresponding to the geometries of flights 1 and 2. At other ϑ_v , ϑ_0 , and ϕ it may be substantially larger (see, e.g., Fig. 5). As a result, for a coupled atmosphere–surface system, the uncertainties of surface BRDF models may yield values of $|\delta_{R_I}^{top}(\vartheta_v, \vartheta_0, \phi)| > |\delta_{R_I}^{sur}(\vartheta_v, \vartheta_0, \phi)|$. In the same range $-60^\circ < \vartheta_v < 40^\circ$ and at ϑ_0 and ϕ corresponding to the geometries of flights 1 and 2, $|\delta_{R_I}^{top}(\vartheta_v, \vartheta_0, \phi)|$ may be up to 4%–5% (see Fig. 7). The largest errors in $|\delta_{R_I}^{top}(\vartheta_v, \vartheta_0, \phi)|$ occur for $0.1 \leq \tau_0 \leq 0.7$ and albedo values $a(\lambda, \vartheta_0) \geq 10\%$. This is due to a significant contribution of the surface–atmosphere radiative interaction to the top-of-atmosphere reflectance (for details, see Litvinov et al. 2011).

Figure 8 shows $\delta_{R_I}^{top}$ and $\delta_{R_I}^{sur}$ for vegetated surfaces. For such surfaces we have found that the fitted BRDF parameters are less dependent on the illumination

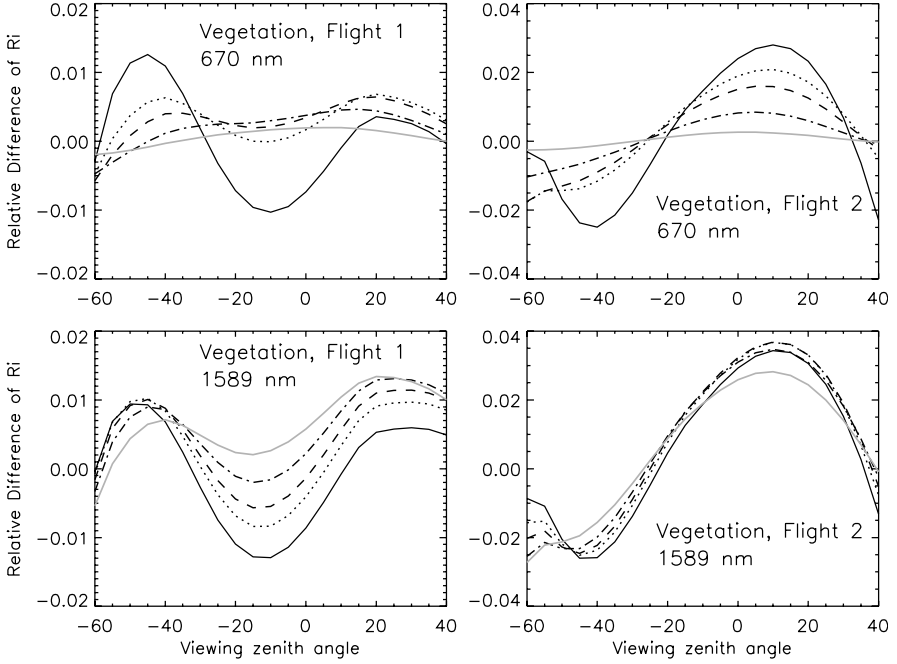


Fig. 8. As in Fig. 7, but for vegetated surfaces.

and viewing geometry, and the uncertainties in the top-of-atmosphere reflectances $|\delta_{R_l}^{top}(\mathcal{G}_v, \mathcal{G}_0, \phi)|$ are caused mainly by $|\delta_{R_l}^{sur}(\mathcal{G}_v, \mathcal{G}_0, \phi)|$, which is mostly $\leq 2\%$. As a result, $|\delta_{R_l}^{top}(\mathcal{G}_v, \mathcal{G}_0, \phi)|$ decreases with increasing atmospheric scattering optical thickness and in all cases is smaller than $|\delta_{R_l}^{sur}(\mathcal{G}_v, \mathcal{G}_0, \phi)|$ (see Fig. 8).

Figure 9 presents the results of calculations of $\Delta_{R_p}^{top}$ and $\Delta_{R_p}^{sur}$ for $\lambda = 670$ nm and for different values of the atmospheric optical thicknesses (see figure captions). Here, R_l^{sur} corresponds to the RPV model, while R_p^{sur1} and R_p^{sur2} correspond to the modified Fresnel BPDF model, Eq. (30), and the Nadal–Bréon model, Eq. (25), respectively.

Both for soil and vegetated surfaces, the top-of-atmosphere differences in the degree of linear polarization $\Delta_{R_p}^{top}$ are determined primarily by the errors in the surface polarized reflectance parameterization $\Delta_{R_p}^{sur}$ for the illumination and viewing geometries at which the BPDF models were fitted to the RSP data (see Fig. 9). This is due to the fact that the surface polarized reflectance is small both for soil and vegetated surfaces, and surface–atmosphere interactions play a minor role in the top-of-atmosphere uncertainty $\Delta_{R_p}^{top}$. As was shown in Fig. 6, the modified Fresnel model is able to describe the surface polarized reflectance much better than the Nadal–Bréon model. Thus, $\Delta_{R_p}^{sur}$ as well as $\Delta_{R_p}^{top}$ in Fig. 9 are mainly caused by the inability of the Nadal–Bréon model to fit the observations. This means that for the widely used Nadal–Bréon model, errors in calculations of the

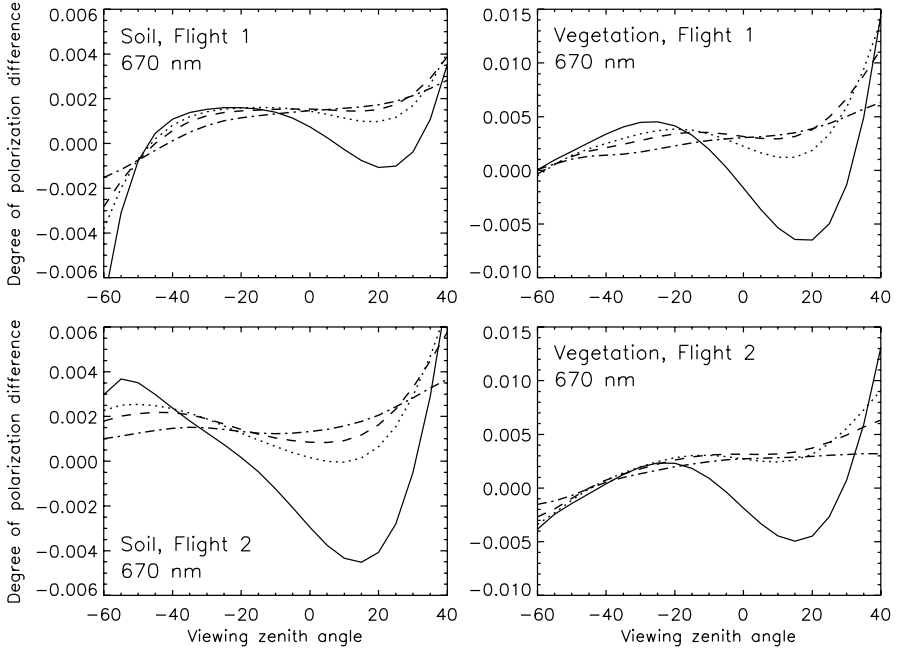


Fig. 9. Angular profiles of the relative differences $\Delta_{R_p}^{sur}$ (solid curves) and $\Delta_{R_p}^{top}$ (other curves) (see Eq. (46)) for soil and vegetated surfaces according to the Nadal–Bréon and modified Fresnel BPDF models and the RPV BRDF model. The results are shown for $\lambda = 670$ nm. The dotted, dashed, and dash-dotted curves are obtained for $\tau_0 = 0.239$, $\tau_0 = 0.417$, and $\tau_0 = 0.774$, respectively.

top-of-atmosphere degree of linear polarization may be as large as 0.005–0.015. This exceeds, for example, the degree of linear polarization uncertainties of the RSP and APS instruments (Cairns et al. 1999; Mishchenko et al. 2004; Mishchenko et al. 2007b). A similar conclusion can be drawn for the one-parameter model (Eq. (29)) of surface polarized reflectance. In general, $|\Delta_{R_p}^{top}| \sim |\Delta_{R_p}^{sur}|$ for low optical thicknesses ($\tau_0 \leq 0.5$) and decreases with growing τ_0 (see Fig. 9).

5. Summary and discussion

In this chapter, we considered the applicability of the existing semi-empirical BRDF and BPDF models to aerosol-property retrievals over land. Three aspects of the BRDF and BPDF models were discussed: (i) constraints on the spectral and angular dependences of BRDF and BPDF models (Section 4.1), (ii) the capability to correctly describe the surface total and polarized reflectance for a given geometry of illumination and measurement (Section 4.2), and (iii) the capability to correctly describe the surface total and polarized reflectance at all other geometries of

illumination and viewing (Section 4.3).

It can be concluded from Section 4.1 that for media with either small multiple-scattering contributions or a similar geometric dependence of the single and multiple scattering, for most geometries the surface total reflectance can be represented as a product of a wavelength-dependent term and a term depending on illumination and viewing geometries only. This provides constraints on the angular and spectral dependences of the BRDF models. Some soil and vegetated surfaces can provide these conditions, as was demonstrated in this chapter as well as in (Litvinov et al. 2010). For such media the spectral invariance of the geometric term can be exploited in aerosol retrieval algorithms over land (Flowerdew and Haigh 1996; Veeffkind et al. 1998; Diner et al. 2005; Kokhanovsky and de Leeuw 2009). However, it should be noted that this constraint of spectral independence of the angular scattering characteristics is not necessarily valid for all geometries and scattering media. Therefore, care must be taken when applying it to the BRDF.

For BPDF models, in most cases it is possible either to neglect the spectral dependence of the BPDF model parameters or to estimate these parameters using measurements in spectral channels where the aerosol contribution is not significant. In the latter case, the BPDF model parameters should not differ substantially for other spectral channels (Deuze et al. 2001; Waquet et al. 2009a).

In Section 4.2 it was demonstrated that the semi-empirical BRDF and BPDF models demonstrate a good ability to fit measured data for a given geometry of illumination and measurement. If they do not fit the data well then it is possible to find a simple way to modify them, as was demonstrated by Maignan et al. (2004), Maignan et al. (2009), and Litvinov et al. (2011).

It follows from Section 4.3 that in most cases, the semi-empirical BRDF and BPDF models are unable to describe the surface total and polarized reflectance at all possible illumination and viewing geometries. This leads to angular dependence uncertainties in the BRDF and BPDF models that can manifest themselves at the top of the atmosphere and thereby contribute to uncertainties in the retrieved aerosol properties. The BRDF model uncertainties may lead to uncertainties of 2%–5% in the top-of-atmosphere total reflectance. This may have a significant impact on the retrieval of aerosol properties and in particular on the retrieved real and imaginary parts of the refractive index (Hasekamp and Landgraf 2007). The uncertainties in the angular dependence of the polarized reflectance can be reduced by introducing a semi-empirical BPDF model providing a better fit to observations for a particular geometry. It is more difficult to reduce the uncertainties in the total reflectance since for this purpose the BRDF model must accurately describe the measured angular dependences for all possible geometries, as the total reflectance is more strongly affected by atmosphere–surface interactions than the polarized reflectance.

Overall, we conclude that the use of semi-empirical BRDF models is expected to cause significant uncertainties in the retrieved aerosol properties over land. For semi-empirical BPDF models this effect is much weaker. To reduce the uncertainties caused by BRDF models, either a reliable algorithm for the separation of the

atmospheric and surface signals (see, e.g., Dubovik et al. 2010) or a more physically based model of the BRDF is required. In such physically-based models, the BRDF and BPDF model parameters would be related to each other, and physical constraints would be imposed by the surface structure and composition (see, e.g., the discussion in Mishchenko et al. 2011).

Acknowledgments

We are grateful to O. Dubovik and K. Knobelspiesse for useful discussions. This research was supported by the Dutch User Support Program under project GO-AO/03, the NASA Glory Mission project, and the NASA Radiation Sciences Program managed by Hal Maring.

References

- Bréon, F.-M., D. Tanré, P. Lecomte, and M. Herman, 1995: Polarized reflectance of bare soil and vegetation: measurements and models. *IEEE Trans. Geosci. Remote Sens.* **33**, 487–499.
- Cairns, B., E. E. Russell, and L. D. Travis, 1999: The Research Scanning Polarimeter: calibration and ground-based measurements. *Proc. SPIE* **3754**, 186–196.
- Chowdhary, J., B. Cairns, M. I. Mishchenko, et al., 2005: Retrieval of aerosol scattering and absorption properties from photo-polarimetric observations over the ocean during the clams experiment. *J. Atmos. Sci.* **62**, 1093–1117.
- Dubovik, O., M. Herman, A. Holdak, et al., 2010: Statistically optimized inversion algorithm for enhanced retrieval of aerosol properties from spectral multi-angle polarimetric satellite observations. *Atmos. Meas. Tech. Discuss.* **3**, 4967–5077.
- Deuze, J. L., F. M. Bréon, C. Devaux, et al., 2001: Remote sensing of aerosols over land surfaces from POLDER-ADEOS-1 polarized measurements. *J. Geophys. Res.* **106**, 4913–4926.
- Diner, D. J., J. V. Martonchik, R. A. Kahn, et al., 2005: Using angular and spectral shape similarity constraints to improve MISR aerosol and surface retrievals over land. *Remote Sens. Environ.* **94**, 155–171.
- Flowerdew, R. J., and J. D. Haigh, 1996: Retrieval of aerosol optical thickness over land using the ATSR-2 dual-look satellite radiometer. *Geophys. Res. Lett.* **23**, 351–354.
- Hansen, J., L. Nazarenko, R. Ruedy, et al., 2005: Earth's energy imbalance: confirmation and implications. *Science* **308**, 1431–1435.
- Hapke, B. B., 1981: Bidirectional reflectance spectroscopy. 1. Theory. *J. Geophys. Res.* **86**, 3039–3054.
- Hasekamp, O. P., and J. Landgraf, 2002: A linearized vector radiative transfer model for atmospheric trace gas retrievals. *J. Quant. Spectrosc. Radiat. Transfer* **75**, 221–238.
- Hasekamp, O. P., and J. Landgraf, 2005a: Linearization of vector radiative transfer with respect to aerosol properties and its use in satellite remote sensing. *J. Geophys. Res.* **110**, D04203.
- Hasekamp, O. P., and J. Landgraf, 2005b: Retrieval of aerosol properties over the ocean from multispectral single-viewing angle measurements of intensity and polarization: Retrieval approach, information content, and sensitivity study. *J. Geophys. Res.* **110**, D20207.
- Hasekamp, O. P., and J. Landgraf, 2007: Retrieval of aerosol properties over land surfaces: capabilities of multiple-viewing-angle intensity and polarization measurements. *Appl. Opt.* **46**, 3332–3344.

- Hovenier, J. W., C. van der Mee, and H. Domke, 2004: *Transfer of Polarized Light in Planetary Atmospheres* (Springer, Berlin).
- Kaufman, Y. J., and D. Tanre, 1992: Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. *IEEE Trans. Geosci. Remote Sensing*, **30**, 261–270.
- Knobelspiesse, K. D., B. Cairns, C. B. Schaaf, et al., 2008: Surface BDRF estimation from an aircraft compared to MODIS and ground estimates at the Southern Great Plains site. *J. Geophys. Res.* **113**, D20105.
- Kokhanovsky, A. A., and G. de Leeuw (Eds.), 2009: *Satellite Remote Sensing Over Land* (Springer, Berlin).
- Li, X., and A. H. Strahler, 1992: Geometrical-optical bidirectional reflectance modeling of the discrete crown vegetation canopy: effect of crown shape and mutual shadowing. *IEEE Trans. Geosci. Remote Sens.* **30**, 276–292.
- Litvinov, P., O. Hasekamp, B. Cairns, and M. Mishchenko, 2010: Reflection models for soil and vegetation surfaces from multiple-viewing angle photopolarimetric measurements. *J. Quant. Spectrosc. Radiat. Transfer* **111**, 529–539.
- Litvinov, P., O. Hasekamp, and B. Cairns, 2011: Models for surface reflection of radiance and polarized radiance: comparison with airborne multi-angle photopolarimetric measurements and implications for modeling top-of-atmosphere measurements. *Remote Sens. Environ.* **115**, 781–792.
- Maignan, F., F. M. Bréon, and R. Lacaze, 2004: Bidirectional reflectance of Earth targets: evaluation of analytical models using a large set of spaceborne measurements with emphasis on the Hot Spot. *Remote Sens. Environ.* **90**, 210–220.
- Maignan, F., F.-M. Bréon, E. Fedele, and M. Bouvier, 2009: Polarized reflectances of natural surfaces: Spaceborne measurements and analytical modeling. *Remote Sens. Environ.* **113**, 2642–2650.
- Martonchik, J. V., D. J. Diner, B. Pinty, et al., 1998: Techniques for the retrieval of aerosol properties over land and ocean using multiangle imaging. *IEEE Trans Geosci Remote Sens.* **36**, 1212–1227.
- Minnaert, M., 1941: The reciprocity principle in lunar photometry. *Astrophys. J.* **93**, 403–410.
- Mishchenko, M. I., and L. D. Travis, 1997: Satellite retrieval of aerosol properties over the ocean using polarization as well as intensity of reflected sunlight. *J. Geophys. Res.* **102**, 16989–17013.
- Mishchenko, M. I., B. Cairns, J. E. Hansen, et al., 2004: Monitoring of aerosol forcing of climate from space: analysis of measurement requirements. *J. Quant. Spectrosc. Radiat. Transfer* **88**, 149–161.
- Mishchenko, M. I., L. D. Travis, and A. A. Lacis, 2006: *Multiple Scattering of Light by Particles* (Cambridge University Press, Cambridge, UK).
- Mishchenko, M. I., I. V. Geogdzhayev, B. Cairns, et al., 2007a: Past, present, and future of global aerosol climatologies derived from satellite observations: a perspective. *J. Quant. Spectrosc. Radiat. Transfer* **106**, 325–347.
- Mishchenko, M. I., B. Cairns, G. Kopp, et al., 2007b: Accurate monitoring of terrestrial aerosols and total solar irradiance: introducing the Glory mission. *Bull. Amer. Meteorol. Soc.* **88**, 677–691.
- Mishchenko, M. I., V. P. Tishkovets, L. D. Travis, et al., 2011: Electromagnetic scattering by a morphologically complex object: fundamental concepts and common misconceptions. *J. Quant. Spectrosc. Radiat. Transfer* **112**, 671–692.
- Muñoz, O., H. Volten, J. W. Hovenier, et al., 2007: Scattering matrix of large Saharan dust particles: experiments and computations. *J. Geophys. Res.* **112**, D13215.

- Nadal, F., and F.-M. Bréon, 1999: Parameterization of surface polarized reflectance derived from POLDER spaceborne measurements. *IEEE Trans. Geosci. Remote Sens.* **37**, 1709–1718.
- Rahman, H., B. Pinty, and M. M. Verstraete, 1993: Coupled Surface-Atmosphere Reflectance (CSAR) model 2. Semiempirical surface model usable with NOAA Advanced Very High Resolution Radiometer Data. *J. Geophys. Res.* **98**, 20791–20801.
- Remer, L. A., Y. J. Kaufman, D. Tanré, et al., 2005: The MODIS aerosol algorithm, products, and validation. *J. Atmos. Sci.* **62**, 947–973.
- Rondeaux, G., and M. Herman, 1991: Polarization of light reflected by crop canopies. *Remote Sens. Environ.* **38**, 63–75.
- Ross, J. K., 1981: *The Radiation Regime and Architecture of Plant Stands* (Dr. W. Junk Publishers, The Hague, The Netherlands).
- Savenkov, S. N., R. S., Mutiah, and Y. A. Oberemok, 2003: Transmitted and reflected scattering matrices from an English oak leaf. *Appl. Opt.* **42**, 4955–4962.
- Roujean, J.-L., M. Leroy, and P.-Y. Deschamps, 1992: A bidirectional reflectance model of the Earth's surface for the correction of remote sensing data. *J. Geophys. Res.* **97**, 20455–20468.
- Schaepman-Strub, G., M. E. Schaepman, T. H. Painter, et al., 2006: Reflectance quantities in optical remote sensing—definitions and case studies. *Remote Sens. Environ.* **103**, 27–42.
- Spurr, R. J. D., 2004: A new approach to the retrieval of surface properties from earthshine measurements. *J. Quant. Spectrosc. Radiat. Transfer* **83**, 15–46.
- Stier, P., J. Feichter, S. Kinne, et al., 2005: The aerosol-climate model ECHAM5-HAM. *Atmos. Chem. Phys.* **5**, 1125–1156.
- Strahler, A. H., J.-P. Muller, and MODIS Science Team, 1999: MODIS BRDF/albedo product: algorithm theoretical basis document version 5. MODIS product ID: MOD43 Version 5.0, http://modis.gsfc.nasa.gov/data/atbd/land_atbd.php (accessed 15 November 2010).
- Tanré, D., Y. J. Kaufman, M. Herman, and S. Mattoo, 1997: Remote sensing of aerosol properties over oceans using the MODIS/EOS spectral radiances. *J. Geophys. Res.* **102**, 971–988.
- Tsang, L., J. A. Kong, and R. T. Shin, 1985: *Theory of Microwave Remote Sensing* (Wiley, New York).
- Tishkovets, V. P., E. V. Petrova, and K. Jockers, 2004: Optical properties of aggregate particles comparable in size to the wavelength. *J. Quant. Spectrosc. Radiat. Transfer* **86**, 241–265.
- Veefkind, J. P., G. de Leeuw, P. Durkee, 1998: Retrieval of aerosol optical depth over land using two-angle view satellite radiometry during TARFOX. *Geophys. Res. Lett.* **25**, 3135–3138.
- Wanner, W., X. Li, and A. H. Strahler, 1995: On the derivation of kernels for kernel-driven models of bidirectional reflectance. *J. Geophys. Res.* **100**, 21 077–21 089.
- Waquet, F., B. Cairns, K. Knobelspiesse, et al., 2009a: Polarimetric remote sensing of aerosols over land. *J. Geophys. Res.* **114**, D01206.
- Waquet, F., J.-F. Leon, B. Cairns, et al., 2009b: Analysis of the spectral and angular response of the vegetated surface polarization for the purpose of aerosol remote sensing over land. *Appl. Opt.* **48**, 1228–1236.
- Woolley, J. T., 1971: Reflectance and transmittance of light by leaves. *Plant Physiol.* **47**, 656–662.

An estimation of surface albedo from the SEVIRI/MSG observing system by using POLDER BRDF measurements

Igor Pokrovsky¹, Oleg Pokrovsky^{1*}, and Jean-Louis Roujean²

¹ Voeikov Main Geophysical Observatory, 7 Karbyshev Str., St.-Petersburg, 194021, Russia

² CNRM/GAME (Météo-France/CNRS), 42, avenue Gaspard Coriolis, 31057 Toulouse Cedex, France

Abstract. In this chapter we develop a concept related to the significant information extracted from the Bi-directional Reflectance Distribution Function (BRDF) of terrestrial targets. The main issues are the choice of the BRDF model, the solution of the inverse problem, and the accuracy assessment of the estimated albedo. The present concept is based on the fact that the exact solution to the inverse problem belongs to a statistically significant region centered on the least squares solution (LSS). Nonetheless, the LSS may be useless if the matrix inversion yields an ill-posed problem. It is then recommended to seek an alternative solution, which will yield a similar confidence interval, but will be more physically sound. A list of 15 kernels entering in a basic model is examined by means of factor analysis performed in vector space, which spans all known kernels. The application is carried out with synthetic angular data generated for the SEVIRI/MSG observing system. Models are evaluated based on statistical results, minimum of squared sum of residuals (SSR) and maximum of explained variance, after adjustment on reflectance data corresponding to a wide set of land cover types. Since the matrix of the model is almost singular, we identify an optimal subset model consisting of 8 kernels, which has a higher conditioned index and falls within the 95% confidence interval. It is found that the reflectance predicted by a multi-kernel model is consistent with measurements. The idea in opting for a multi-kernel approach comes from the necessity to perform a higher angular resolution for the BRDF retrieval. Inversion experiments confirm an advantage of the composite model over conventional three-parameter models in accuracy assessment of reflectance and albedo in the case of uniform and restricted angular samplings. Three methods are considered: statistical inversion (provided by the LSS), ridge regression, and statistical regularization. The two latter methods are recommended to solve the ill-conditioned inverse problem. Statistical regularization uses *a priori* statistical information. The inversion numerical experiment with SEVIRI/MSG angular geometry shows that only ridge regression provides a reasonable solution when a

* Corresponding author. E-mail: pokrov_06@mail.ru

composite model is used. In addition, ridge regression and statistical regularization methods provide physically acceptable solutions in terms of BRDF and albedo predictability, even for 3-parameters models. It is advised that the LSS be implemented only in the middle of the summer season in the Northern Hemisphere. Otherwise, the use of ridge regression and statistical regularization is recommended to retrieve BRDF and albedo at other time periods in extra-tropical latitudes.

Keywords: remote sensing, land surface albedo, Bi-directional Reflectance Distribution Function

1. Introduction

The implementation of assimilation schemes describing the surface properties in numerical weather prediction (NWP) models (Giard and Bazile 2000) requires fields of land surface conditions to be frequently updated. In this respect, the potential of remote-sensing observations has been widely exploited, since they yield the only means to get a frequent global coverage of continental areas. Until now, the Advanced Very High Resolution Radiometer (AVHRR) on board the NOAA (National Oceanic and Atmospheric Administration) satellite series has been the most popular observing system for a wide range of applications, in particular, in meteorology. An argument sustaining the use of the AVHRR time series is the long-term acquisition, which allows perpetuating the investment of the producers in methodological developments and to supply the user community with multi-year observations. However, the AVHRR is flown onboard a polar system, which means that for most latitudes, a given point on the Earth is observed by the afternoon satellite at most twice every day assuming cloud-free sky conditions. In fact, the frequency of clear scenes may be rather low for certain regions, which places a severe limit on the extent at which the AVHRR can be relied upon in order to retrieve surface properties of a relatively low temporal inertia. The surface albedo, which is clearly of first priority in NWP models, varies generally slowly in time, typically on a weekly scale. Nevertheless, users' requirements can be more stringent, in particular considering snow conditions for which surface information may be required on a daily basis. In this respect, the advent of geo-stationary satellites like Meteosat Second Generation (MSG) is of crucial interest for its potential to contribute to advanced surveys over the continents and circumvent the lack of data due to high cloudiness in certain regions.

The determination of a surface albedo product requires the implementation of a number of data-processing steps of which cloud screening and the removal of atmospheric effects, namely water vapor content and aerosols loadings, are the most important. Furthermore, satellite systems provide a sparse angular sampling of the Bi-directional Reflectance Distribution Function (BRDF), while a hemispherical knowledge of this latter is necessary to estimate surface albedo. Flown on a sun-synchronous platform, the AVHRR scans the same target daily under changing viewing conditions but similar sun geometry, unless data from the morn-

ing satellite are also employed. After a few days, it can provide a sampling of the BRDF close to the principal plane, the plane containing the sun, the sensor and the target, in both forward and backscattering directions. On the other hand, since the SEVIRI/MSG mission is a geo-stationary sensor system, it will cover a large range of solar angular measurements but at fixed-view zenith angles and for various sets of relative azimuths. In this latter case, the BRDF sampling will be a warping of the perpendicular plane toward the backscattering area, away from the tropical belt, with the exception of the summer season. It is anticipated that the lack of data in the principal plane, where angular effects are amplified, will lead to biased estimates of the BRDF and thereby surface albedo. When only sparse angular measurements are available, it is difficult to calibrate correctly BRDF models. In many cases the ill-conditioned index (ICI), the ratio of minimum to maximum eigenvalues of a matrix to be inverted, lies in an interval of small values. Therefore, the retrieved BRDF model coefficients are sensitive to small perturbations of reflectance values registered at the satellite level, and unphysical solutions may arise. In some respect, this situation might be improved by the selection of optimal angular subsets providing higher ICI values than initially (Pokrovsky and Roujean 2003b). Since the surface signal is contaminated through the atmospheric pathway, it contains residual noise that hampers the ability to segregate optimum geometries in a systematic manner. Note that partial cloudiness is still not properly detected by existing remote-sensing methods (see Bicheron and Leroy 2000) and cannot be described by radiation-transfer methods.

This chapter is devoted to developing the issues described above with an application to SEVIRI/MSG geometry. The consideration of scope questions and their formulation are presented in Section 2. The theoretical background for the solution to an ill-posed inverse problem is developed in Section 3. Selection of kernels for composite BRDF model and its substantiation is discussed in Section 4. An application of the proposed method to the acquisition of regularized solutions is described in Section 5. We show that albedo estimation and its uncertainty may be derived directly from reflectance data without explicit BRDF model inversion in Section 6. Finally, Section 7 contains some recommendations for future studies.

2. Statement of problem

The accuracy assessment of land surface albedo from SEVIRI/MSG measurements depends on the quality of the data with respect to these two issues: i) the development of a computer screening technique for quality control of input data sets; and ii) the development of an advanced surface BRDF model inversion technique to be more resistant to atmospheric effects and model performance. Our approach aims at implementing an operational procedure working as an heuristic system. In this regard, the POLDER experiment (see Deschamps et al. 1994) actually yields the most consistent global surface angular data sets. The POLDER BRDF database is considered here to initiate this training effort based on the 16

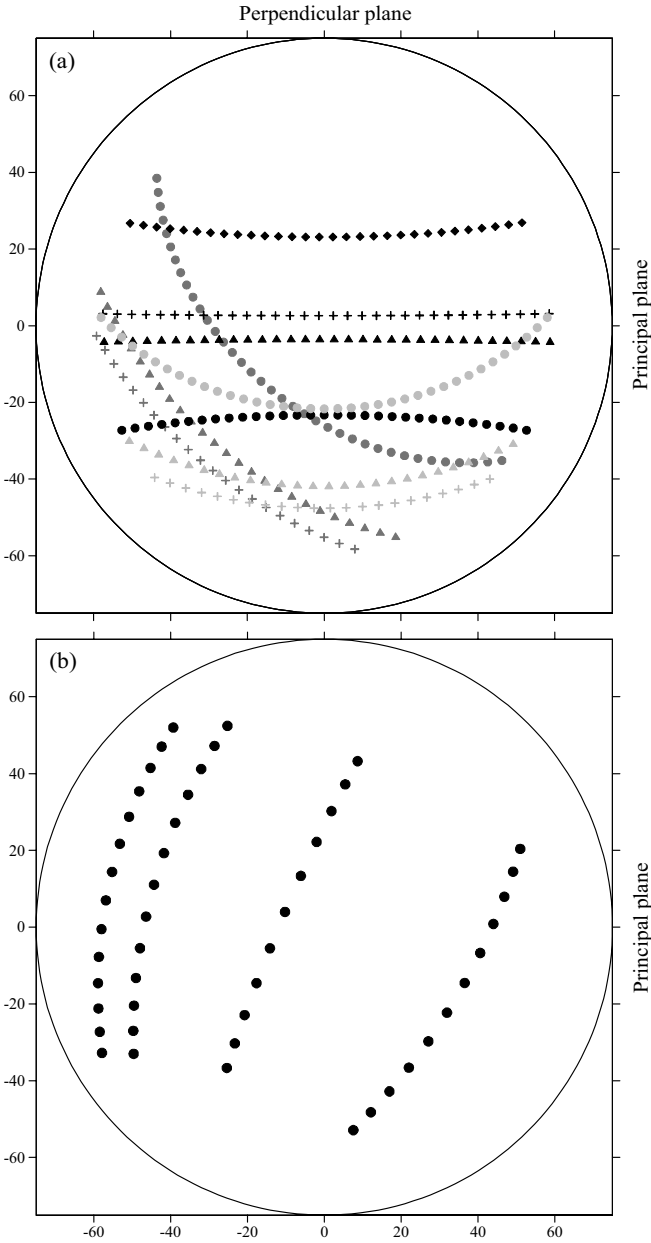


Fig. 1. (a) Polar graph of SEVIRI/MSG angular sampling geometry for points with the coordinates 45°N, 45°E (black), 45°N, 0°E (light grey), and 0°N, 0°E (dark grey) on 3 different dates: March 15 (+), June 15 (●) and September 15 (▲). (b) Examples of POLDER angular geometries for 4 consecutive days.

IGBP (International Geosphere and Biosphere Programme) land cover classes. POLDER measurements provide a better BRDF sampling (Fig. 1b) than SEVIRI data (Fig. 1a). Thus, the POLDER experiment appears suitable to accumulate the *a priori* knowledge on BRDF for major landscapes and biome types. Among the several thousands of POLDER BRDF data sets prepared by the Centre National d'Etudes Spatiales, four hundred sites were considered with the following objectives:

- To test the prototype of the operational procedure;
- To examine the visual and control procedure, rejecting angular data contaminated by residual atmospheric contamination;
- To simulate geometrical and reflectance data for the SEVIRI/MSG sensor;
- To perform inversion experiments with simulated data to evaluate expected error bars for albedo retrieval for different seasons and latitudes;
- To acquire *a priori* statistics for albedo values (black and white sky albedo-BSA and WSA, respectively), BRDF model coefficients and reflectance attributed to each of the 16 IGBP classes.

Kernel-driven models have been used extensively to adjust the observed surface BRDF in order to provide further estimates of key surface parameters like albedo (Strahler and Muller 1999; Roujean et al. 1992; Wanner et al. 1995). This class of models assumes that the BRDF can be developed using a few angular functions describing elementary mechanisms of matter-photon interactions, the weight of each of them being determined by empirical coefficients. The purpose here is to get a better accuracy on these retrieved BRDF coefficients by imposing constrained intervals. For this, we form *a priori* statistics for each of the IGBP classes within the POLDER data set (Bicheron and Leroy 2000; Pokrovsky et al. 2003a). The volume of POLDER BRDF data sets – up to several thousands of samples – supports the accumulation of BRDF coefficient statistics for each class. Since an interactive device was not possible in this case, an automatic control procedure had to be developed. Note this procedure should allow selection of the BRDF model as well as the combination of data subsets. The operational procedure should then incorporate the following modules:

1. Input detection of outliers to reject unusual angular reflectance measurements;
2. BRDF model inversion using *a priori* information on model coefficients;
3. Output control of retrieved model coefficients and estimated albedo;
4. Updating of *a priori* knowledge of albedo, BRDF model coefficients and reflectance in the form of relevant statistics.

The present work is focused on the four above modules, with the exception of the third one that is developed in a companion paper (Pokrovsky et al. 2003a,b). Recent advances in computer techniques for data selection serve to accumulate statistical information in order to compensate for the lack of detailed sun-view satellite geometries. The proposed operational procedure will be efficient in excluding any inappropriate albedo estimates with the trade-off of a greater processing

time due to modules 1 and 4. Module 1 uses Fisher statistics criteria (see Pokrovsky and Roujean 2003a), which permits the rejection of unusual measurements given a significant level of probability. We find that this level should be in the range 90%–95%. Higher levels are less robust to the occurrence of outliers; whereas, lower levels are unacceptable as some relevant information may be lost.

In fact, module 3 incorporates more sophisticated criteria than module 1. A first control criterion is based on T -multivariate statistics (see Anderson 1958), which permits the detection of outliers in regard to their cross-correlation. The four criteria are based on radiometric indices: the bowl shape and dome shape indices (BSI and DSI), WSA and BSA variances introduced in Li et al. (2001). In fact, we demonstrate that only three are independent (Pokrovsky et al. 2003b). This permits the performance of robust control techniques for the Li-Sparse (Wanner et al. 1997) and Roujean (Roujean et al. 1992) BRDF models.

For module 2, we carry out a thorough preliminary comparison of different inversion techniques: (i) statistical inversion method (Pokrovsky and Roujean 2003a), (ii) SVD (singular value decomposition), (iii) ridge regression, and (iv) statistical regularization (Pokrovsky 1984). The last method uses *a priori* statistics (mean values, covariance matrix) on BRDF model coefficients (Li et al. 2001). Preliminary experiments rely on ground-based data (Kimes 1983) and POLDER space-borne data (Bicheron and Leroy 2000), including both uniform and restricted-view angle geometries. These experiments reveal that for a uniform angular sampling, all the methods perform similarly. However, given a limited angular sampling corresponding to MSG observations, the results differ. The best accuracy is obtained with the regularization technique, which provides the smallest error bars. The next favourite is ridge regression, which gives slightly worse results. The statistical inversion method is ranked third, its difference with the ridge regression being noticeable for very sparse angular samplings observed in autumn and wintertime. The SVD method appears as the least appropriate method. It is based on the truncation of eigenvector expansion of the solution to the ill-posed inverse problem. Numerical experiments reveal that in the case of the SEVIRI/MSG geometry the exclusion of one eigenvector from the 3 by 3 matrix expansion associated with the BRDF model is prejudicial. As a result, the dependence of albedo derived by the SVD method on sun geometry is too flat to be eligible. Such a result may be explained by a decrease of the BRDF model angular resolution with only two parameters. Conversely, our approach assumes a potential enhancement in angular resolution with model expansion beyond three kernels. Indeed, accumulating *a priori* statistics is an important part of a self-learning system. Thanks to statistical POLDER information, keeping in mind that each of selected IGBP classes should be regularly updated, modules 2 and 4 should be closely linked.

3. Generalization of inversion methodology

Considerable attention is currently paid to the biased estimate of the parameters of a linear regression applied to kernel BRDF models (Li et al. 2001). Such

concerns are justified by the inability of the classical least squares (LS) method to provide reasonable point estimates when a BRDF matrix inversion is an ill-posed problem. The goal is to obtain minimum variance in the class of linear unbiased estimators in the Gauss-Markov sense. The LS estimators can have very large variances when the arrays of the BRDF matrix are collinear. An alternative solution is to consider the biased estimators instead of the LS estimator. Hoerl and Kennard (1970) have proposed the use of a ridge regression to overcome those difficulties. In fact, deviation from LS is a well-known problem. Whenever one uses a subset selection procedure to obtain an equation with a reduced number of variables, one compromise of the LS principle is to set some coefficients equal to zero (Aitken 1974). Whether or not LS is used for the reduced problem, a deviation from the LS estimates for the full model has been presented (Arvesen and McCabe 1975).

The purpose of this paper is not to propose or justify the use of a particular method for constructing alternatives to the LS, since the literature on statistics abounds with such material. Instead, the aim is to provide a framework for evaluation of such estimates (McDonalds and Schwing 1973; McDonald 1975) and to demonstrate their application to the problem of BRDF model inversion in an ill-conditioned numerical environment. The statistical model is the standard multiple regression model given by

$$\mathbf{y} = \mathbf{A} \cdot \mathbf{b} + \boldsymbol{\varepsilon}, \quad (1)$$

where $\mathbf{y}^T = (y_1, \dots, y_n)$ is the vector of the observed variable (T stands for “transposed”), \mathbf{A} is the n by $p + 1$ model matrix of known values with the first column standing for the unity vector, and $\mathbf{b}^T = (b_1, \dots, b_{p+1})$ is the vector of unknown coefficients. The error vector $\boldsymbol{\varepsilon}^T = (\varepsilon_1, \dots, \varepsilon_n)$ is assumed to be normally distributed with zero mean and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} = \sigma^2 \mathbf{I}$, where σ^2 is unknown, i.e., $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$. Here $N(*, **)$ is a conventional notation of the normal (Gaussian) distribution with two background parameters: $*$ is the vector of mean values, and $**$ is the covariance matrix (see Anderson 1958). In short, we can write:

$$\mathbf{y} \sim N(\mathbf{A} \cdot \mathbf{b}, \sigma^2 \mathbf{I}). \quad (2)$$

The LS estimator of \mathbf{b} from Eq. (1) is given by

$$\hat{\mathbf{b}}_{\text{LS}} = (\mathbf{A}^T \cdot \mathbf{A})^{-1} \cdot \mathbf{A}^T \cdot \mathbf{y}. \quad (3)$$

For Eq. (3), the sum of squares of residuals (SSR) is given by

$$R_{\text{SSR}}(\hat{\mathbf{b}}_{\text{LS}}) = \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (4)$$

where \hat{y}_i is the i -th component of the vector $\hat{\mathbf{y}} = \mathbf{A} \cdot \hat{\mathbf{b}}$. Note that Eq. (4) has an alternative representation:

$$R_{\text{SSR}}(\hat{\mathbf{b}}_{\text{LS}}) = \mathbf{y}^T [\mathbf{I} - \mathbf{A} \cdot (\mathbf{A}^T \cdot \mathbf{A})^{-1} \cdot \mathbf{A}^T] \cdot \mathbf{y}. \quad (5)$$

The usual unbiased estimate of the parameter σ^2 is given by the sum of squared residuals divided by its degrees of freedom (Behnken and Draper 1972; Draper and Smith 1998):

$$s^2 = R_{SSR}(\hat{\mathbf{b}}_{LS}) / (n - p - 1). \tag{6}$$

A $(1 - \alpha) \times 100\%$ confidence region for parameter \mathbf{b} is given by

$$S_\alpha = \{\mathbf{b} : (\hat{\mathbf{b}}_{LS} - \mathbf{b})^T \cdot (\mathbf{A}^T \cdot \mathbf{A}) \cdot (\hat{\mathbf{b}}_{LS} - \mathbf{b}) < s^2(p + 1)F_{p+1, n-p-1, 1-\alpha}\}, \tag{7}$$

where $F_{p+1, n-p-1, 1-\alpha}$ is the upper $\alpha \in (0, 1)$ quantile of the F distribution with $p + 1$ and $n - p - 1$ degrees of freedom (Anderson 1958).

For any $p + 1$ dimensional vector \mathbf{b} , let

$$D(\mathbf{b}) = (\hat{\mathbf{b}}_{LS} - \mathbf{b})^T \cdot (\mathbf{A}^T \cdot \mathbf{A}) \cdot (\hat{\mathbf{b}}_{LS} - \mathbf{b}). \tag{8}$$

It is clear that $D(\mathbf{b})$ is a square distance, which indicates how far \mathbf{b} is from $\hat{\mathbf{b}}_{LS}$ in appropriate metric. Since $S_\alpha = \{\mathbf{b} : D(\mathbf{b}) \leq d_\alpha\}$, where

$$d_\alpha = s^2(p + 1)F_{p+1, n-p-1, 1-\alpha}, \tag{9}$$

the confidence region may be viewed as sets of \mathbf{b} close to the LS solution $\hat{\mathbf{b}}_{LS}$ (Obenchain 1977; McCabe 1979). If α is fixed and \mathbf{b} is the true parameter vector, then $P_b\{D(\mathbf{b}) \leq d_\alpha\} = 1 - \alpha$, where P_b denotes the probability calculated under the assumption that \mathbf{b} is the true parameter vector. A value of \mathbf{b} for which $D(\mathbf{b}) > d_\alpha$ can therefore be rejected or viewed as unacceptable. Such values of \mathbf{b} are too far from $\hat{\mathbf{b}}_{LS}$ to be viewed as reasonable candidate for the true solution. Also for any \mathbf{b} , the value of α , for which $D(\mathbf{b}) = d_\alpha$, may be considered as an acceptable confidence level (Obenchain 1977).

The above considerations might be applied to any estimate $\hat{\mathbf{b}}$ obtained for \mathbf{b} in Eq. (1). More precisely, the corresponding value of $D(\hat{\mathbf{b}})$ (see Eq. (8)), gives an idea of how far the estimate is from the LS solution. The acceptance level of the estimate gives a different quantification of this distance, which can be readily interpreted. In fact, when α is small, that is, in the range $[0.001, 0.01]$, the estimate of $\hat{\mathbf{b}}$ may be far from $\hat{\mathbf{b}}_{LS}$ to be compatible with data measurements. In contrast, for large values of α , the estimate $\hat{\mathbf{b}}$ is close to $\hat{\mathbf{b}}_{LS}$ in the metric of Eq. (8). The most appropriate interval for α is conventionally $[0.01, 0.1]$. This supports the idea of building classes of accessible estimates with the possibility of having proper estimates for each class. In particular, if the estimate has to be matched with *a priori* statistics (Li et al. 2001; Pokrovsky et al. 2003a), it can be confidently used given that it is not statistically different from the LS estimate $\hat{\mathbf{b}}_{LS}$.

The confidence region indicated in Eq. (7) describes an hyper-ellipsoid in the p -dimensional space (Draper and Smith 1998). In practice, it is useless for high p values in regard to the difficulty of computation. Therefore, a more convenient form of Eq. (7) linking the components of the vector \mathbf{b} has been proposed (Pokrovsky and Roujean 2003a). We discuss below another alternative characterization of Eq. (7). Let us suppose again that $\hat{\mathbf{b}}$ gives an estimate of \mathbf{b} in Eq. (1). Thus, one

can obtain $R_{SSR}(\hat{\mathbf{b}})$ from Eq. (4) or (5). It is straightforward to show that $D(\hat{\mathbf{b}}) = R_{SSR}(\hat{\mathbf{b}}) - R_{SSR}(\hat{\mathbf{b}}_{LS})$. Therefore, $D(\hat{\mathbf{b}})$ and $R_{SSR}(\hat{\mathbf{b}})$ are minimized by $\hat{\mathbf{b}}_{LS}$ and the estimate $\hat{\mathbf{b}}$ is fairly close to $\hat{\mathbf{b}}_{LS}$ for a relative increase of $R_{SSR}(\hat{\mathbf{b}})$ with respect to $R_{SSR}(\hat{\mathbf{b}})$. For any estimate $\hat{\mathbf{b}}$ one can introduce the multiple correlation coefficient (MCC) (Anderson 1958):

$$R^2(\hat{\mathbf{b}}) = 1 - R_{SSR}(\hat{\mathbf{b}})/T_{SS}, \tag{10}$$

where $T_{SS} = \|\mathbf{y} - \bar{\mathbf{y}}\|^2$ (see Pokrovsky and Roujean 2003b; Pokrovsky et al. 2003a). For LS estimates, R^2 is the usual MCC (Draper and Smith 1998). Note that sometimes the MCC can have a negative value, which is the case when the estimate $\hat{\mathbf{b}}$ performs the vector $\hat{\mathbf{y}}$, this latter deviating from vector of measurement data \mathbf{y} beyond the vector of mean values $\bar{\mathbf{y}}$. It is worth outlining also that for any α the estimate $\hat{\mathbf{b}}$ belongs to the confidence region (7), if and only if $\hat{\mathbf{b}}$ satisfies the inequality:

$$R^2(\hat{\mathbf{b}}) \geq R^2(\hat{\mathbf{b}}_{LS}) - [1 - R^2(\hat{\mathbf{b}}_{LS})] \frac{p+1}{n-p-1} F_{p+1, n-p-1, 1-\alpha}. \tag{11}$$

The criterion (11) means that the estimate $\hat{\mathbf{b}}$ is close to $\hat{\mathbf{b}}_{LS}$ if the MCC $R^2(\hat{\mathbf{b}})$ is sufficiently large. Here, one can note that Eq. (11) might be rewritten in another form:

$$\frac{R_{SSR}^2(\hat{\mathbf{b}})}{R_{SSR}^2(\hat{\mathbf{b}}_{LS})} \leq 1 + \frac{p+1}{n-p-1} F_{p+1, n-p-1, 1-\alpha}. \tag{12}$$

The inequality (12) might be interpreted in the way that the relative increase in the SSR is a fundamental feature of any biased estimate $\hat{\mathbf{b}}$ deviated from the LS $\hat{\mathbf{b}}_{LS}$. In the inequalities (11)–(12), we consider for the cases of reference only equality signs. On one hand, these equations determine the exact criterion level or the boundary of confidence region S_α presented by Eq. (7) when $\mathbf{b} = \hat{\mathbf{b}}$. On the other hand, each of these equations can be resolved with respect to α if all other variables are fixed. Therefore, we can evaluate a significance level α for any given estimate $\hat{\mathbf{b}}$ of true solution \mathbf{b} for Eq. (1). Finally, it is important to note that the unknown true solution of Eq. (1) belongs to S_α with a probability $(1 - \alpha) \times 100\%$. Based on physical considerations, one could select within S_α the more appropriate solution, which is as α -accessible as the unknown true solution \mathbf{b} .

4. Selection of kernels for BRDF composite model

In two recent studies, it has been shown there is no evident favourite BRDF kernel model from a strict statistical point of view. Relevant studies (Pokrovsky and Roujean 2003a; Gao et al. 2001) differ whether the covariance matrix Σ_ϵ of uncertainty measurements and model are known and that the estimate of $\Sigma_\epsilon = \sigma^2 \mathbf{I}$ is made by means of Eq. (6). In practice, the lack of information about the source of uncertainty in data measurement is a major concern. The rationale of the AMBRALS code (see Wanner et al. 1995) is to reach a best approximation of

Table 1. Ill-conditioning indices of the BRDF model matrix for the SEVIRI/MSG angular geometry

Location (latitude, longitude)	Julian day	Roujean model	LSR model
45°N, 0°E	74	0.0251	0.0466
	166	0.1680	0.5396
	258	0.0817	0.0278
45°N, 45°E	74	4.4711	4.4635
	166	7.2222	28.763
	258	6.6143	7.5611
0°N, 0°E	74	0.1013	0.2222
	166	0.0541	0.1473
	258	0.1024	0.2257
	349	0.0439	0.1133

Index values are multiplied by 10^4 .

the angular reflectance measurements based on testing several candidate models. The set of ill-conditioning indices corresponding to different candidate models and spatial points is presented in Table 1. This suggests some *a priori* preferences for selection among candidate models.

The general form of a BRDF kernel-driven model is as follows:

$$\rho(\theta_s, \theta_v, \phi, \lambda) = \sum_i k_i(\lambda) f_i(\theta_s, \theta_v, \phi), \quad (13)$$

where $\rho(\theta_s, \theta_v, \phi, \lambda)$ is the measured reflectance at the wavelength λ , for solar zenith θ_s , view zenith θ_v , and azimuth angles ϕ . The analytical kernel functions $f_i(\theta_s, \theta_v, \phi)$ mimic the different mechanisms of solar-radiation interactions within the soil-vegetation layer. Classically, a zero term in the right side of (13) is denoted as a Lambertian reflection. Therefore, $f_0(\theta_s, \theta_v, \phi) = 1$ for any angle values. The other terms in Eq. (13), so-called volumetric and geometric, correspond to solar radiation transformations by reflection from plant stands. Along with the Roujean (Roujean et al. 1992) and Li-Sparse reciprocal (LSR; Wanner et al. 1997) model kernels, which are the most widely used now, there are some others that are applicable in particular cases. These are the Ross-Thick, Ross-Thin, Li-Dense, and Li-Transit kernels, as well as some of their variants (Wanner et al. 1995; Gao et al. 2000, 2001; Li et al. 2001): Li-Dense LO, Li-Dense LP, Li-Dense HO, Li-Dense HP, Li-Dense MODIS, Li-Sparse LO, Li-Sparse LP, Li-Sparse HO, Li-Sparse HP, and Li-Sparse MODIS. Ross-Thick and Ross-Thin describe the behaviour with large and small values of the leaf area index (LAI), respectively. All the above modifications of Li-Sparse and Li-Dense are distinguished by two parameter values: b/r and h/b . These parameters describe crown shape and relative height. The

Table 2. Status of the BRDF kernels entered in basic composite models

No.	Name	Notation	Basic collection 1	Basic collection 2
1	Lambert	L	+	+
2	Ross-Thick - MODIS	RTK	+	+
3	Li-Sparse - MODIS	LS	+	+
4	Ross - Thin	RTN	+	+
5	Li-Dense - MODIS	LD	+	
6	Roujean volumetric	RV	+	
7	Roujean geometric	RG	+	+
8	Li-Transit	LT		
9	Li-DenseLO	LDLO	+	+
10	Li-DenseLP	LDLP	+	+
11	Li-DenseHO	LDHO		
12	Li-DenseHP	LDHP		
13	Li-SparseLO	LSLO	+	+
14	Li-SparseLP	LSLP		
15	Li-SparseHO	LSHO	+	+
16	Li-SparseHP	LSHP	+	+
17	Li-TransitLO	LTLO	+	+
18	Li-TransitLP	LTLP		
19	Li-TransitHO	LTHO	+	+
20	Li-TransitHP	LTHP	+	+

chosen values are $b/r = 0.75$ and 2.5 for crown shape, representing a slightly oblate crown (O), and a prolate (P) crown for which $h/b = 1.5$ and 2.5 , representing low (L) and high (H) cases. The BRDF model used in the MODIS processing has the following parameter values: $b/r = 1$ and $h/b = 2$. The same values are adopted for the Li-Transit kernel. A complete list of candidate models is presented in [Table 2](#). Some kernels are cross-correlated, but for the BRDF approximation attributed to arbitrary land surface pixels, each of them might be potentially useful.

The rationale for describing a BRDF model as an extensive combination of kernels is to better capture the multi-scale properties of complex vegetation structures that are observed by wide field-of-view sensors (Pokrovsky et al. 2003a,b). The use of a 3-parameter model relies on the assumption that plant canopies should be randomly distributed within the pixel of a coarse scale resolution sensor. Further, retrieved structural variables (distance, height, orientation, distribution) correspond to mean estimates or modes. In reality, most vegetation yields several levels of structure or clumping (leaf, branch, stand, patch). The aim here is to obtain a description of the various scales with a model having a high degree of free-

dom that follows an unconstrained inversion procedure. The coexistence of Ross-thin and Ross-thick kernels is justified to mimic the radiation regime for vegetation entities that dominate different spatial scales like leaf and patch entities.

Our aim is to develop a composite model in the general form of Eq. (13), including all the above concurrent kernels. Our approach is twofold: on one hand, we retain the best model adjustment of measured reflectance; on the other hand, we allow more flexibility in selecting model subsets. The theoretical background developed in Section 2 permits us to vary the model selection based on statistical criterion (10). Some additional notation is required to address the kernel subset selection problem. Let us assume that all kernels taken from the above individual BRDF models enter in the composite model (13). The composite model of (13) might be formed from all known kernels. The entire list consists of 19 kernels plus the isotropic term. To circumvent the collinearity problem of kernels, the selection of the subset of kernels will proceed as follows: (i) consider a linear space spanned over the matrix columns, which is close to that of the full kernel set, and (ii) consider fewer, but more linear, independent columns.

For the sake of simplicity, we consider the case of two subset models: \mathbf{A}_1 and \mathbf{A}_2 . We assume that \mathbf{A} is a full composite model, that is $\mathbf{A} = (\mathbf{A}_1, \mathbf{A}_2)$. When considering a subset model, we assume that the selected variables are the first k and the eliminated variables are the last $p-k$. Then, instead of Eq. (1), we have a model rewritten in the form:

$$\mathbf{y} = \mathbf{A}_1 \cdot \mathbf{b}_1 + \mathbf{A}_2 \cdot \mathbf{b}_2 + \boldsymbol{\varepsilon}, \quad (14)$$

where $\mathbf{b} = (\mathbf{b}_1^T, \mathbf{b}_2^T)^T$. Note that the Lambertian term is comprised of the sub-vector \mathbf{b}_1 of dimension $(k+1)$. The estimate $\hat{\mathbf{b}}$ of \mathbf{b} corresponds to the subset model if $\mathbf{b}_2 = 0$. By denoting $\hat{\mathbf{b}}_s$ any estimator for the subset model, then $\hat{\mathbf{b}}_s = (\hat{\mathbf{b}}_1^T, \mathbf{0}^T)^T$ is the closest to $\hat{\mathbf{b}}_{LS}$ in the sense defined in Section 2 where $\hat{\mathbf{b}}_1$ is the LS solution in the form of Eq. (3) but for the subset matrix \mathbf{A}_1 . In other words, the subset estimate closer to the unrestricted LS estimate is the LS vector for the restricted problem. Therefore, we come to the conclusion that the most acceptable subset estimate for any given subset size is the one with the largest $R^2(\hat{\mathbf{b}}_s)$ defined by Eq. (10). The largest $R^2(\hat{\mathbf{b}}_s)$ corresponds to the smallest $R_{SSR}(\hat{\mathbf{b}}_s)$ value. The number of possible subsets is in fact very large, about $2^p - 1$, that is 32767 for $p=15$. Rank-order of subsets might be in regard to $R^2(\hat{\mathbf{b}}_s)$ or $R_{SSR}(\hat{\mathbf{b}}_s)$ values and their determination for best BRDF adjustment for each pixel as suggested by Wanner et al. (1995, 1997). From a practical point of view, it is advisable to restrict the number of kernels to p due to time constraints. Based on the approach developed in Section 2, we suggest a simple method to overcome such difficulties. Instead of researching the best fit by a set of kernel models, we examine if each candidate subset model satisfies a given α confidence requirement given in Eqs. (11)–(12). The set of α -acceptable subset models consists of all the subset models for which the corresponding LS estimate is within S_α (7). Each subset model provides the LS or any other estimate $\hat{\mathbf{b}}_s$. By assuming that we selected some confidence level α and substitute $\hat{\mathbf{b}}_s$ instead of \mathbf{b}

in Eqs. (11)–(12), then we can examine if this subset model and its solution $\hat{\mathbf{b}}_s$ are α -acceptable. The final step is to construct a collection of best subset models with a given prescribed probability $(1-\alpha)\times 100\%$.

Aitken (1971) used another definition of criteria variables and obtained another form of a test for subset model examination:

$$\frac{R_{SSR}^2(\hat{\mathbf{b}}_s)}{R_{SSR}^2(\hat{\mathbf{b}}_{LS})} \leq 1 + \frac{p}{n-p-1} F_{p, n-p-1, 1-\alpha} \quad (15)$$

This test provides a slightly narrower significance region S_α (7) than those determined by Eq. (12), leading to a reduced selection of subset models than in Eq. (12). Candidate models from Eq. (13) or (14) belong to significance region S_α with a probability $(1-\alpha)\times 100\%$. Keeping in mind some physical background, one can select within S_α a most appropriate BRDF model, which is as α -accessible as the unknown true model. The selected collection of subset models cannot be predetermined as indicated in a recent paper (Gao et al. 2001, p. 59). The reason seems to be that all the above statistics based on s^2 in Eq. (6) depend on reflectance vector \mathbf{y} in Eq. (1) attributed to a given land surface pixel. Therefore, the kernel selection is a rather flexible procedure aiming at achieving the best adjustment of measured reflectance within the significance region. The ill-conditioning feature of the inverse matrix signifies the latter.

The kernels selected typically show similar directional signatures. To get more insight into kernel similarity, we formed a cross-correlation matrix of rank 19 for a uniform angular grid. This grid was created by means of equal increments in view zenith angle (VZA) (15°) and in relative azimuth angle (RAA) (45°) in the intervals $[0^\circ, 75^\circ]$ and $[0^\circ, 360^\circ]$, respectively. This correlation matrix turned out to be singular and to contain many values equal to unity, or close to it. We applied factor analysis tools to extract the most important common factors in this kernel set (Lawly and Maxwell 1963). Generally, factor analysis is applicable in the case when it is necessary to deal with two characteristics of sample variables: similarity and variance. The first one is regressive because the greater the similarity between variables, the less the number of common factors. The second, in contrast, is progressive as the greater the variance, the more factors are involved in its description. Results of double procedure (i.e., principal component analysis – centroid method and varimax rotations) implementation for 2 solar zenith angles (SZA) values are presented in Fig. 2. It was found that only three significant factors might be derived in the case of high Sun (SZA = 25°) and only two factors were revealed in the case of low Sun (SZA = 60°). We may conclude that, in general, the former angular range is more preferable for reflectance measurements. The critical level of 0.7 is accepted in factor analysis to discover significant variables. This level is also apparent in Fig. 2. The two kernels LS and LSLP are the more efficient with only two factor values beyond the critical level. Others kernels exceed the critical level for only one factor. The following pairs of kernels (LTLP, LTHP), (LSLO, LSHP), (LDLP, LDHP), and (LDHO, LTLP) behave similarly;

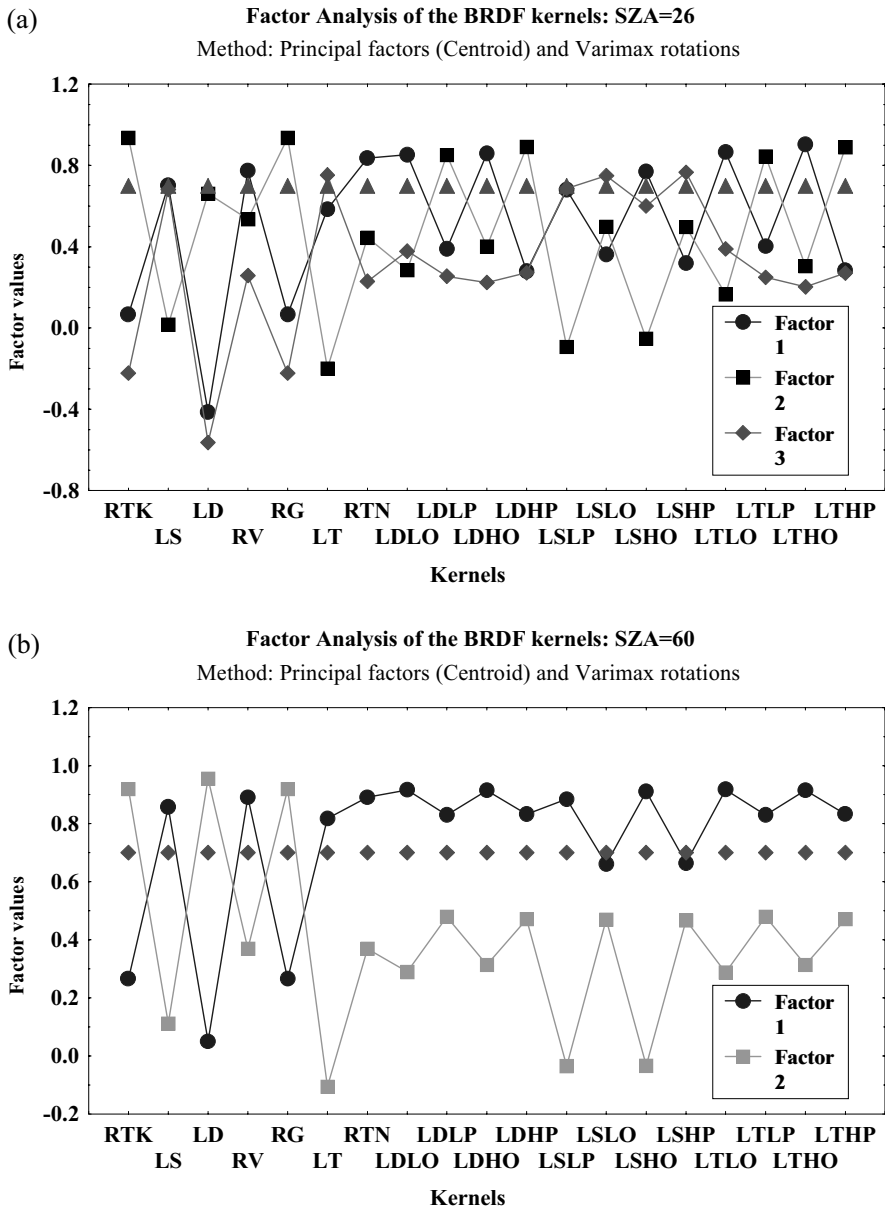


Fig. 2. Factor analysis results for BRDF candidate kernels: (a) SZA = 26°, (b) SZA = 60°.

some could therefore be eliminated. Considering a factor score, some kernels are in fact inefficient due to their great similarity.

We removed four kernels from the list and compounded a basic collection 1, designated C1 (see third column of Table 2), and a basic collection 2, designated C2 (fourth column in Table 2), containing only Li-Sparse-Dense-Transit kernel series and the Roujean geometric kernel. The basic model C2 was introduced due to its homogeneity. Both sets of kernels provide a very high level of explained variance (0.98–0.99) and a low level of RSS (3%–5%) in reflectance value for most land cover types (LCT). Unfortunately, this kernel set leads to extremely ill-conditioned problems when obtaining the LS solutions. To illustrate this issue and to isolate acceptable subset models, we considered ten subset models between C1 and C2 (Table 3). Calculations were performed for pine forest ground-based reflectance measurements (Kimes 1983). Comparison of its acceptance levels shows that C1 is less sensitive to a reduction of the subset model size. The removal of 8 kernels from C1, which consists of 15 kernels, leads to a 5% reduction in acceptance level. In contrast, the elimination of 6 kernels from C2, containing 13 kernels, leads to a 6% or even more reduction in acceptance level. A parallel survey of acceptance levels and explained variance permits the conclusion that the former is more sensitive to changes in subset models. It is worth noting that the conventional subset models Li-Sparse and Roujean have a low acceptance level, which is equal to several percents. This fact indicates that there is a considerable potential for increasing the accuracy of the BRDF and albedo retrievals. Certainly, it depends on the success in the development of a robust inversion technique for multi-kernel models. In such cases, the LS solution is not acceptable because it oscillates due to the ill-conditioned character of the inverse matrix, which is about 10^{-12} or less.

The SCAR land data (Ranson et al. 1985) for cereal fields were used for the second round of numerical experiments to derive a composite model (Table 4). For this, we used a uniform subset of reflectance data for an angular grid of 192 cells with the following increments: 5° in VZA and 30° in RAA. Comparing these results with Table 3, one can find many common features even in the small details. A parallel survey of acceptance level and explained variance values allows us to conclude that these characteristics do not behave in full agreement, and that the acceptance-level value is a more sensitive characteristic from the point of view of the subset BRDF model selection. Another important finding is that the best subset models are identical for both LCT. Therefore, the revealed collection of the acceptable subset models does not depend on basic models (C1 and C2 in our study) and LCT. Interestingly, amongst acceptable subsets is model 8 consisting of the following kernel series: 1–2–3–4–9–10–15–19. It contains a minimum number of kernels and attains a high level of acceptance level and explained variances. Therefore, the procedure implementation allows an optimal low-dimensional BRDF model to be obtained, which does not lie far away from the basic model of 13 or 15 parameters. The next interesting subset model, No. 6, includes 9 kernels series: 1–2–3–4–7–9–10–15–19. It differs from the previous one in adding the

Table 3. Acceptable BRDF subset models obtained for pine forest BRDF measurements in visible band at SZA = 59° (Kimes 1983)

Basic model	Subset models									
	1	2	3	4	5	6	7	8	9	10
Parameters										
Collection 1:	1 2 3 4 5 6	1 2 3 4 5 6	1 2 3 4 5	1 2 3 4 5 6	1 2 3 4 5 6	1 2 3 4 5 6	1 2 3 4 5 7	1 2 3 4 7	1 2 3 4 9	1 3 4 13
Kernel	7 9 10 13	7 9 10 13	6 7 9 10	7 9 10 15	7 9 10 15	9 10 15 16	9 10 15 16	9 10 15	10 15 19	10 13 15
collection	15 16 17	15 16 17	13 15 17	16 17	16 17 20	17 19	19			15
(numbers)	20									
19 20										
Acceptance level (%)	100	100	98.4	100	100	100	99.99	99.92	86.15	78.63
Explained variance	0.999	0.994	0.96	0.9943	0.981	0.9948	0.9934	0.99	0.989	0.988
Collection 2:	1 2 3 4 7 9	1 2 3 4 7 9	1 2 3 4 7	1 2 3 4 7	1 2 3 4 5 6	1 2 3 4 5 7	1 2 3 4 7	1 2 3 4 9	1 2 3 4 9	1 3 4 13
Kernel	10 13 15	10 13 15	9 10 13	10 13 15	7 9 10 15	9 10 15 16	9 10 15	10 15 19	10 13 15	15
collection	16 17 19	16 19	15 19	19	16 17 20	17 19	19			
(numbers)	20									
20										
Acceptance level (%)	99.99	99.99	94.67	89.29	99.97	99.95	94.32	87.81	38.24	29.58
Explained variance	0.995	0.995	0.993	0.993	0.981	0.9948	0.9934	0.99	0.989	0.988

Table 4. Acceptable BRDF subset models obtained for crop field BRDF measurements in visible band at SZA = 60° (Tsay)

Basic model	Parameters	Subset models											
		1	2	3	4	5	6	7	8	9	10		
Collection 1: 1 2 3 4 5 6 7 9 10 13 15 16 17 19 20	Kernel	1 2 3 4 5 6	1 2 3 4 5 6	1 2 3 4 5 6	1 2 3 4 5 6	1 2 3 4 5 6	1 2 3 4 5 6	1 2 3 4 5 7	1 2 3 4 7	1 2 3 4 9	1 2 3 4 9	1 2 3 4 9	1 3 4 13
	collection	7 9 10 13	7 9 10 13	6 7 9 10	7 9 10 15	7 9 10 15	9 10 15 16	9 10 15 16	9 10 15	10 15 19	10 15 19	10 13 15	15
	(numbers)	15 16 17	15 16 17	13 15 17	16 17	16 17 20	17 19	17 19	19				
		20											
	Acceptance level (%)	99.99	99.94	99.79	98.959	99.99	100	99.99	99.99	99.95	25.00	1.29	
Explained variance		0.9301	0.92 63	0.945	0.900	0.931	0.944	0.943	0.942	0.934	0.927		
	Kernel	1 2 3 4 7 9	1 2 3 4 7 9	1 2 3 4 7	1 2 3 4 9	1 2 3 4 5 6	1 2 3 4 5 7	1 2 3 4 7	1 2 3 4 9	1 2 3 4 9	1 2 3 4 9	1 3 4 13	
Collection 2: 1 2 3 4 7 9 10 13 15 16 17 19 20	collection	10 13 15	10 13 15	9 10 13	10 13 15	7 9 10 15	9 10 15 16	9 10 15	10 15 19	10 15 19	10 13 15	15	
	(numbers)	16 17 19	16 19	15 19	19	16 17 20	17 19	19					
	Acceptance level (%)	100	99.94	99.94	99.58	99.90	100.00	99.94	99.57	13.25	0.42		
	Explained variance	0.944	0.943	0.943	0.942	0.931	0.944	0.943	0.942	0.934	0.927		

Roujean geometric kernel. The data of Table 3 show that these kernels are really important only in the case of the forest coverage and with respect to one of the basic models. Updating the subset model with Li-Sparse HP does not have a consequential impact on the acceptance level. For conventional models, the acceptance level attains very small values, 0.07 (Roujean) and 10^{-7} (LSR).

In a final step, PARABOLA data from FIFE (Deering et al. 1992) were used in the analysis of subset BRDF models (Table 5). This case differs from the previous one by the large size of the angular measurement set ($n = 466$) and more confined angular data. It leads to very high sensitivity of acceptance level to small variations in SSR, which are due to changes in the kernel set used. Another distinction is that the PARABOLA data were not uniformly distributed over the hemisphere. This latter point makes them more similar to the satellite data than the two data sets described above. In general, the acceptance levels are less than in Table 3 but the explained variance attains very high levels for most subset models. Despite a decrease in acceptance level, there is good agreement between characteristics obtained for the same subsets for both basic models. An important result of these numerical experiments is that an acceptable subset model should include at least between 8 and 10 kernels.

It can be concluded at this stage that existing kernels are adequate to describe the essentials of the BRDF variance for most LCT. Increasing the model size provides a considerable reduction in the regression residual, and thereby the retrieval accuracy of BRDF model coefficients. On the other hand, there is some flexibility in kernel selection for the BRDF model. It is important to stress that our approach suggests a theoretical and not an empirical background for selection of an optimal subset model with a minimal number of parameters and an appropriate acceptance level. A decrease of ICI when the number of kernels entering in the BRDF model increases is a serious issue, which would delay the implementation of composite models in inversion algorithms.

5. Advanced inversion methods

Hoerl and Kennard (1970) suggested using ridge regression in the case of multi-collinear columns of matrix \mathbf{A} in a regression model like Eq. (1). Their ridge estimator of the standardized regression model coefficient vector reads as follows:

$$\hat{\mathbf{b}}(\beta) = (\mathbf{A}^T \cdot \mathbf{A} + \beta \mathbf{I})^{-1} \cdot \mathbf{A}^T \cdot \mathbf{y}, \quad (16)$$

where β is a positive constant and $\mathbf{A}^T \cdot \mathbf{A}$ is in correlation form (Anderson 1958). The selection of parameter β is the major issue. Most studies on this subject suggest more or less empirical rules, according to the problem to be solved. Here, we use the above acceptability concept as a background for an advanced selection rule. Based on results of Hoerl and Kennard (1970), the solution of Eq. (16) requires that $\hat{\mathbf{b}}(\beta)$ moves away from the LSS when β increases. Therefore, the value of β corresponding to $\hat{\mathbf{b}}(\beta)$, when it leaves the acceptable area of Eq. (7), might be used as an appropriate value for two reasons: 1. $\hat{\mathbf{b}}(\beta)$ is still an accept-

able solution, and 2. $\hat{\mathbf{b}}(\beta)$ is the most stable solution of Eq. (1) within Eq. (7). Hence, the advantage of the above approach requires some theoretical background for this parameter acquisition.

The main result of Section 2 is that all solutions should be examined with criteria (7). Returning to the function $D(\mathbf{b})$ defined in Eq. (8), its minimum value is attained at $\mathbf{b} = \hat{\mathbf{b}}_{\text{LS}}$. On the other hand, $\hat{\mathbf{b}}(\beta) = \hat{\mathbf{b}}_{\text{LS}}$ when $k = 0$. It is easy to show that $D(\hat{\mathbf{b}}(\beta))$ is a monotonically increasing function of k (Obenchain 1977). Therefore, one can conclude that for each significant level value $\alpha \in (0, 1)$, there is a positive number $\beta = \beta(\alpha)$ such that $\hat{\mathbf{b}}(\beta) \in S_\alpha$, if and only $\beta \leq \beta(\alpha)$ (Obenchain 1977). The latter may be considered as a means for setting a reasonable upper bound on the ridge regression parameter β . Moreover, the concept of an acceptable solution set developed in Section 2 might be used also for ridge regression parameter determination. If the acceptance level is high, then the estimate may be used with the knowledge that it is really not very far from the LS solution in a statistical sense. If the acceptance level is small, however, considerably more faith is required in the ridge method to justify the chosen value of β .

Obenchain (1977) defined the associated probability of a ridge estimate as a value of α for which the following equality is valid

$$D(\hat{\mathbf{b}}(\beta)) = s^2 p F_{p, n-p-1, 1-\alpha}. \quad (17)$$

According to Eq. (9) we obtain a slightly different expression in changing the degree of freedom:

$$D(\hat{\mathbf{b}}(\beta)) = s^2 (p+1) F_{p+1, n-p-1, 1-\alpha}. \quad (18)$$

The above equalities also might be considered as equations formulated with respect to β in the case of all other parameters being determined. It is the case, and both of these equations have single solutions for β . It is valid because $D(\hat{\mathbf{b}}(\beta))$ is a monotonically increasing function of β .

Li et al. (2001) suggested using *a priori* information for the BRDF model inversion in the cases of poor angular sampling. They obtained mean values and covariance matrices for Li-Sparse model coefficients. In an allied paper (Pokrovsky et al. 2003a) we obtained the same statistics for Li-Sparse and Roujean model, but for each of 16 IGBP classes provided by the authors of the POLDER data set and that of Bicheron and Leroy (2000). These statistics were used in inversion experiments with the simulated SEVIRI/MSG angular data. It is known that the LS solution of Eq. (1) can be obtained by minimization of the cost function $\min_{\mathbf{b}} (\mathbf{A} \cdot \mathbf{b} - \mathbf{y})^T \cdot \Sigma_{\mathbf{e}}^{-1} \cdot (\mathbf{A} \cdot \mathbf{b} - \mathbf{y})$ (Draper and Smith 1998) with $\Sigma_{\mathbf{e}} = \sigma^2 \mathbf{I}$. Bayesian formulas for conditional probability permit the realization of the generalized form for the cost function when we dispose of *a priori* statistics:

$$\min_{\mathbf{b}} \{ (\mathbf{A} \cdot \mathbf{b} - \mathbf{y})^T \cdot \Sigma_{\mathbf{e}}^{-1} \cdot (\mathbf{A} \cdot \mathbf{b} - \mathbf{y}) + (\mathbf{b} - \bar{\mathbf{b}})^T \cdot \Sigma_{\mathbf{b}}^{-1} \cdot (\mathbf{b} - \bar{\mathbf{b}}) \}. \quad (19)$$

The minimum of Eq. (19) is a regularized solution of the problem (1). Details on numerical algorithms related to Eq. (19) can be found in Pokrovsky (1984).

Based on the above theory, we developed numerical algorithms for three inversion techniques: statistical inversion (SI – see Pokrovsky and Roujean 2003a), ridge regression (RR) and statistical regularization (SR) and applied them for comparative analysis in experiments with the BRDF model inversion. In these experiments, we used two land reflectance data sets: Kimes (1983) and Tsay (Ranson et al. 1985). We carried out three series of inversion experiments. The first series was performed with uniform angular geometry and different kernel models to assess their ability to retrieve the main features of the BRDF for primary LCT. Two other series were targeted to obtain expected estimates of the BRDF retrieval accuracy in the case of SEVIRI/MSG geometry. The second series was based on simulated measurements for the least number of SZA values occurring in the middle of summer. The cases of spring and autumn were considered in a third series of experiments.

For the first series we implemented two inversion techniques: SI (Pokrovsky and Roujean 2003a) and RR. Several examples of retrieved results for the principal plane are presented in Fig. 3. We considered several BRDF models from the previous section: basic models (C1 and C2) and some of the acceptable models (see Tables 3–5). Figure 3a shows that SI is unstable when the basic BRDF model C2 consisting of 13 kernels is used in the case of the lawn grass; whereas, RR solution for the same model is stable and adjusts very well to the measurement data. All other plots (Figs. 3b–f) illustrate that the LSR model could not satisfactorily restore major features of the BRDF (e.g., in the hot spot area) and the basic model C2 (Fig. 3b) and optimal minimum acceptance model (No. 7, see Table 3) consisting of 8 kernels provides much better results in the BRDF fitting for all LCT (Figs. 3c–f). These illustrations also show that the deviation between the solutions obtained by the SI for the LSR model from the measurement curve is so considerable for all LCT that it could also impact albedo estimates. We performed the same study with SCAR data obtained for two LCT: cereal field and forest (Figs. 4a, b). These data sets were derived with a very high angular resolution: 1° for both VZA and RAA. Nonetheless, we obtained similar results. A good fit of the BRDF in the hot spot area is obtained by means of the SI and RR solutions applied to optimal minimum acceptance model (No. 7) with a systematic deviation in the SI solution from measurements for the LSR model. Hence, the main conclusion of the first experiment series is that the LSR model gives a considerable deviation in BRDF adjustment while the optimal composite model can be very helpful in obtaining a more accurate solution.

In the second series of the retrieval experiments, we applied the SI, RR and SR techniques to the land data set of Kimes (1983). Keeping in mind that our goal is to test these algorithms in the case of the SEVIRI/MSG geometry, we carried out an interpolation of the measurement data in the simulated angular coordinate sets corresponding to 5 geographic locations ($0^\circ\text{N}, 0^\circ\text{E}$; $45^\circ\text{N}, 0^\circ\text{E}$; $45^\circ\text{N}, 45^\circ\text{N}$; $45^\circ\text{S}, 0^\circ\text{E}$; $45^\circ\text{S}, 45^\circ\text{E}$). Simulated geometry designs have been presented in Fig. 1. It was assumed that $\text{SZA} < 70^\circ$. The value of the view zenith angle (VZA) is fixed for any given pixel because of the geostationary orbit of MSG. Rejection of meas-

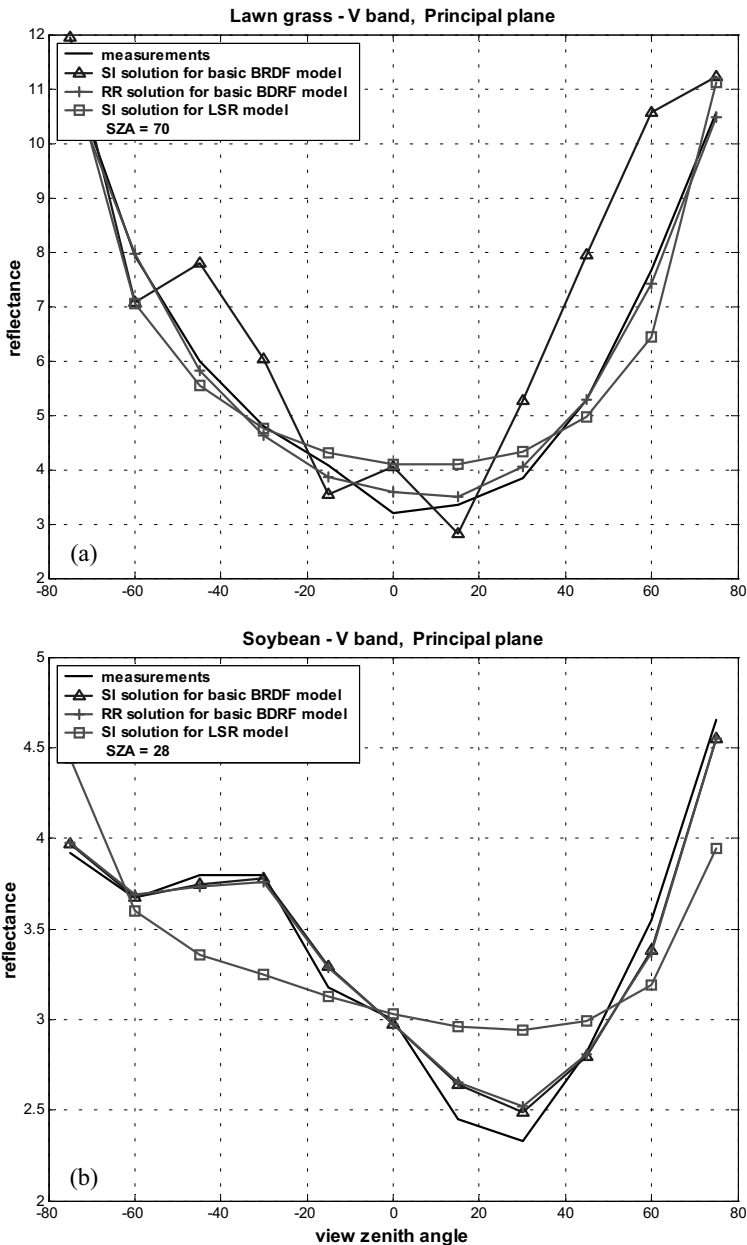


Fig. 3. Comparison of reflectance inversions in the principal plane (visible band) by means of different BRDF models and inversion techniques in the case of the uniform angular measurement distribution over the hemisphere for various LCT: (a) lawn grass, (b) soybean, (c) pine forest, (d) corn, (e) orchard grass, and (f) deciduous forest (from Kimes (1983) measurements).

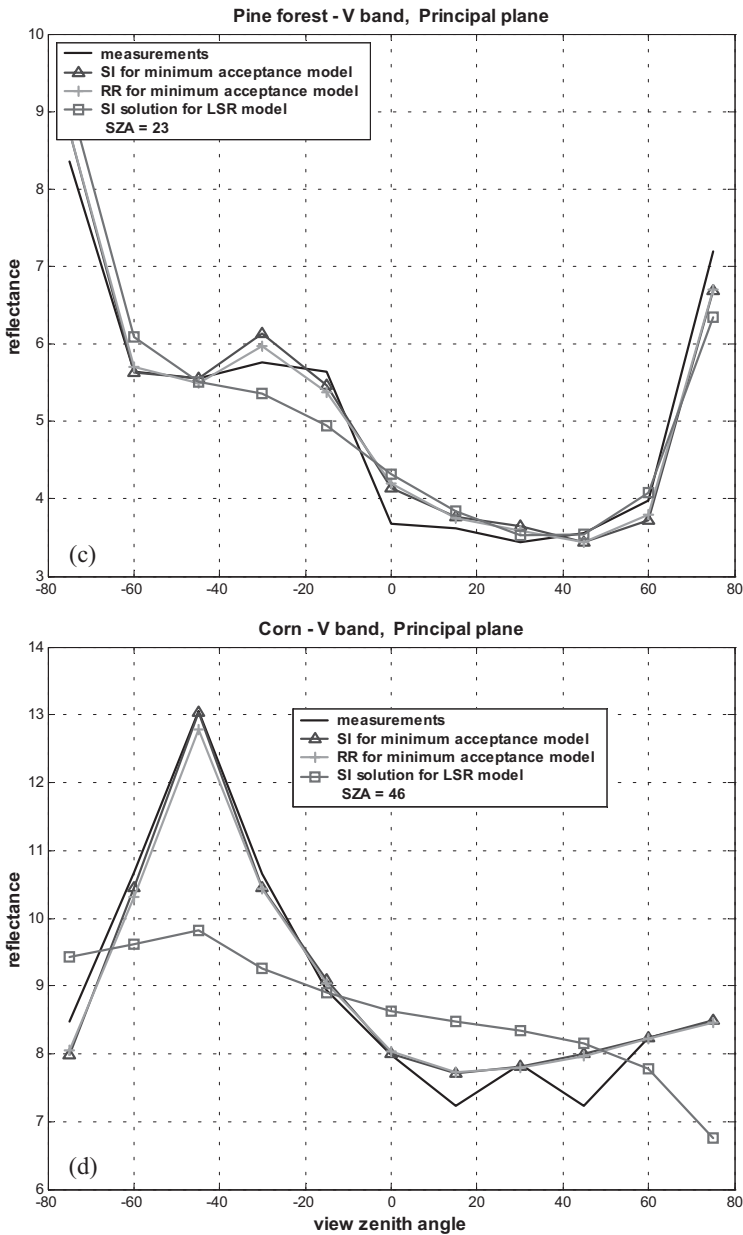


Fig. 3 (continued)

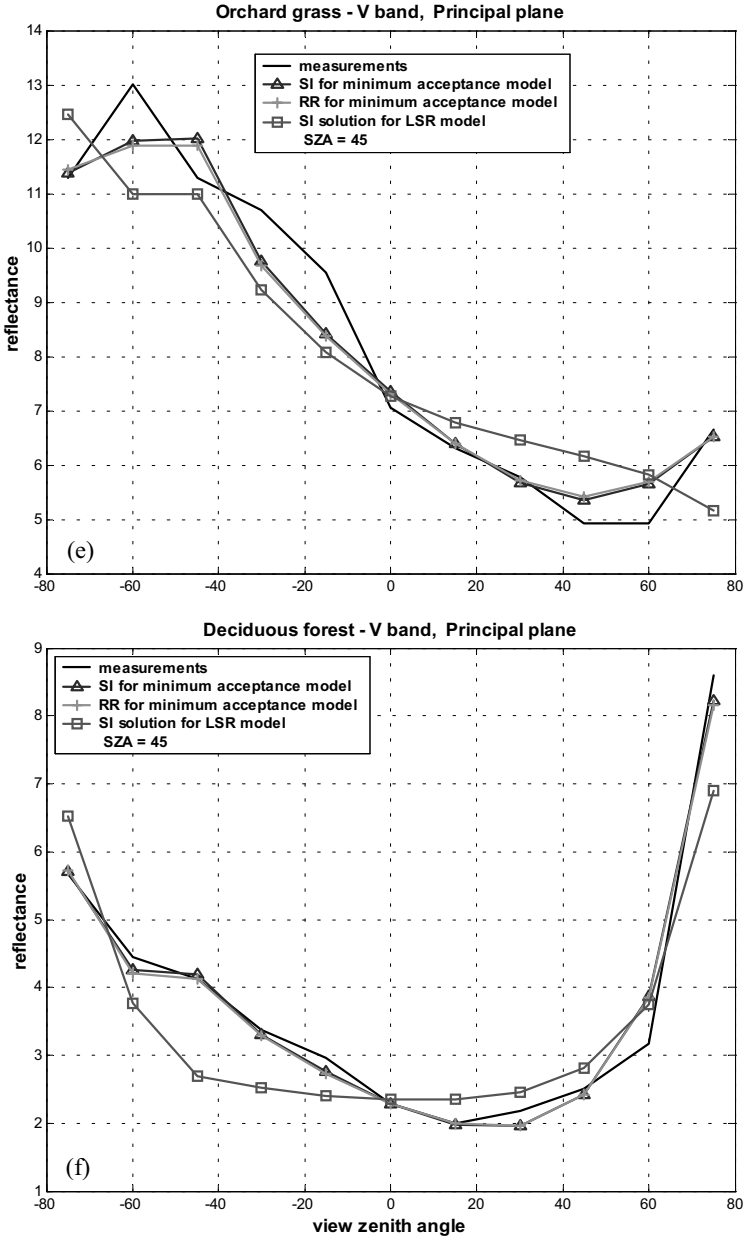


Fig. 3 (continued)

urements for $\text{SZA} > 70^\circ$ is the reason that after December 15 there are no angular measurements available for middle latitude land surface points in the Northern Hemisphere. For convenience and due to the reciprocal feature of all the kernels used, we interchanged the SZA and the VZA; that is, the SZA was fixed and the VZA was varied. Also we generated a reference solution in each of the considered cases. This solution should be most reliable for every considered LCT, as it was obtained from Kimes' data uniformly distributed in the hemispheric angular range by means of SI for every LCT. The differences in the reference solutions acquired by any of the considered inversion techniques were small as shown in the figures.

In the frame of the second series of experiments, we learned that conventional three-parameter BRDF models do not provide satisfactory accuracy in reflectance fitting by any of the considered inversion techniques (Figs. 5a–f) with the exclusion of the SR method using *a priori* statistics. This provides results close to those for multi-kernel models and the RR inversion technique. It is necessary to note that the presented figures are related to the most preferable time of the year (Julian day 166 – see Fig. 1 and Table 1) for inversion, since the SZA values lie in the range that provides the most informative measurements (Pokrovsky and Roujean 2003b). As expected, the deviations from the reference solution and measurements show a systematic bias. More precisely, the retrieved solution for the SEVIRI/MSG geometry sometimes contains overestimated reflectance magnitudes (Figs. 5b, c, e, f). As a consequence, one can expect that the albedo estimates should also exceed the corresponding exact values. Another important point is that these deviations are reduced with increasing SZA. Hence, SZA dependence of the black sky albedo (BSA) becomes underestimated. Therefore, the BRDF estimates derived by means of conventional BRDF model inversion from SEVIRI/MSG geometry data can differ from exact magnitudes even under ideal conditions and without taking into account other regressive factors like signal contamination within the atmosphere or inadequacy of the model for particular pixels. There are two causes that could explain such occurrences. First, three-parameter models are insufficient to describe BRDF angular variance due to the restriction in their angular resolution as illustrated in Figs. 3 and 4. And second, the poor angular sampling of the SEVIRI/MSG geometry with only a single measurement in the principal plane prevents the retrieval of useful information from the reflectance in the principal plane where the BRDF reaches its maximum variance. This problem could be overcome by means of a more sophisticated selection of appropriate kernels for the BRDF model. Figure 3 demonstrates a perceptible improvement in the angular resolution of the retrieved solution after consideration of novel multi-kernel models. We have to remember that the three-parameter models are unacceptable with respect to the basic models (C1 and C2 – Tables 3–5) from a statistical point of view. Carrying out similar experiments with multi-kernel models, the attempt to use the SI technique failed due to instability problems. In fact, the ICI magnitudes for matrices to be inverted lies in the interval $[10^{-3}, 10^{-4}]$, which excludes any possibility of obtaining a stable solution. Otherwise, we obtained very encouraging results in applying the ridge regression algorithm (see Figs. 5a–f).

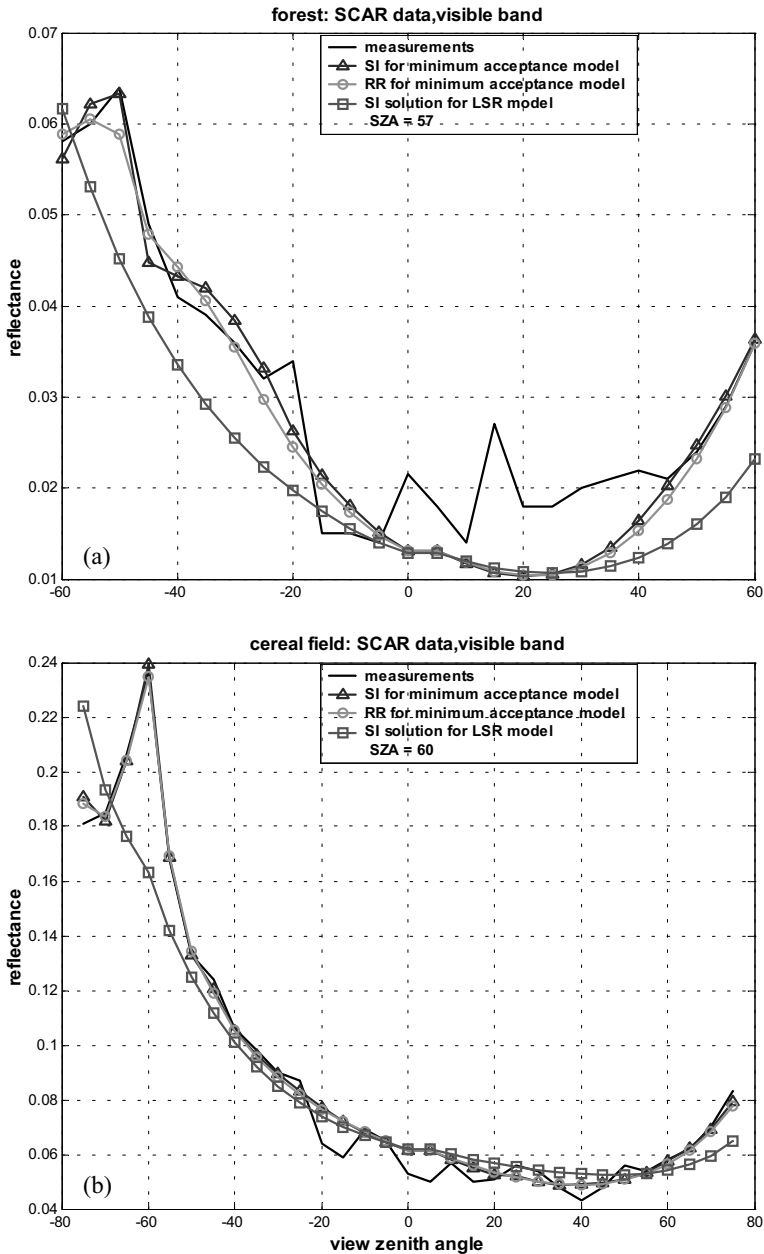


Fig. 4. Comparison of reflectance inversions in the principal plane (visible band) by means of different BRDF models for various LCT: (a) forest, (b) cereal field (from Tsay measurements given in Ranson et al. (1985)).

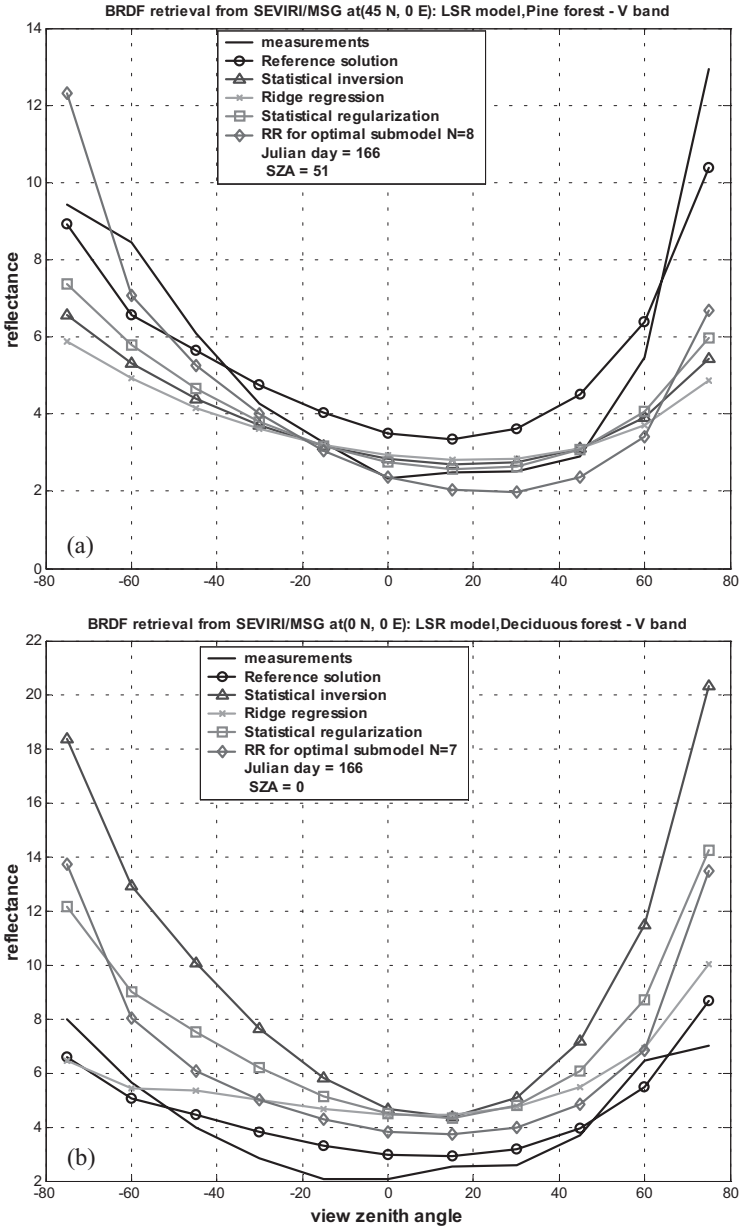


Fig. 5. Comparison of reflectance inversions in the principal plane (visible band) by means of different BRDF models and inversion techniques in the case of the SEVIRI/MSG angular sampling geometry for 166 Julian day and for various LCT: (a) pine forest lawn grass, (b) deciduous forest, (c) irrigated wheat, (d) plowed field, (e) lawn grass, and (f) orchard grass (from Kimes (1983) measurements).

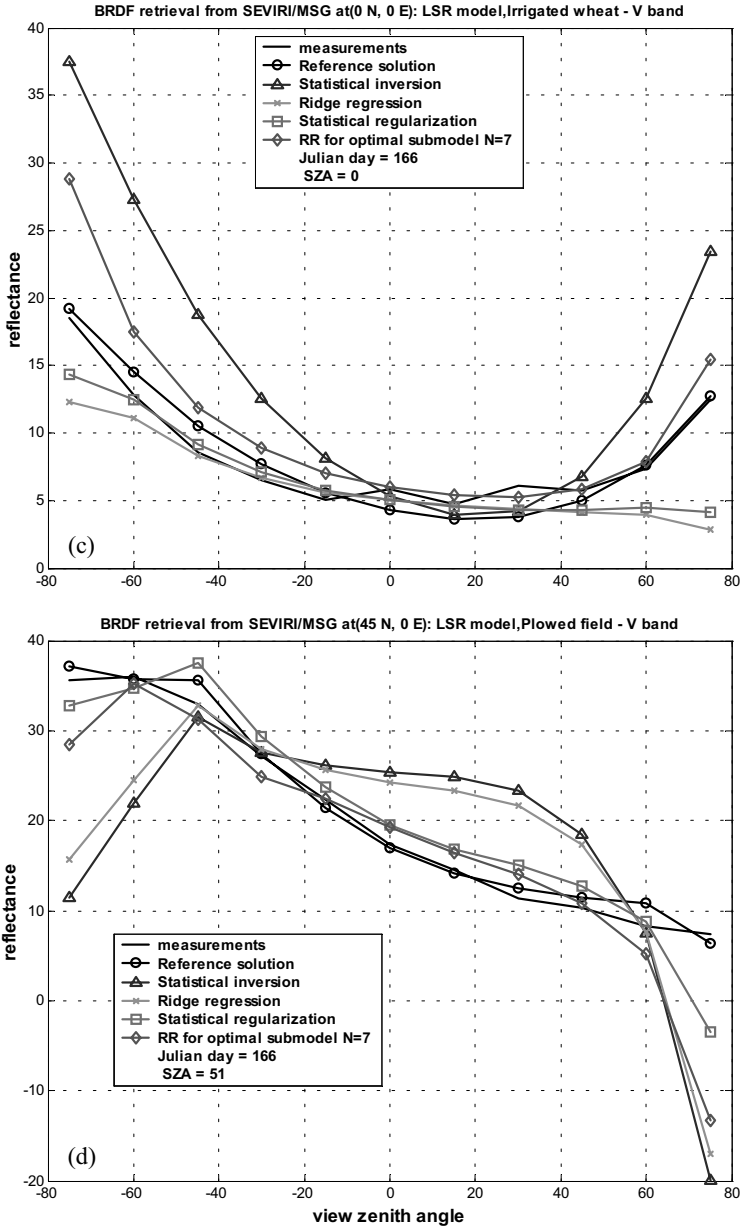


Fig. 5 (continued)

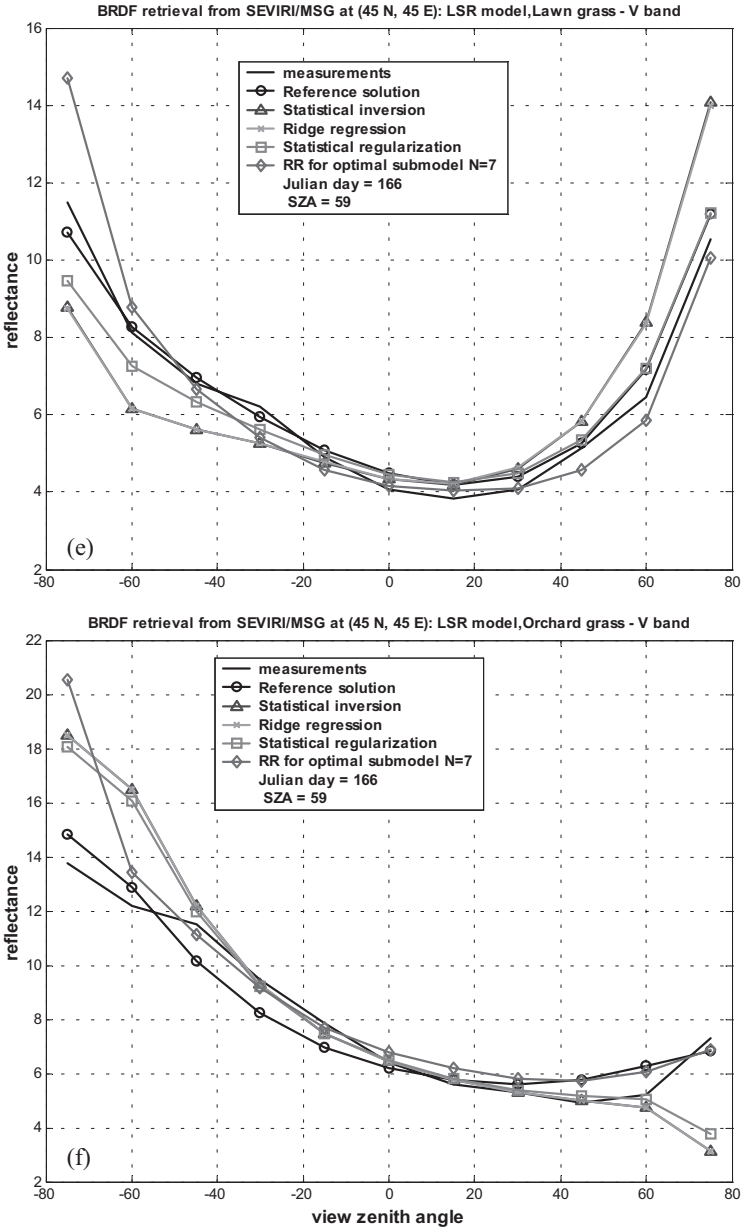


Fig. 5 (continued)

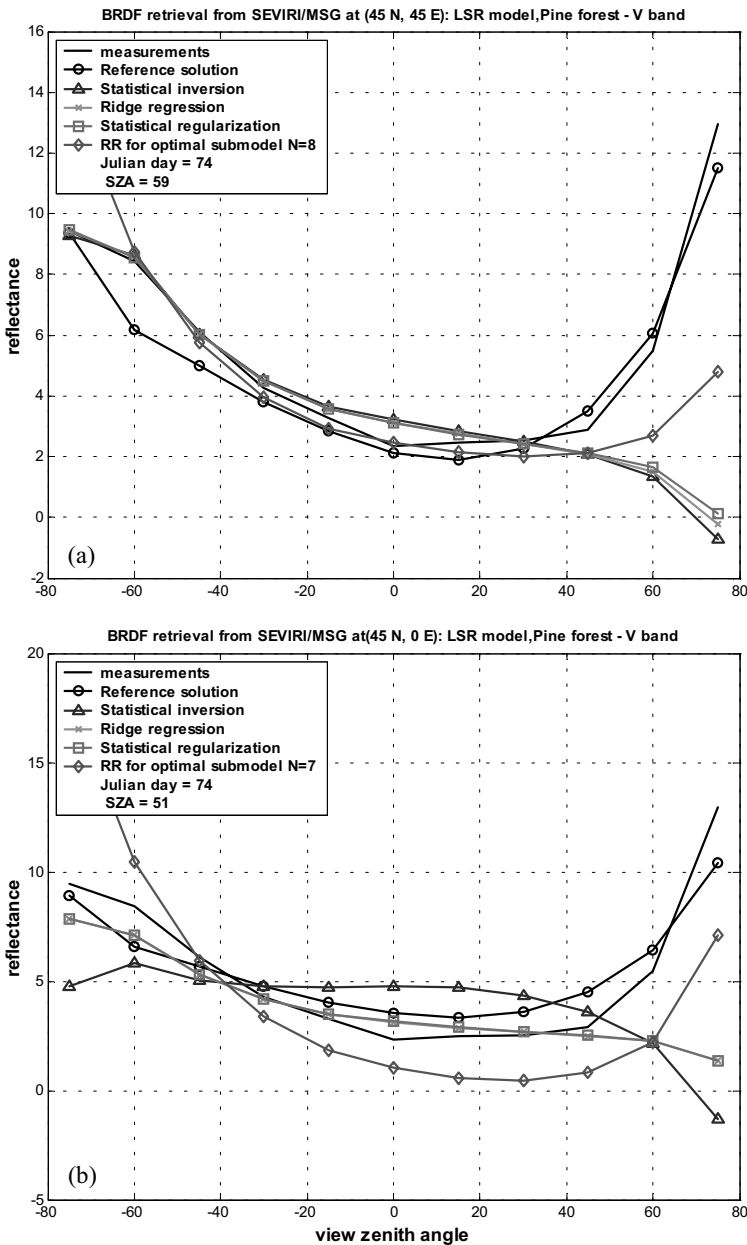


Fig. 6. Comparison of reflectance inversions in the principal plane (visible band) by means of different BRDF models and inversion techniques in the case of the SEVIRI/MSG angular sampling geometry for two geographic locations (45°N, 0°E and 45°N, 45°E), for Julian days 74 and 258, and for various LCT: (a,b) pine forest, (c,d) deciduous forest, (e,f) soybean (from Kimes (1983) measurements).

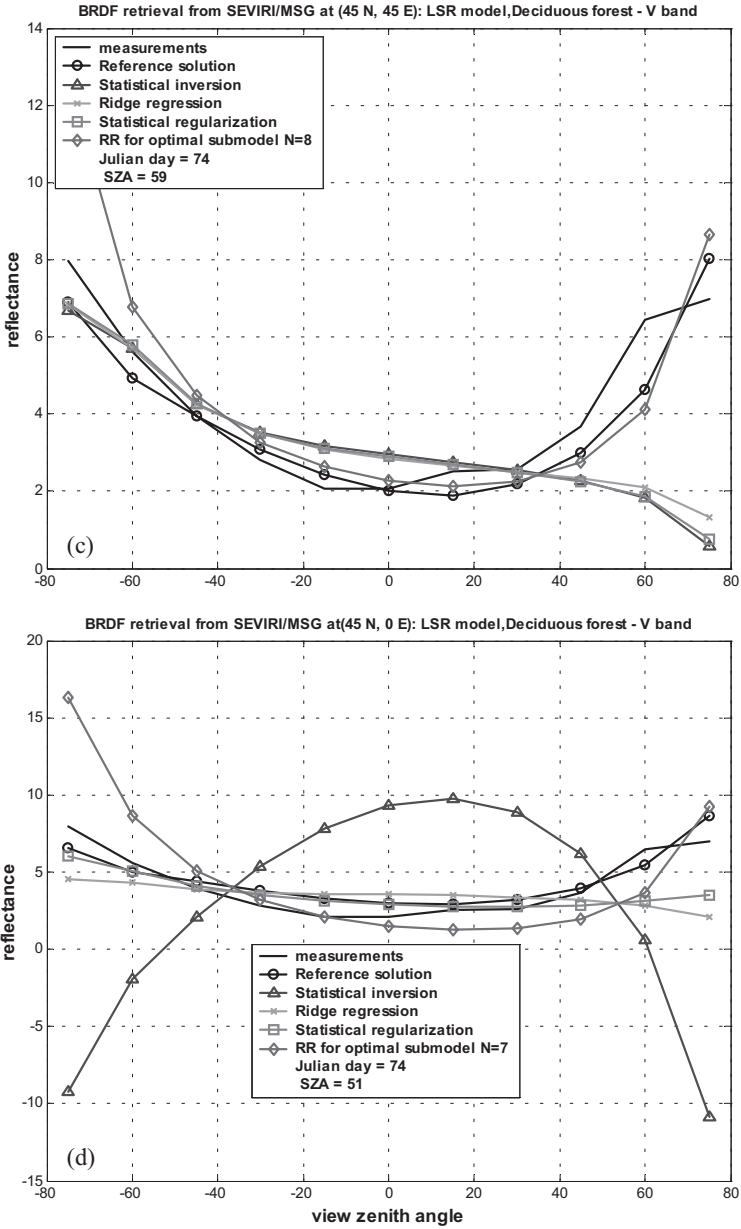


Fig. 6 (continued)

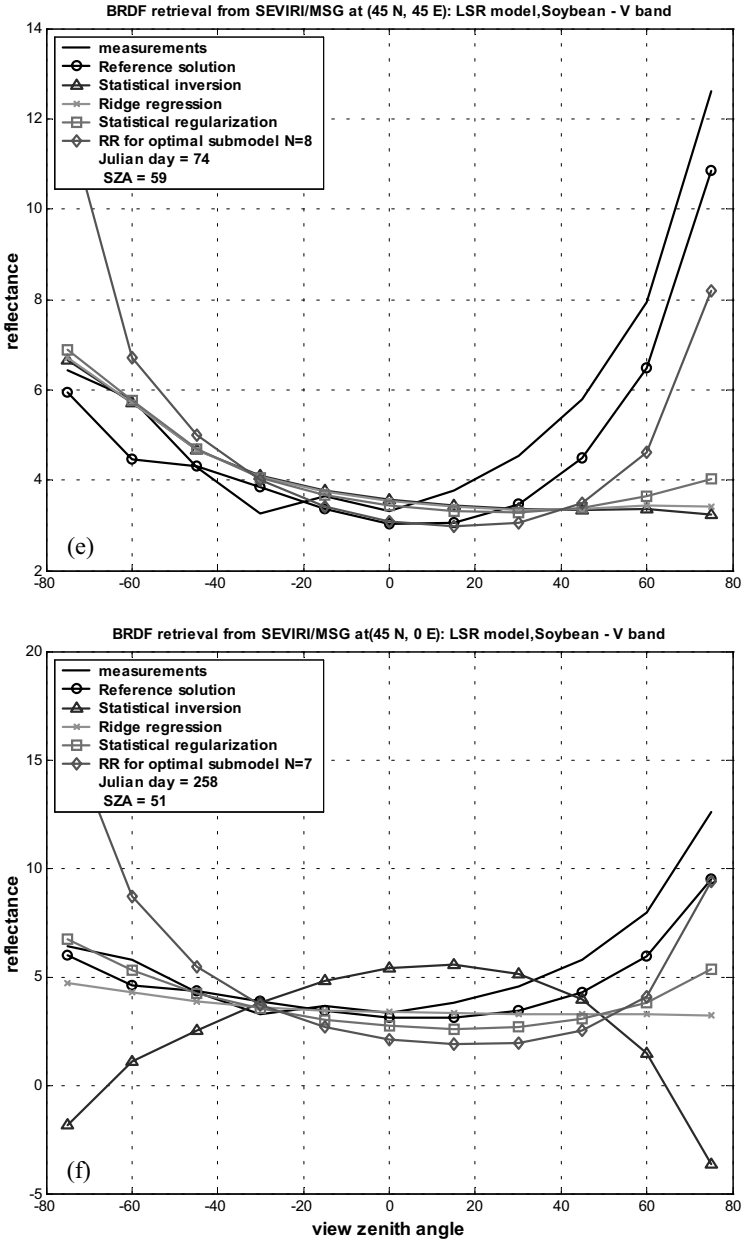


Fig. 6 (continued)

The solution does not reproduce some specific features of the BRDF angular dependence, but rather provides a much better approximation. SR and RR techniques are clearly the favorites of the second series of experiments. The achieved results permit us to continue efforts to improve the accuracy assessment of the solution to the inverse problem.

In a third step related to the retrieval experiments, we again applied the SI, RR and SR techniques to the data set of Kimes (1983) interpolated in SEVIRI/MSG sites. The results are even worse due to a lower Sun position in spring and autumn seasons (Fig. 6). With the LSR model, the unstable solutions were obtained for all LCT (Figs. 6b,d,f) in applying the SI method. The implementation of the RR and SR techniques provides a stable inversion in the same cases (Figs. 6b,d,f), but the solutions are smoothed in the backscattering direction and underestimated in the forward-scattering range. Inversion results when the RR was applied to a composite model demonstrate an improved BRDF adjustment (Figs. 6a,c,e). It is necessary to note that in this case, the retrieved reflectance is overestimated in the backscattering range and slightly underestimated in the forward-scattering range. The third series of experiments showed that inversion techniques might be ordered by achieved accuracy in the following way: (i) RR (composite model), (ii) SR (LSR), (iii) RR (LSR), (iv) SI (LSR). In fact, the last method may be considered as inapplicable in the case of the SEVIRI/MSG geometry. The first and second techniques provide similar results, but the SR for LSR gives a smoothed solution. Therefore, there is some potential for inversion improvement based on implementation of the SR method to the composite model.

6. Albedo uncertainty evaluation

It is worth emphasizing that the concept developed in Section 2 is directly applicable to the problem of albedo estimation. The aim of this section is the further development of linking the angular reflectance measurements with the black sky albedo (BSA) magnitudes as suggested by Gao et al. (2001). As above we assume that the covariance matrix standing for noise in Eq. (1) is unknown and we estimate it by means of relationship (6). That is a major difference in our approach with respect to the work of Gao et al. (2001). Nonetheless, we also employ the linearity of the integral in the expression for BSA. Hence, for each value of SZA there is a vector formed by a hemispheric integral of the kernels. Let us denote this vector-row by $\mathbf{a}_0(\theta_s) = [a_0^1(\theta_s), \dots, a_0^p(\theta_s)]$. In this case the estimate of the BSA, $\hat{h}(\theta_s)$, might be expressed in a form of the scalar product of two vectors: $\hat{h}(\theta_s) = \mathbf{a}_0(\theta_s) \cdot \hat{\mathbf{b}}$. It is important to note that vector \mathbf{a}_0 belongs to the same vector space as do the matrix \mathbf{A} rows. We suppose that matrix \mathbf{A} is of full rank; that is, all columns of \mathbf{A} are linearly independent. Otherwise, matrix $\mathbf{A}^T \cdot \mathbf{A}$ is singular and does not possess the usual inverse. Hence, \mathbf{a}_0 might be expressed by a linear combination of matrix \mathbf{A} rows. This fact helps us to derive the expression for confidence area of the estimated albedo. Now we obtain the analog of Eq. (7) for measured reflectance. It follows from Eqs. (7) and (9) that for any $\hat{\mathbf{b}} \in S_\alpha$, the ine-

quality $D(\hat{\mathbf{b}}) \leq d_\alpha$ is valid. Here we assume that $D(\hat{\mathbf{b}}) = \|\hat{\mathbf{y}}_{LS} - \hat{\mathbf{y}}\|^2$, where $\hat{\mathbf{y}} = \mathbf{A} \cdot \hat{\mathbf{b}}$, $\hat{\mathbf{y}}_{LS} = \mathbf{A} \cdot \hat{\mathbf{b}}_{LS}$. From this and Eq. (7), we obtain:

$$\|\hat{\mathbf{y}}_{LS} - \hat{\mathbf{y}}\|^2 \leq s\gamma_\alpha, \quad (20)$$

where $\gamma_\alpha = [(p+1)F_{p+1, n-p-1, 1-\alpha}]^{1/2}$. The inequality (20) might be useful in two respects. On one hand it provides insight in the evaluation of $\hat{\mathbf{y}}$ corresponding to a given estimate $\hat{\mathbf{b}}$ from the point of view of acceptability. On the other hand, having some physically reasonable solution of Eq. (1) $\hat{\mathbf{b}}$, one can obtain appropriate values of the significance level α from Eq. (20). Having in mind Eq. (20) and after minimization of the SSR functional by the Lagrange multipliers method, one can obtain a modification of inequality (7) for albedo estimate $\hat{h}(\theta_s) = \mathbf{a}_0(\theta_s) \cdot \hat{\mathbf{b}}$. This result should be interpreted as follows: if $\hat{\mathbf{b}} \in S_\alpha$ then, for any vector \mathbf{a}_0 , the following inequality is valid:

$$\mathbf{a}_0 \cdot \mathbf{b}_{LS} - s\gamma_\alpha [\mathbf{a}_0 \cdot (\mathbf{A}^T \cdot \mathbf{A})^{-1} \cdot \mathbf{a}_0^T]^{1/2} \leq \mathbf{a}_0 \cdot \hat{\mathbf{b}} \leq \mathbf{a}_0 \cdot \mathbf{b}_{LS} + s\gamma_\alpha [\mathbf{a}_0 \cdot (\mathbf{A}^T \cdot \mathbf{A})^{-1} \cdot \mathbf{a}_0^T]^{1/2}. \quad (21)$$

It is very important that Eq. (21) is preserved for any arbitrary vector \mathbf{a}_0 . Hence Eq. (21) might be used for the evaluation of uncertainty bands for BSA at any SZA. Also, Eq. (21) is uniformly valid for a given significance level.

The establishment of a significance region S_α is of practical importance as it generalizes a known rule based on the standard deviation (SD) or sigma interval, which is widely used in statistical applications (Draper and Smith 1998). In fact, the SD confidence interval is obtained when parameter α attains one particular value 0.33. Therefore, Eq. (21) provides a more general estimation rule than might be achieved by means of the covariance matrix technique suggested by Gao et al. (2001).

Results of the previous section demonstrate that one could more or less retrieve the BRDF of a land surface pixel from SEVIRI/MSG data only in the middle of the summer season when the Sun position is high above the horizon. Otherwise, systematic deviations from the reference solution were obtained, which explains differences in the derived albedo from the exact solution. The aim of this section is to acquire some estimates of albedo uncertainty retrieved from SEVIRI/MSG data. To achieve this aim we carried out twin numerical experiments with two angular geometries: uniform angular distribution and SEVIRI/MSG. The albedo obtained from the former is considered as the reference solution. Kimes' reflectance data and simultaneous albedo measurements at the same SZA values were used in these experiments. Albedo measurements as well as reference solutions were used to evaluate retrieval estimates obtained from restricted-geometry angular data. To attain albedo estimates based on composite-model coefficients, black sky albedo (BSA) integrals for all kernels entering in the composite model were calculated.

The inversion results presented in Fig. 5 allows the presumption that albedo retrieval should be more or less satisfactory in tropical zone pixels and in extra-

tropical pixels in the middle of summer, when the Sun position is high. The only available reflectance in the principal plane in extra-tropical pixels lies in the interval of $45^\circ - 60^\circ$ for VZA in other seasons of the year, assuming the reciprocity principle. This interval includes hotspot measurements, which may explain the observed biases in albedo estimates as well as in reflectance values. In fact, for some LCT, we obtained overestimated albedo values (Figs. 7a,c,e,g), and underestimated values for others (Figs. 7f,h). Application of the reciprocity principle helps to understand causes of the above phenomenon. In fact, the cases with overestimated albedo (Fig. 7 a) occurred when the geometric trajectory (Fig. 1) intersects the principal plane at $VZA=59^\circ$ (SZA) lying in the hotspot area for a given LCT, where reflectance values exceed corresponding mean values. The opposite situation occurs when the geometric trajectory intersects the principal plane away from the hotspot (Julian day 74, site (45°N , 0°E), $VZA=48^\circ$; Fig. 7d). Albedo underestimation for plowed field, orchard grass and hard wheat can be explained by other causes. For these LCT, the reflectance corresponding to the MSG geometry decreases with SZA in the range $[45^\circ, 60^\circ]$ in the perpendicular plane. Therefore, it is not surprising that corresponding estimates lie systematically below reference values. In the case of uniform geometry the reflectance variance is smoothed by means of averaging over the entire angular hemisphere and, as a consequence, the resulting magnitudes are not biased.

Another point is that the ICI of the inverse matrix is about 10^{-3} at Julian day (JD) 166 and about 10^{-4} at JDs 74, 258 (Table 1). The LS inversion solution is stable at JD 166 and reveals unstable features at JDs 74, 258 (Figs. 7b,d,f,g). Therefore, one can suppose that the instability threshold lies at about 10^{-4} . Other solutions (RR and SR) deliver stable solutions for any site and season of the year (Figs. 7a–h). It is necessary to note that in the case of the LSR model, the retrieved BRDF is smoothed with respect to the reference solution (Fig. 7) and the corresponding albedo has a weak SZA dependence. In contrast, the RR solution for the multi-kernel (composite) model is in better agreement with a corresponding reference solution (Figs. 6a–f), as well as the related albedo magnitudes (Figs. 7a–f). The latter demonstrates an even stronger SZA dependence than those for the reference solution. The cause of this is the overestimation of reflectance magnitudes in the RR solution at higher VZA values.

Let us consider the differences between retrieved albedo and background albedo measurements provided by Kimes at the same site and period. It is interesting also to consider corresponding deviations from reference solutions. As indicated above, the systematic deviation is a major component in differences between the MSG and reference solutions. Those attain maximum values for the LSR model and are equal to 20%–30% of relative magnitude (Fig. 7). Retrievals obtained by means of the composite model do not contain any systematic components and their deviations from reference solutions are lower. In this case the SZA dependence of retrieval albedo is stronger than in the reference solution. Therefore, approximate albedo is lower than the reference solution in the interval of the SZA from $0^\circ - 50^\circ$, and greater in the complementary range of the SZA from $50^\circ - 75^\circ$.

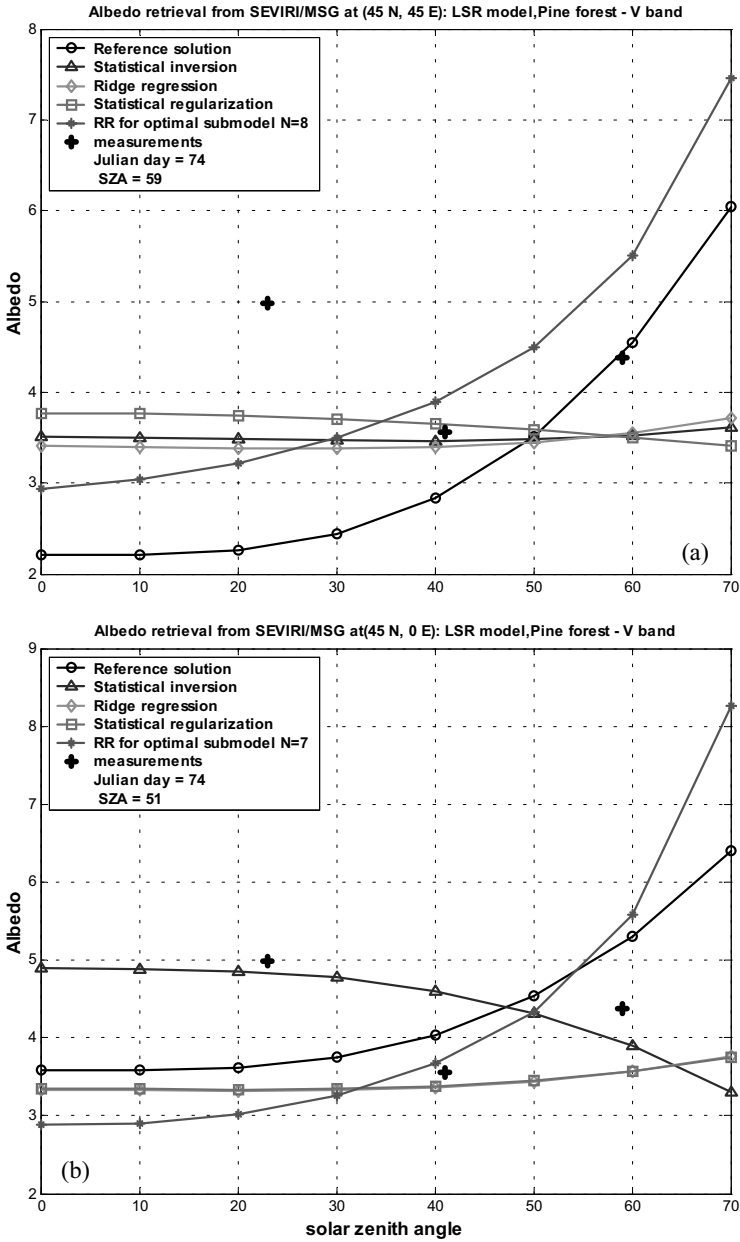


Fig. 7. Comparison of BSA (albedo) estimates derived with three inversion techniques in the case of the SEVIRI/MSG angular sampling geometry for two sites (45°N, 0°E and 45°N, 45°E), for 74 and 258 Julian days, and for various LCT: (a,b) pine forest, (c,d) deciduous forest, (e,f) soybean, and (g,h) lawn grass (from Kimes (1983) measurements).

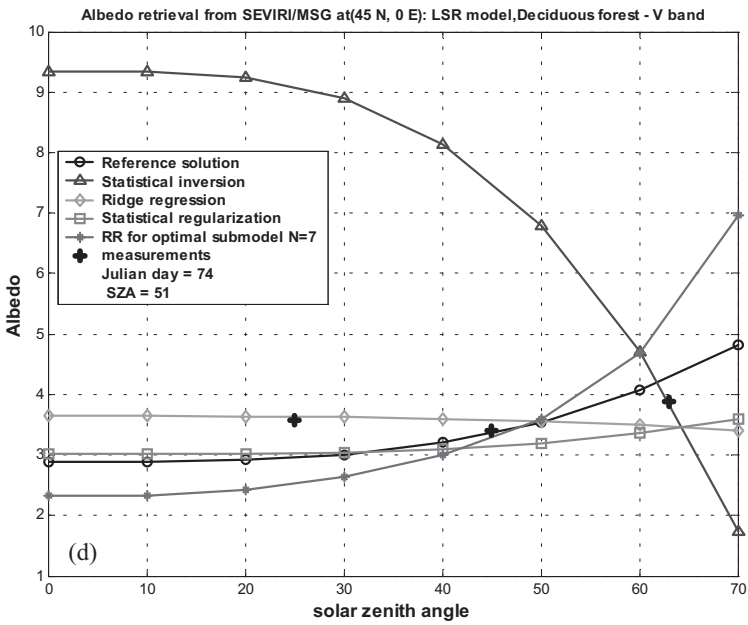
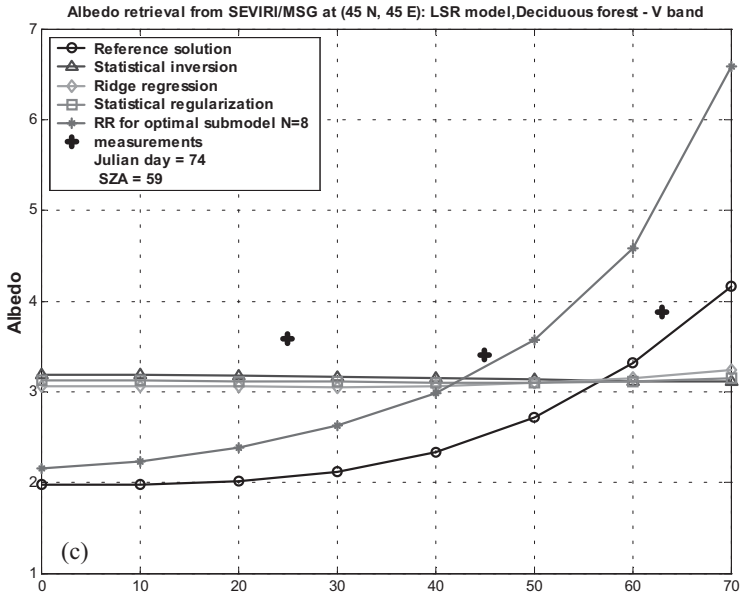


Fig. 7 (continued)

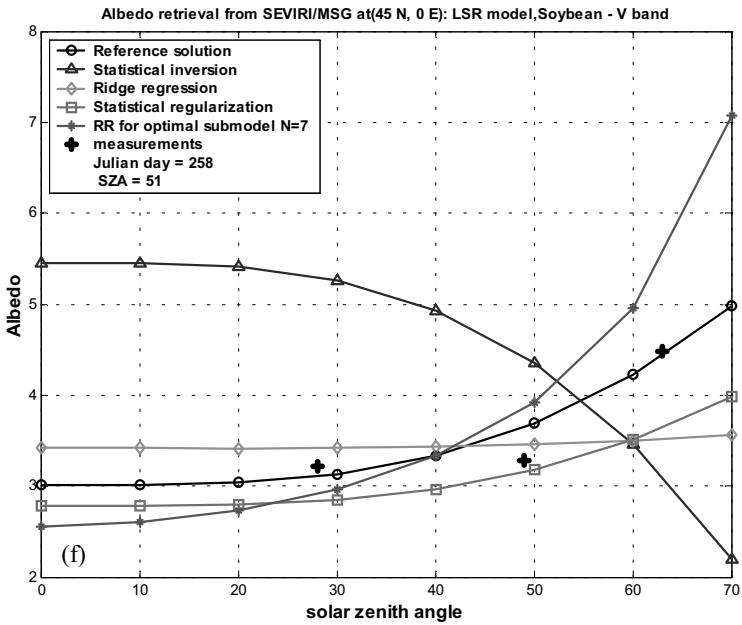
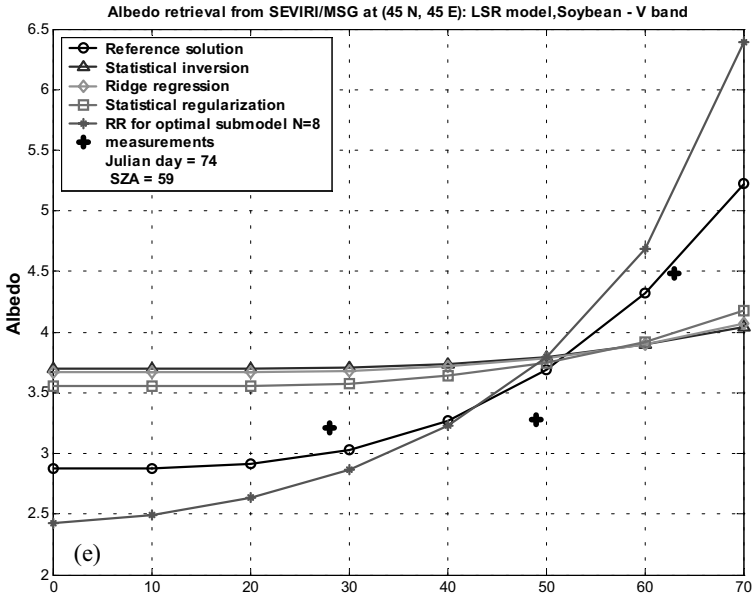


Fig. 7 (continued)

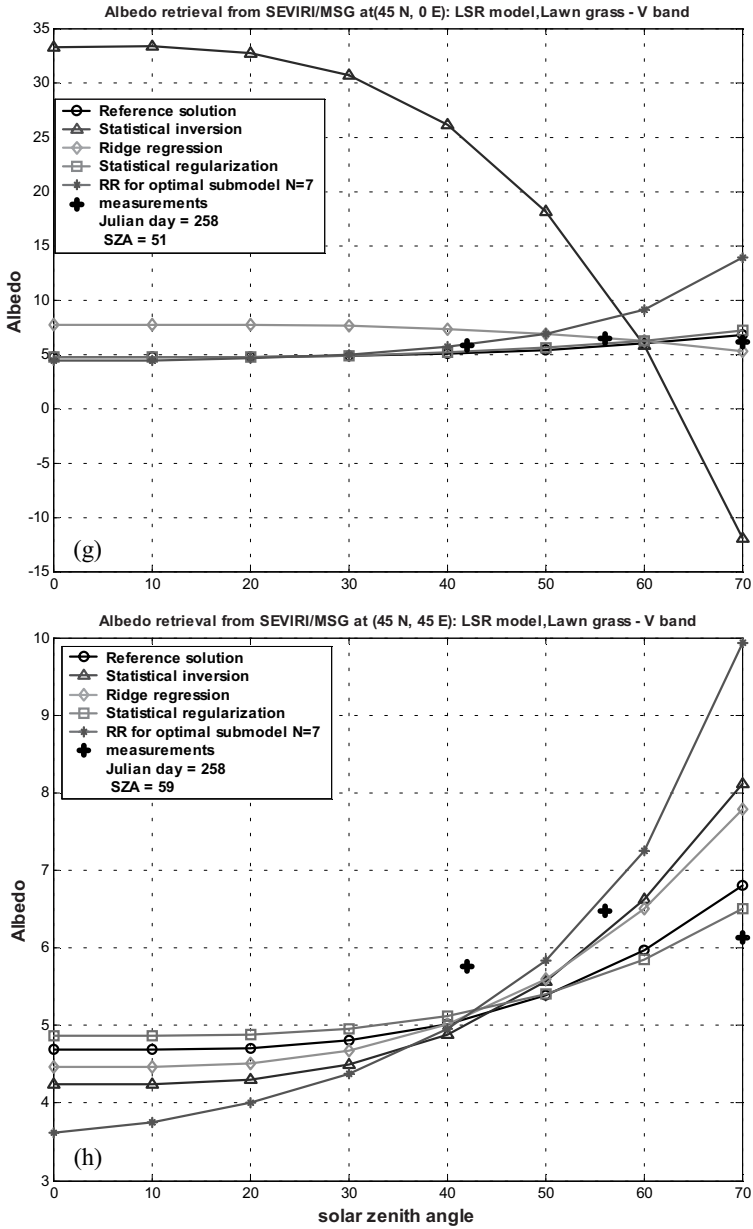


Fig. 7 (continued)

Comparison of the retrievals with the albedo measurement data (denoted by crosses) shows that the albedo field observations do not follow the monotonic albedo dependence on the SZA assumed in any BRDF kernel model. The best agreement is observed for deciduous and pine forest where two of three available albedo values are located within 5% – 10% intervals centered by reference solutions or by RR solutions for the composite model. In the case of orchard, lawn grass, and soybean one can find that only one of three points is close to the best approximate solutions. The best agreement is achieved in the SZA range of $40^\circ - 60^\circ$ and in the middle of the summer season. The experimental points are randomly distributed with respect to the reference and RR solutions for the composite model, but those are systematic with respect to the solution obtained by means of the LSR model. These solutions demonstrate a weaker SZA dependence than those for the reference solutions in every considered case. The latter occurs because of underestimation of retrieved reflectance in the forward direction in the principal plane when the LSR model is used (see Fig. 6 and Section 5).

7. Conclusion

In the present work, the definition of confidence intervals has been suggested as an alternative tool to conventional LS estimation of the coefficients in BRDF models. It was discovered that this concept is of value as it highlights several key issues. First, it led to the selection of the minimum subset of kernels based on the achievement of the best BRDF adjustment. Second, it showed the existence of a bias in the results of the BRDF model inversion when the model matrix is ill-conditioned and *a priori* statistical information on the solution is available. Third, it provides an evaluation of the uncertainty bands for albedo estimates. We considered a simple and attractive approach for the implementation of the above concept in relation to the optimal design problem (see Pokrovsky and Roujean 2003b). This approach assumes that if the true but unknown solution of the inverse problem lies within the significance region S_α centered on the LS solution, then it is reasonable to adopt the assumption that a physically acceptable solution might be sought within the same domain. It allows us to vary the BRDF model, the inversion techniques and albedo estimates within this domain. We developed several techniques to solve the above particular problems, which were based on this general concept. Certainly, the attached illustrations cannot be considered as complete, and additional efforts are required to acquire more reliable results in all three mentioned directions.

In fact, the perspective of future studies looks very promising because it may lead to generalization and extension of existing BRDF models. If the solution is restricted by the ill-conditioning of the matrix to be inverted, then implementation of ridge regression and other biased estimates permits avoidance of this obstacle and increases the accuracy in the acquisition of albedo and other biophysical parameters. It is necessary to note that the ill-conditioning feature appears not only in the case of restricted angular sampling. Our study shows that even for uniform angular sampling, the BRDF model matrix is ill-conditioned if the number of ker-

nels approaches 10 because of their similarity, which leads to linear dependence of model kernels. Therefore, further progress in BRDF modeling and inversion is related to these model extensions and implementation of advanced inversion techniques.

Further improvement in albedo retrieval can be achieved in two ways. The first way is to perform an optimal selection of kernels for a composite model. In this study the selection was carried out from an arbitrary kernel set. The second way is through implementation of the SR method to a composite model. To achieve this aim it is necessary to accumulate *a priori* statistics for several alternative multi-kernel models.

References

- Aitken, M. A., 1974: Simultaneous inference and the choice of variable subsets in multiple regression. *Technometrics* **16**, 221–227.
- Anderson, T. W., 1958: *An Introduction to Multivariate Statistical Analysis* (Wiley, New York).
- Arvesen, J. N., and G. P. McCabe, 1975: Subset selection problems for variances with applications to regression analysis. *J. Amer. Statist. Assoc.* **70**, 166–170.
- Behnken, D. W., and N. Draper, 1972: Residuals and their variance patterns. *Technometrics* **14**, 101–111.
- Bicheron, P., and M. Leroy, 2000: BRDF signatures of major biomes observed from space. *J. Geophys. Res.* **105**, 26 669–26 681.
- Deering, D. W., T. F. Eck, and T. Grier, 1992: Shinnery oak bidirectional reflectance properties and canopy model inversion. *IEEE Trans. Geosci. Remote Sens.* **30**, 339–348.
- Deschamps, P. Y., F. M. Breon, M. Leroy, et al., 1994: The POLDER mission: instrument characteristics and scientific objectives. *IEEE Trans. Geosci. Remote Sens.* **32**, 598–615.
- Draper, N. R., and H. Smith, 1998: *Applied Regression Analysis* (Wiley, New York).
- Gao, F., X. Li, A. Strahler, and C. Schaaf, 2000: Evaluation of the LiTransit kernel for BRDF modeling. *Remote Sens. Rev.* **19**, 205–224.
- Gao, F., C. B. Schaaf, A. H. Strahler, and W. Lucht, 2001: Using a multi-kernel least-variance approach to retrieve and evaluate albedo from limited bi-directional measurements. *Remote Sens. Environ.* **76**, 57–66.
- Giard, D., and E. Bazile, 2000: Implementation of a new assimilation scheme for soil and surface variables in a global NWP model. *Mon. Weather Rev.* **128**, 997–1015.
- Hoerl, A. E., and R. W. Kennard, 1970: Ridge regression: biased estimation for non-orthogonal problems. *Technometrics* **12**, 55–67.
- Kimes, D. S., 1983: Dynamics of directional reflectance factor distribution for vegetation canopies. *Appl. Opt.* **22**, 1364–1372.
- Lawly, D. N., and A. E. Maxwell, 1963: *Factor Analysis as a Statistical Method* (Butterworths, London).
- Li, X., F. Gao, J. Wang, and A. H. Strahler, 2001: A priori knowledge accumulation and its application to linear BRDF model inversion. *J. Geophys. Res.* **106**, 11 925–11 935.
- McCabe, G. P., 1979: Evaluation of regression coefficient estimates using α -acceptability. *Technometrics* **20**, 131–139.
- McDonald, G. C., 1975: Discussion of “Ridge analysis following a preliminary test of a shrunken hypothesis”. *Technometrics* **17**, 443–445.

- McDonald, G. C., and R. C. Schwing, 1973: Instability of regression estimates relating air pollution to mortality. *Technometrics* **15**, 463–481.
- Obenchain, R. L., 1977: Classical F-tests and confidence regions for ridge regression. *Technometrics* **19**, 429–439.
- Pokrovsky, O. M., 1984: *Optimization of the Meteorological Remote Sensing* (Hydrometeorizdat, Leningrad) (in Russian).
- Pokrovsky, O. M., and J. L. Roujean, 2003a: Land surface albedo retrieval via kernel-based BRDF modeling: I. Statistical inversion method and model comparison. *Remote Sens. Environ.* **84**, 100–119.
- Pokrovsky, O. M., and J. L. Roujean, 2003b: Land surface albedo retrieval via kernel-based BRDF modeling: II. An optimal design scheme for the angular sampling. *Remote Sens. Environ.* **84**, 120–142.
- Pokrovsky, I. O., O. M. Pokrovsky, and J.-L. Roujean, 2003a: Development of an operational procedure to estimate surface albedo from the SEVIRI/MSG observing system by using POLDER BRDF measurements: I. Data quality control and accumulation of information corresponding to the IGBP land cover classes. *Remote Sens. Environ.* **87**, 198–214.
- Pokrovsky, I. O., O. M. Pokrovsky, and J.-L. Roujean, 2003b: Development of an operational procedure to estimate surface albedo from the SEVIRI/MSG observing system by using POLDER BRDF measurements: II. Comparison of several inversion techniques and uncertainty in albedo estimates. *Remote Sens. Environ.* **87**, 215–242.
- Ranson, K. J., L. L. Biehl, and M. E. Bauer, 1985: Variation in spectral response of soybean with respect to illumination, view and canopy geometry. *Int. J. Remote Sens.* **6**, 1827–1842.
- Roujean, J.-L., M. Leroy, and P. Y. Deschamps, 1992: A bi-directional reflectance model of the Earth's surface for the correction of remote sensing data. *J. Geophys. Res.* **97**, 20455–20468.
- Strahler, A. H., and J. P. Muller (Eds.), 1999: MODIS BRDF/Albedo Product: Algorithm Theoretical Basis Document, Version 5.0.
- Wanner, W., X. Li, and A. H., Strahler, 1995: On the derivation of kernels for kernel-driven models of bidirectional reflectance. *J. Geophys. Res.* **100**, 21077–21090.
- Wanner, W., A. H. Strahler, B. Hu, et al., 1997: Global retrieval of bidirectional reflectance and albedo over land from EOS MODIS and MISR data: theory and algorithm. *J. Geophys. Res.* **102**, 17143–17162.



From left to right: Ivan Andronov, Kirill Antoniuk, and Oleg Pokrovsky.

Plasmonic spectroscopy of 2D densely packed and layered metallic nanostructures

A. N. Ponyavina^{1*} and S. M. Kachan²

¹ Institute of Physics, National Academy of Sciences of Belarus, Nezavisimosti Ave. 68, Minsk, 220072, Belarus

² Belarusian National Technical University, Khmel'nitskogo Str. 9, Minsk, 220013, Belarus

Abstract. This chapter is an overview of size and concentration effects on electrodynamic coupling in two-dimensional densely packed arrays of metallic nanospheres in the frequency range of the surface plasmon resonance (SPR). Our theoretical analysis is based on the statistical theory of multiple scattering of waves. We show that concentration effects, such as the enhanced long-wavelength transmission of light and the strong resonance quenching of transmission, are effectively interpreted in terms of constructive and destructive interference of waves incident on and scattered by a monolayer of closely-packed submicrometer plasmonic particles. The concentration SPR red shift observed in the case of dipole metal nanoparticles is highly sensitive to the matrix refractive index and results from lateral near-field couplings. We also demonstrate phenomena caused by a strong plasmonic–photonic confinement in multilayered metal–dielectric nanostructures consisting of densely packed monolayers. For example, we show that employing the size and/or concentration gradient of dipole metallic nanoparticles in a quarter-wavelength multilayered system allows one to achieve an almost total absorbance.

Keywords: metallic nanostructures, surface plasmon resonance, electrodynamic coupling

1. Introduction

Recent progress in nanotechnologies and plasmonics promotes the development of new types of spectrally selective nanocomponents and devices needed in non-linear optics, laser physics, optoelectronics, energy and high-density data storage, information technologies, life sciences, and security [6,20,29,36,37].

Disperse plasmonic nanostructures, that is, materials containing noble-metal nanoparticles, are highly attractive owing to their unique optical properties caused by localized surface plasmon resonances (LSPR) of light attenuation accompanied by strongly enhanced optical fields near metal nanoparticles in this spectral range. The LSPR appears upon photoexcitation in the visible and near IR regions and re-

* Corresponding author. E-mail: ponyavina@imaph.bas-net.by

sults from collective electron oscillations confined by a metal-nanoparticle surface [7,21].

Among the most promising passive plasmonic nanomaterials are self-organized and densely packed metal–dielectric nanostructures [27,30,33] as well as layered metal–dielectric nanocomposites [3,5,24,25]. Their optical properties are further defined by collective effects resulting from the electronic and electrodynamic couplings and hence can be effectively tailored by an appropriate choice of material and nanostructure parameters.

In order to establish stable correlations between optical and topological properties of these plasmonic nanomaterials, one still needs to study and clarify the nature and manifestations of collective effects in more detail. This chapter is a brief review of results and outstanding problems in this novel field of plasmonics related to metal–dielectric nanocomposites with partially-ordered structures.

2. Localized surface plasmon resonance: spectral and near-field features

It is already well known that the fundamental difference of the optical properties of metal nanoparticles from both the properties of bulk samples and the characteristics of individual atoms is caused by resonance collective oscillations of conduction electrons near the surface of a metal particle induced by a light wave with an appropriate frequency ω_0 , called the Frolich frequency, defined approximately by the following conditions:

$$\operatorname{Re}[\varepsilon(\omega_0)] = -2\operatorname{Re}[\varepsilon_0(\omega_0)], \quad \operatorname{Im}[\varepsilon(\omega_0)] \approx 0, \quad \operatorname{Im}[\varepsilon_0(\omega_0)] \approx 0. \quad (1)$$

Here $\varepsilon(\omega_0)$ and $\varepsilon_0(\omega_0)$ are the dielectric permittivities of the metal particle and the surrounding medium, respectively.

In the electrostatic limit, it is easily seen that these conditions maximize both the scattering (Q_{sca}) and the absorption (Q_{abs}) efficiency factor of a spherical particle:

$$Q_{\text{sca}} = \frac{8k^4 r^4}{3} \left| \frac{\varepsilon - \varepsilon_0}{\varepsilon + 2\varepsilon_0} \right|^2, \quad Q_{\text{abs}} = 4kr \operatorname{Im} \left[\frac{\varepsilon - \varepsilon_0}{\varepsilon + 2\varepsilon_0} \right].$$

Here, k is the wave number and r is the particle radius.

LSPRs of individual metal particles defined in the visible through their extinction efficiency are highly sensitive to the nanoparticle material, size, shape, and internal structure as well as to the dielectric properties of the surrounding medium. One of the main features of an LSPR, which manifests itself as a change of color of a sparse metal colloid, is the red shift of the LSPR with increasing nanoparticle size (see Fig. 1) or increasing refractive index of the surrounding medium. This feature is caused by retardation effects and the Coulomb screening, both reducing the surface plasmon frequency [21].

It is also important to note that increasing the particle size up to tens or hundreds of nanometers also leads to drastic changes in the modal structure of the extinction spectrum of non-interacting nanoparticles, viz., higher-order modes ap-

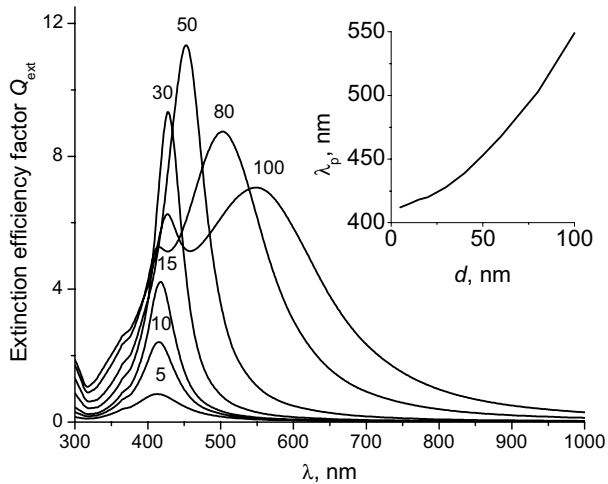


Fig. 1. Spectral extinction efficiency factors of silver nanoparticles dispersed in a medium with a refractive index of 1.5 for different particle diameters (as indicated near the respective curves). The insert shows the dependence of the LSPR peak wavelength on the particle diameter.

pear, while the modes' relative weights change. For low-order modes, the type of plasmon-excitation decay changes from non-radiative to radiative, while particles become strong scatterers. In the respective spectral regions, the SPRs in the extinction spectra of nanoparticles essentially become scattering resonances (see Fig. 2).

To calculate individual-nanosphere LSPRs, one can use the Mie theory, while other advanced single-scattering numerical techniques (such as the DDA, VIEF, and T -matrix methods) can be applied to nonspherical nanoparticles and nanoparticle clusters [26,32].

As a rule, experimentally measured LSPR bands are much broader than the theoretically calculated ones, even for arrays with a high degree of nanoparticle monodispersity. The nature of the strong widening of bands is still under discussion [23]. One of the main causes is the so-called internal size effect resulting in a modified nanoparticle permittivity ε . Most easily understood is the explanation of the size-dependent permittivity ε in the framework of the ballistic model, also known as the limitation of the electron mean free path (LEMFP) model [21]. An attractive feature of this approach is that it yields the size-dependent ε of metal nanospheres and nanoshells [11] in a closed analytical form.

According to the LEMFP model, the scattering of conduction electrons at the nanoparticle surface leads to a reduction in the effective mean free path of electrons compared to that in a bulk material. The longer the mean free path of electrons in the bulk material, the wider the range of nanoparticle sizes for which the internal size effect is observed. Among noble metals, this range is the widest for

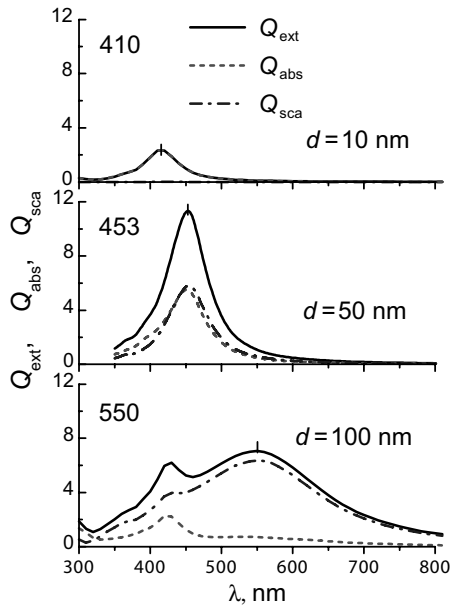


Fig. 2. Spectral dependence of extinction, absorption and scattering efficiency factors for silver nanospheres with different diameters d , embedded in the matrix with a refractive index of 1.5.

silver nanoparticles. Despite this factor, the LSPR of Ag nanoparticles is the most promising from the practical standpoint, owing to the ready availability of this material. Furthermore, radiative losses and high-order modes appear for silver nanoparticles at smaller sizes than for copper and gold nanoparticles. All of these traits are a consequence of the difference in the spectral interval between the LSPR and interband absorption for these noble metals.

However, the knowledge of scattering and/or absorption efficiency factors is not sufficient to define the optical spectra of densely packed plasmonic arrays which depend, via strong electrodynamic coupling, on both far-field and near-field characteristics of single scattering.

As to plasmonic nanoparticles, their near-field patterns are still poorly known in comparison with their far-field spectral and scattering counterparts. We can only mention the detailed study of the near-field enhancement (“hot spots”) for fractal plasmonic nanoarrays based on the electrostatic approximation [36] and limited data on single plasmonic nanoparticles (see references in [6]) obtained with the Mie theory and the DDA technique. These data demonstrate a high sensitivity of the near-field distribution to particle sizes, shapes, and internal structures as well as to optical properties of the surrounding medium and the polarization state of the incident light. For example, it has been recognized that the maximal

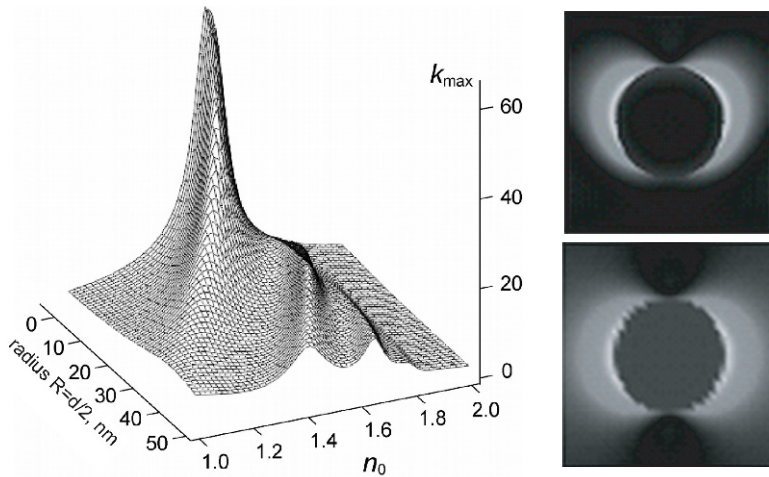


Fig. 3. On the left: dependence of the coefficient of maximal local electric field enhancement near a silver nanosphere on its size and the refractive index of the surrounding medium. On the right: near-field distributions for silver nanospheres with $d = 50$ nm (upper panel) and 20 nm (lower panel). The wavelength is 400 nm (after [2]).

coefficient of the local field enhancement for silver nanospheres exhibits a non-monotonous dependence on nanoparticle sizes and the matrix refractive index (see Fig. 3).

The most important feature essential for the understanding of the near-field contribution to electrodynamic coupling is that the hot-spot distribution strongly expands along the direction of the incident light if the particle size and/or the matrix refractive index increase [2,28].

3. Statistical approach to 2D densely packed metallic nanoparticle arrays

Electrodynamic coupling manifests itself as multiple coherent re-scattering and, as a consequence, is highly sensitive to the type of plasmonic nanoparticles and their packing density. For these reasons, the existing effective-medium theories are not always convenient and/or adequate [7]. Modeling and studies of coupled nanoarrays are considered to be one of the frontiers of modern theoretical plasmonics. In the framework of the electrostatic approximation, the features of SPRs in fractal arrays of metal nanoparticles have been studied in [36]. Two-dimensional (2D) and three-dimensional (3D) regular plasmonic nanoarrays were analyzed with the use of the plane-wave expansion method [31], the finite-difference time-domain method, and the transfer-matrix method [39]. In [38], certain principles of the theory of multiple scattering of waves were combined with the Bloch wave's formalism, which allows one to calculate both the band structure and the transmission and reflection coefficients for perfectly regular particle ar-

rays. However, topological and/or size disorders introduced in plasmonic nano-materials during their fabrication with one of the existing technologies allows one to classify these materials as not perfectly regular, but rather as highly-ordered arrays of nanoparticles.

In our opinion, an effective way to describe the features of electrodynamic coupling in partially ordered plasmonic nanoarrays is to use the approach based on the statistical theory of multiple scattering of waves (STMSW) [9]. According to the STMSW, the resulting field $\mathbf{E}(\mathbf{r})$ at any point in space \mathbf{r} (either inside or outside a disperse medium) and the effective field $\mathbf{E}_{\text{eff}}(\mathbf{r})$ at any particle are both superpositions of the incident and multiply scattered waves:

$$\begin{aligned}\mathbf{E}(\mathbf{r}) &= \mathbf{E}_0(\mathbf{r}) + \sum_j \mathbf{t}(\mathbf{r}, \mathbf{r}_j) \mathbf{E}_{\text{eff}}(\mathbf{r}_j), \\ \mathbf{E}_{\text{eff}}(\mathbf{r}_j) &= \mathbf{E}_0(\mathbf{r}_j) + \sum_{j \neq i} \mathbf{t}(\mathbf{r}_j, \mathbf{r}_i) \mathbf{E}_{\text{eff}}(\mathbf{r}_i).\end{aligned}\quad (2)$$

Here $\mathbf{t}(\mathbf{r}, \mathbf{r}_j)$ is the single-particle scattering operator.

The STMSW accounts for electrodynamic coupling in spatially correlated arrays through the superposition and subsequent averaging of fields scattered by all nanoparticles. The STMSW operates with the field moments, which are the values averaged over all possible configurations of a given particle array. The main characteristics of the STMSW are the mean (or coherent) field $\langle \mathbf{E}(\mathbf{r}) \rangle$, the coherent intensity $I_c = |\langle \mathbf{E}(\mathbf{r}) \rangle|^2$, and the field covariance $\langle \mathbf{E}(\mathbf{r}) \cdot \mathbf{E}^*(\mathbf{r}') \rangle$.

Lax [22] proposed to average Eqs. (2) over different configurations of a particle ensemble with the assumption that the fixation of any particle specifies the spatial configuration of the whole assembly. It is clear that the more regular the particle array, the better this assumption. This scheme was called the quasicrystal-line approximation (QCA). More recently, Hong [8] considered 2D particles ensembles in the framework of the QCA and derived the following system of two equations defining the mean field at a \mathbf{r} :

$$\begin{aligned}\langle \mathbf{E}(\mathbf{r}) \rangle &= \mathbf{E}_0(\mathbf{r}) + p_0 \int d\mathbf{R} \int d\mathbf{r}' \Gamma(\mathbf{r}, \mathbf{r}' + \mathbf{R}) \langle \mathbf{E}(\mathbf{r}' + \mathbf{R}) \rangle_{\mathbf{R}}, \\ \langle \mathbf{E}(\mathbf{r} + \mathbf{R}) \rangle_{\mathbf{R}} &= \mathbf{E}_0(\mathbf{r} + \mathbf{R}) + \int d\mathbf{r}' \Gamma(\mathbf{r} + \mathbf{R}, \mathbf{r}' + \mathbf{R}) \langle \mathbf{E}(\mathbf{r}' + \mathbf{R}) \rangle_{\mathbf{R}} \\ &\quad + p_0 \int d\mathbf{R}' g(|\mathbf{R} - \mathbf{R}'|) \int d\mathbf{r}' \Gamma(\mathbf{r} + \mathbf{R}, \mathbf{r}' + \mathbf{R}') \langle \mathbf{E}(\mathbf{r}' + \mathbf{R}') \rangle_{\mathbf{R}, \mathbf{R}'},\end{aligned}\quad (3)$$

where p_0 is the number of particles per unit area; $\mathbf{E}_0(\mathbf{r})$ is the incident field, $\langle \mathbf{E}(\mathbf{r}) \rangle_{\mathbf{R}}$ is the average field with one particle fixed at a point \mathbf{R} ; $\langle \mathbf{E}(\mathbf{r}' + \mathbf{R}') \rangle_{\mathbf{R}, \mathbf{R}'}$ is the average field with two particles fixed at points \mathbf{R} and \mathbf{R}' ; $\Gamma(\mathbf{r}, \mathbf{r}')$ is the tensorial Green function; and $g(|\mathbf{R} - \mathbf{R}'|)$ is the radial distribution function defined as the probability for two particles to be separated by a distance $R = |\mathbf{R} - \mathbf{R}'|$.

Characteristic features of the radial distribution function depend on the type of disperse medium. For a sparse medium, when interparticle distances are considerably greater than the particle size, $g(|\mathbf{R} - \mathbf{R}'|) = 1$ for any $|\mathbf{R} - \mathbf{R}'| \geq d$. This

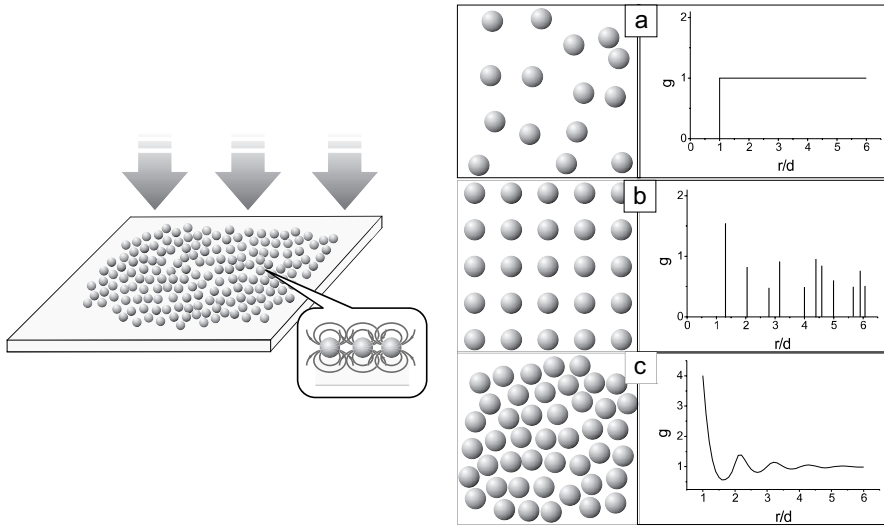


Fig. 4. The structure of a densely packed monolayer (on the left) and radial distribution functions (on the right) of (a) a sparse random as well as (b) regularly and (c) densely packed partially ordered disperse layers.

means that the particles are randomly distributed and the scattered wave phases are random. On the other hand, regular structures like “nanoparticle crystals” are characterized by long-distance ordering. Particles embedded at specific locations in the “crystal lattice” are separated by well-defined distances. For a perfect regular particle array, the $g(|\mathbf{R} - \mathbf{R}'|)$ is a set of delta functions corresponding to the fixed lattice nodes \mathbf{R}_i . In this case, the scattered wave phases are perfectly deterministic.

A partially ordered structure occupies an intermediate place between the extreme cases of a sparse group and a regular system. For example, as can be seen in Fig. 4, densely packed arrays are characterized by short-range ordering. The locations of neighboring particles are partially correlated due to finite particle sizes and the high particle concentration. As a consequence, $g(|\mathbf{R} - \mathbf{R}'|)$ has pronounced maxima corresponding to the most probable distances between particles. For particles located in the range of short ordering, the scattered waves are partially coherent, and the corresponding phase relations are not random. The stronger the spatial ordering owing to an increased particle concentration, the more significant the interference of the scattered waves. The denser the nanoparticle packing, the stronger the near-field coupling.

The $g(|\mathbf{R} - \mathbf{R}'|)$ function for a densely packed system of particles can be calculated by using the model of solid spheres in the framework of the Percus–Yevick approximation [42].

When the particles are spheres, it is convenient to solve Eqs. (3) by expanding the electromagnetic fields and the tensorial Green function in vector spherical harmonics. Then one obtains the following relations for the coherent transmission T_m and reflection R_m coefficients of a partially ordered monolayer of monodisperse particles [1,34,35]:

$$\begin{aligned} T_m &= \left| 1 - \frac{\pi}{k^2} p_0 \sum_l (2l+1)(b_{lM} + b_{lE}) \right|^2, \\ R_m &= \left| -\frac{\pi}{k^2} p_0 \sum_l (-1)^l (2l+1)(b_{lM} - b_{lE}) \right|^2. \end{aligned} \quad (4)$$

Here, the coefficients b_{lM} and b_{lE} are determined from the following system of equations:

$$\begin{aligned} b_{lM} &= b_l + p_0 b_l \sum_{l'} (P_{ll'} b_{l'M} + Q_{ll'} b_{l'E}), \\ b_{lE} &= a_l + p_0 a_l \sum_{l'} (P_{ll'} b_{l'M} + Q_{ll'} b_{l'E}), \end{aligned}$$

where a_l and b_l are the Mie coefficients. The terms containing the functions $P_{ll'}$ and $Q_{ll'}$ depend on $g(|\mathbf{R} - \mathbf{R}'|)$ in a complicated manner [34]. When the individual-particle permittivity becomes size-dependent, one must take this into account while calculating the Mie coefficients.

It can easily be shown that neglecting the lateral electrodynamic coupling in Eqs. (3) yields the following simple analytical expressions for the coherent transmission and reflection coefficients:

$$T_m = 1 - \eta Q_{\text{ext}} + \frac{4\pi\eta^2}{\rho^2} x(0) \Lambda Q_{\text{ext}}, \quad R_m = \frac{4\pi\eta^2}{\rho^2} x(0) \Lambda Q_{\text{ext}}. \quad (5)$$

Here, Q_{ext} , $x(\gamma)$, and ΛQ_{ext} are the single-particle extinction efficiency factor, phase function, and scattering efficiency factor, respectively; $\rho = \pi d / \lambda$ is the size parameter; d is the particle diameter; and $\eta = p_0 \pi d^2 / 4$ is the monolayer overlap parameter defined as the fraction of a monolayer area occupied by nanoparticles. Note, that these expressions, called the coherent single-scattering approximation (CSSA), can be quite accurate if the single-particle phase function is strongly elongated in the forward-scattering direction.

4. Electrodynamic coupling and spectral properties of 2D densely packed plasmonic nanostructures

Surface plasmon photoexcitation in nanoarrays consisting of densely packed or partially ordered metal nanoparticles has a collective nature. Small interparticle distances and short-range ordering of particles cause strong electrodynamic coupling. Its optical manifestations strongly depend on sizes of plasmonic nanoparticle.

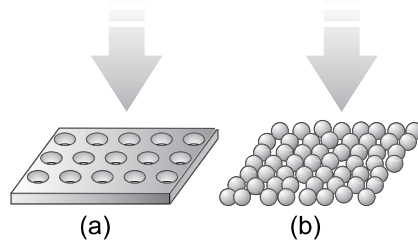


Fig. 5. Two-dimensional metal–dielectric nanostructures: (a) a perforated thin metal film and (b) a densely packed array of metal nanoparticles (after [15]).

4.1. Enhanced transmission through monolayers with extremely dense packing of submicron plasmonic nanoparticles

The recent discovery of enhanced resonant transmission of light through optically thick metal films perforated with an array of subwavelength holes [4] has attracted great attention, and its plasmon-induced origin in the case of *perforated* structures of different geometries is now actively explored. An analogous phenomenon is expected to arise in 2D densely packed *disperse* metal structures with the scale of inhomogeneities comparable to that in perforated metal films. One particular example is a 2D array of densely packed, submicrometer metal nanoparticles, which is essentially the reverse of an array of holes in a metal film (compare Fig. 5b with Fig. 5a). Substantial progress in the synthesis of such 2D arrays of submicrometer metal nanoparticles by a self-assembling organization has recently been reported [20].

We analyzed theoretically the feasibility of enhanced transmission of visible light through 2D random, densely packed arrays of submicrometer noble-metal nanoparticles and established its electrodynamic nature [15].

Figure 6 shows spectra of the direct transmission of light through a 2D densely packed nanoparticle array calculated in the framework of the QCA, wherein the lateral electrodynamic coupling between particles is considered as the interference of multiply scattered waves in partially-ordered arrangement of scatterers. The transmission spectrum of an equivalent sparse system composed of non-interacting nanoparticles was calculated from the standard Bouguer law: $T_B = \exp(-\eta Q_{\text{ext}})$. As one can see, the densely packed array is characterized by an enhanced long-wavelength transmission (ET) with a peak value of 87% (and even higher for larger particles) within a lower-order surface plasmon extinction band inherent to the same but uncoupled particles. Thus, the 2D disperse metal–dielectric layers of submicron silver nanoparticles can exhibit enhanced transparency in the LSPR spectral region (see also [19,27]). The ET is a result of constructive interference coupling between the incident light and radiative delocalized surface plasmon modes of lower orders. This effect becomes stronger with increasing density, since the scale of surface plasmon localization increases due to increasing range of nanoparticle correlations and near-field coupling.

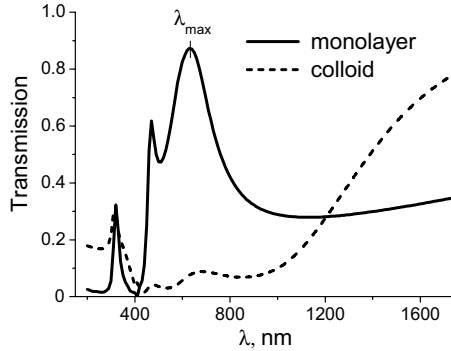


Fig. 6. Transmission spectrum of a 2D densely packed array of silver particles ($d = 200$ nm; $\eta = 0.7$; $n_m = 1.5$) and that of a sparse system consisting of the same particles (after [15]).

It should be noted that 2D disperse nanostructures have obvious advantages over perforated metal films since they provide additional possibilities for tuning their optical response. While the change in the shape of inhomogeneities and the type of arrangement is applicable to both structures considered, the inhomogeneity composition and the controlling interior structure are realized much easier with densely packed nanoparticle arrays.

4.2. Strong light quenching by monolayers of submicron plasmonic nanoparticles

With an intermediate concentration of submicrometer plasmonic particles in a monolayer, a strong resonance quenching (SRQ) may arise. The SRQ was observed experimentally for the first time by Chumanov with colleagues for monolayers consisting of 100-nm Ag nanoparticles and having an overlap parameter of 0.3 [27]. It is worthwhile to note that their experimental results are in excellent agreement with our simulations based on the QCA (see Fig. 7).

Among the reasons leading to the SRQ, specific plasmonic coupling and near-field effects have been suggested in [27]. However the fact that the most pronounced SRQ was realized not with the highest but rather with intermediate concentrations necessitates searching for other mechanisms rooted in coherent scattering.

In order to answer the question concerning the nature of this effect, let us consider the simplest scheme of multiple coherent scattering (namely the CSSA), which takes into account coherent coupling only as the interference of waves singly scattered by individual particles (see Eqs. (5)). Re-scattering between particles is not taken into account. Each individual particle is illuminated by the incident wave only. This model yields simple conditions of the total interference quenching (TIQ) caused by a destructive interference of the incident and scattered waves. This happens when the amplitudes of the incident wave and the total scattered

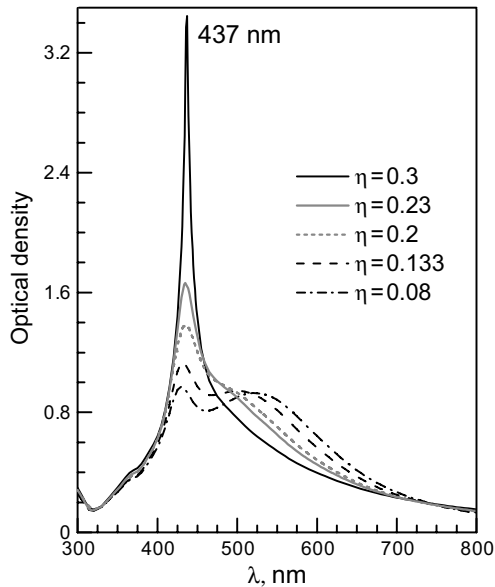


Fig. 7. Optical density of 2D arrays of $d=100$ nm silver particles embedded in polydimethylsiloxane, for different surface particle concentrations. The calculation are based on the QCA for the experimental conditions of [27].

wave are equal, while their phases are opposite [10]. According to Eq. (5), $T_m = 0$ if

$$\frac{16 |S(0)|^2}{\rho^4 Q_{\text{ext}}} = Q_{\text{ext}}, \quad \eta_0 = \frac{\rho^4 Q_{\text{ext}}}{8 |S(0)|^2}.$$

Here, $S(\gamma)$ is the single-scattering amplitude.

Importantly, the conditions of the TIQ based on the CSSA are analogous to the conditions of the SRQ in the framework of the QCA. Detailed information is given in Fig. 8; as one can see, both approximations yield qualitatively similar dependences of the optimal overlap parameter and the spectral position of the SRQ on the particle size. Furthermore, for submicrometer particle sizes these results agree numerically. Residual quantitative differences are related to the fact that the CSSA does not take into account coherent re-scattering, which intensifies with decreasing size parameter.

Thus, the main results are the following. The interference quenching is never realized for monolayers of dipole plasmonic nanoparticles, even at the highest particle concentration. The optimal overlap parameter changes non-monotonically when particle sizes grow. Furthermore, the spectral position of the SRQ gradually shifts toward the spectral position of high-order modes with increasing particle sizes.

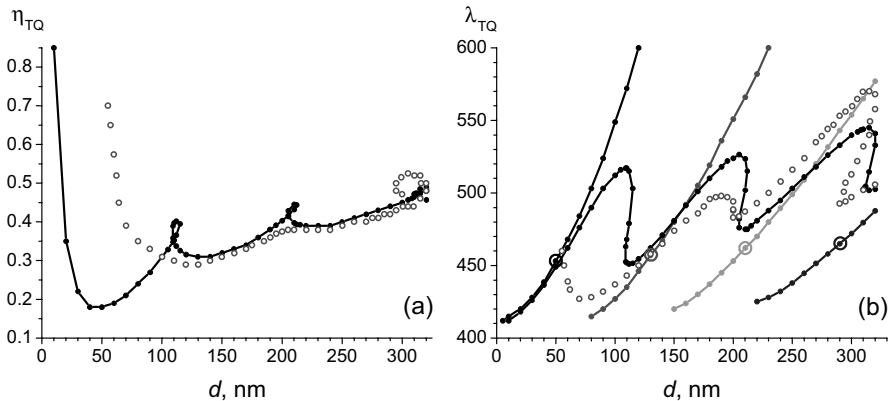


Fig. 8. (a) Optimal overlap parameter and (b) spectral position of the SRQ peak for a monolayer of silver nanoparticles embedded in a matrix with $n_m = 1.5$. Filled circles: calculations based on the CSSA (quenching until $T_m = 10^{-3} - 10^{-4}$); open circles: calculations based on the QCA ($10^{-4} - 10^{-5}$). “Branches” in (b) show spectral positions of the LSPR modes.

4.3. Concentration red shift of the collective plasmon resonance frequency

It was already noted that nanoparticle positions in the range of short ordering and the respective phases of scattered waves are strongly correlated in densely packed structures. Furthermore, for dipole nanoparticles the hot spots are concentrated in lateral directions, which supports strong near-field coupling. As a consequence, a new scale of surface plasmon localization arises. It may be quantitatively identified with the existence of “effective plasmonic particles” with a size comparable to the scale of short-range ordering. Increasing concentration leads to a growing scale of the surface plasmon localization and increasing size of the effective plasmonic particles (see Fig. 9). Thus, the close proximity and near-range ordering of dipole plasmonic nanoparticles are accompanied by the appearance of attenuation bands attributed to the coupled (collective) plasmon absorbance resonance [13,14].

It is worth noting that there are two ways of controlling the particle concentration. The first one is to adjust the system volume for a fixed number of particles (compare the solid and dot-dashed curves for $\eta = \text{const}$ in Fig. 10). The second way is to change the number of particles for the same system volume (compare, e.g., solid curves for different η in Fig. 10).

As can be seen from Fig. 10, in either case the resulting increase in the particle packing density causes a red shift of the SPR frequency calculated based on the QCA. Furthermore, when interparticle distances decrease, the enhancement of electrodynamic coupling leads to strengthening of the plasmon resonance. On the other hand, for sparse systems of the same particles (i.e., calculations based on the

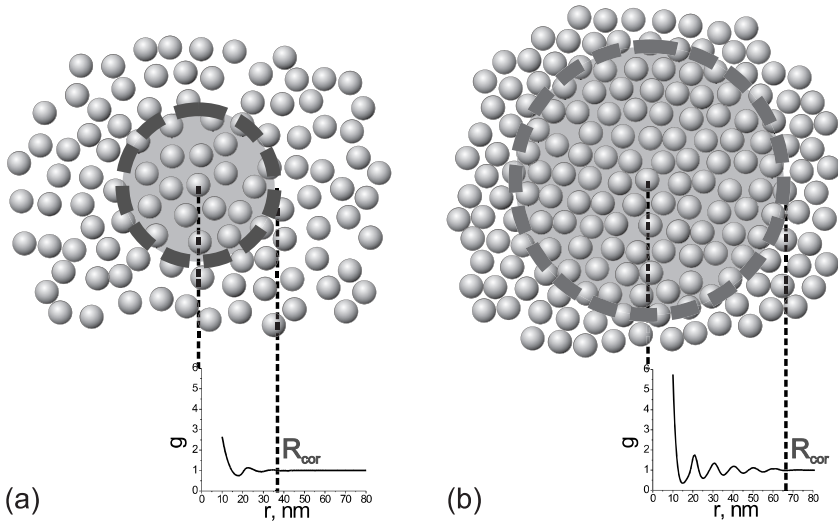


Fig. 9. Concentration-caused enlargement of the surface plasmon localization scale in 2D densely packed arrays: (a) $\eta = 0.4$ and (b) $\eta = 0.7$.

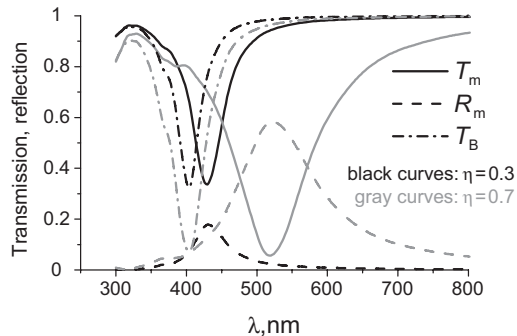


Fig. 10. Transmission T_m and reflection R_m spectra of monolayers made of silver nanoparticles with $d = 2$ nm, for different overlap parameters η (the matrix refractive index $n_m = 1.4$). The transmission spectrum T_B corresponds to a sparse system of the same nanoparticles.

Bouger law), there is no plasmon frequency change with increasing overlap parameter.

The second case, wherein the particle density is modified by changing the number of particles, was realized experimentally, for example, by Kreibig et al. [21]. They noted that the interference of waves scattered by densely packed nanoparticle aggregates acts, apparently, in the same manner as the growth of an

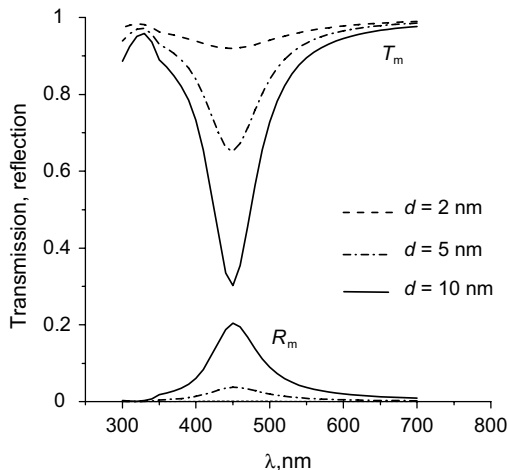


Fig. 11. Transmission T_m and reflection R_m spectra of monolayers made of silver nanoparticles with different diameters d ($\eta = 0.4$ and $n_m = 1.4$).

isolated particle size. The aggregation of particles is somewhat similar to their growth. In the framework of the QCA, the appearance of densely packed nanoparticle aggregates is defined by increasing short-range ordering.

However, the red concentration shift may be caused not only by the short-range electrodynamic interactions related to the particle aggregation and, consequently, to a change in the nearest particle surrounding. In addition, it may be caused by the effective field modification as a whole when the particle concentration grows [13,14]. Apparently, a single inclusion is excited by a wave propagating in some effective medium associated with the average coherent field and characterized by the effective refractive index.

Then, by analogy with the Frolich frequency ω_0 determining the LSPR spectral position via the conditions (1), we introduce the effective Frolich frequency ω'_0 which determines the collective SPR frequency:

$$\text{Re}[\varepsilon(\omega'_0)] = -2 \text{Re}[\varepsilon_{\text{eff}}(\omega'_0)], \quad \text{Im}[\varepsilon(\omega'_0)] \approx 0, \quad \text{Im}[\varepsilon_{\text{eff}}(\omega'_0)] \approx 0.$$

The effect of the particle size on characteristics of the collective SPR is shown in Fig. 11. The growth of particle diameters over the considered range of Rayleigh scattering ($d < 10$ nm) is accompanied by the amplification and narrowing of the plasmon attenuation peak, with the resonance spectral position being defined only by the particle concentration.

It is interesting that selective reflectance in the SPR spectral region is enhanced significantly when the particle size increases. For example, for monolayers of silver nanoparticles with $\eta = 0.4$, the maximal reflectance increases from 5% to 20% as the particle diameter changes from 5 to 10 nm (see Fig. 11).

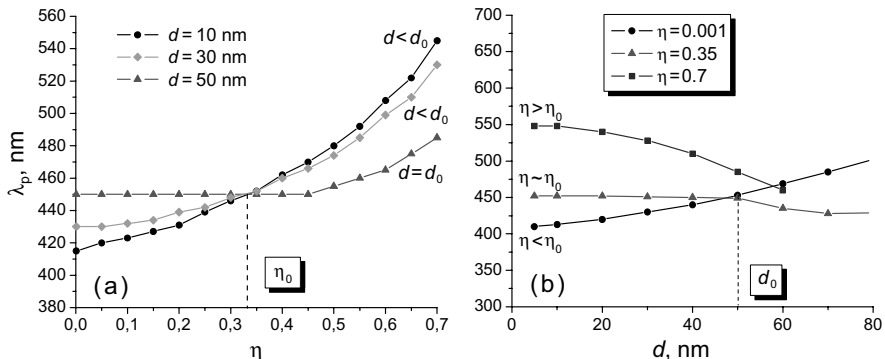


Fig. 12. (a) Concentration and (b) size dependences of the SPR peak wavelength λ_p for a silver nanoparticle monolayer [16].

Thus, the experimental data and their numerical simulation based on the QCA demonstrate that the coherent collective effects in densely packed monolayers of dipole plasmonic nanoparticles result in a red shift (with respect to that of isolated particles) of the SPR frequency as well as to the SPR amplification and broadening. The lateral electrodynamic coupling transforms the structure of plasmon resonances and effects their spectral positions.

The results on the size and concentration dependences of the dipole SPR maxima are summarized in Fig. 12. One can use these data to retrieve nanoparticle sizes and concentrations from experimental spectroscopic data [16].

4.4. Enhanced response of 2D densely packed arrays of metal nanoparticles to environmental changes

Nanostructures consisting of noble-metal nanoparticles have become increasingly popular in various sensor applications [6,36,37]. One of the most often utilized effects is the spectral shift of the SPR in response to local changes of the ambient refractive index n_m .

We analyzed theoretically the SPR spectra of 2D densely packed arrays of plasmonic nanospheres for different values of n_m , nanoparticle diameter d , and surface concentration c and revealed a dramatic growth of the sensitivity factor $S = \Delta\lambda_{\max}/\Delta n_{\max}$ with increasing η for small d (see Fig. 13) [15,41]. This growth can be explained in terms of a large effective scatterer formed by correlated nanoparticles in an array with a short-range ordering.

However, we found the largest sensitivity factors (≥ 400 nm/RIU) for 2D densely packed arrays of submicrometer noble-metal nanospheres with $d \geq 200$ nm. The spectral position of the ET band is very sensitive to changes in the surrounding medium (see Fig. 14).

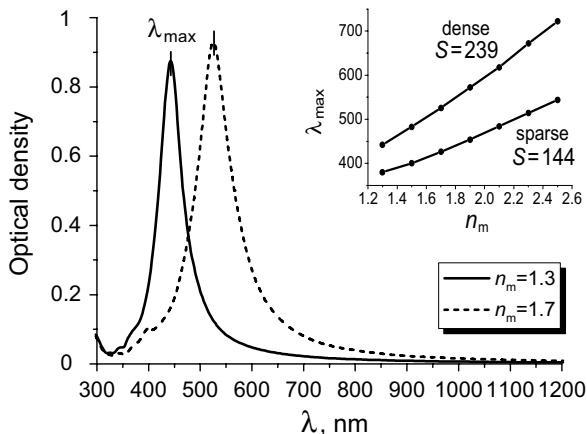


Fig. 13. Optical density spectra of a 2D densely packed array of dipole silver nanospheres ($d = 10$ nm, $\eta = 0.4$) in various dielectric environments. The inset shows the SPR peak position as a function of n_m for different packing densities [41].

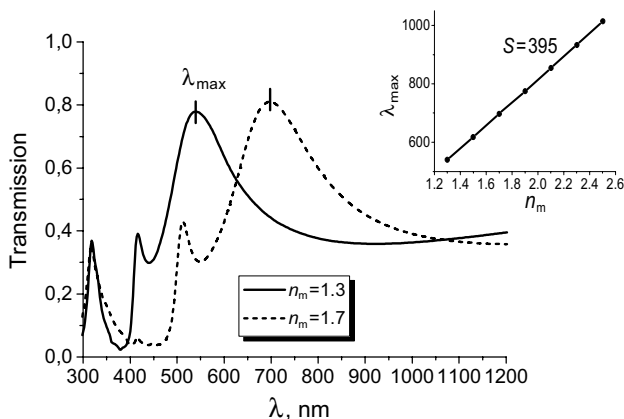


Fig. 14. Transmission spectra of a 2D densely packed array of submicrometer silver spheres ($d = 200$ nm, $\eta = 0.4$) in various dielectric environments. The inset shows the ET peak position as a function of n_m [15].

Figures 15 and 16 show the dependences of the SPR sensitivity on nanoparticle material and packing density. As one can see, monolayers consisting of silver nanoparticles are much more effective than those containing gold and copper ones.

4.5. Controlling light absorption in a monolayer of metal nanoparticles

Let us now consider in more detail the spectral absorbance of densely packed monolayers of metal nanoparticles [17]. It was already noted that the maximum of

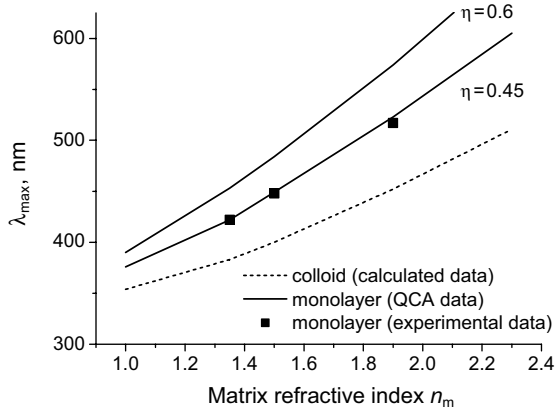


Fig. 15. Dependence of the SPR peak position on the matrix refractive index for single silver nanospheres ($d = 5$ nm) and densely packed monolayers of such nanoparticles with different surface concentrations η (after [41]).

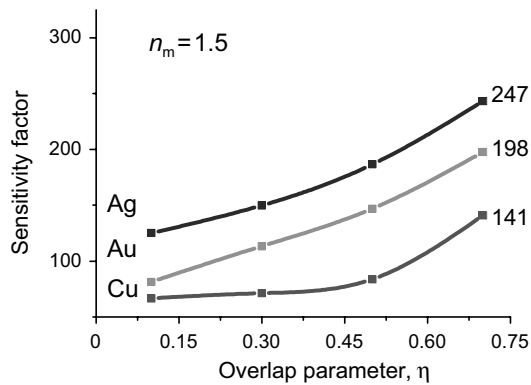


Fig. 16. Dependence of the nanoparticle monolayer sensitivity factor S on the overlap parameter for different noble metals ($d = 10$ nm; after [41]).

the absorption band is observed at the wavelength of the collective plasmon resonance λ_p . Over the Rayleigh range of particle sizes (less than 20 nm for silver), the plasmon frequency of a monolayer hardly depends on the nanoparticle diameter d and, in fact, is primarily determined by the nanoparticle surface concentration (see Fig. 17a). In contrast to λ_p , the value of maximal absorption of light at the frequency of the plasmon resonance depends strongly on the nanoparticle diameter. For non-negligible surface concentrations (i.e., when the lateral electrodynamic interactions between the particles become significant), there is an optimal diameter d_0 maximizing the monolayer absorption for a given η

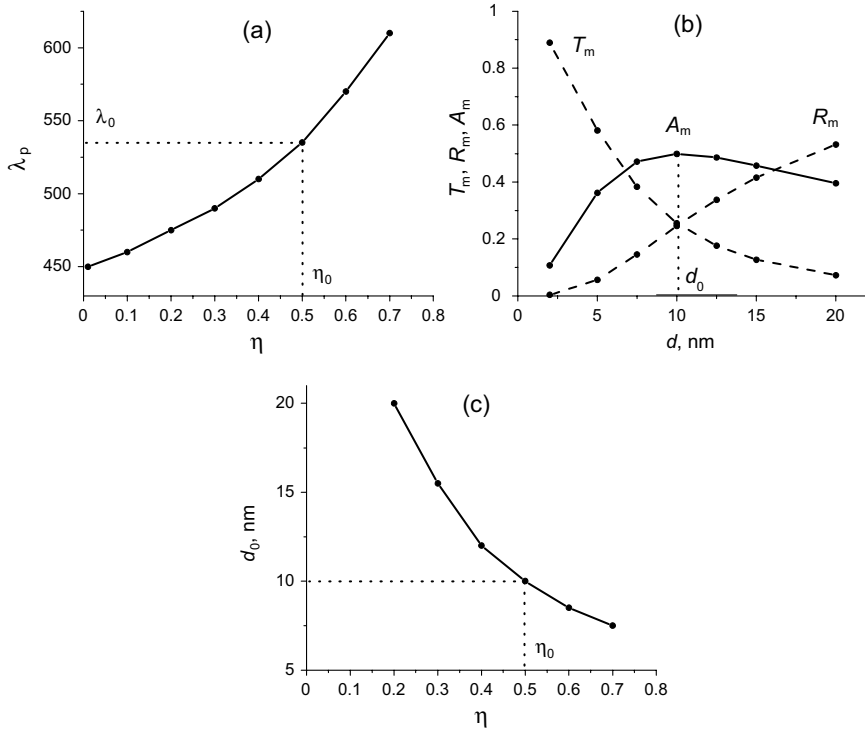


Fig. 17. Characteristics of a monolayer of silver nanoparticles in a Al_2O_3 matrix ($n_m = 1.75$): (a) the dependence of the plasmon resonance wavelength λ_p on the nanoparticle concentration; (b) the dependence of the resonant values of the spectral coefficients on the nanoparticle diameter for $\eta = 0.5$ ($\lambda_p = 530$ nm); (c) the dependence of the nanoparticle diameter d_0 maximizing the absorption in a monolayer on the nanoparticle concentration.

(see Fig. 17b). The dependence of d_0 on the particle concentration is shown in Fig. 17c. Since the diffuse scattering of light by a densely packed monolayer of small metal nanoparticles is negligible, the absorption of light is strictly determined by the direct transmittance, T_m , and specular reflection, R_m , coefficients: $A = 1 - T_m - R_m$.

Thus, we can propose a methodology for choosing the parameters of a monolayer that maximize the absorbance in a single monolayer of metal nanoparticles at a given wavelength.

It is interesting to note that, as follows from our numerical calculations, the maximal absorption at the plasmon frequency corresponds to the condition $T_m \approx R_m$, with T_m and R_m being equal to 25% with high accuracy. Thus, the maximal absorption in a densely packed monolayer of metal nanoparticles is equal to $A_m^{\max} \approx 50\%$. The high accuracy of these conditions in a wide range of parameters

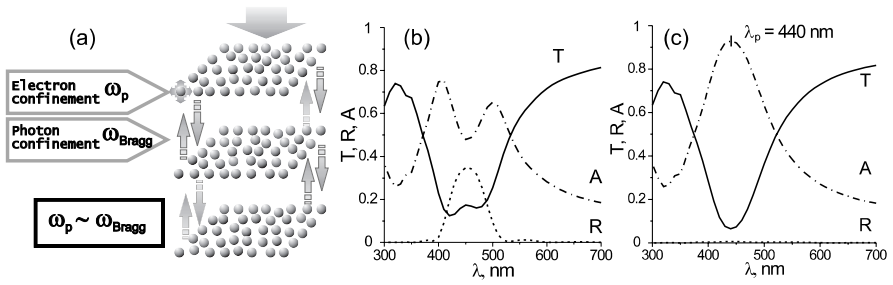


Fig. 18. (a) The general scheme and (b,c) calculated spectra of a multilayer stack composed of silver nanoparticle monolayers ($d = 3.5$ nm; $\eta = 0.4$; number of monolayers $N = 7$) separated by dielectric films with different optical thicknesses τ : (b) $\tau = \lambda_p/2$; (c) $\tau = \lambda_p/4$.

evidently implies some fundamental properties of monolayers and deserves a more detailed study.

5. Plasmonic–photonic confinement in multilayered metal nanostructures

An effective way to control spectral properties of metal nanostructures is their subwavelength periodicity leading to strong photonic confinement and causing the appearance of photonic band gaps (PBGs). Spatial ordering may be realized in different ways. Technologically, one of the simplest approaches is to form a stack of metal nanoparticle monolayers separated by solid intermediate films. A system thus formed is a kind of one-dimensional (1D) photonic crystal owing to the 1D periodicity of the dielectric function. When the PBG and the SPR are realized at similar frequencies, the photonic and plasmonic modes are strongly coupled, and the hybrid resonances are extremely sensitive to all nanocomposite parameters (see Fig. 18) [12].

The transfer matrix–QCA (TM–QCA) scheme that we used for numerical simulations defines the coherent fields transmitted and reflected by an N -monolayer stack as follows [35]:

$$\begin{aligned} \langle \mathbf{E}(\mathbf{z}) \rangle &= \exp(ikz) \left(\mathbf{e} + \sum_{j=1}^N \mathbf{G}_j^+ \right), \\ \langle \mathbf{E}(-\mathbf{z}) \rangle &= \exp(ikz) \sum_{j=1}^N \mathbf{G}_j^- \exp[(j-1)2ikl_m]. \end{aligned} \quad (6)$$

Here, $i = (-1)^{1/2}$, $\mathbf{G}_j^\pm = \mathbf{G}(\pm \mathbf{z})$ are the scattering amplitudes in the forward and backward directions for the j th monolayer in the presence of the other monolayers of the system, and l_m is the interlayer distance, i.e., the spatial interval between the centres of adjacent monolayers.

The method of self-consistent field leads to the following system of equations:

$$\begin{aligned} \mathbf{G}_j^+ &= \mathbf{F}^+ + \mathbf{F}^+ \sum_{p=1}^{j-1} \mathbf{G}_p^+ + \sum_{p=j+1}^N \mathbf{G}_p^- \exp[(p-j)2ikl_m], \\ \mathbf{G}_j^- &= \mathbf{F}^- + \mathbf{F}^- \sum_{p=1}^{j-1} \mathbf{G}_p^+ + \sum_{p=j+1}^N \mathbf{G}_p^- \exp[(p-j)2ikl_m]. \end{aligned} \quad (7)$$

Here, $\mathbf{F}^\pm = \mathbf{F}(\pm \mathbf{z})$ are the vector scattering amplitudes of the monolayer obtained in the framework of the QCA.

The sums in these equations take into account the coherent irradiation of the j -th monolayer by the other monolayers. Solving this set of equations with respect to \mathbf{G}_j^\pm and substituting the result in Eqs. (6) yields the coherent fields of the multilayer structure composed of equidistant monodisperse layers.

Note that for a planar geometry, the spectral coefficients of coherent transmission and reflection can be expressed in the coherent fields as $T = |\langle \mathbf{E}(\mathbf{z}) \rangle|^2$ and $R = |\langle \mathbf{E}(-\mathbf{z}) \rangle|^2$, respectively. This approach allows one to take into account both the lateral electrodynamic coupling in each densely packed monolayer and the intermonolayer interference, i.e., the 1D photonic confinement.

Returning to Fig. 18b, for the case of intermonolayer optical distance equal to half the plasmon peak wavelength λ_p , one can see a strong and narrow reflection peak as well as a broadening and a doublet structure of the transparency spectra in the vicinity of the SPR [12,40].

The case of quarter-wavelength dielectric films (see Fig. 18c) corresponds to the PBG spectral position at the collective plasmon resonance frequency. In this case the PBG formation is accompanied by light localization in a stack, while the destructive intermediate multibeam interference minimises both the transmission and the reflection, thereby resulting in a significant absorption increase over the SPR. The absorption in a quarter-wavelength multilayer exceeds substantially the total absorption by the same number of non-interacting monolayers [40].

Thus, 1D ordering of plasmonic nanoparticle arrays allows for the tailoring of multilayer optical properties by matching the thickness of separating dielectric films to the SPR wavelength of the metal nanoparticle monolayer. The effectiveness of these manipulations based on the intermediate multibeam interference depends on the monolayer reflection and transmission spectra or, in other words, on the nanoparticle material, sizes, and surface density.

As an example, Fig. 19 shows the conditions which allow one to achieve the requisite absorption at the plasmon frequency. It is seen from Fig. 19a that like for a single monolayer (see Section 4.5), there is an optimal size of nanoparticles maximizing the quarter-wavelength-stack absorbance for a certain number and packing density of monolayers in a stack. Now, however, the optimal size of nanoparticles in a stack is determined by two factors (see Figs. 19b,c): (i) the rate of absorption saturation with increasing the number of monolayers N (this is determined by the fraction of light transmitted through individual monolayers, which decreases with nanoparticle sizes and surface concentrations), and (ii) by the value

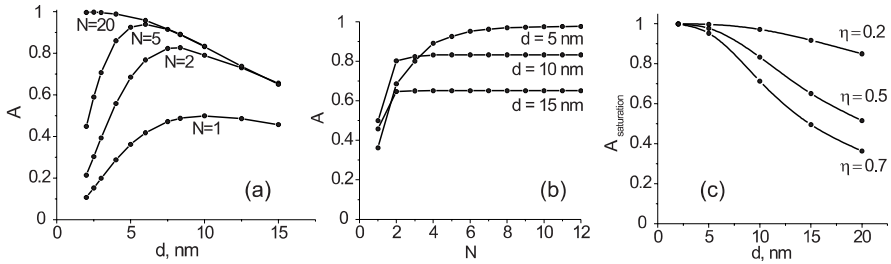


Fig. 19. Calculations for a quarter-wavelength Ag/Al₂O₃ stack. (a, b) Dependence of light absorption near the SPR frequency on (a) the diameter of silver nanoparticles in a stack of N monolayers ($\eta = 0.5$) and (b) the number of monolayers of nanoparticles with different diameters ($\eta = 0.5$). (c) Saturated absorbance as a function of particle diameter for different concentrations η .

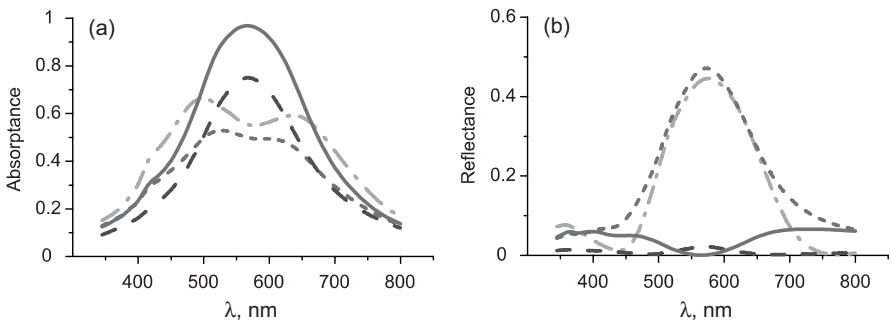


Fig. 20. Calculated (a) absorption and (b) reflection spectra for three different quarter-wavelength structures composed of two densely packed silver nanoparticle monolayers ($\eta = 0.4$). Solid curves correspond to a size-gradient system with d increasing from 5 nm to 15 nm. Dotted curves correspond to the case of the opposite-direction incidence of light on the size-gradient system. Dashed and dot-dashed curves pertain to non-gradient systems with $d = 5$ nm and $d = 15$ nm, respectively.

of the saturated absorption (which is determined by the fraction of light reflected from a monolayer and, therefore, increases with decreasing particle sizes and surface concentrations). The choice of surface concentration η is defined by the requirement of the equality of the collective plasmon resonance frequency and the frequency at which maximal absorptivity is needed.

It is worth noting that the efficiency of the destructive interference in quarter-wave systems can be enhanced by achieving comparable intensities of the beams reflected from the adjacent monolayers. To this end, in order to compensate for the increasing attenuation of light as it penetrates into a multilayer, one should maximize the reflectance of each subsequent monolayer. This can be achieved by using a gradual change of monolayer parameters [18], for example, by (i) chang-

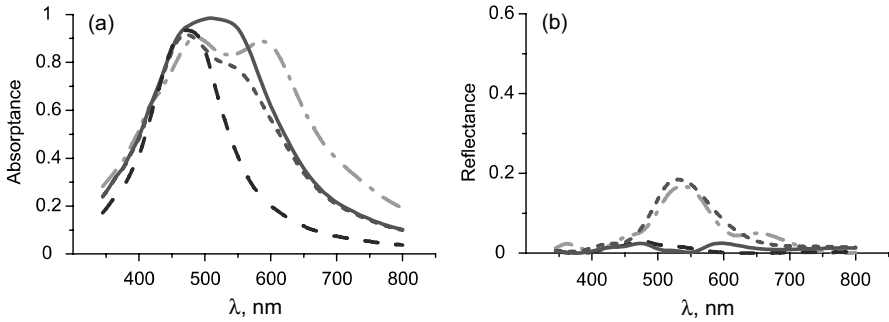


Fig. 21. Calculated (a) absorption and (b) reflection spectra for three different quarter-wavelength structures composed of five densely packed silver nanoparticle monolayers ($d = 10$ nm). Solid curves correspond to a concentration-gradient system with η increasing from 0.13 to 0.33. Dotted curves correspond to the case of the opposite-direction incidence of light on the concentration-gradient system. Dashed and dot-dashed curves pertain to non-gradient systems with $\eta = 0.13$ and $\eta = 0.33$, respectively.

ing nanoparticle sizes from monolayer to monolayer in so-called size-gradient nanostructures, and (ii) changing nanoparticle concentrations from monolayer to monolayer in so-called concentration-gradient nanostructures. For instance, in order to realize an efficient destructive interference in a quarter-wave size-gradient system, one can increase the nanoparticle size in each subsequent layer along the incidence direction.

Frequency matching of separate monolayers in a size-gradient multilayer enables one to achieve extraordinarily strong absorption with just a few monolayers in a stack (see Fig. 20). As one can see, if the incident beam first impinges on monolayer 1 then the peak absorption in this gradient two-monolayer stack is appreciably higher (97%) than for both non-gradient structures (75% and 55%). For the opposite direction of the incident beam (i.e., impinging first on monolayer 2), there is an abrupt change in the optical response (A , R) which reaches almost half the incident intensity (47%). Concentration-gradient stacks (see Fig. 21) combine the notable advantage in absorption with a negligibly small reflectance over the SPR range.

6. Conclusions

Employing spatially arranged ensembles of nanoparticles opens new ways for controlling optical properties of metallic nanostructures. Optical effects of a collective nature are dominated by different size-dependent coherent electromagnetic interactions between particles, including the near-field coupling.

Spectral manifestations of electrodynamic coupling in 2D densely packed metal–dielectric nanostructures include an enhanced long-wavelength transmission of light through a dense layer of multimode plasmonic nanoparticles; strong

resonance quenching at intermediate concentrations of submicron metal particles; and the SPR red concentration shift and enhancement of its sensitivity to the matrix refractive index in the case of dipole nanoparticles.

1D ordering of metal nanoparticle arrays combined with an appropriate choice of solid intermediate film thickness(es) gives rise to a strong photonic confinement in the vicinity of a SPR. Owing to this fact, multilayered plasmonic nanostructures enable an easy tailoring of their absorption and optical responses, which become dependent on the incidence direction for gradient-composition multilayers.

Numerical calculations based on the QCA of the STMSW along with the LEMFP model of internal size effects offer a promising way of modeling basic spectral features of 2D densely packed plasmonic nanostructures and multilayers. The knowledge of stable spectral–structural correlations is an effective tool for initial estimation of nanoparticle materials and topological parameters ensuring desirable spectral properties of planar nanostructures. Furthermore, the strong SPR sensitivity to structural parameters of 2D densely packed metal–dielectric arrays can be used to develop improved characterization techniques based on plasmonic spectroscopy.

References

1. Bogomolov, V. N., S. V. Gaponenko, I. N. Germanenko, et al., 1997: Photonic band gap phenomenon and optical properties of artificial opals. *Phys. Rev. E* **55**, 7619–7625.
2. Dunich, R. A., and A. N. Ponyavina, 2008: Effect of metallic nanoparticle sizes on the local field near their surface. *J. Appl. Spectrosc.* **75**, 832–838.
3. Dyachenko, P. N., and Yu. V. Miklyaev, 2007: One-dimensional photonic crystal based on a nanocomposite “metal nanoparticles – dielectric”. *Kompyuternaya Optika* **31**, 31–34 (in Russian).
4. Ebbesen, T. W., H. J. Lezec, H. F. Graemi, et al., 1998: Extraordinary optical transmission through sub-wavelength hole array. *Nature* **391**, 667–669.
5. Fan, Sh., P. R. Villeneuve, and J. D. Joannopoulos, 1996: Large omnidirectional band gaps in metallodielectric photonic crystals. *Phys.Rev. B* **54**, 11245–11251.
6. Gaponenko, S., 2010: *Introduction to Nanophotonics* (Cambridge University Press, Cambridge, UK).
7. Gehr, R. J., and R. W. Boyd, 1996: Optical properties of nanostructured optical materials. *Chem. Matter* **8**, 1807–1819.
8. Hong, K. M., 1980: Multiple scattering of electromagnetic waves by a crowded monolayer of spheres: application to migration imaging films. *J. Opt. Soc. Am.* **70**, 821–826.
9. Ishimaru, A., 1978: *Propagation and Scattering of Waves in Randomly Inhomogeneous Media* (Academic Press, New York).
10. Ivanov, A. P., V. A. Loiko, and V. P. Dik, 1988: *Light Propagation in Close-packed Disperse Media* (Nauka i Tekhnika, Minsk, in Russian).
11. Kachan, S. M., and A. N. Ponyavina, 2001: Resonance absorption spectra of composites containing metal coated nanoparticles. *J. Mol. Struct.* **267**, 563–564.

12. Kachan, S. M., and A. N. Ponyavina, 2002: Spectral characteristics of confined photonic and plasmonic nanostructures. *Proc. SPIE* **4705**, 88–94.
13. Kachan, S. M., and A. N. Ponyavina, 2002: Spectral properties of close-packed monolayers consisted of metal nanospheres. *J. Phys. Condens. Matter* **14**, 103–111.
14. Kachan, S. M., and A. N. Ponyavina, 2002: The spatial ordering effect on spectral properties of close-packed metallic nanoparticle monolayers. *Surf. Sci.* **507–510**, 603–608.
15. Kachan, S. M., and A. N. Ponyavina, 2007: Enhanced optical sensitivity of close-packed arrays of noble-metal nanoparticles at environmental changes. *CD International conference on Coherent and Nonlinear Optics (ICONO/LAT–2007)* (Minsk, Belarus), 102/V–6.
16. Kachan, S. M., and A. N. Ponyavina, 2007: Optical diagnostics of 2D self-assembled silver nanoparticles arrays. In V. E. Borisenko, S. V. Gaponenko, and V. S. Gurin, Eds., *Physics, Chemistry and Application of Nanostructures* (World Scientific, Singapore), pp. 165–168.
17. Kachan, S. M., and A. N. Ponyavina, 2007: Total light absorption in ultrathin size-gradient metal-dielectric nanostructures. *Proc. SPIE* **6728**, 672838.
18. Kachan, S., O. Stenzel, and A. Ponyavina, 2006: High-absorbing gradient multilayer coatings with silver nanoparticles. *Appl. Phys. B* **84**, 281–287.
19. Khlebtsov, B. N., V. A. Khanadeyev, J. Ye, et al., 2008: Coupled plasmon resonances in monolayers of metal nanoparticles and nanoshells. *Phys. Rev. B* **77**, 035440.
20. Kim, B., S. L. Tripp, and A. Wei, 2001: Self-organization of large gold nanoparticle arrays. *J. Am. Chem. Soc.* **123**, 7955–7956.
21. Kreibig, U., and M. Volmer, 1995: *Optical Properties of Metal Clusters* (Springer, Berlin).
22. Lax, M., 1952: The effective field in dense systems. *Phys. Rev.* **58**, 621–629.
23. Lerme, J., 2000: Introduction of quantum finite-size effects in the Mie’s theory for a multilayered metal sphere in the dipolar approximation: application to free and matrix-embedded noble metal clusters. *Eur. Phys. J. D* **10**, 265–277.
24. Li, J., G. Sun, and C. T. Chan, 2006: Optical properties of photonic crystals composed of metal-coated spheres. *Phys. Rev. B* **73**, 075117.
25. Liao, H., W. Lu, S. Yu, et al., 2005: Optical characteristics of gold nanoparticle-doped multilayer thin film. *J. Opt. Soc. Am. B* **22**, 1923–1926.
26. Mackowski, D. W., and M. I. Mishchenko, 1996: Calculation of the T matrix and the scattering matrix for the ensembles of spheres. *J. Opt. Soc. Am. A* **13**, 2266–2278.
27. Malynych, S., and G. Chumanov, 2003: Light-induced coherent interactions between silver nanoparticles in two-dimensional arrays. *J. Am. Chem. Soc.* **125**, 2896–2898.
28. Messinger, B. J., K. U. von Raben, R. K. Chang, and P. W. Barber, 1981: Local fields at the surface of noble-metal microspheres. *Phys. Rev. B* **24**, 649–657.
29. Nalwa, H. S., Ed., 2001: *Nanostructured Materials and Nanotechnology* (Academic Press, New York).
30. Oraevsky, A. N., and I. E. Protsenko, 2001: Optical properties of heterogeneous media. *Quantum Electron.* **31**, 252–256.
31. Pendry, J. B., 1994: Photonic Band Structures. *J. Mod. Opt.* **41**, 209–229.
32. Penttila, A., E. Zubko, K. Lumme, et al., 2007: Comparison between discrete dipole implementations and exact techniques. *J. Quant. Spectrosc. Radiat. Transfer.* **106**, 417–436.
33. Pileni, M. P., 2001: Self-assemblies of nanocrystals: fabrication and collective properties. *Appl. Surf. Sci.* **171**, 1–14.

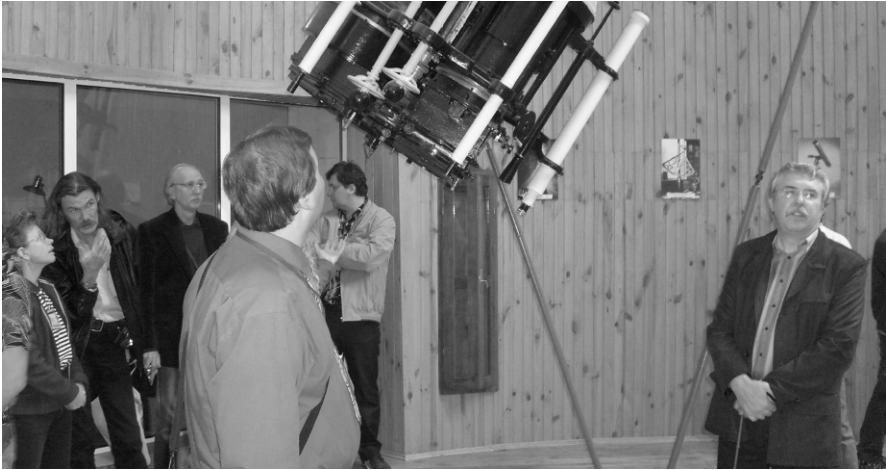
34. Ponyavina, A. N., and N. I. Sil'vanovich, 1994: Interference effects and spectral characteristics of multilayer scattering media. *Opt. Spektrosk.* **76**, 648–655 (in Russian).
35. Ponyavina, A. N., S. M. Kachan, and N. I. Silvanovich, 2004: Statistical theory of multiple scattering of waves applied to 3D photonic crystals. *J. Opt. Soc. Am. B* **21**, 1866–1875.
36. Shalaev, V. M., and S. Kawata, Eds., 2007: *Nanophotonics with Surface Plasmons* (Elsevier, Amsterdam).
37. Shipway, A. N., E. Katz, and I. Willner, 2000: Nanoparticle arrays on surfaces for electronic, optical and sensoric applications. *Chem. Phys. Chem.* **1**, 18–52.
38. Stefanou, N., G. Gantzounis, and C. Tserkezis, 2009: Multiple-scattering calculations for plasmonic nanostructures. *Int. J. Nanotechnol.* **6**, 137–163.
39. Wang, Z., C. T. Chan, W. Zhang, et al., 2001: Three-dimensional self-assembly of metal nanoparticles: possible photonic crystal with a complete gap below the plasma frequency. *Phys. Rev. B* **64**, 113108.
40. Zamkovets, A. D., S. M. Kachan, A. N. Ponyavina, and N. I. Silvanovich, 2003: Optical spectra of metal-dielectric nanocomposites with a layered subwave structure. *J. Appl. Spectrosc.* **70**, 593–598.
41. Zamkovets, A. D., S. M. Kachan, and A. N. Ponyavina, 2008: Concentration-related enhancement of the sensitivity of surface plasmon resonance of metallic nanoparticles to the characteristics of a dielectric environment. *J. Appl. Spectrosc.* **75**, 588–592.
42. Ziman, J., 1979: *Models of Disorder* (Cambridge University Press, Cambridge, UK).



Nikolai Voshchinnikov and Alina Ponyavina.



Maxim Yurkin (left) and Adrian Doicu.



At the Museum of the Main Astronomical Observatory.



From left to right: Godern Videen, Karri Muinonen, and Klim Churyumov.



Carl Codan and Larissa Nazarenko enjoying a coffee break.

Opposition optical phenomena in planetary astrophysics: observational results

Vera K. Rosenbush^{1*} and Michael I. Mishchenko²

¹Main Astronomical Observatory of the National Academy of Sciences of Ukraine,
27 Akademika Zabolotnoho St., 03680 Kyiv, Ukraine

²NASA Goddard Institute for Space Studies, 2880 Broadway, New York, NY 10025, USA

Abstract. The photometric phenomenon in the form of a nonlinear increase of brightness at small phase angles and a negative branch of polarization are typical of the majority of Solar System bodies. In addition, some high-albedo objects reveal a backscattering polarization feature in the form of a spike-like negative-polarization minimum called the polarization opposition effect. These optical phenomena are important tests of our theoretical descriptions of light scattering by regolith planetary surfaces. In this chapter we review the recent progress in the study of optical opposition phenomena in planetary astrophysics. We primarily focus on the results of polarimetric observations of asteroids, the Galilean satellites of Jupiter, the Saturnian satellite Iapetus, Centaurs, and trans-Neptunian objects at backscattering geometries including phase angles approaching zero.

Keywords: polarimetry, opposition effects, asteroids, satellites, Centaurs, trans-Neptunian objects

1. Introduction

Polarimetry and photometry are especially sensitive diagnostic tools helping to infer the physical properties of bodies whose observational characteristics are governed by small scatterers, e.g., dust and regolith grains. It is these methods that have provided convincing indications of the presence of regolith on planetary surfaces long before the advent of space missions. Measurements of the intensity and polarization of scattered light as functions of the phase angle and wavelength have been used to gain an improved understanding of the microphysical properties of the surfaces of many Solar System bodies. The phase curves of brightness and polarization and their spectral dependences are controlled by the fundamental phenomenon of light scattering and are intimately related to the physical properties of the scattering media such as the size, morphology, composition, and packing of the constituent particles. These characteristics can often be inferred by solving theoretically the inverse remote-sensing problem.

* Corresponding author. E-mail: rosevera@mao.kiev.ua

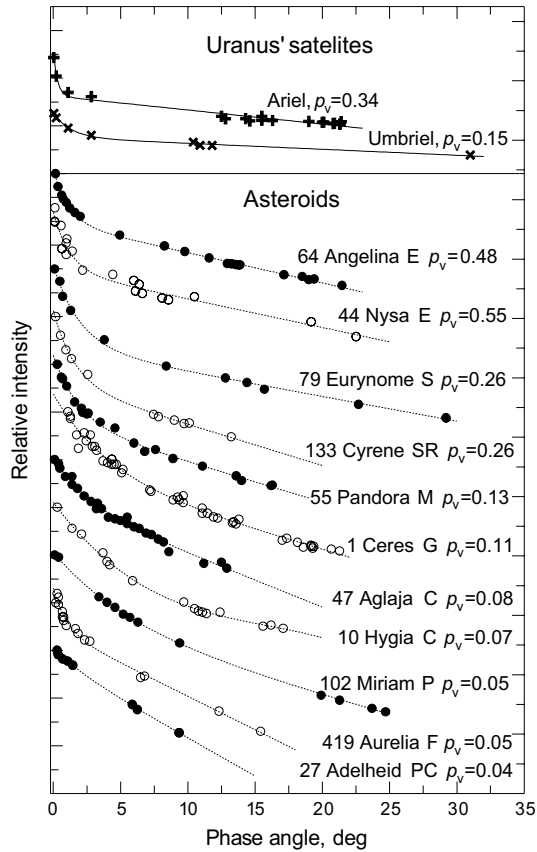


Fig. 1. Phase-angle dependences of brightness for asteroids of different types and albedos and major satellites of Uranus (after Rosenbush et al. 2006 and Avramchuk et al. 2007). Solid curves represent the best fits to the data with an exponential–linear function (Rosenbush et al. 2002).

Much attention is currently paid to the behavior of the radiation scattered at small phase angles, $\alpha < \sim 20^\circ$. Two interesting optical phenomena are observed for many bodies of quite varying nature, such as atmosphereless planets, planetary satellites and rings, and asteroids as well as cometary and interplanetary dust. The first phenomenon is a nonlinear increase in brightness at phase angles approaching zero, i.e., the brightness opposition effect (BOE). [Figure 1](#) illustrates phase-angle dependences of brightness for asteroids of different types and albedos and major satellites of Uranus. It is seen that there can be a sharp surge of brightness at phase angles less than 2° (e.g., for 44 Nysa, 64 Angelina, and Ariel) or a rather smooth angular variation. The second phenomenon is negative linear polarization at small phase angles (meaning that the electric field vector component parallel to the scat-

tering plane dominates the perpendicular component), first discovered through lunar observations by Lyot (1929). The angular dependence of negative polarization for different Solar System bodies and laboratory samples can exhibit a wide, almost parabolically shaped negative polarization branch (NPB) with a minimum near 5° – 12° (as, e.g., for asteroids in Fig. 2), and/or a sharp asymmetric minimum of polarization centered at about 0.5° – 2° , called the polarization opposition effect (POE). There are different opinions as to the shape of the polarization phase dependence at small phase angles, ranging from a sharply asymmetric NPB to a secondary minimum of negative polarization distinctly separated from the main NPB minimum. The secondary minimum has been detected for Saturn's rings (Dollfus 1984; see also Rosenbush et al. 1997), E-type asteroids (Rosenbush et al. 2005, 2009), and bright Galilean satellites (Rosenbush and Kiselev 2005). A very asymmetric phase-angle curve is detected for some Centaurs and trans-Neptunian objects (TNOs) (Boehnhardt et al. 2004; Rousselot et al. 2005; Bagnulo et al. 2006, 2008; Belskaya et al. 2010). The parameters of the NPB, namely the degree of polarization at the minimum P_{\min} , the corresponding phase angle α_{\min} , and the inversion angle α_{inv} at which polarization changes sign from negative to positive, can differ significantly even for members of the same class of objects, which is clearly seen in Fig. 2.

Bernard Lyot was the first to measure in the laboratory an extremely narrow backscattering polarization minimum for a particulate MgO surface (Lyot 1929). Subsequently, both types of negative polarization and a narrow backscattering intensity peak have also been detected for many particulate laboratory samples (e.g., Geake and Geake 1990; Shkuratov et al. 2002).

The specific “physical explanations” of the opposition effects and the NPB have not been completely agreed upon. Whatever their actual practical worth is (cf. Mishchenko et al. 2011), a number of optical “mechanisms” have been proposed to explain the BOE and negative polarization resulting from the interaction of light with porous, powder-like surface layers or rough surfaces of atmosphereless Solar System bodies (ASSBs) (Hapke 1993; Shkuratov et al. 1994; Muinonen et al. 2002; Mishchenko et al. 2002, 2006a, 2009a,b, 2010). Presently, the interference effect of coherent backscattering (CB), the purely geometric effect of mutual shadowing (MS), and the single-particle scattering are mentioned as the primary candidates to “explain” the observed opposition phenomena. Some features of the polarization and brightness phase dependences are also predicted by the so-called near-field theory (Petrova et al. 2007; Tishkovets 2008; Mishchenko et al. 2010, and references therein). CB is caused by the constructive interference of reciprocal trajectories of light multiply scattered by a particulate surface at small phase angles. This interference can contribute to the BOE as well as to the POE. CB is more pronounced for bright objects, whereas the single-particle scattering can cause a NPB for objects with different albedos. MS is more effective in the formation of the BOE for low-albedo bodies. However, this mechanism does not yield negative polarization. From comparisons of observational data and the results of theoretical modeling and laboratory measurements, the most likely size,

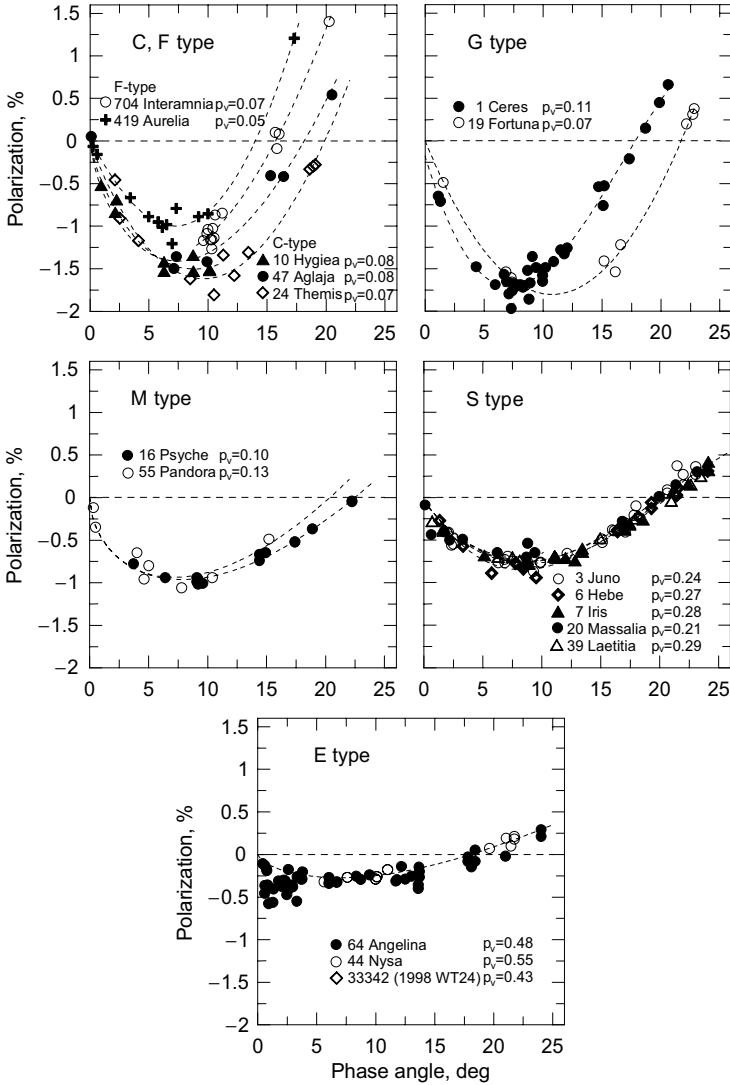


Fig. 2. NPB for different types of asteroids. The polarimetric data are taken from APD, Rosenbush et al. (2005), Kiselev et al. (2002), Belskaya et al. (2005), and Zellner and Gradie (1976). The curves show approximations of the NPB data with a trigonometric polynomial introduced by Lumme and Muinonen (1993).

composition, and structure of dust particles and some other properties of ASSB surfaces can be inferred. Such comparisons, however, require sufficiently complete phase curves of brightness and polarization covering wide ranges of phase angles and wavelengths.

Currently, photometric and polarimetric data for various objects at the smallest phase angles ($\alpha < 2^\circ$) and in different spectral bands are rare, and the behavior of the negative polarization is still unknown in detail. This is due to the fact that the absolute value of polarization at these angles is often small (several tenths of a percent), thereby necessitating very high measurement accuracy and making most of the available data poorly suitable for studies of the opposition effects. Because polarimeters are available at only a handful of telescopes, the visibility of ASSBs is typically limited, and poor weather often reduces the already limited allocation of observation time, detailed sampling of the polarization curve at the smallest phase angles becomes extremely difficult. As a consequence, despite the significant recent progress in theoretical and modeling studies (e.g., Mishchenko et al. 2009a,b, 2010, and references therein), the cumulative polarimetric data for ASSBs are still limited.

To replenish the polarimetric database, we have been carrying out a specialized program of polarimetric observations near opposition of selected ASSBs since 1995. The main purpose of this program is obtaining measurements of polarization at the smallest accessible phase angles with the highest achievable accuracy. The ultimate objective is to study in detail the behavior of polarization near opposition and record the shape of the secondary polarization minimum initially predicted by the theory of CB (Mishchenko 1993) and subsequently discovered experimentally by Rosenbush et al. (1995, 1997). The list of targets includes high-albedo bodies, such as the Galilean satellites Io, Europa, Ganymede, and Callisto, Saturnian satellites Iapetus and Enceladus, E-type asteroids, including 44 Nysa and 64 Angelina, and several other objects.

In this chapter we summarize the results of photopolarimetric observations of different ASSBs at small phase angles. In particular, we compare polarimetric properties of objects located in the inner part of the Solar System with those observed recently with the largest ground-based telescopes for objects residing in the outer part of the Solar System (such as Centaurs and trans-Neptunian objects).

2. Telescopes, polarimeters, and data reduction

To measure the linear polarization and intensity of light scattered by selected ASSBs, we use the 2.6-m and 1.25-m telescopes of the Crimean Astrophysical Observatory (CrAO) and the 0.7-m telescope of the Chuguev Observation Station of the Institute of Astronomy of the Kharkiv National University (KhNU). All three telescopes are equipped with aperture photopolarimeters (see chapter 2 of Mishchenko et al. 2010 for details).

Two single-channel polarimeters with rapidly rotating polarizers (~ 33 rotations per second) are mounted at the 2.6-m and 0.7-m telescopes. The polarizer in the former instrument is an achromatic quarter-wave phase plate (followed by a fixed Polaroid; see Shakhovskoy et al. 2001), whereas that of the latter instrument is a simple (rotating) Polaroid (Kiselev and Velichko 1997). Usually the measurements are taken in the standard B, V, and R bands (central wavelengths 0.44,

0.53, and 0.69 μm , respectively) or in the broadband red filter WR (0.69–0.93 μm). The corresponding passbands are close to those of the standard BV (Johnson) and R (Cousins) systems, thereby allowing an accurate conversion of the photometric data from the instrumental to the standard system.

A calcite prism serves as the analyzer and splits the incoming beam into two components with orthogonal linear polarizations in the UBVR photopolarimeter mounted at the 1.25-m CrAO telescope (Piirola 1988; Berdyugin and Shakhovskoy 1993). The corresponding images are formed inside two diaphragms in the focal plane. The intensities of two images are measured by a single photomultiplier using a chopping technique. For the measurement of linear polarization, the signals are modulated using a rotating half-wave phase plate. During the course of the measurement, the phase plate rotates with 22.5° steps in front of a fixed calcite plate, which allows one to detect the intensities of the two orthogonal states of polarization corresponding to the eight position angles of the retarder. One measurement consists of the integration over eight different orientations of the retarder. Using four dichroic filters, the instrument yields measurements of the intensity and polarization for the object in the standard U, B, V, R, and I bands (with the corresponding effective wavelengths 0.36, 0.44, 0.53, 0.69, and 0.83 μm) simultaneously. The method of synchronic detections is used in all three polarimeters. This technique ensures a quasi-simultaneous measurement of the normalized Stokes parameters q and u . Subsequently, the degree of linear polarization and the position angle of the polarization plane are determined using the following standard expressions:

$$P = \sqrt{q^2 + u^2}, \quad \theta = \frac{1}{2} \arctg \frac{u}{q}.$$

The respective errors

$$\sigma_P = \sqrt{(\sigma_q^2 + \sigma_u^2)/2} \quad \text{and} \quad \sigma_\theta = 28.65 \frac{\sigma_P}{P}$$

are calculated using the mean square errors of the individual measurements of parameters q and u which, in turn, are estimated using both the statistics of recorded photoelectrons and their internal dispersion (Shakhovskoy and Efimov 1972). The greater of these two errors is adopted as an accuracy measure of the weighted mean values of the polarization parameters P and θ for each set of measurements during the night. The errors of the parameters P and θ averaged over the entire night are determined using the variance of the individual values. Typical errors in the degree of linear polarization in the R filter vary from 0.03% to 0.07%.

The weakness of polarization at very small phase angles ($|P| \sim 0.1\% - 0.3\%$) necessitates measurement accuracies better than $\sim 0.03\%$, which means that the instrumental polarization must be known to better than 0.02%. In fact, the errors in the parameters of instrumental polarization determined from observations of unpolarized standard stars did not exceed 0.01%, as was the case during our polarimetric observations of the Galilean satellites (Rosenbush and Kiselev 2005).

The calibration of the position angle of the polarization plane is performed using observations of standard stars with well-known large interstellar polarization. The zero-point of position angles is determined with an accuracy of $\pm 0.7^\circ$. All polarimetric standard stars are usually taken from the database compiled in Heiles (2000) and references therein. The cumulative error in our measurements of the degree of linear polarization includes the mean square error averaged over each night and the error of the instrumental system. In most cases the cumulative error did not exceed 0.03% – 0.07%. The zero-point of position angles was stable to within $\sim 1^\circ$.

3. Polarimetric data for asteroids near opposition

Lyot (1934) was the first to discover polarization of light scattered by asteroids. He found that asteroids 1 Ceres and 4 Vesta exhibit negative polarization at small phase angles, similarly to the Moon, Mars, and Mercury. An extensive program of polarimetric observations has been carried out by Zellner, Gradie, and Gehrels (Zellner et al. 1974; Zellner and Gradie 1976), and the results were summarized by Dollfus and Zellner (1979). In the subsequent years, polarimetric observations of asteroids have been actively pursued, and a large volume of data (mostly UBVR polarimetry) has been archived in the Asteroid Polarimetric Database (APD; Lupishko and Vasil'ev 2008, <http://sbn.psi.edu/pds/resource/apd.html>). As a result, the NPB for many asteroids has been studied in detail, and their polarimetric albedos were estimated. The parameters P_{\min} , α_{\min} , and α_{inv} have been found to be different for the main compositional types of asteroids (F, C, M, S, A, and E types) and, therefore, have been widely used for the classification of asteroids (Mishchenko et al. 2010, and references therein).

Currently, the APD contains all published and the majority of unpublished results of polarimetric observations of about 280 asteroids. However, only for six asteroids the behavior of polarization has been investigated in detail near opposition, at very small phase angles ($< 2^\circ$). Among them are the high-albedo E-type asteroids 64 Angelina (Rosenbush et al. 2005) and 44 Nysa (Rosenbush et al. 2009), V-type 4 Vesta (Rosenbush 2006), S-type 20 Massalia (Belskaya et al. 2003), and F-type 1021 Flammario and 419 Aurelia (Fornasier et al. 2006b).

3.1. High-albedo E-type asteroids

Harris et al. (1989) found that the phase curve of brightness for the high-albedo asteroids 64 Angelina and 44 Nysa exhibits a strong and unusually narrow opposition spike, or brightening, of about 0.25^m at small phase angles. Rosenbush et al. (2002) estimated the half-width at half-maximum (HWHM) of the opposition surge to be 0.85° in the V band. According to the theoretical prediction by Mishchenko (1993), the sharp opposition effect of brightness should be accompanied by the POE at very small phase angles provided that the BOE is caused by a fine-grained regolith surface. If so, both effects could be spectacular manifesta-

tions of the same light-scattering phenomenon, namely, CB. Therefore, our observational program has been designed to capture minimal achievable phase angles in search of the POE.

Asteroids 44 Nysa and 64 Angelina. Asteroid 44 Nysa is the largest member of the class of high-albedo E-type asteroids; its albedo and diameter, estimated from IRAS data, are 0.55 and 70.6 km, respectively (Tedesco et al. 2002). The geometric albedo and effective size of 64 Angelina estimated from radiometric data are 0.43 and 60 km, respectively. The albedo derived from polarimetric data is 0.48 (Lupishko and Mohamed 1996), which is greater than that estimated from radiometry.

Zellner and Gradie (1976) were the first to carry out polarimetric observations of 44 Nysa and 64 Angelina in the B and G (V) spectral bands within the range of phase angles 3.7° – 24° . During the 1987 apparition, Kiselev and Lupishko measured polarization of 44 Nysa in the U, B, V, R, and I filters at phase angles 7.6° – 11° (see APD). Thus, the existence of the NPB for 44 Nysa and 64 Angelina had been a well-known fact. However, no polarimetric observations had been performed at phase angles smaller than 3.7° .

The first near-opposition polarimetric observations at phase angles from 0.43° to 13.02° were carried out for 64 Angelina in the U, B, V, R, and I filters with the 1.25-m CrAO telescope during the oppositions of 1995, 1999, 2000/01 (Rosenbush et al. 2005), and 2008 (unpublished data). These observations have shown that in all the spectral bands, there is a secondary minimum of negative polarization (i.e., the POE) with an amplitude of about 0.4% in the R band centered at a phase angle $\alpha_{\min} \sim 1.8^\circ$. This minimum along with a regular NPB forms a bimodal phase curve of polarization. The amplitude of the POE appears to be apparition-dependent. The POE as well as the regular NPB depend on the wavelength, but in different ways. As an example, Fig. 3 shows the phase dependence of polarization for 64 Angelina observed in the R filter; these measurements have the highest accuracy ($\sim 0.03\%$ – 0.07%).

The first V-filter polarimetric observations of 44 Nysa were carried out at phase angles ranging from 0.41° to 7.49° on 10–14 August 2005 (Rosenbush et al. 2009). The most important result of these observations was the detection of a bimodal phase-angle dependence of polarization, which consists of a POE with parameters $\alpha_{\min, \text{POE}} \sim 0.8^\circ$ and $P_{\min, \text{POE}} \approx -0.31\% \pm 0.04\%$ superposed on a much broader NPB with $P_{\min, \text{NPB}} \approx -0.30\%$ at $\alpha_{\min, \text{NPB}} \sim 5.8^\circ$ (Fig. 4). Polarimetric observations of 44 Nysa near opposition in February 2011 with the 2.6-m CrAO telescope and the 0.7-m KhNU telescope (unpublished data) confirmed the previously established existence of the POE (see Fig. 4).

Near-opposition photometric observations of 64 Angelina and 44 Nysa in different spectral bands were also carried out during the 1999 and 2005 apparitions, respectively (Rosenbush et al. 2005, 2009). Our photometric data confirmed the earlier detection by Harris et al. (1989) of a very strong and unusually narrow opposition spike, i.e., the BOE, for both asteroids in the B, V, R, and I spectral

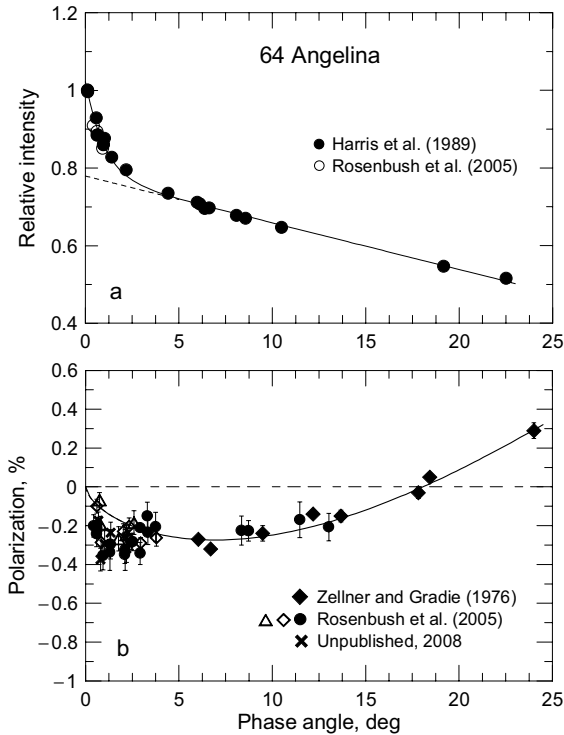


Fig. 3. (a) BOE and (b) POE for asteroid 64 Angelina. Polarization measurements in the R filter were obtained during four apparitions: 1995, 1999, 2000/01 (Rosenbush et al. 2005), and 2008 (unpublished data by Kiselev, Rosenbush, and Zaitsev). The solid curves represent the best fits to the data.

bands. The composite phase curves obtained in the B and V bands for both asteroids showed that the parameters of the BOE for 44 Nysa are similar to that for 64 Angelina within the measurement accuracy at comparable phase angles. It appears that all color indices increase as the phase angle decreases from $\sim 2^\circ$ to zero. Although the amplitude of this increase for 44 Nysa is only 0.03^m , there may be some evidence of the wavelength dependence of the BOE.

Asteroid 214 Aschera. This rather faint E-type asteroid has an estimated diameter of 23 km and a geometric albedo of 0.52 (Tedesco et al. 2002). Near-opposition polarimetric observations of this object were performed during the 1996–98 apparitions using the 1.25-m CrAO telescope equipped with the UBVRI photopolarimeter (Beskaya et al. 2003). The measurements were taken at phase angles 0.72° , 5.3° , and 11.5° . Substantial measurement errors for Aschera, comparable to its own polarization, in combination with poor phase-angle sampling do not allow one to record the NPB minimum and identify a secondary polarization minimum

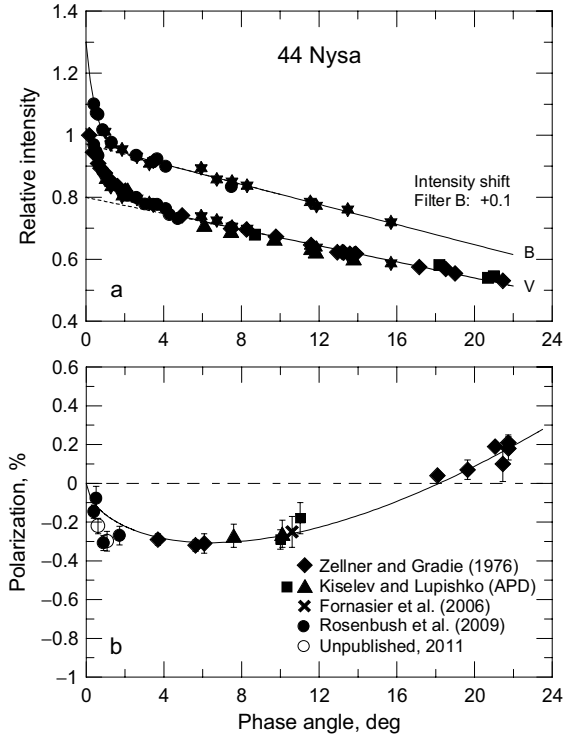


Fig. 4. The composite phase curves of (a) brightness and (b) polarization for asteroid 44 Nysa. In the top panel, triangles, asterisks, diamonds, squares, and circles represent the data obtained during the 1979 (Birch et al. 1983), 1982 (Tupieva 2003), 1986 (Harris et al. 1989), 1987 (Shevchenko et al. 1992), and 2005 apparitions, respectively. Unpublished polarization data obtained by Kiselev, Rosenbush, Velichko, and Zaitsev during the 2011 apparition are depicted by open circles in panel (b). The solid curves represent the best fits to the data.

in individual spectral bands or to characterize the spectral variability of polarization. Therefore, in order to determine the phase-angle dependence of polarization for Aschera more accurately, we averaged data for this asteroid in the B, V, R, and I filters using the respective errors σ_P as weights. As a result, we have found that Aschera exhibits a rather significant degree of polarization, $P = -0.38\% \pm 0.09\%$, at a phase angle $\alpha = 0.72^\circ$.

3.2. Comparison of opposition effects for E-type asteroids

Comparison of the photometric phase curves and the NPBs for 44 Nysa and 64 Angelina (Mishchenko et al. 2010) reveals certain differences potentially attributable to differences in the morphology and/or composition of the reflecting

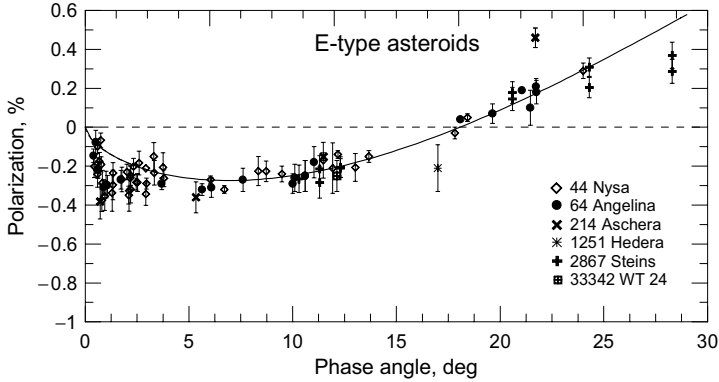


Fig. 5. Composite phase-angle dependence of polarization for E-type asteroids based on all available data (Zellner and Gradie 1976; Kiselev et al. 2002; Belskaya et al. 2003; Rosenbush et al. 2005, 2009; Fornasier et al. 2006a,b; data by Kiselev and Lupishko from APD; Cellino, personal communication).

surfaces. In particular, Nysa exhibits a significantly narrower POE than Angelina. The difference between the measured polarization and extrapolated NPB value in the minimum of POE is $\sim 0.12\% - 0.25\%$ for Angelina and $\sim 0.12\%$ for Nysa. One can also notice differences between the respective NPBs: $P_{\min} = -0.30\% \pm 0.02\%$ and $\alpha_{\min} = 5.8^\circ \pm 0.1^\circ$ for Nysa and $P_{\min} = -0.26\% \pm 0.02\%$ and $\alpha_{\min} = 6.9^\circ \pm 0.1^\circ$ for Angelina.

Figure 5 summarizes all polarization measurements for E-type asteroids in the V/G and R filters for different oppositions (APD; Zellner and Gradie 1976; Belskaya et al. 2003; Fornasier et al. 2006a,b; Rosenbush et al. 2005, 2009; Cellino, private communication). In addition to our results, polarimetric data for the E-type asteroids 214 Aschera, 1251 Hedera, and 2867 Steins (Fornasier et al. 2006a,b) as well as NEA 33342 WT24 (Kiselev et al. 2002) are shown. The polarization values measured for these asteroids are characteristic of the E type. Moreover, one can see that the value of negative polarization for 214 Ashera at a phase angle $\sim 0.8^\circ$ coincides with the secondary minimum of negative polarization for Nysa and is close to that for Angelina. Thus, the composite phase-angle dependence of polarization for high-albedo asteroids reliably establishes the presence of the POE in the form of a sharp minimum of negative polarization centered at a phase angle of $\sim 1^\circ$ and having an amplitude of $\sim 0.35\%$. Of course, additional high-accuracy polarimetric observations of E-type asteroids are required to confirm these conclusions, because there remains a non-zero probability that the phase-angle dependence of polarization for Ashera has a very asymmetrical shape of the NPB. It is important, however, that for at least two asteroids the characteristic angular widths of the respective BOEs and POEs are essentially the same, which is consistent with a common CB origin of these opposition phenomena. Unfortunately, we have no reliable and detailed data for other asteroids.

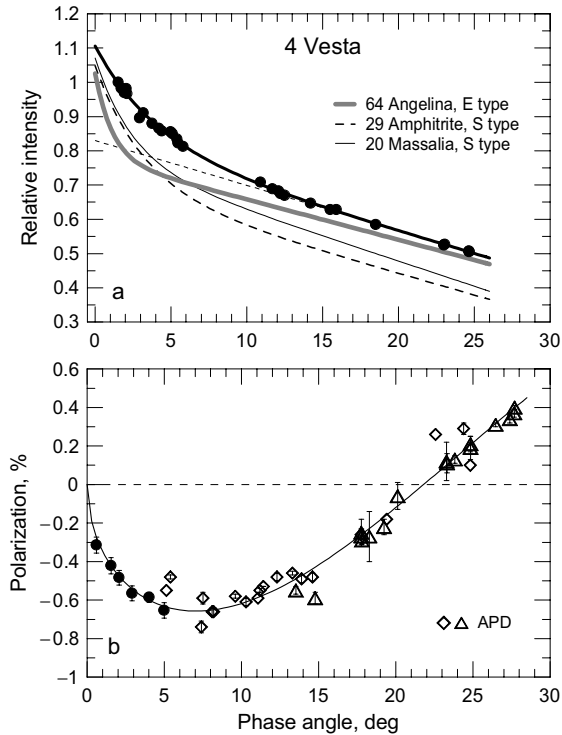


Fig. 6. Phase-angle dependence of (a) brightness and (b) polarization for asteroid Vesta in the V filter. The data obtained during the 2005–06 apparition are depicted by filled circles (Rosenbush 2006). Open symbols depict APD data. The approximation of the phase-angle dependence of brightness in panel (a) is defined by the empirical function for E-type asteroid 64 Angelina and two S-type asteroids 20 Massalia and 29 Amphitrite (Rosenbush et al. 2002). The curve in panel (b) shows the approximation of the NPB data with a trigonometric polynomial according to Lumme and Muinonen (1993).

3.3. Near-opposition polarimetric observations of other types of asteroids

V-type asteroid 4 Vesta. According to the IRAS data, the albedo of this object is $p_V = 0.423$ (Tedesco et al. 2002), i.e., is quite comparable to those of E-type asteroids. However, using polarimetric data, Lupishko and Mohamed (1996) inferred a much lower value, $p_V = 0.255$. To resolve this discrepancy, and also since all previous polarimetric observations of 4 Vesta had been performed at phase angles $\alpha > 5.13^\circ$ and all photometric observations at phase angles $\alpha > 1^\circ$, the obvious objective of our observations was to measure Vesta's brightness and polarization at very small phase angles, analyze the potential opposition effects, and determine the albedo of this asteroid. Therefore during the 2005–06 apparition, photopolarimetric observations in the U, B, V, R, and I filters were performed using the

1.25-m CrAO telescope. The resulting range of phase angles was from 0.6° to 5.0° , which yielded a more detailed and accurate polarization phase dependence (Rosenbush 2006).

One can see from Fig. 6a that within the measurement errors ($\pm 0.02\% - 0.04\%$), the V-filter polarimetry of Vesta shows no secondary minimum at opposition. The absence of this feature may actually imply that Vesta is not a high-albedo asteroid. We have compiled a detailed polarization phase curve from all observations of Vesta during several previous oppositions (Zellner and Gradie 1976; APD) along with our more recent observations to obtain improved estimates of the parameter P_{\min} and polarimetric slope h at the inversion phase angle, which usually correlate with the surface albedo. The values $P_{\min} = -0.66\% \pm 0.05\%$ and $h = 0.066 \pm 0.002$ were used to estimate the geometric albedo of Vesta at $p_V = 0.24$, which is smaller than the IRAS-based albedo by a factor of almost two. This low albedo is indirectly confirmed by photometric data. It can be seen from Fig. 6a that the average phase-angle curve of brightness for Vesta is distinctly different from that for E-type asteroid 64 Angelina, which shows a very strong and narrow BOE. Obviously, the photometric data for Vesta are consistent with those for typical moderate-albedo S-type asteroids 20 Massalia and 29 Amphitrite. Thus, the absence of a POE for asteroid 4 Vesta may not, after all, be surprising.

S-type asteroid 20 Massalia. Gehrels (1956) was the first to discover the BOE for asteroids, specifically, for the S-type asteroid 20 Massalia, from photometric observations near opposition at phase angles down to 0.5° . This asteroid has been estimated to have a diameter of 145 km and a geometric albedo of 0.21 (Tedesco et al. 2002). As one can see in Fig. 7a, Massalia shows a pronounced opposition effect at small phase angles, with a rather large amplitude (Rosenbush et al. 2002).

Polarimetric observations of Massalia at extremely small phase angles (down to 0.08°) were carried out during the 1998 apparition by Belskaya et al. (2003). The phase-angle dependence of polarization is shown in Fig. 7b. Here, the degree of polarization in each of the B, V, R, and I filters as well as filter-averaged values are given. The measurement errors are within $0.02\% - 0.05\%$. One can see that this asteroid has a significant scatter of data points at a phase angle of 0.6° which is close to the phase angle of the secondary minimum of negative polarization for E-type asteroids. This scatter, along with a strong BOE, precludes a reliable conclusion regarding the presence or absence of a secondary minimum of polarization for Massalia near the opposition. Further detailed observations of this asteroid as well as other asteroids of the S type at different wavelengths are needed to derive a definitive conclusion.

F-type asteroids 419 Aurelia and 1021 Flammario. According to the IRAS data, the geometric albedo of 419 Aurelia is 0.046 and that of 1021 Flammario is 0.0458 (Tedesco et al. 2002). Belskaya et al. (2005) and Fornasier et al. (2006b) measured the degree of polarization for these asteroids at extremely small phase angles, down to 0.1° , in different spectral bands. These were the first detailed polarimetric observations near opposition for dark asteroids. Measured values of po-

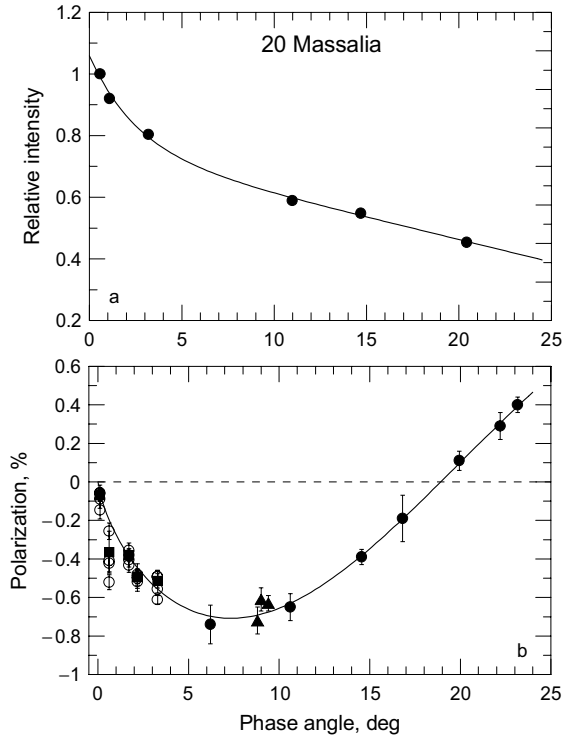


Fig. 7. S-type asteroid 20 Massalia. (a) Phase-angle dependence of brightness in the V filter according to data taken from Gehrels (1956); the solid curve shows the approximation of data according to Rosenbush et al. (2002). (b) Phase-angle dependence of polarization based on data taken from Belskaya et al. (2003), Zellner and Gradie (1976), and Kiselev (APD). Open circles depict the data separately in the B, V, R, and I filters, while the filled squares show the respective average values. The filled circles in panel (b) depict the data by Zellner and Gradie (1976). The solid curves represent the best fits to the data.

larization are shown in Fig. 8. Small positive polarization at phase angles $\alpha < 1^\circ$ was observed for Flammario in all filters, while negative polarization was observed for Aurelia. Given the rather high accuracy of polarization measurements (0.04%–0.05%) for Flammario and large deviations of the polarization plane from the scattering plane (approximately from 14° to 40°), these authors proposed that positive polarization in the phase angle range 0.1° – 0.6° can be related to nonstandard scattering geometry caused by the surface structure of the asteroid, e.g., an anisotropy caused by grooves, such as the Phobos grooves. Certain laboratory measurements (e.g., Geake and Dollfus 1986; Shkuratov et al. 2002) imply that such an explanation is physically sound. However, it is obvious that these effects can not be caused by multiple scattering of light.

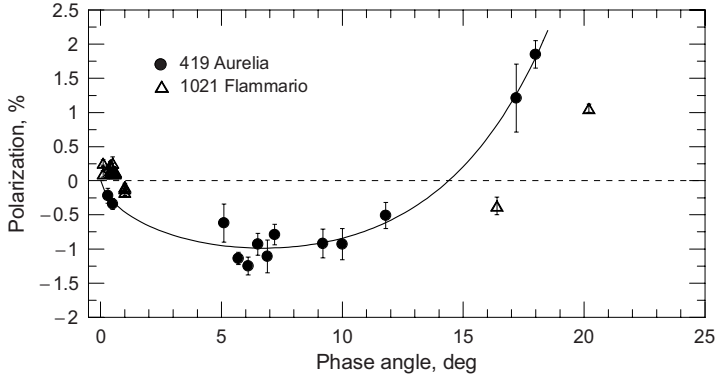


Fig. 8. Degree of linear polarization for low-albedo F-type asteroids 419 Aurelia and 1021 Flammario as measured by Belskaya et al. (2005) and Fornasier et al. (2006b).

4. Opposition effects for planetary satellites

Systematic polarimetric observations of the Galilean satellites of Jupiter in the U, B, V, and R filters and in the entire accessible range of phase angles from $\sim 11.8^\circ$ down to almost zero as well as sparse polarization measurements for three Saturnian satellites Dione, Rhea, and Iapetus have been carried out by different authors (see Rosenbush et al. 2002, and references therein). The phase-angle, orbital, and spectral dependencies of polarization have been studied. There are certain systematic differences in the depth of the NPB for Io, Europa, Ganymede, and Callisto. Large orbital variations in the degree of polarization were found for Callisto and Io, whereas smaller longitudinal effects have been detected for Ganymede and Europa.

For the first time, during the 1988–91 oppositions, the presence of a POE centered at a very small phase angle of $\alpha \approx 0.5^\circ - 0.7^\circ$ and superposed on a regular NPB was detected for Io, Europa, and Ganymede (Rosenbush et al. 1997). This phase angle is comparable to the width of the brightness peak observed for Europa (Thompson and Lockwood 1992), thereby indicating that both opposition phenomena are likely to be caused by CB (Mishchenko 1993).

However, the previous polarimetric data were often obtained with low accuracy even for the bright satellites of Jupiter; they were rather limited and even mutually inconsistent. Furthermore, up until then there were no sufficient and reliable polarization data at phase angles smaller than 2° . Stimulated by the theoretical prediction of the POE by Mishchenko (1993), we initiated a long-term program of low-phase-angle observations of the Galilean satellites of Jupiter and the Saturnian satellite Iapetus in order to investigate in detail the behavior of polarization near opposition. In this section, we present and analyze the most recent results.

Galilean satellites. For the first time, the polarization phase dependence for the Galilean satellites near opposition was studied in detail and with high accuracy

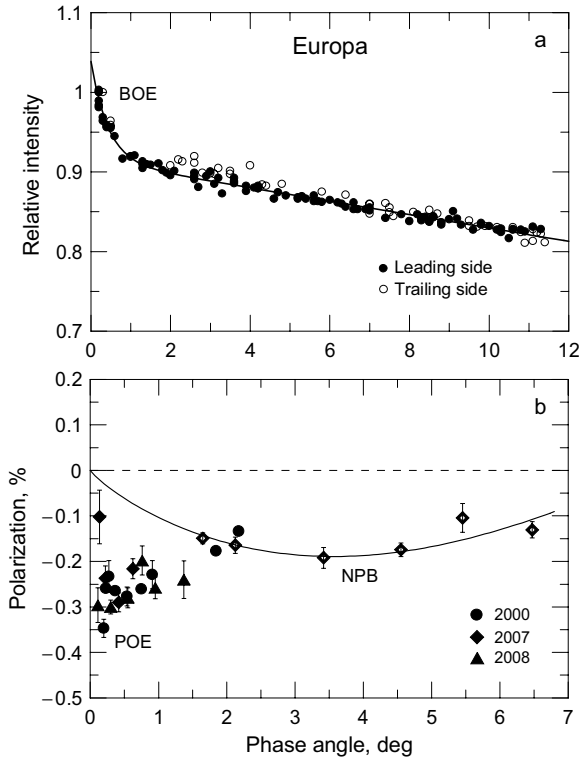


Fig. 9. (a) BOE and (b) POE for Europa in the V band. The brightness data are fitted with an exponential–linear function according to Rosenbush et al. (2002), while the polarization data are fitted with a trigonometric polynomial according to Lumme and Muinonen (1993).

(about $\pm 0.02\%$) during the 2000 opposition (Rosenbush and Kiselev 2005). The minimal phase angle reached was as small as 0.20° . Also, high-accuracy polarimetric observations were carried out during 2007–08 in different spectral bands (Kiselev et al. 2009; Mishchenko et al. 2009a). The results of these observations fully confirmed our previous conclusion (Rosenbush et al. 1997) that Io, Europa, and Ganymede exhibit a pronounced and narrow POE at very small phase angles (0.4° – 0.7°) superposed on a regular NPB. We have also established that Callisto exhibits only the regular NPB for both the leading and the trailing hemisphere, with no traces of a POE (Rosenbush 2002).

Figure 9 summarizes the results of observations of Europa, which is believed to be covered by almost pure particulate water ice. The top panel shows the photometric results from Thompson and Lockwood (1992), while the bottom panel depicts the results of the 2000, 2007, and 2008 polarimetric observations (Rosenbush and Kiselev 2005; Kiselev et al. 2009). The 2000 and 2008 polarimetric data obtained with several spectral filters are averaged according to the proce-

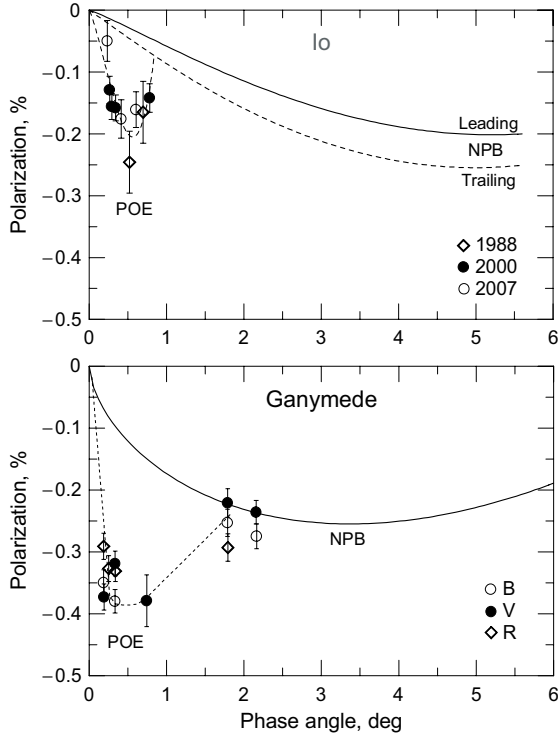


Fig. 10. POE for the Galilean satellites Io and Ganymede. The respective NPBs are shown schematically for the leading and trailing hemispheres of Io and the whole disk of Ganymede.

ture described in Rosenbush et al. (2009). It is quite remarkable that Europa exhibits a strong BOE (see also Helfenstein et al. 1998) as well as a pronounced and asymmetric POE with an amplitude of about 0.35% at a phase angle of $\sim 0.2^\circ$. The latter is superposed on a wide, nearly parabolic NPB typical of most ASSBs (Muinonen et al. 2002). Furthermore, the phase angle of the POE is comparable to the angular semi-width of the BOE, in agreement with the theory of CB (Mishchenko 1993; Mishchenko et al. 2000, 2006b).

Figures 10a and 10b illustrate the behavior of polarization for Io and Ganymede near opposition. By comparing these figures with Fig. 9 we can identify significant differences between the POEs for the three Galilean satellites. The POE for Europa has the shape of a sharp and asymmetric minimum with a depth of $\sim 0.3\%$ centered at $\alpha_{\min, \text{POE}} \approx 0.2^\circ$. The POE for Ganymede is wider ($\alpha_{\min, \text{POE}} \approx 0.6^\circ$), and, apparently, slightly deeper ($P_{\min, \text{POE}} \approx -0.4\%$). The POE for Io appears to be more symmetric and shallow ($P_{\min, \text{POE}} \approx -0.2\%$) and is centered at $\alpha_{\min, \text{POE}} \approx 0.5^\circ$. In addition, the POEs for Europa and Ganymede deepen with decreasing wavelength.

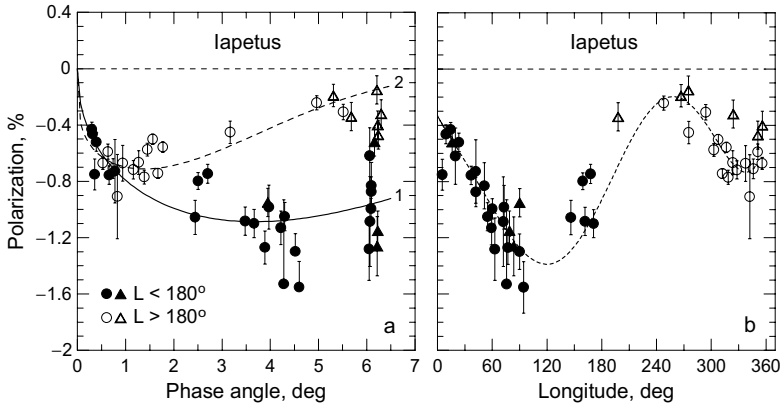


Fig. 11. Degree of linear polarization *versus* (a) phase angle and (b) longitude for Iapetus in the R filter. The respective NPBs are shown schematically for the leading (1 – filled symbols) and trailing (2 – open symbols) hemispheres. Triangles show polarization data in the V filter taken from Zellner (1972) and Veverka (1977).

In summary, multiple polarimetric observations of the Galilean satellites confirm unambiguously the existence of the POE.

Saturnian satellite Iapetus. The two thoroughly different hemispheres of Iapetus are of exceptional interest from the standpoint of microphysical diagnostics of its surface as well as the study of pertinent scattering effects. The leading hemisphere of the satellite is dominated by dark materials consisting of organic PAH and aliphatic compounds and has an albedo of $\sim 4\%$, whereas the bright trailing side is covered with water ice and has a geometric albedo exceeding 60% (Goguen et al. 1983; Squyres et al. 1984). High-resolution images of Iapetus suggest the presence of fine-grained and porous surface materials.

A nonlinear surge of brightness was found for both sides of the satellite (Franklin and Cook 1974; Noland et al. 1974). Prior to our observations, there were only several sparse polarization measurements for Iapetus by Zellner (Zellner 1972; Veverka 1977), which were sufficient to estimate the depth of the NPB, but the detailed phase-angle dependence of polarization could not be determined. It was found that the NPB for Iapetus is much deeper for the dark leading hemisphere ($P_{\min} \approx -1.3\%$) than for the bright trailing side ($P_{\min} \approx -0.2\%$).

We have been carrying out UBVR polarimetric observations of Iapetus since 1998 using the 2.6-m and 1.25-m CrAO telescopes and the 2-m telescope of the Pick Terskol Observatory (Northern Caucasus, Russia). The minimal phase angle reached is as small as 0.3° . The results of these observations in the R filter, for which the measurement accuracy is the best, along with the V-filter data by Zellner (1972) are shown in Fig. 11. The observed phase-angle dependences of polarization (Fig. 11a) for the leading dark side ($0^\circ \leq L \leq 180^\circ$) and the trailing bright

side ($180^\circ \leq L \leq 360^\circ$) of Iapetus as well as the longitudinal dependence of polarization (Fig. 11b) are depicted. It is seen that the observed absolute value of polarization reaches a maximum at $L \approx 130^\circ$ for the leading hemisphere and at $L \approx 250^\circ$ for the trailing hemisphere, thereby indicating that the dark and bright sides of Iapetus do not coincide precisely with its leading and trailing hemispheres, respectively. It should be noted that the correction for the orbital latitudinal variations was not taken into account in plotting the NPBs for either hemisphere.

One can readily see significant NPB differences between the two hemispheres. The leading-hemisphere polarization (Fig. 11a, curve 1) changes abruptly from about -0.4% to about -0.8% within the phase-angle range from $\sim 0.3^\circ$ to $\sim 1^\circ$. At $\alpha \approx 6^\circ$ the polarization is close to that at $\alpha \approx 1^\circ$. Therefore, the NPB for the leading hemisphere of Iapetus in the phase-angle range $0.3^\circ - 6.1^\circ$ can be represented by an asymmetric parabola with a minimum at $\alpha \approx 3^\circ - 4^\circ$ without any hint of a secondary minimum. It is worth noticing that asteroids with albedos as low as that of the leading hemisphere of Iapetus have minimal polarization at $\alpha_{\min} \approx 10^\circ$.

The bright-side polarization of Iapetus measured at phase angles $5^\circ - 6^\circ$ is close to -0.2% , which is a typical value for the regular NPB of high-albedo asteroids such as 64 Angelina and 44 Nysa or ice-covered satellites such as Europa and Ganymede at the corresponding phase angles. On the other hand, within the phase angle range 1.5° to 0.3° , the phase dependence of polarization for the bright side of Iapetus shows strong negative polarization, about -0.7% at $\alpha \approx 1^\circ$. Taking into account the value of polarization $P = -0.5\%$ measured in April 2010 at the phase angle $\sim 3^\circ$, we cannot exclude a strongly asymmetric polarization phase curve with a $\sim 0.7\%$ minimum at $\sim 1.5^\circ$ similar to those observed in the laboratory measurements for particulate MgO (Lyot 1929) and Al_2O_3 (Geake and Geake 1990) surfaces.

Further polarimetric observations of Iapetus are obviously needed, especially within the phase-angle range $2^\circ - 5^\circ$, to understand whether the POE for the bright hemisphere has the shape of a strongly asymmetric curve or is a separate secondary minimum of polarization near opposition.

5. Trans-Neptunian objects and Centaurs

TNOs and Centaurs are believed to represent one of the oldest and possibly most primitive populations of Solar System bodies. For observations of these distant and faint ($\sim 20^m$) bodies, large telescopes with mirrors in the 8 – 10-m diameter range are required. Because of the geometric conditions of ground-based observations, the maximal possible phase angle that can be reached for Centaurs is $7^\circ - 8^\circ$, while that for the TNOs is less than 2° . Nevertheless, even at such small phase angles the measured polarization, especially when combined with results from other observations, can yield valuable information about the optical and physical properties of the most primitive surface materials preserved from early stages of evolution of the Solar System.

The first polarimetric observations of TNOs, with the exception of (134340) Pluto, were carried out in 2002 (Boehnhardt et al. 2004). At present, the list of TNOs and Centaurs, for which polarimetric observations have been carried out, includes eight TNOs [(28978) Ixion, (29981) 1999 TD₁₀, (50000) Quaoar, (136199) Eris, (26375) 1999 DE₉, (38628) Huya, (136108) Haumea, and (20000) Varuna] and three Centaurs [(2060) Chiron, (5145) Pholus, and (1019) Chariklo] (Bagnulo et al. 2011, and references therein). There are also polarimetric data for Pluto, which is now classified as a TNO (Fix and Kelsey 1973; Breger and Cochran 1982; Avramchuk et al. 1992).

Summarizing the studies by Boehnhardt et al. (2004), Rousselot et al. (2005), Bagnulo et al. (2006, 2008), and Belskaya et al. (2010), one can conclude that the main observational results of polarimetry of the selected Centaurs and TNOs are the following (see Fig. 12):

- the degree of polarization varies from -0.3% to -2% in the phase angle range from 0.1° to 4.4° . The strong polarization observed for some Centaurs and TNOs at a phase angle of 1° is unique among ASSBs;
- there are two classes of TNOs with different polarimetric behavior: (i) TNOs with diameters smaller than 1000 km (Ixion, Huya, DE₉, Varuna, and TD₁₀) exhibit negative polarization which rapidly increases (in absolute value) with phase angle, reaching about -1% at a phase angle of 1° , and (ii) larger TNOs (Eris, Haumea, Quaoar, and Pluto) exhibiting relatively weak negative polarization ($\leq 0.7\%$ in absolute value) which is nearly constant in the observed phase-angle range;
- the Centaurs observed (Chiron, Pholus, and Chariklo) exhibit different phase-angle dependences of polarization with a minimum value varying from about -1% to about -2.1% at phase angles of $1.5^\circ - 2.0^\circ$. It appears that the polarization minima are shifted toward the smallest angles as compared to any other ASSB observed so far. Unlike the TNOs, all three Centaurs show a noticeable diversity in the behavior of the polarization phase-angle dependence.

As one can see from Fig. 12, the behavior of the polarization phase dependence for objects residing in the inner part of the Solar System (i.e., at distances from the Sun limited by radius of the Saturnian orbit) differs significantly from that exhibited by more distant objects. All high-albedo objects in the left-hand panel (E-type asteroids and the Galilean satellites) show similar polarization phase curves: there is a secondary polarization minimum (POE) at phase angles of $0.2^\circ - 0.7^\circ$ superposed on a shallow NPB. There is, however, a considerable distinction between the POEs exhibited by silicate asteroids and water-ice-covered satellites. At the same time, much greater differences in the polarization phase dependences are observed among TNOs and Centaurs. In particular, there is no distinct secondary minimum of polarization for these distant objects. However, deep and very asymmetric phase curves with minima at phase angles of about $1^\circ - 2^\circ$ are observed for some dark (e.g., Chiron, Huya, and DE₉) and bright (e.g., the

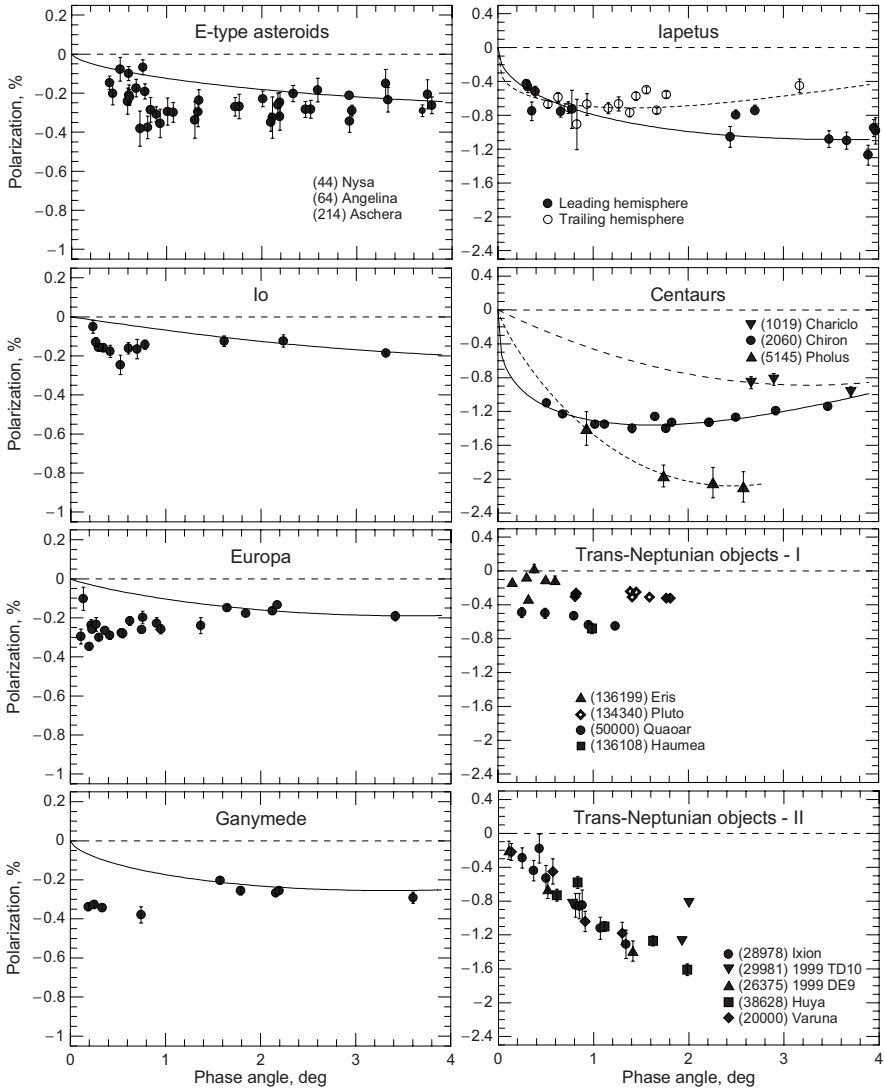


Fig. 12. The degree of linear polarization *versus* phase angle for different objects. Secondary local minima of polarization near opposition are observed for the E-type asteroids and Jovian satellite Io, Europa, and Ganymede. The polarization data for Io, Europa, and Ganymede at phase angles greater than 1° were obtained by averaging data from Dollfus (1975), Rosenbush et al. (1997), and Rosenbush and Kiselev (2005) over 1° intervals with equal weights assigned to all observations. The solid curves represent the best fits to the data.

trailing side of Iapetus) objects as well as moderate-albedo bodies (e.g., Ixion and Varuna). It is obvious that this great diversity in the observed polarization phase

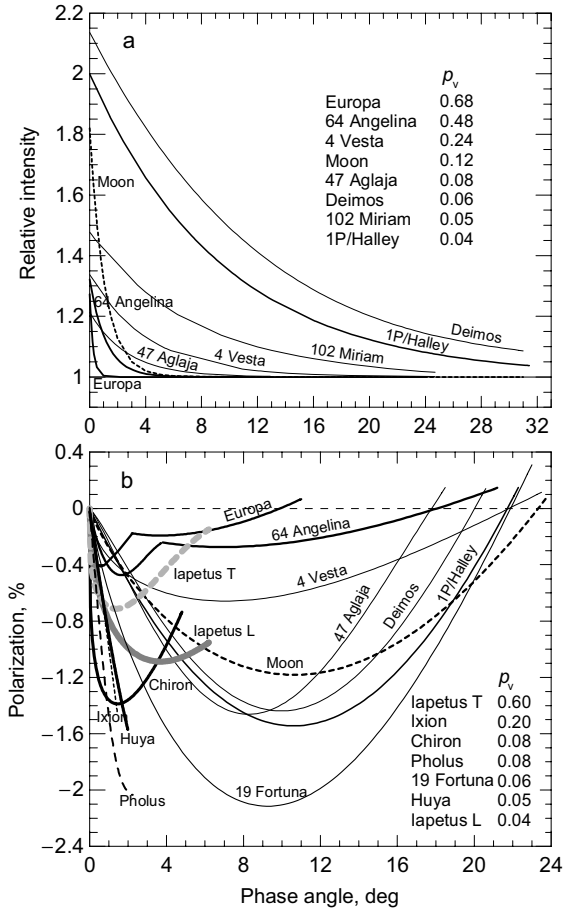


Fig. 13. Comparison of (a) photometric phase curves after the subtraction of the linear part and (b) polarization phase curves for different objects. A second local minimum of polarization near opposition is shown for the Jovian satellite Europa and asteroid 64 Angelina.

curves cannot be explained as resulting solely from differences in the respective surface albedos.

6. Summary

The photometric and polarimetric behavior of different ASSBs near opposition is highly variable, as illustrated in Fig. 13 (as well as in other figures, e.g., Figs. 1, 2, and 12). Here, the results of photometric and polarimetric observations of several objects are summarized in order to demonstrate the following features. Certain high-albedo objects (e.g., asteroid 64 Angelina and the Jovian satellite Eu-

ropa) exhibit a very narrow opposition brightness peak (see Figs. 3, 4, 9, and 13). For some objects (e. g., for the low-albedo Deimos and comet Halley), the curvature of the photometric phase curve changes rather slowly over a broad range of phase angles, which does not allow one to separate a linear part (see Fig. 13). Some dark objects (e.g., C- and P-type asteroids) show a very weak increase of intensity at phase angles approaching zero, and one can identify only a linear part of the brightness phase curve (see Fig. 1). At the same time, the backscattering intensity peaks are extremely narrow for moderate-albedo major satellites of Uranus (e.g., the darkest satellite Umbriel with an albedo of $p_V = 0.19$ and the brightest satellite Ariel with $p_V = 0.35$; see Fig. 1 and Avramchuk et al. 2007). According to our studies (Rosenbush et al. 2002), the amplitude (i.e., the brightness increase relative to the background intensity) of the BOE lies in the range 1.1–5.4, the HWHM of the BOE varies from 0.08° to 6.1° , the phase angle at which the nonlinear BOE begins varies from 0.4° to $\sim 30^\circ$, and the phase coefficient varies from 0.0005 to 0.06 mag/deg.

The shape of the polarization phase curve is also quite variable (see Figs. 3, 12, and 13). For certain high-albedo objects, for which a spike-like BOE has been detected, the phase-angle curves of polarization consist of a POE in the form of a separate peak of negative polarization at phase angles $\alpha < 2^\circ$ and a shallow regular NPB. Our most recent observations confirm the existence of a class of high-albedo ASSBs (including asteroids 44 Nysa and 64 Angelina, the Jovian satellites Io, Europa, and Ganymede, and the A and B rings of Saturn) with unique opposition properties: a spike-like BOE is accompanied by an equally narrow POE. These observational results are in a reasonable quantitative agreement with the existing theory (Mishchenko et al. 2006a,b, 2009a,b, 2010).

For dark objects (e.g., C- and P-type asteroids and Martian satellites), deep regular NPBs are observed. An especially deep NPB with a minimum of $P_{\min} \approx 2.1\%$ at $\alpha_{\min} \approx 9.5^\circ$ is found for asteroid Fortuna (APD). However, no polarization feature at small phase angles has been observed for moderate- and low-albedo objects. The corresponding NPB parameters typically vary within the following ranges: $0.2\% - 2.1\%$ for $|P_{\min}|$, $2^\circ - 12^\circ$ for α_{\min} , and $\sim 7^\circ - 29^\circ$ for α_{inv} .

As one can see from Figs. 12 and 13, the phase dependences of polarization for objects belonging to the inner part of the Solar System (within the Saturnian orbit) differ considerably from those for objects residing in the outer part. Unexpectedly asymmetric and deep phase curves of polarization with very small inversion angles are observed for the bright hemisphere of Iapetus (about -0.7% at $\alpha \approx 1.1^\circ$), Chiron (about -1.4% at $\alpha \approx 1.5^\circ$), and Pholus (about -2.3% at $\alpha \approx 2.5^\circ$) (Belskaya et al. 2010). Extremely sharp and deep negative polarization features are exhibited by some TNOs (such as Ixion, Huya, DE₉, and Varuna), although other TNOs (Eris and Pluto) show much smaller absolute polarization values at the same phase angles, only $0.1\% - 0.2\%$ (Bagnulo et al. 2008; Mishchenko et al. 2010, and references therein; Bagnulo et al. 2011). Even though the range of phase angles covered does not allow one to determine unambiguously the full shape of the respective NPBs, we can conclude that the NPBs for objects in the

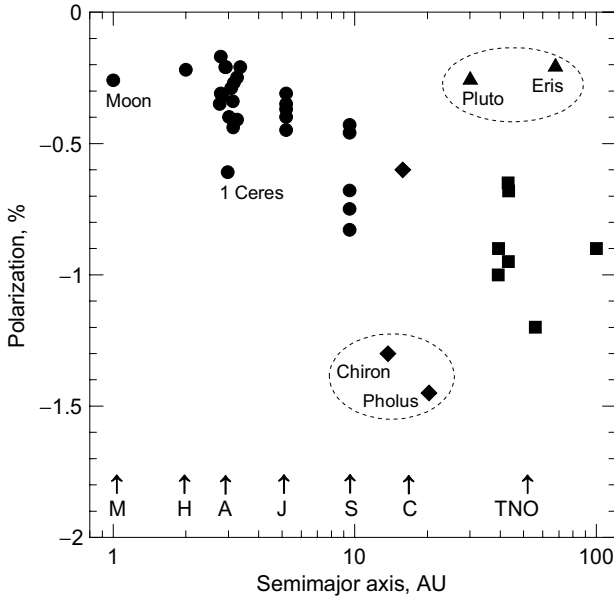


Fig. 14. The degree of polarization at the phase angle 1° for different objects as a function of the distance from the Sun. Triangles show TNOs with shallow NPBs, while the other TNOs are depicted by squares. Centaurs are shown by diamonds. Notation: M – the Moon; H – comet 1P/Halley; A – asteroids; J – Jovian satellites; S – Saturnian satellites and rings; C – Centaurs (diamonds); TNO – trans-Neptunian objects (squares and triangles).

inner and outer parts of the Solar System are very different. It appears that the observed polarization of objects from the inner part is generally related to their albedos, whereas the polarization observed for distant objects cannot be explained as being controlled solely by the brightness of the surface material.

This conclusion may be confirmed by Fig. 14, in which the degree of polarization at the phase angle 1° versus the distance from the Sun (semi-major orbital axis) is shown for different objects. Despite the small sample of objects, there is a rather significant correlation with a correlation coefficient of -0.737 . It is quite natural to associate different types of behavior of polarization phase dependences with different compositions and/or morphologies of ASSB surfaces. In particular, one can expect a more pronounced evolution of the surface owing to greater insolation for objects residing closer to the Sun.

Large-amplitude BOEs and an unusually strong polarization for TNOs are indicative of a specific composition and/or microstructure of their surfaces. These special properties can be caused by a relatively weak insolation of primitive materials typical of the outer regions of the Solar System. An example would be irradiated ices enriched by organic matter. It is natural to assume increasingly primitive and unprocessed surface materials with increasing distance from the Sun. In other

words, the composition and structure of the surfaces of small bodies should vary with their distance from the Sun rather than being determined solely by the bodies' evolutionary history.

A more detailed quantitative analysis of the observational data in terms of specific physical parameters of the regolith layer is hardly possible at this time given the limited nature of the observational dataset, the constrained theoretical ability to compute all photometric and polarimetric characteristics of the opposition phenomena for realistic polydisperse particle models, and the extreme morphological complexity and heterogeneity of the surfaces of the ASSBs.

Acknowledgments

V. Rosenbush appreciates a grant from the NATO ASI. The authors are grateful to N. Kiselev, S. Zajtsev, and F. Velichko for sharing their unpublished polarimetric data for 44 Nysa, 64 Angelina, and Iapetus which supplement our data. The authors also thank N. Kiselev for useful discussions and assistance in improving the original manuscript. This research was partly funded by the NASA Radiation Sciences Program managed by Hal Maring.

References

- Avramchuk, V. V., V. Yu. Rakhimov, G. P. Chernova, and V. I. Shavlovskij, 1992: Photometry and polarimetry of Pluto near its perihelion position. I. *Kinem. Phys. Celest. Bodies* **8**, 30–37.
- Avramchuk, V. V., V. K. Rosenbush, and T. P. Bul'ba, 2007: Photometric study of the major satellites of Uranus. *Solar Syst. Res.* **41**, 186–202.
- Bagnulo, S., H. Boehnhardt, K. Muinonen, et al., 2006: Exploring the surface properties of trans-Neptunian objects and Centaurs with polarimetric FORS1/VLT observations. *Astron. Astrophys.* **450**, 1239–1248.
- Bagnulo, S., I. Belskaya, K. Muinonen, et al., 2008: Discovery of two distinct polarimetric behaviours of trans-Neptunian objects. *Astron. Astrophys.* **491**, L33–L36.
- Bagnulo, S., I. Belskaya, H. Boehnhardt, et al., 2011: Polarimetry of small bodies of the solar system with large telescopes. *J. Quant. Spectrosc. Radiat. Transfer*, in press.
- Belskaya, I. N., V. G. Shevchenko, N. N. Kiselev, et al., 2003: Opposition polarimetry and photometry of S- and E-type asteroids. *Icarus* **166**, 276–284.
- Belskaya, I. N., Yu. G. Shkuratov, Yu. S. Efimov, et al., 2005: The F-type asteroids with small inversion angles of polarization. *Icarus* **178**, 213–221.
- Belskaya, I. N., S. Bagnulo, M. A. Barucci, et al., 2010: Polarimetry of Centaurs (2060) Chiron, (5145) Pholus and (10199) Chariklo. *Icarus* **210**, 472–479.
- Berdugun, A. V., and N. M. Shakhovskoy, 1993: A method of simultaneous measurements of four Stokes parameters in the UVBRI bands. *Izv. Krym. Astrofiz. Obs.* **87**, 122–134 (in Russian).
- Birch, P. V., E. F. Tedesco, R. C. Taylor, et al., 1983: Lightcurves and phase function of asteroid 44 Nysa during its 1979 apparition. *Icarus* **54**, 1–12.
- Boehnhardt, H., S. Bagnulo, K. Muinonen, et al., 2004: Surface characterization of 28978 Ixion (2001 KX76). *Astron. Astrophys.* **415**, L21–L25.

- Breger, M., and W. D. Cochran, 1982: Polarimetry of Pluto. *Icarus* **49**, 120–124
- Dollfus, A., 1975: Optical polarimetry of the Galilean satellites of Jupiter. *Icarus* **25**, 416–431.
- Dollfus, A., 1984: The Saturn ring particles from optical reflectance polarimetry. In *CNES Planetary Rings* (SEE N85-26473 15–91), pp. 121–143.
- Dollfus, A., and B. Zellner, 1979: Optical polarimetry of asteroids and laboratory samples. In T. Gehrels, Ed., *Asteroids* (University of Arizona Press, Tucson, AZ), pp. 170–183.
- Fix, L. A., and J. D. Kelsey, 1973: Polarimetry of Pluto. *Astrophys. J.* **184**, 633–636.
- Fornasier, S., I. Belskaya, M. Fulchignoni, et al., 2006a: First albedo determination of 2867 Steins, target of the Rosetta mission. *Astron. Astrophys.* **449**, L9–L12.
- Fornasier, S., I. Belskaya, Yu. G. Shkuratov, et al., 2006b: Polarimetric survey of asteroids with the Asiago telescope. *Astron. Astrophys.* **455**, 371–377.
- Franklin, F. A., and A. F. Cook, 1974: Photometry of Saturn’s satellites: the opposition effect of Iapetus at maximum light and the variability of Titan. *Icarus* **23**, 355–362.
- Geake, J. E., and A. Dollfus, 1986: Planetary surface texture and albedo from parameter plots of optical polarization data. *Mon. Not. R. Astron. Soc.* **218**, 75–91.
- Geake, J. E., and M. Geake, 1990: A remote sensing method for sub-wavelength grains on planetary surfaces by optical polarimetry. *Mon. Not. R. Astron. Soc.* **245**, 46–55.
- Gehrels, T., 1956: Photometric studies of asteroids. V. The light-curve and phase function of 20 Massalia. *Astrophys. J.* **123**, 331–335.
- Goguen, J., M. Tripicco, and D. Morrison, 1983: Voyager feature-fixed photometry of the surface of Iapetus. *Bull. Am. Astron. Soc.* **15**, 855.
- Hapke, B., 1993: *Theory of Reflectance and Emittance Spectroscopy* (Cambridge University Press, Cambridge, UK).
- Harris, A. W., J. W. Young, L. Contreiras, et al., 1989: Phase relations of high albedo asteroids: the unusual opposition brightening of 44 Nysa and 64 Angelina. *Icarus* **81**, 365–374.
- Heiles, C., 2000: 9286 stars: an agglomeration of stellar polarization catalogs. *Astron. J.* **119**, 923–927.
- Helfenstein, P., N. Currier, B. E. Clark, et al., 1998: Galileo observations of Europa’s opposition effect. *Icarus* **135**, 41–63.
- Kiselev, N. N., and F. P. Velichko, 1997: Aperture polarimetry and photometry of comet Hale–Bopp. *Earth Moon Planets* **78**, 347–352.
- Kiselev, N. N., V. K. Rosenbush, K. Jockers, et al., 2002: Polarimetry of near-Earth asteroid 33342 (1998 WT24). Synthetic phase angle dependence of polarization for the E-type asteroids. In B. Warmbein, Ed., *Asteroids, Comets, Meteors* (ACM 2002, ESA SP–500), pp. 887–890.
- Kiselev, N., V. Rosenbush, F. Velichko, and S. Zaitsev, 2009: Polarimetry of the Galilean satellites and Jupiter near opposition. *J. Quant. Spectrosc. Radiat. Transfer* **110**, 1713–1718.
- Lumme, K., K. Muinonen, 1993: A two-parameter system for linear polarization of some solar system objects. In *IAU Symposium 160: Asteroids, Comets, Meteors* (Houston, TX), LPI 810, pp. 194–197.
- Lupishko, D. F., and R. A. Mohamed, 1996: A new calibration of the polarimetric albedo scale of asteroids. *Icarus* **119**, 209–213.
- Lupishko, D. F., and S. V. Vasil’ev, Eds., 2008: Asteroid Polarimetric Database V6.0. EAR-A-3-RDR-APD-POLARIMETRY-V6.0. NASA Planetary Data System.

- Lyot, B., 1929: Recherches sur la polarisation de la lumière des planetes et de quelques substances terrestres. *Ann. Observ. Paris, Sect. Meudon.* **8**. English translation: Research on the polarization of light from planets and from some terrestrial substances. NASA Tech. Transl. NASA TT F-187 (National Aeronautics and Space Administration, Washington, DC, 1964).
- Lyot, B., 1934: Polarisation des petites planètes. *C. R. Acad. Sci. Paris* **199**, 774–782.
- Mishchenko, M. I., 1993: On the nature of the polarization opposition effect exhibited by Saturn's rings. *Astrophys. J.* **411**, 351–361.
- Mishchenko, M. I., J.-M. Luck, and Th. M. Nieuwenhuizen, 2000: Full angular profile of the coherent polarization opposition effect. *J. Opt. Soc. Am. A.* **17**, 888–891.
- Mishchenko, M. I., L. D. Travis, and A. A. Lacis, 2002: *Scattering, Absorption, and Emission of Light by Small Particles* (Cambridge University Press, Cambridge, UK) (<http://www.giss.nasa.gov/staff/mmishchenko/books.html>).
- Mishchenko, M. I., L. D. Travis, and A. A. Lacis, 2006a: *Multiple Scattering of Light by Particles: Radiative Transfer and Coherent Backscattering* (Cambridge University Press, Cambridge, UK).
- Mishchenko, M. I., V. K. Rosenbush, and N. N. Kiselev, 2006b: Weak localization of electromagnetic waves and opposition phenomena exhibited by high-albedo atmosphereless solar system objects. *Appl. Opt.* **45**, 4459–4463.
- Mishchenko, M. I., Dlugach J. M., Liu Li, et al., 2009a: Direct solutions of the Maxwell equations explain opposition phenomena observed for high-albedo solar system objects. *Astrophys. J.* **705**, L118–L122.
- Mishchenko, M. I., Dlugach J. M., Liu Li, 2009b: Azimuthal asymmetry of the coherent backscattering cone: theoretical results. *Phys. Rev. A* **80**, 053824.
- Mishchenko, M. I., V. K. Rosenbush, N. N. Kiselev, et al., 2010: *Polarimetric Remote Sensing of Solar System Objects* (Akademperiodyka, Kyiv) (arXiv:1010.1171).
- Mishchenko, M. I., V. P. Tishkovets, L. D. Travis, et al., 2011: Electromagnetic scattering by a morphologically complex object: fundamental concepts and common misconceptions. *J. Quant. Spectrosc. Radiat. Transfer* **112**, 671–692.
- Muononen, K., J. Piironen, Yu. G. Shkuratov, et al., 2002: Asteroid photometric and polarimetric phase effects. In W. F. Bottke, Jr., A. Cellino, P. Paolicchi, R. P. Binzel, Eds., *Asteroids III* (University of Arizona Press, Tucson, AZ), pp. 123–138.
- Noland, M., J. Veverka, D. Morrison, et al., 1974: Six-color photometry of Iapetus, Titan, Rhea, Dione and Tethys. *Icarus* **23**, 334–354.
- Petrova, E. V., V. P. Tishkovets, and K. Jockers, 2007: Modeling of opposition effects with ensembles of clusters: interplay of various scattering mechanisms. *Icarus* **188**, 233–245.
- Pirola, V., 1988: Simultaneous five-colour (UBVRI) photopolarimeter. In G. V. Coyne et al., Eds., *Polarized Radiation of Circumstellar Origin* (Vatican Observatory/University of Arizona Press, Vatican City/Tucson), pp. 735–746.
- Rosenbush, V. K., 2002: The phase-angle and longitude dependence of polarization for Callisto. *Icarus* **159**, 145–155.
- Rosenbush, V. K., 2006: Light-scattering properties of small Solar System bodies. Habilitation dissertation (Main Astronomical Observatory, Kyiv).
- Rosenbush, V. K., and N. N. Kiselev, 2005: Polarization opposition effect for the Galilean satellites of Jupiter. *Icarus* **179**, 490–496.
- Rosenbush, V. K., V. V. Avramchuk, and A. E. Rosenbush, 1995: The polarization opposition effects of the Galilean satellites of Jupiter. *Bull. Am. Astron. Soc.* **27**, 1170.

- Rosenbush, V. K., V. V. Avramchuk, A. E. Rosenbush, and M. I. Mishchenko, 1997: Polarization properties of the Galilean satellites of Jupiter: observations and preliminary analysis. *Astrophys. J.* **487**, 402–414.
- Rosenbush, V., N. Kiselev, V. Avramchuk, and M. Mishchenko, 2002: Photometric and polarimetric phenomena exhibited by Solar System bodies. In G. Videen, and M. Kocifaj, Eds., *Optics of Cosmic Dust* (Kluwer Academic Publishers, Dordrecht, The Netherlands), pp. 191–226.
- Rosenbush, V. K., N. N. Kiselev, V. G. Shevchenko, et al. 2005: Polarization and brightness opposition effects for the E-type asteroid 64 Angelina. *Icarus* **178**, 222–234.
- Rosenbush, V., N. Kiselev, and V. Avramchuk, 2006: Similarity and diversity in photometric and polarimetric opposition effects of small Solar System bodies. *J. Quant. Spectrosc. Radiat. Transfer* **100**, 325–339.
- Rosenbush, V., V. Shevchenko, N. Kiselev, et al., 2009: Polarization and brightness opposition effects for E-type asteroid 44 Nysa. *Icarus* **201**, 655–665.
- Rousselot, P., A. C. Lvasseur-Regourd, K. Muinonen, and J.-M. Petit, 2005: Polarimetric and photometric phase effects observed on transneptunian object (29981) 1999 TD10. *Earth Moon Planets* **97**, 353–364.
- Shakhovskoy, N. M., and Yu. S. Efimov, 1972: Polarization observations of unstable stars and extragalactic object. I: Instrumentation, method of observation and reduction. *Izv. Krym. Astrofiz. Obs.* **45**, 90–110 (in Russian).
- Shakhovskoy, N. M., I. L. Andronov, S. V. Kolesnikov, and A. V. Khalevin, 2001: Methods of measurements of circular polarization by using a single-channel photopolarimeter. *Izv. Krym. Astrofiz. Obs.* **97**, 91–100 (in Russian).
- Shevchenko, V. G., V. G. Chernyi, and Yu. N. Kruglyi, 1992: Photometry of seventeen asteroids. *Icarus* **100**, 295–306.
- Shkuratov, Yu. G., K. Muinonen, E. Bowell, et al., 1994: A critical review of theoretical models of negatively polarized light scattered by atmosphereless solar system bodies. *Earth Moon Planets* **65**, 201–246.
- Shkuratov, Yu., A. Ovcharenko, E. Zubko, et al., 2002: The opposition effect and negative polarization of structural analogs of planetary regoliths. *Icarus* **159**, 396–416.
- Squyres, S. W., B. Buratti, J. Veverka, and C. Sagan, 1984: Voyager photometry of Iapetus. *Icarus* **59**, 426–435.
- Tedesco, E. F., P. V. Noah, M. Noah, and S. D. Price, 2002: The supplemental IRAS minor planet survey. *Astron. J.* **123**, 1056–1085.
- Thompson, D. T., and G. W. Lockwood, 1992: Photoelectric photometry of Europa and Callisto 1976–1991. *J. Geophys. Res.* **97**, 14761–14772.
- Tishkovets, V. P., 2008: Light scattering by closely packed clusters: shielding of particles by each other in the near field. *J. Quant. Spectrosc. Radiat. Transfer* **109**, 2665–2672.
- Tupieva, F. A., 2003: UVB photometry of the asteroid 44 Nysa. *Astron. Astrophys.* **408**, 379–385.
- Veverka, J., 1977: Polarimetry of satellite surfaces. In J. A. Burns, Ed., *Planetary Satellites*. (University of Arizona Press, Tucson, AZ), pp. 171–209.
- Zellner, B. H., 1972: On the nature of Iapetus. *Astrophys. J.* **174**, L107–L109.
- Zellner, B., and J. Gradie, 1976: Minor planets and related objects. XX. Polarimetric evidence for the albedos and compositions of 94 asteroids. *Astron. J.* **81**, 262–280.
- Zellner, B., T. Gehrels, and J. Gradie, 1974: Minor planets and related objects. XVI. Polarimetric diameters. *Astron. J.* **79**, 1100–1110.

Mueller-matrix characterization of biological tissues

Sergey N. Savenkov*

Department of Radiophysics, Kyiv Taras Shevchenko University, Kyiv, Ukraine

Abstract. Various kinds of abnormalities and pathologies result in changes of the structural properties of collagen and other fibrous biological tissues, thereby leading to significant alterations of their morphological and anisotropic properties. The Mueller matrix contains all the optical information that one can obtain from light scattered by a medium and, therefore, shows great promise for the efficient use of polarization parameters in the characterization of sizes, shapes, and orientations of tissue structural elements as well as the birefringence, dichroism, depolarization, etc. of biological tissues. The extraction of this information via the interpretation of experimental results of Mueller-matrix measurements represents an important challenge, especially when the various tissue properties must be characterized simultaneously. The purpose of this chapter is to (i) present, in a systematic way, the main properties of Mueller and Jones matrices that can be experimentally or numerically derived, and (ii) provide a review of the information content of Mueller matrices for biological tissues in the framework of models according to which a tissue is a discrete ensemble of scatterers or a continuous distribution of optical parameters.

Keywords: Mueller matrix, biological tissue, linear and circular birefringence and dichroism, depolarization, polar decomposition

1. Introduction

To develop polarimetric methods for biological tissue diagnostics and therapeutics, one needs to understand the relation between polarimetric and biological properties of tissues. The understanding of the nature of light interaction with biological tissues is relevant to a number of noninvasive medical diagnostic techniques for cancer and other tissue pathologies. As light propagates through a biological tissue, its polarization state changes as does its intensity. Polarization properties of the scattered light contain extensive information on morphological and functional properties of the tissue. For example, since depolarization of scattered polarized light depends on the morphological and optical parameters of scatterers (i.e., density, size, distribution, shape, refractive index etc.) present in the

* Corresponding author. E-mail: sns@univ.kiev.ua

tissue (Bickel et al. 1976; Tuchin et al. 2006), this information can be utilized for tissue diagnostic techniques.

Many constituents of a tissue also exhibit anisotropic properties such as birefringence and dichroism. For example, collagen is a structural protein present in tissue that has linear birefringence property due to its oriented fibrous structure. The structural properties of collagen and other fibrous structures present in tissues with various kinds of abnormalities such as keratosis, neurofibroma, psoriasis, erythema, and the numerous types of carcinomas (e.g., melanomas, basal cell, squamous cell carcinoma, etc.) are characterized by changes leading to significant alterations in their linear birefringence (see, e.g., Maitland et al. 1997; de Boer et al. 1997, 2002; Everett et al. 1998; Simonenko et al. 2000). The presence of sucrose or glucose possessing intrinsic circular birefringence due to their chiral structure results in the rotation of linear polarization about the direction of light propagation (Applequist 1987; Barron 1982; Studinski and Vitkin 2000; Hadley and Vitkin 2002; Manhas et al. 2006).

The enormous importance of the Mueller matrix is that it contains all the information (birefringence, dichroism, and depolarization) that one can obtain from light scattered by a medium (Bohren and Huffman 1983; Brosseau 1998; Azzam and Bashara 1987; Collett 1993; Shurcliff 1962). The information contained in the Mueller matrix has many useful applications in such diverse fields as interaction with various optical systems (Shurcliff 1962; Azzam and Bashara 1977; Collett 1993; Brosseau 1998), cloud diagnostics (van de Hulst 1957; Bohren and Huffman 1983; Mishchenko et al. 2000, 2002; Kokhanovsky 2003b), remote sensing of the ocean, atmosphere, and planetary surfaces (Boerner 1992; Kokhanovsky 2001, 2003a,b; Muttiah 2002; Mishchenko et al. 2010), and biological tissue optics (Priezzhev et al. 1989; Tuchin 2002; Tuchin et al. 2006).

Hence, the Mueller matrix shows great promise for the efficient quantification of polarization parameters, i.e., sizes, shapes and orientation of the tissue structural elements as well as birefringence, dichroism, depolarization, etc. of biological tissues. To extract this information and interpret experimental results on light scattering, one needs an appropriate polarimetric matrix model for the particular tissue in question. There are two approaches which are currently used for tissue modeling. First, a tissue can be modeled as a discrete ensemble of scatterers (Wang et al. 2002; Tuchin et al. 2006); and second, the tissue can be modeled as a medium with a continuous distribution of optical parameters (Tuchin 1994; Schmitt et al. 1996; Fanjul-Vélez and Arce-Diego 2010). The choice of approach is fully determined by both the structural features of the tissue in question and the type of light-scattering characteristics that are to be obtained. In the first case, the Mueller matrix contains information related to the optical properties, size, shape, and composition of the constituent scatterers (van de Hulst 1957; Bohren and Huffman 1983), while in the second case, the Mueller matrix contains information related to the anisotropy of the medium, viz., linear and circular dichroism and birefringence (i.e., amplitude and phase anisotropies) (Landau et al. 1984; Berry and Dennis 2003). In both cases the Mueller matrix can contain information on depo-

larization (Chipman 1995; Mishchenko and Travis 2000; Gil 2007). The methods of interpretation of the Mueller matrices have been developed by many authors (Hurwitz and Jones 1941; Whitney 1971; Cloude 1986; Gil and Bernabeu 1987; Lu and Chipman 1994, 1996; Mar'enko and Savenkov 1994; Savenkov et al. 2005, 2006, 2007b).

Our main focus in this chapter is on the Mueller-matrix interpretation intended for the characterization of biological tissues. Many important polarimetric applications which do not involve the measurement of the Mueller matrix are beyond the scope of our discussion. Hence, our reference list should by no means be considered exhaustive and is merely intended to provide initial reference points for the interested reader.

2. Mueller matrix calculus

In the Mueller-matrix calculus, the polarization state of light can be completely characterized by a Stokes vector, while the polarization transforming properties of a medium can be completely characterized by a Mueller matrix:

$$\mathbf{S}^{out} = \mathbf{M}\mathbf{S}^{inp}, \quad (1)$$

where the four-component Stokes column vector (with “out” and “inp” denoting the Stokes vectors of the output and input light, respectively) consists of the following parameters:

$$\mathbf{S} = \begin{pmatrix} I \\ Q \\ U \\ V \end{pmatrix} = \begin{pmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \end{pmatrix} = \begin{pmatrix} \langle |E_x|^2 + |E_y|^2 \rangle \\ \langle |E_x|^2 - |E_y|^2 \rangle \\ \langle E_x^* E_y + E_x E_y^* \rangle \\ i \langle E_x^* E_y - E_x E_y^* \rangle \end{pmatrix}, \quad (2)$$

with $i = (-1)^{1/2}$. Among the pioneering contributions to this field of research, we note those by Soleillet (1929), Perrin (1942), Mueller (1948), and Parke (1948, 1949).

The Stokes parameter I is proportional to the total energy flux of the light beam. The Stokes parameters Q and U represent the differences between two components of the flux in which the electric vectors oscillate in mutually orthogonal directions. The Stokes parameter V is the difference between two oppositely circularly polarized components of the flux. As indicated by the angular brackets, the Stokes parameters s_i are ensemble averages (or time averages in the case of ergodic, stationary processes). This implies that no coherence effects are considered.

The Stokes vectors and Mueller matrices represent operations on intensities and their differences, i.e., incoherent superpositions of light beams; they are not adequate to describe either interference or diffraction effects. However, they are well suited to describe partially polarized and unpolarized light. Extensive lists of various Mueller matrices have been presented by several authors (e.g., Shurcliff 1962; Kligler et al. 1990; Gerrard and Burch 1975).

The Stokes parameters obey the inequality

$$s_1^2 \geq s_2^2 + s_3^2 + s_4^2. \quad (3)$$

This inequality is called the Stokes–Verdet criterion and is a consequence of the Schwartz (or Cauchy–Buniakovski) theorem (Barakat 1963). The degree of polarization p is defined by

$$p = \frac{\sqrt{s_2^2 + s_3^2 + s_4^2}}{s_1}. \quad (4)$$

In Eq. (3), the equality holds for a completely polarized (pure) beam of light. In this case, $p=1$. Another limiting case, $p=0$, occurs when $s_2^2 + s_3^2 + s_4^2 = 0$, i.e., when the electric vector vibrates in all directions randomly and with no preferential orientation. An intermediate case, $0 < p < 1$, implies that light contains both polarized and depolarized components and is, therefore, called partially polarized.

The inequality (3) plays an important role in polarimetry because it allows one to classify the character of the light–medium interaction. Assume first that the input light is completely polarized. In this case, the equality in Eq. (3) implies that the medium is non-depolarizing. Note that the terms “non-depolarizing” and “deterministic” or “pure” are not, in general, identical. The term “deterministic” means that the Mueller matrix describing such a medium can be derived from the corresponding Jones matrix (Simon 1982; Gil and Bernabeu 1985; Anderson and Barakat 1994; Gopala Rao et al. 1998b). This condition is more rigid than the condition of a non-depolarizing medium (Savenkov and Yushtin 2000). Hereinafter, we call this class of matrices pure Mueller matrices (Hovenier 1994). If the output light results in an inequality in Eq. (3) then the scattering medium is not deterministic. If, in addition, the transformation matrix in Eq. (1) can be represented as a convex sum of deterministic Mueller matrices (Cloude 1986; Gil 2000, 2007) then the result is a depolarizing Mueller matrix (hereinafter Mueller matrix); otherwise, the result is a Stokes transformation matrix, i.e., the transformation matrix ensures the fulfillment of the Stokes–Verdet criterion only. The properties of matrices transforming Stokes vectors into Stokes vectors, i.e., those satisfying the Stokes–Verdet criterion, have been studied by many authors (Xing 1992; van der Mee 1993; van der Mee and Hovenier 1992; Sridhar and Simon 1994; Nagirner 1993; Givens and Kostinski 1993; Gopala Rao et al. 1998a).

Any pure Mueller matrix \mathbf{M} can be traced back to the corresponding Jones matrix \mathbf{T} using the following relation (Parke III 1949; Azzam and Bashara 1977):

$$\mathbf{M} = \mathbf{A}(\mathbf{T} \otimes \mathbf{T}^*)\mathbf{A}^{-1}, \quad (5)$$

where the asterisk denotes the complex-conjugate value,

$$\mathbf{T} = \begin{pmatrix} t_1 & t_4 \\ t_3 & t_2 \end{pmatrix}, \quad (6)$$

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 \\ 0 & i & -i & 0 \end{pmatrix}, \quad (7)$$

the t_i are, in general complex, and \otimes is the tensorial (Kronecker) product.

Since the element m_{11} is a gain for unpolarized incident light, it must satisfy the following inequality:

$$m_{11} > 0. \quad (8)$$

Furthermore, the elements of the Mueller matrix must obey the following conditions:

$$m_{11} \geq |m_{ij}|, \quad (9)$$

$$\text{Tr}(\mathbf{M}) \geq 0, \quad (10)$$

$$\mu \mathbf{T} \sim |\mu|^2 \mathbf{M}, \quad (11)$$

where Tr denotes the trace operation and μ is an arbitrary real or complex constant.

Note that the last relation defines the ability of the Jones and Mueller matrices to represent a “physically realizable” medium (Lu and Chipman 1994; Anderson and Barakat 1994; Gil 2007) and implies the physical restriction according to which the ratio g of the intensities of the emerging and incident light beams (the gain or intensity transmittance) must always be in the interval $0 \leq g \leq 1$. This condition is called the gain or transmittance condition and can be written in terms of the elements of the Mueller matrix as follows (Barakat 1987):

$$\begin{aligned} m_{11} + (m_{12}^2 + m_{13}^2 + m_{14}^2)^{1/2} &\leq 1, \\ m_{11} + (m_{21}^2 + m_{31}^2 + m_{41}^2)^{1/2} &\leq 1. \end{aligned} \quad (12)$$

While a Jones matrix has generally 8 independent parameters, the absolute phase is lost in Eq. (5), yielding only seven independent elements for a pure Mueller matrix. Evidently, this results in the existence of interrelations for the elements of a general pure Mueller matrix. This fact was pointed out for the first time, although without a derivation of their explicit form, by van de Hulst (1957). Since then this subject has been studied by many authors (e.g., Abhyankar and Fymat 1969; Fry and Kattawar 1981; Hovenier et al. 1986). In the most complete and refined form these interrelations are presented in Hovenier (1994).

In particular one can derive the following important equation for the elements of a pure Mueller matrix:

$$\sum_{i=1}^4 \sum_{j=1}^4 m_{ij}^2 = 4m_{11}^2. \quad (13)$$

This equality was obtained for the first time by Fry and Kattawar (1981). However, the question of whether this is a sufficient condition for \mathbf{M} to be a pure Mueller matrix has been the subject of extensive discussions (see, e.g., Simon 1982, 1987; Hovenier 1994; Kim et al. 1987; Kostinski 1992; Kostinski et al. 1993; Gil and Bernabeu 1985; Anderson and Barakat 1994; Brosseau 1990; Brosseau et al. 1993). Under the premise that the Mueller matrix in question can be represented as a convex sum of pure Mueller matrices, Eq. (13) is both a necessary and a sufficient condition for \mathbf{M} to be a pure Mueller matrix (Gil 2007).

In addition to the equalities presented above, a set of inequalities can be derived to characterize the structure of the pure Mueller matrix, as follows (Hovenier et al. 1986):

$$\begin{aligned}
 m_{11} + m_{22} + m_{12} + m_{21} &\geq 0, \\
 m_{11} - m_{22} - m_{12} + m_{21} &\geq 0, \\
 m_{11} + m_{22} - m_{12} - m_{21} &\geq 0, \\
 m_{11} - m_{22} + m_{12} - m_{21} &\geq 0, \\
 m_{11} + m_{22} + m_{33} + m_{44} &\geq 0, \\
 m_{11} + m_{22} - m_{33} - m_{44} &\geq 0, \\
 m_{11} - m_{22} + m_{33} - m_{44} &\geq 0, \\
 m_{11} - m_{22} - m_{33} + m_{44} &\geq 0.
 \end{aligned} \tag{14}$$

Equation (5) can be used to derive interrelations between the structures of a Jones matrix and the corresponding pure Mueller matrix. For example, the successive application of transposition and sign reversal for the off-diagonal elements of the Jones matrix in Eq. (6) yields

$$\begin{pmatrix} t_1 & -t_3 \\ -t_4 & t_2 \end{pmatrix} \leftrightarrow \begin{pmatrix} m_{11} & m_{21} & -m_{31} & m_{41} \\ m_{12} & m_{22} & -m_{32} & m_{42} \\ -m_{13} & -m_{23} & m_{33} & -m_{43} \\ m_{14} & m_{24} & -m_{34} & m_{44} \end{pmatrix}. \tag{15}$$

Physical reasons for the above relations are quite clear. Indeed, Eq. (15) originates from the operation of interchanging the incident and emerging light beams, the principle of reciprocity (Saxon 1955; Sekera 1966; Vansteenkiste et al. 1993; Potton 2004), and mirror symmetry (Hovenier 1969; 1970).

It is important to note that although analyses of the internal structure of a general pure Mueller matrix have historically been carried out in the framework of light scattering by a single particle, the results obtained can also be used in the more general context of properties of a pure Mueller matrix.

If the incident light is fully polarized and the output light is characterized by an inequality in Eq. (3) then the equalities for matrix elements obtained in Hovenier (1994) and Eq. (13), which determine the structure of the Mueller matrix as a deterministic matrix, are lost. In this case the output light is composed of several incoherent contributions, and the medium as a whole cannot be represented by a

Jones matrix. However, the medium can be considered as a parallel set of deterministic media, each one being described by a well-defined Jones matrix, in such a way that the light beam is shared among these different media. It is important to point out that the same result could be obtained by considering the medium as an ensemble (Kim et al. 1987) so that each realization i , characterized by a well-defined Jones matrix \mathbf{T}^i , occurs with a probability p^i .

If a Mueller matrix can be represented by a weighted sum of deterministic Mueller matrices (Cloude 1986; Simon 1987; Cloude and Pottier 1995; Gil 2007) then it is called a depolarizing Mueller matrix. It is important to note that this class of matrices does not coincide with the class of matrices, called Stokes matrices, satisfying the Stokes–Verdet criterion, i.e., matrices transforming Stokes vectors into Stokes vectors, see Eq. (3). Any physical Mueller matrix is a Stokes matrix, but the converse is not, in general, true (Gil 2007). On the other hand, no method has been quoted to physically realize a Stokes matrix that cannot be represented as a weighted sum of deterministic Mueller matrices.

Linear inequalities for the elements of a pure Mueller matrix are also valid for a depolarizing Mueller matrix \mathbf{M} (Hovenier and van der Mee 2000), in particular:

$$\begin{aligned} m_{11} &\geq 0; & m_{11} &\geq |m_{ij}|, \\ m_{11} + m_{22} + m_{12} + m_{21} &\geq 0, \\ m_{11} + m_{22} - m_{12} - m_{21} &\geq 0, \\ m_{11} - m_{22} + m_{12} - m_{21} &\geq 0, \\ m_{11} - m_{22} - m_{12} + m_{21} &\geq 0. \end{aligned} \tag{16}$$

In this case the following quadratic inequalities are also valid (Fry and Kattawar 1981):

$$\begin{aligned} (m_{11} + m_{12})^2 - (m_{21} + m_{22})^2 &\geq (m_{31} + m_{32})^2 + (m_{41} + m_{42})^2, \\ (m_{11} - m_{12})^2 - (m_{21} - m_{22})^2 &\geq (m_{31} - m_{32})^2 + (m_{41} - m_{42})^2, \\ (m_{11} + m_{21})^2 - (m_{12} + m_{22})^2 &\geq (m_{13} + m_{23})^2 + (m_{14} + m_{24})^2, \\ (m_{11} - m_{21})^2 - (m_{12} - m_{22})^2 &\geq (m_{13} - m_{23})^2 + (m_{14} - m_{24})^2, \\ (m_{11} + m_{22})^2 - (m_{12} + m_{21})^2 &\geq (m_{33} + m_{44})^2 + (m_{34} - m_{43})^2, \\ (m_{11} - m_{22})^2 - (m_{12} - m_{21})^2 &\geq (m_{33} - m_{44})^2 + (m_{34} + m_{43})^2, \end{aligned} \tag{17}$$

while Eq. (13) becomes an inequality as well:

$$\sum_{i=1}^4 \sum_{j=1}^4 m_{ij}^2 \leq 4m_{11}^2.$$

The study and characterization of depolarization is of considerable importance owing to the fact that depolarization phenomena are encountered in many theoretical and experimental applications of polarimetry to discrete random media and media with bulk and surface inhomogeneities. Note that the light–medium inter-

action with depolarization is heretofore studied in considerably less detail than the problem described by Mueller–Jones matrices discussed above.

Depolarization is the result of decorrelation of the phases and the amplitudes of the electric field vectors and/or selective absorption of polarization states (Brosseau 1998). Depolarization can be observed in both single and multiple light scattering and depends on geometrical and physical characteristics of the scattering particles: shape, morphology, refractive index, size parameter (ratio of the particle circumference to the wavelength of the incident light), and orientation with respect to the reference frame (Mishchenko and Travis 2000). Furthermore, multiple scattering results in depolarization of the output light even in the case of a collection of spherically symmetric particles and often reinforces depolarization caused by particle nonsphericity. Our purpose here is to consider the depolarization phenomenon using the Mueller-matrix formalism; in particular, we intend to discuss single-number depolarization metrics and Mueller matrices of depolarization. The notion of depolarization Mueller-matrix factorization has long existed in the literature and will be considered in Section 4.

Depolarization metrics provide a single scalar number that varies from zero, thereby corresponding to a totally depolarized output light, to a certain positive number corresponding to a totally polarized output light. All intermediate values are associated with partial polarization.

The depolarization index was introduced by Gil and Bernabeu (1985, 1986):

$$DI(\mathbf{M}) = \sqrt{\sum_{i,j=1}^4 m_{ij}^2 - m_{11}^2} / (\sqrt{3} m_{11}). \quad (18)$$

The depolarization index is bounded according to $0 \leq DI(\mathbf{M}) \leq 1$. The extreme values of $DI(\mathbf{M})$ correspond to the case of unpolarized and totally polarized output light, respectively.

An “analog” to the degree of polarization, Eq. (4), for linearly polarized input light in terms of Mueller matrix elements, the so-called index of linear polarization, was introduced by Bueno (2001):

$$G_L = \frac{\sqrt{3}}{2m_{11}} \left(m_{21}^2 + m_{31}^2 + \frac{1}{3} \sum_{i=1}^4 (m_{2i}^2 + m_{3i}^2) \right)^{1/2}. \quad (19)$$

It can be seen that G_L is the ratio of the mean of the sum of the squares of matrix elements corresponding to linear polarization of the output light and the value of the corresponding averaged intensity normalized by the maximum value of this ratio which occurs for a linear polarizer: $(G_L)_{\max} = 2/3^{1/2}$. The former implies the following range of variation: $0 \leq G_L \leq 1$.

The average degree of polarization was defined by Chipman (2005):

$$\text{Average DoP}(\mathbf{M}) = \frac{1}{4\pi} \int_0^\pi \int_{-\pi/2}^{\pi/2} p[\mathbf{MS}(\varepsilon, \zeta)] \cos \varepsilon d\varepsilon d\zeta. \quad (20)$$

The term $\cos \varepsilon d\varepsilon d\zeta$ scans the incident polarization state over the Poincaré sphere, with the latitude ε and longitude ζ . The Stokes vector $\mathbf{S}(\varepsilon, \zeta)$ is a function of ellipticity and orientation azimuth of the polarization ellipse of light:

$$\mathbf{S}(\varepsilon, \zeta) = [1 \quad \cos 2\varepsilon \cos 2\zeta \quad \cos 2\varepsilon \sin 2\zeta \quad \sin 2\varepsilon]^T, \quad (21)$$

where T stands for “transposed”.

The so-called $Q(\mathbf{M})$ metric is defined as follows (Espinosa-Luna and Bernabeu 2007):

$$Q(\mathbf{M}) = \frac{\sum_{i=2}^4 \sum_{j=1}^4 m_{ij}^2}{\sum_{j=1}^4 m_{1j}^2} = \{3[DI(\mathbf{M})]^2 - [D(\mathbf{M})]^2\} / \{1 + [D(\mathbf{M})]^2\}, \quad (22)$$

where $D(\mathbf{M}) = (m_{12}^2 + m_{13}^2 + m_{14}^2)^{1/2}$ is the diattenuation parameter and $0 \leq D(\mathbf{M}) \leq 1$. The metric $Q(\mathbf{M})$ is bounded according to $0 \leq Q(\mathbf{M}) \leq 3$. Specifically, $Q(\mathbf{M}) = 0$ corresponds to a totally depolarizing medium; $0 < Q(\mathbf{M}) < 1$ describes a partially depolarizing medium; $1 \leq Q(\mathbf{M}) < 3$ represents a partially depolarizing medium if, in addition, $0 < DI(\mathbf{M}) < 1$; otherwise, it represents a non-depolarizing diattenuating medium; finally, $Q(\mathbf{M}) = 3$ for a non-depolarizing non-diattenuating medium.

Thus, the depolarization metrics provide a summary of the depolarizing property of a medium via a single number. The depolarization index $DI(\mathbf{M})$ and the $Q(\mathbf{M})$ metrics are directly related to the Mueller matrix elements only and, in contrast to the average degree of polarization *Average DoP*, require no scan of the whole Poincaré sphere of the input polarizations. Furthermore, $Q(\mathbf{M})$ provides more detailed information about depolarization properties of a medium.

Quantities referring to the intrinsic depolarization properties of light have had wide applications in polarimetry as well. These are the linear, δ_L , and circular, δ_C , depolarization ratios defined according to (Mishchenko and Hovenier 1995; Mishchenko and Travis 2000)

$$\delta_L = (s_1 - s_2) / (s_1 + s_2), \quad (23)$$

$$\delta_C = (s_1 + s_4) / (s_1 - s_4). \quad (24)$$

The interest in these parameters is explained by the fact that they are susceptible to particle nonsphericity. Indeed, for spherical particles both ratios are equal to zero identically, whereas for nonspherical scatterers both δ_L and δ_C can substantially deviate from zero (Mishchenko and Hovenier 1995). The former means that if the incident light is linearly polarized then the backscattered light is completely linearly polarized in the same plane, whereas if the incident light is circularly polarized then the backscattered light is completely circularly polarized in the opposite sense. For nonspherical particles this is generally not the case.

Chipman (1995) introduced somewhat different versions of the degrees of linear and circular polarization:

$$DoLP = \sqrt{s_2^2 + s_3^2} / s_1, \quad (25)$$

$$DoCP = s_4/s_1. \quad (26)$$

These parameters turn out to be very useful for applications in meteorology, astronomy, ophthalmology, optical fibers, etc. (e.g., Bueno 2001, and references therein).

Some media depolarize all polarization states equally. Other depolarizing media partially depolarize most polarization states but may not depolarize one or two particular incident states. A single-number metric providing a summary of depolarization by a medium cannot give detailed information about all features of depolarization. Such information can only be obtained from Mueller-matrix models of depolarization. The case when for all polarizations of the input light the degree of polarization p of the output light is the same is called isotropic depolarization. When the degree of polarization of the output light is a function of parameters of the input polarization, one speaks of anisotropic depolarization.

There seems to be a consensus regarding the form of the Mueller matrix model describing isotropic depolarization (Brosseau 1998; Chipman 1999):

$$\text{diag}[1 \quad p \quad p \quad p]. \quad (27)$$

It can be seen that the properties of this type of depolarization are the following:

- (i) the transmittance is the same for all polarizations of the incident light;
- (ii) p of the output light is the same for all input polarizations.

On the other hand, there is no agreement in the literature concerning the Mueller matrix for the case of anisotropic depolarization. Apparently, one of the most accepted forms of the Mueller matrix describing the dependence of p of the output light on the incident polarization is the following (Shindo 1995; Brosseau 1998; Chipman 1999; Ossikovski 2009):

$$\text{diag}[1 \quad a \quad b \quad c]. \quad (28)$$

The elements a , b , and c are interpreted physically in the following manner: a and b are the degrees of linear depolarization, while c is the degree of circular depolarization. If $a = b = c = 0$ then the Mueller matrix represents an ideal depolarizer. Bicout et al. (1994) discussed the depolarization arising in multiple scattering of light by spherical scatterers in the Rayleigh regime and given by Eq. (28) with $a = b$.

The most general expression for the Mueller matrix describing depolarization was suggested by Lu and Chipman (1996):

$$\begin{pmatrix} 1 & \vec{\mathbf{O}}^T \\ \vec{\mathbf{P}}_\Delta & \mathbf{m}_\Delta \end{pmatrix}, \quad \mathbf{m}_\Delta^T = \mathbf{m}_\Delta, \quad (29)$$

where $\vec{\mathbf{P}}_\Delta$ denotes the so-called polarizance vector. The polarizance vector describes the state of polarization generated by this Mueller matrix from unpolarized incident light. The Mueller matrix of Eq. (29) has nine degrees of freedom, and this situation is of interest because this matrix along with a generalized determi-

nistic Mueller matrix are jointly characterized by 16 degrees of freedom. This means that in this way one obtains the generalized Mueller matrix of an arbitrary medium that has 16 degrees of freedom and linearly interacts with polarized light.

3. Polarization properties of tissues: the case of a discrete ensemble of scatterers

In this section we consider the structures and information content of Mueller matrices in the framework of the model according to which a tissue is a discrete ensemble of scatterers. The effect of the symmetry of the individual scatterers and collections of scatterers on the structure (number of independent parameters) of the Mueller matrix has been considered by van de Hulst (1957). In particular, he demonstrated the following structural features of the Mueller matrices:

(i) The collection contains one kind of particles, and for each particle in one position there is a particle in the reciprocal position:

$$\begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{12} & m_{22} & m_{23} & m_{24} \\ -m_{13} & -m_{23} & m_{33} & m_{34} \\ m_{14} & m_{24} & -m_{34} & m_{44} \end{pmatrix} \quad (10 \text{ parameters}) \quad (30)$$

with $m_{11} - m_{22} + m_{33} - m_{44} = 0$.

(ii) The collection contains equal numbers of particles and their mirror particles:

$$\begin{pmatrix} m_{11} & m_{12} & 0 & 0 \\ m_{21} & m_{22} & 0 & 0 \\ 0 & 0 & m_{33} & m_{34} \\ 0 & 0 & m_{43} & m_{44} \end{pmatrix} \quad (8 \text{ parameters}). \quad (31)$$

If, in addition, $t_3 = t_4$ then $m_{11}^2 - m_{12}^2 - m_{33}^2 - m_{34}^2 = 0$, $m_{22} = m_{11}$, $m_{33} = m_{44}$, and $m_{34} = -m_{43}$.

The Mueller matrix of Eq. (31) plays a key role in many light-scattering applications. Some of them will be discussed later in this section. The structure of Eq. (31) can be caused by a symmetry of individual particles and a collection of particles in single and multiple scattering (van de Hulst 1957; Mishchenko and Travis 2000) as well as by illumination-observation geometries for backward (Zubko et al. 2004) and forward (Savenkov et al. 2007a) scattering.

The model of a medium described by the Mueller matrix of Eq. (31) has been used in studies of optical characteristics of oceanic water (Voss and Fry 1984; Kokhanovsky 2003a); ensembles of identical, but randomly oriented fractal particles (Kokhanovsky 2003c); dense spherical particle suspensions in the multiple-scattering regime (Kaplan et al. 2001); ice clouds consisting of nonspherical ice crystals in the multiple-scattering regime (Lawless et al. 2006); polydisperse, randomly oriented ice crystals modeled by finite circular cylinders with different size

distributions (Xu et al. 2002); cylindrically shaped radially inhomogeneous particles (Manickavasagam and Mengüç 1998); and small spherical particles (ranging in diameter from 0.2 to 1.5 μm) sparsely seeded on the surface of a crystalline silicon c-Si wafer (Kaplan and Drevillon 2002). Other applications included measurements of the complex refractive index of isotropic materials as matrices of isotropic and ideal metal mirror reflections (Deibler and Smith 2001); the development of a symmetric three-term product decomposition of a Mueller–Jones matrix (Ossikovski 2008); and the description of very general and practically important cases of (i) randomly oriented particles with a plane of symmetry (Hovenier and van der Mee 2000) and/or (ii) equal numbers of particles and their mirror particles (Mishchenko et al. 2002). This list of applications can be extended significantly.

An example of the situation in which the Mueller matrix has the structure of Eq. (31) and contains information on the strong dependence of depolarization and depolarization on the polarization state of the input light is the exact forward scattering of polarized light by a slab of inhomogeneous linear birefringent medium (Savenkov et al. 2007a).

If for any particle described by Eq. (6) there is a particle in the position corresponding to the transposition of the off-diagonal elements of the Jones matrix in Eq. (6) then the resulting Mueller matrix has the form

$$\begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{12} & m_{22} & m_{23} & m_{24} \\ m_{13} & m_{23} & m_{33} & m_{34} \\ -m_{14} & -m_{24} & -m_{34} & m_{44} \end{pmatrix} \quad (10 \text{ parameters}). \quad (32)$$

(iii) The collection contains equal numbers of particles in positions described by Eqs. (6) and (15) and those corresponding to the transposition and sign reversal of the off-diagonal elements of the Jones matrix (6); then the resulting Mueller matrix reads:

$$\begin{pmatrix} m_{11} & m_{12} & 0 & 0 \\ m_{12} & m_{22} & 0 & 0 \\ 0 & 0 & m_{33} & m_{34} \\ 0 & 0 & -m_{34} & m_{44} \end{pmatrix} \quad (6 \text{ parameters}). \quad (33)$$

The scattering angles 0° (exact forward direction) and 180° (exact backward direction) deserve special attention owing to their importance in numerous practical applications, including the scattering by biological tissues. For the first time the general form of Mueller matrices for these scattering angles was derived by van de Hulst (1957). Hu et al. (1987) presented a comprehensive study of forward and backward scattering by an individual particle in a fixed orientation. For forward scattering, they distinguished sixteen different symmetry shapes which were classified into five symmetry classes; for backward scattering, four different symmetry shapes were identified and classified into two symmetry classes. A large number of relations were derived in this way. Below we present the struc-

tures of Mueller matrices for various collections of particles in the cases of forward and backward scattering (van de Hulst 1957; Hovenier and Mackowski 1998).

For the *forward scattering* we have the following scenarios:

(i) There is one kind of asymmetric particles in the collection, and the assumption of rotational symmetry is made. Then

$$\begin{pmatrix} m_{11} & 0 & 0 & m_{14} \\ 0 & m_{22} & m_{23} & 0 \\ 0 & -m_{23} & m_{22} & 0 \\ m_{41} & 0 & 0 & m_{44} \end{pmatrix} \text{ (6 parameters).} \quad (34)$$

(ii) There is one kind of asymmetric particles. The assumption of rotational symmetry is combined with the assumption that equal numbers of particles occur in reciprocal positions. Then

$$\begin{pmatrix} m_{11} & 0 & 0 & m_{14} \\ 0 & m_{22} & m_{23} & 0 \\ 0 & -m_{23} & m_{22} & 0 \\ m_{14} & 0 & 0 & m_{44} \end{pmatrix} \text{ (5 parameters).} \quad (35)$$

(iii) Particles and their mirror counterparts occur in equal numbers. The assumption of rotational symmetry is combined with the assumption that any particle is accompanied by its mirror particle with respect to any plane through the axis. Then

$$\begin{pmatrix} m_{11} & 0 & 0 & 0 \\ 0 & m_{22} & 0 & 0 \\ 0 & 0 & m_{22} & 0 \\ 0 & 0 & 0 & m_{44} \end{pmatrix} \text{ (3 parameters)} \quad (36)$$

with $m_{44} \geq 2m_{22} - m_{11}$. In the special case of homogeneous, optically inactive, spherical particles, $m_{11} = m_{22} = m_{44}$ (Hovenier and Mackowski 1998).

(iv) Particles and mirror particles occur in equal numbers. The assumption of rotational symmetry is combined with the assumption of mirror symmetry with respect to a plane perpendicular to the axis. Then

$$\begin{pmatrix} m_{11} & 0 & 0 & m_{14} \\ 0 & m_{22} & 0 & 0 \\ 0 & 0 & m_{22} & 0 \\ -m_{14} & 0 & 0 & m_{44} \end{pmatrix} \text{ (4 parameters).} \quad (37)$$

For the *backward scattering* we consider the following scenarios:

(i) No assumption about the kind of particles is made:

$$\begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{12} & m_{22} & m_{23} & m_{24} \\ -m_{13} & -m_{23} & m_{33} & m_{34} \\ m_{14} & m_{24} & -m_{34} & m_{44} \end{pmatrix} \quad (10 \text{ parameters}). \quad (38)$$

(ii) There is one kind of asymmetric particles in the collection, and the assumption of rotational symmetry is made. Then

$$\begin{pmatrix} m_{11} & 0 & 0 & m_{14} \\ 0 & m_{22} & 0 & 0 \\ 0 & 0 & -m_{22} & 0 \\ m_{14} & 0 & 0 & m_{44} \end{pmatrix} \quad (4 \text{ parameters}) \quad (39)$$

with $m_{11} \geq m_{22} \geq 0$; $m_{44} = m_{11} - 2m_{22}$, and $m_{22} - m_{11} \leq m_{44} \leq m_{11} - m_{22}$.

(iii) The collection contains particles and their mirror counterparts in equal numbers. The assumption of rotational symmetry is combined with the assumption that any particle is accompanied by its mirror particle with respect to any plane through the axis. Then

$$\begin{pmatrix} m_{11} & 0 & 0 & 0 \\ 0 & m_{22} & 0 & 0 \\ 0 & 0 & -m_{22} & 0 \\ 0 & 0 & 0 & m_{44} \end{pmatrix} \quad (3 \text{ parameters}) \quad (40)$$

In the special case of homogeneous, optically inactive, spherical particles, one has $m_{11} = m_{22} = -m_{44}$ (Hovenier and Mackowski 1998).

Multiple scattering can significantly effect the process of polarization (the occurrence of at least one non-zero Stokes parameter other than the total intensity for an unpolarized input Stokes vector) and depolarization. In particular, polarization can be considerably diminished by multiple scattering. At the same time, multiple scattering causes non-zero depolarization values even for spherically-symmetric particles and often reinforces depolarization caused by particle nonsphericity (Mishchenko and Travis 2000; Mishchenko et al. 2006).

Depolarization depends significantly on the polarization state of the input light in the multiple-scattering regime as well (Bicout et al. 1994; Rojas-Ochoa et al. 2004; Kim et al. 2006, and references therein). In particular, Bicout et al. (1994) studied numerically and experimentally how depolarization evolves for linear and circular input polarizations as the size of the particles increases from very small (Rayleigh regime) to large (Mie regime) in the case of a forward scattering geometry.

Based on the above results, the following features can be pointed out which allow one to classify certain parameters of scatterers, e.g., tissue or other biomedical structural components, by interpreting their Mueller matrices.

The ratio $-m_{21}/m_{11}$ is called the degree of linear polarization for unpolarized incident light (Mishchenko and Travis 2000; Kokhanovsky 2003c). It is often

negative for most scattering angles. This means that the vibrations of the electric field vector occur predominantly in the plane perpendicular to the scattering plane. For single spheres, the ratio $-m_{21}/m_{11}$ at scattering angles 0 and π is identically equal to zero. Mishchenko et al. (2002) pointed out that the most prominent feature of nonspherical scattering appears to be the bridge of positive linear polarization at scattering angles near 120° .

The ratio m_{22}/m_{11} is equal to unity for spherical particles at any scattering angle (Mishchenko et al. 2002). Therefore, the deviation of m_{22}/m_{11} from unity can be used as an indicator of particle nonsphericity sensitive, at least in some cases, to the particle size and aspect ratio (Mishchenko and Hovenier 1995; Quinby-Hunt et al. 2000). The angular features of this ratio have been studied for different kinds of pollen (Bickel et al. 1976) and marine organisms (Voss and Fry 1984; Lofftus et al. 1992). However, similar features can be caused by multiple scattering in turbid collections of spherical particles (e.g., Mishchenko et al. 2006).

He et al. (2010) showed that characteristic scattering features of the skeletal muscle (bovine *Sternomandibularis*) can be described by sphere–cylinder scattering because the element m_{22} has an asymmetry in the image scattering pattern characteristic of the scattering by cylinders. The Mueller-matrix element patterns of the skeletal muscle found by He et al. are similar to those reported in other studies (Ranasinghesagara and Yao 2007; Li et al. 2008).

The ratio m_{34}/m_{11} describes the transformation of the linearly polarized incident light into the circularly polarized output light. The m_{34}/m_{11} decreases with the particle refractive index. Bickel et al. (1976) and Bickel and Stafford (1980) have found a high specificity of the normalized element m_{34}/m_{11} for every type of biological scatterers. Strong distinctions are revealed in the values of m_{34}/m_{11} for spores of two mutant varieties of bacteria, which are distinguished by variations in their structures not detectable by traditional scattering techniques. The sensitivity of other matrix elements to these types of scatterers appears to be significantly weaker.

If scattering is well described by the Rayleigh–Gans approximation then $m_{34} = 0$. Thus, a non-zero value for this element can be related to a deviation of the particle characteristics from those satisfying the Rayleigh–Gans scattering regime (Bohren and Huffman 1983).

As noted in Bickel et al. (1976), Bickel and Stafford (1980), Hoekstra and Sloot (1993, 2002), Bronk et al. (1995), and Van de Merwe et al. (1989, 1997), the element m_{34} is sensitive to properties of various biological microorganisms, in particular, to small morphological alterations in scatterers. It has been shown that m_{34} is affected by a small surface roughness of a sphere (Hoekstra and Sloot 1993). It has also been demonstrated that measurements of m_{34} can be a basis for determining the diameters of rod-shaped bacteria (*Escherichiacoli* cells) which are difficult to characterize using other techniques (Bronk et al. 1995). The angular dependences of the normalized element m_{34}/m_{11} for different bacteria turn out to be oscillating functions whose maxima positions are very sensitive to varying

sizes of the bacteria (Bronk et al. 1995; Van de Merwe et al. 1989, 1997) thus allowing bacterial growth to be traced. Mueller matrix measurements have been used to examine the formation of liposome complexes with plague capsular antigens (Tuchin 1993; Tuchin et al. 2006) and various particle suspensions, e.g., those of spermatozoid spiral heads (Johnston 1988; Salzman et al. 1990).

The ratio m_{33}/m_{11} describes the reduction of the degree of circular polarization for circularly polarized incident light. Like for m_{22}/m_{11} , the behavior of the element ratios m_{33}/m_{11} and m_{44}/m_{11} also represents asymmetric particle shapes, i.e., the deviation of m_{44}/m_{11} from m_{33}/m_{11} indicates a nonspherical morphology of particles (Mishchenko et al. 2002; Xu et al. 2002). The polarization characteristics of suspensions of biological particles have been described by Lopatin et al. (2004) who analyzed the sensitivity of different matrix elements to variations in scatterer shapes and sizes. It was noted that the values of the elements m_{33} and m_{44} in the backward scattering direction may serve as indicators of biological particle nonsphericity.

By experimentally measuring backscattering Mueller-matrix patterns, it was demonstrated in Antonelli et al. (2010) that for the tumoral and healthy parts of a sample of *ex vivo* human colon tissue the absolute value of m_{22} and m_{33} was larger for the diseased tissue than in healthy regions. The same difference between cancerous and healthy regions was observed in the m_{44} element. For both kinds of tissue, the following inequality holds:

$$|m_{22}| = |m_{33}| > |m_{44}|. \quad (41)$$

Note that this trend seems to be quite general. Indeed, it has been observed for healthy and cancerous cell suspensions by Hielscher et al. (1997) and Sankaran et al. (2002) for a variety of tissues (fat, tendon, arterial wall, myocardium, blood) in transmission. Only whole blood displayed the opposite trend,

$$|m_{22}| = |m_{33}| < |m_{44}|, \quad (42)$$

with lower depolarization for circularly polarized incident light.

Many tissues are characterized by optical activity manifesting itself in circular dichroism and circular birefringence. The optical activity of tissues may be caused by the optical activity of the substances they are formed from and by their structural features. Circular intensity differential scattering is the difference between scattered intensities for left- and right-hand circularly polarized incident light. Circular intensity differential scattering effects can be studied by measuring the matrix element m_{14} (Bohren and Huffman 1983). The so-called “form circular intensity differential scattering” is an anisotropy resulting from the helical structure of a scatterer (Hoekstra and Slood 2002). Dorman and Maestre (1973) were likely the first to point out that the matrix element m_{14} could be useful in studies of helical structures. Indeed, Maestre et al. (1982) reported large m_{14} values for octopus sperm heads of the species *Eledone cirrhosa*. In Shapiro et al. (1994a), the Mueller matrix elements for a model of DNA plectonemic helix were calculated using

the coupled dipole approximation. The calculations of the Mueller matrix elements were carried out for two and four turn helices versus scattering angle. It was shown that the elements m_{12} and m_{14} exhibit the largest change with respect to the number and position of nodes, maxima, and minima for the calculation performed at $\lambda = 20$ nm and allow one to discriminate between the cases of two and four turn helices. In Shapiro et al. (1994b), the matrix element m_{14} was used for the determination of the average DNA orientation of this scatterer.

In Maksimova et al. (1992), it was shown that for small particles the effects of multiple scattering become evident as a broken symmetry between the Mueller matrix elements, Eq. (33), i.e., $m_{12} \neq m_{21}$, $m_{33} \neq m_{44}$, and as a significant reduction of linear polarization of the light scattered at angles close to $\pi/2$.

4. Polarization properties of tissues: the case of a continuous distribution of optical parameters

In this section, we consider the problem of Mueller matrix interpretation in the framework of the approach wherein the tissue is modeled as a medium with a continuous (and possibly random) distribution of optical parameters. It is important to note that the results for the internal structure of a pure Mueller matrix, the symmetry relations caused by interchanging the incident and emerging light beams, and the principle of reciprocity presented earlier and obtained in the framework of light scattering by discrete particles are also relevant to pure Mueller matrices in the continuous-medium approximation.

The polarization of light changes if the amplitudes and phases of the components of the electric field vector \mathbf{E} change separately or simultaneously (Shurcliff 1962; Azzam and Bashara 1977; Brosseau 1998). It is, therefore, customary to distinguish between the corresponding classes of anisotropic media: dichroic (or possessing amplitude anisotropy), influencing only the amplitudes; birefringent (or possessing phase anisotropy), influencing only the phases; and “all other” (possessing both amplitude and phase anisotropy) affecting both the amplitudes and the phases of the components of the electric field vector. Among these classes, four types of anisotropic mechanisms are recognized as basic or, after Jones, elementary (Jones 1941, 1942, 1947, 1956; Hurwitz and Jones 1941): linear and circular phase and linear and circular amplitude anisotropies. Biological tissues are often optically anisotropic (Hunter et al. 1999; Tuchin et al. 2006, and references therein; Benoit et al. 2001; Naoun et al. 2005; Chung et al. 2007; Twietmeyer et al. 2008; Li and Yao 2009; Fanjul-Vélez and Arce-Diego 2010) and, in principle, can exhibit all these elementary types of anisotropy.

Tissues are primarily characterized by the linear birefringence caused by fibrous structures, which are common constituents of many connective tissues. A large variety of tissues, such as eye cornea, tendon, cartilage, eye sclera, dura mater, testis, muscle, nerve, retina, bone, arteries, myocardium, myelin, etc., exhibit birefringence. All of these tissues contain uniaxial and/or biaxial birefringent structures. The refractive index of the medium is greater along the length of the fi-

bers than across the width. Reported birefringence values for tendon, muscle, coronary artery, myocardium, sclera, cartilage, and skin are on the order of 10^{-3} (Maitland et al. 1997; de Boer et al. 1997, 2002; Everett et al. 1998; Simonenko et al. 2000). Linear birefringence is described by the following pure Mueller matrix:

$$\mathbf{M}^{LP} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos^2 2\alpha + \sin^2 2\alpha \cos \Delta & \cos 2\alpha \sin 2\alpha (1 - \cos \Delta) & -\sin 2\alpha \sin \Delta \\ 0 & \cos 2\alpha \sin 2\alpha (1 - \cos \Delta) & \sin^2 2\alpha + \cos^2 2\alpha \cos \Delta & \cos 2\alpha \sin \Delta \\ 0 & \sin 2\alpha \sin \Delta & -\cos 2\alpha \sin \Delta & \cos \Delta \end{pmatrix} \quad (43)$$

where Δ is the phase shift between two orthogonal linear components of the electric field vector and α is the azimuth of the anisotropy.

Linear dichroism (diattenuation), i.e., differential wave attenuation for two orthogonal polarizations, in systems formed by long cylinders or plates is defined by the difference between the imaginary parts of the effective refractive indices. For example, the linear dichroism of the retinal nerve's fiber layer gives rise to partial polarization of an unpolarized laser beam that has passed the retinal fiber layer twice (Benoit et al. 2001; Louis-Dorr et al. 2004; Naoun et al. 2005). Naoun et al. (2005) reported significant differences in dichroism between normal and glaucomatous eyes. The magnitudes of birefringence and diattenuation are related to the density and other properties of the collagen fibers, whereas the orientation of the fast axis indicates the orientation of the collagen fibers. The pure Mueller matrix describing linear dichroism is as follows:

$$\mathbf{M}^{LA} = \begin{pmatrix} 1+P & (1-P)\cos 2\gamma & & & & \\ (1-P)\cos 2\gamma & \cos^2 2\gamma(1+P) + 2\sin^2 2\gamma\sqrt{P} & & & & \\ (1-P)\sin 2\gamma & \cos 2\gamma \sin 2\gamma(1-\sqrt{P})^2 & & & & \\ 0 & 0 & & & & \\ & & (1-P)\sin 2\gamma & 0 & & \\ & & \cos 2\gamma \sin 2\gamma(1-\sqrt{P})^2 & 0 & & \\ & & \sin^2 2\gamma(1+P) + 2\cos^2 2\gamma\sqrt{P} & 0 & & \\ & & 0 & 2\sqrt{P} & & \end{pmatrix}, \quad (44)$$

where P is the relative absorption of two linear orthogonal components of the electric vector and γ is the azimuth of the anisotropy.

In addition to linear birefringence and dichroism, many tissue components exhibit optical activity. There are a number of effects generically called optical activity resulting from the molecule's chirality that stems from its asymmetric structure (Applequist 1987; Barron 1982; Studinski and Vitkin 2000; Hadley and Vitkin 2002; Manhas et al. 2006). A well-known manifestation of optical activity is the ability to rotate the plane of linearly polarized light about the axis of propagation. The angle of rotation depends on the chiral molecular concentration, the pathlength through the medium, and the wavelength of light. Interest in optically

active turbid media is caused by the possibility of noninvasive *in situ* optical monitoring of the glucose in diabetic patients. The Mueller matrix describing circular birefringence is as follows:

$$\mathbf{M}^{CP} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\varphi & \sin 2\varphi & 0 \\ 0 & -\sin 2\varphi & \cos 2\varphi & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (45)$$

where φ is the induced phase shift between two orthogonal circular components of the electric vector.

Many tissues demonstrate effects of optical activity resulting in circular dichroism. An example is the difference between the scattered intensities for left- and right-handed circularly polarized incident light (the so-called circular intensity differential scattering) resulting from the helical structure of scatterers (Bustamante et al. 1984; Zietz et al. 1983; Gross et al. 1991). In terms of the Mueller-matrix calculus, this type of anisotropy is described by the following matrix:

$$\mathbf{M}^{CA} = \begin{pmatrix} 1+R^2 & 0 & 0 & 2R \\ 0 & 1-R^2 & 0 & 0 \\ 0 & 0 & 1-R^2 & 0 \\ 2R & 0 & 0 & 1+R^2 \end{pmatrix}, \quad (46)$$

where R is the magnitude of anisotropy, i.e., the relative absorption of two orthogonal circular components of the electric vector. The six quantities α , Δ , P , γ , φ , and R are called anisotropy parameters.

It can be seen that the matrices describing linear and circular birefringence belong to the class of unitary matrices (in the case of matrices with real-valued elements – orthogonal matrices). The matrices of linear, Eq. (44), and, circular Eq. (46), dichroism belong to the class of Hermitian matrices (in the case of matrices with real-valued elements – symmetric matrices).

The Mueller matrices of Eqs. (43)–(46) represent tissues exhibiting individual types of anisotropy. Experimental measurements of these matrices or of the corresponding informative matrix elements allow one to interpret and characterize anisotropy properties of biological tissues. However, more often two or more types of anisotropy are exhibited by a tissue simultaneously. Evidently, such cases require the development of more sophisticated polarimetric matrix models (Hurwitz and Jones 1941; Cloude 1986; Gil and Bernabeu 1987; Lu and Chipman 1996; Savenkov et al. 2006; Ossikovski 2008, 2009).

Note that the interpretation of the individual contributions of several types of anisotropy occurring simultaneously is made considerably more complicated by multiple scattering in optically thick turbid media (such as most biological tissues), resulting in strong depolarization. Depolarization is controlled by a large number of tissue parameters such as the concentration, size, shape, and refractive index of the scatterers, the detection geometry, and the incident light's state of po-

larization (Sankaran et al. 2002; Ghosh et al. 2003; Bicout et al 1994)

The matrix model that is used most extensively in polarimetry for decoupling constituent polarization properties of biological tissues is the generalized polar decomposition proposed by Lu and Chipman (1996). This model is based on the so-called polar decomposition theorem (Lancaster and Tismenetsky 1985), according to which an arbitrary matrix \mathbf{M} can be represented by a product

$$\mathbf{M} = \mathbf{M}_p \mathbf{M}_R \quad \text{or} \quad \mathbf{M} = \mathbf{M}_R \mathbf{M}'_p, \quad (47)$$

where \mathbf{M}_p and \mathbf{M}'_p are Hermitian matrices and \mathbf{M}_R is a unitary one. The Hermitian matrix is associated with amplitude anisotropy, while the unitary matrix describes phase anisotropy (Whitney 1971). The matrices \mathbf{M}_p and \mathbf{M}_R are called the dichroic and the phase polar form (Whitney 1971; Gil and Bernabeu 1987; Lu and Chipman 1996).

The polar decomposition was first employed by Whitney (1971) without finding explicit expressions for \mathbf{M}_p and \mathbf{M}_R . They were proposed later, independently by Gil and Bernabeu (1987) and Lu and Chipman (1996). Alternatively, the dichroic and phase polar forms can be derived using spectral methods of linear algebra (Azzam and Bashara 1977).

The phase polar form \mathbf{M}_R (using notation from Lu and Chipman 1996) is given by

$$\mathbf{M}_R = \begin{pmatrix} 1 & \vec{0}^T \\ \vec{0} & \mathbf{m}_R \end{pmatrix}, \quad (48)$$

$$(m_R)_{ij} = \delta_{ij} \cos R + a_i a_j (1 - \cos R) + \sum_{k=1}^3 \varepsilon_{ijk} a_k \sin R,$$

where $\vec{0}$ is the 3×1 zero vector; $[1 \ a_1 \ a_2 \ a_3]^T = [1 \ \vec{R}^T]^T$ is the normalized Stokes vector for the fast axis of \mathbf{M}_R ; δ_{ij} is the Kronecker delta; ε_{ijk} is the Levi-Civita permutation symbol, \mathbf{m}_R is the 3×3 submatrix of \mathbf{M}_R obtained by striking out the first row and the first column of \mathbf{M}_R ; and R is the birefringence given by

$$R = \arccos \left(\frac{1}{2} \text{Tr} \mathbf{M}_R - 1 \right). \quad (49)$$

The dichroic polar form \mathbf{M}_p is as follows:

$$\mathbf{M}_p = T_u \begin{pmatrix} 1 & \vec{\mathbf{D}}^T \\ \vec{\mathbf{D}} & \mathbf{m}_p \end{pmatrix}, \quad (50)$$

$$\mathbf{m}_p = \sqrt{1 - D^2} \mathbf{I} + \left(1 - \sqrt{1 - D^2} \right) \hat{\mathbf{D}} \hat{\mathbf{D}},$$

where \mathbf{I} is the 3×3 identity matrix; $\hat{\mathbf{D}} = \vec{\mathbf{D}} / |\vec{\mathbf{D}}|$ is the unit vector in the direction of the diattenuation vector $\vec{\mathbf{D}}$; T_u is the transmittance for unpolarized light; and the value of diattenuation can be obtained as

$$D = \{1 - 4|\det(\mathbf{T})|^2 / [\text{Tr}(\mathbf{T}^* \mathbf{T})]^2\}^{1/2}. \quad (51)$$

The models of anisotropic media based on the polar decomposition contain six independent parameters, three for the phase polar form \mathbf{M}_R and three for the dichroic polar form \mathbf{M}_P . It can be seen that the phase polar form is a unitary (orthogonal) matrix and the dichroic polar form is a Hermitian (symmetric) matrix. Note that unitarity (orthogonality) of the phase polar form, Eq. (48), is in complete agreement with the first Jones' equivalence theorem (Hurwitz and Jones 1941), and is a general model of elliptically birefringent media. The situation with the dichroic polar form is more complex (Savenkov et al. 2005, 2007). Mathematically, the complexity originates from the fact that, in contrast to unitary matrices, the product of Hermitian matrices is generally not a Hermitian matrix (Lancaster and Tismenetsky 1985).

The product of Mueller matrices of the polar forms (48) and (50) and the depolarizing Mueller matrix (29),

$$\mathbf{M} = \mathbf{M}_\Delta \mathbf{M}_R \mathbf{M}_P, \quad (52)$$

is the generalized polar decomposition and a multiplicative matrix model of an arbitrary Mueller matrix (Lu and Chipman 1996; Gil 2000, 2007).

The product of the phase polar form and the depolarizing matrices can then be obtained as

$$\mathbf{M}_\Delta \mathbf{M}_R = \mathbf{M}' = \mathbf{M} \mathbf{M}_P^{-1}. \quad (53)$$

Then

$$\tilde{\mathbf{P}}_\Delta = (\tilde{\mathbf{P}} - \mathbf{m}\tilde{\mathbf{D}})/(1 - D^2), \quad (54)$$

where $\tilde{\mathbf{P}} = (1/m_{11})[m_{21} \ m_{31} \ m_{41}]^T$ and \mathbf{m} is the submatrix of the initial matrix \mathbf{M} . The \mathbf{m}' is the submatrix of \mathbf{M}' and can be written as

$$\mathbf{m}' = \mathbf{m}_\Delta \mathbf{m}_R. \quad (55)$$

The submatrix \mathbf{m}_Δ can be calculated as follows:

$$\begin{aligned} \mathbf{m}_\Delta = \pm & [\mathbf{m}'(\mathbf{m}')^T + (\sqrt{\lambda_1\lambda_2} + \sqrt{\lambda_2\lambda_3} + \sqrt{\lambda_1\lambda_3})\mathbf{I}]^{-1} \\ & \times [(\sqrt{\lambda_1} + \sqrt{\lambda_2} + \sqrt{\lambda_3})\mathbf{m}'(\mathbf{m}')^T + \sqrt{\lambda_1\lambda_2\lambda_3}\mathbf{I}], \end{aligned} \quad (56)$$

where λ_i are the eigenvalues of $\mathbf{m}'(\mathbf{m}')^T$. The sign “+” or “-” is determined by the sign of the determinant of \mathbf{m}' . The net depolarization coefficient Δ can be calculated according to

$$\Delta = 1 - \frac{1}{3} |\text{Tr}(\mathbf{M}_\Delta) - 1|. \quad (57)$$

Using the polar decomposition model, Manhas et al. (2006) showed that the change in the orientation angle of the polarization vector of light propagating through a chiral turbid medium arises not only due to the circular birefringence

property of the medium but also is caused by linear diattenuation and linear birefringence of light scattered at large angles. Measurements were carried out on chiral turbid samples prepared using known concentrations of scatterers and glucose molecules. This research can potentially facilitate the determination of the concentration of chiral substances present in a turbid medium using the measured Mueller matrix.

Structurally each muscle fiber consists of many myofibrils, which appear striated due to the periodic sarcomere structure (Tuchin et al. 2006). Sarcomeres are the fundamental functional unit in each muscle fiber. Using the bovine *Sternomandibularis* muscle as an example, Li and Yao (2009) applied the polar decomposition to study the effect of sarcomere lengths, which is closely related to force generation in skeletal muscles, by comparing polarization images of muscle samples in their original states and after stretching along the muscle fibers.

Many forms of precancerous changes in tissues are difficult to detect using conventional techniques, which require a histological examination of biopsies obtained from visible lesions or random surveillance biopsies. On the other hand, it is known that cancerous tissues depolarize light less than their surrounding tissues, whereas noncancerous tissues have the same depolarizing properties as the surrounding tissue (Smith 2001). This effect is caused by changes in the subsurface structures of cancerous tissues that prevent light from penetrating inside the tissue as deeply as it would in the case of a normal tissue. These facts are indicative of the likely potential of the polarimetric approach to noninvasive cancer detection.

This potential was demonstrated by Chung et al. (2007) based on the polar decomposition of *in vivo* Mueller matrix image patterns of nine cheek pouches in female golden Syrian hamsters (*Mesocricetus auratus*). The histological features in this model have been shown to correspond closely with those of premalignancy and malignancy in human oral mucosa. Figures 1 and 2 show *in vivo* depolarization and birefringence images for normal and precancerous tissues, respectively. These images were taken from three different hamster cheek pouches, and the imaged fields were 2.25×2.25 mm.

Thus, depolarization and birefringence images can be used to identify quantitatively the dysplastic region of a tissue. Apparently, this suggests that depolarization and birefringence results will facilitate earlier and more sensitive diagnosis of precancerous changes and to improve the monitoring of disease progression, the identification of region boundaries, and the assessment of cancer response to therapy. Ghosh et al. (2009a) used the polar decomposition methodology for monitoring regenerative treatments of myocardial infarction. To this end, the Mueller matrices were measured in the transmission mode for 1 mm thick *ex vivo* myocardial samples from Lewis rats harvested after myocardial infarction, both with and without the stem-cell treatment. The measured Mueller matrices were analyzed with the polar decomposition method to obtain the values of linear birefringence. The results are shown in Fig. 3. It can be seen that significant differences in the derived linear birefringence values exist between normal and infarcted regions as well as between infarcted regions with and without the stem-cell treatment. An in-

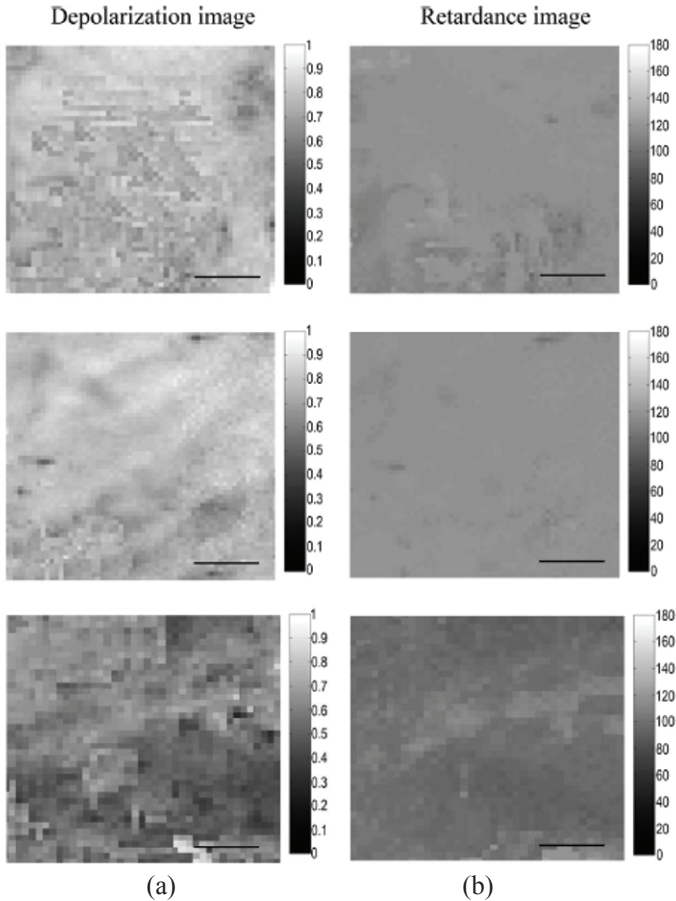


Fig. 1. (a) Depolarization images and (b) birefringence images of the m_{33} element of the Mueller matrix for healthy tissues. The scale bars are 0.5 mm (after Chung et al. 2007).

crease in birefringence in the infarcted regions of the treated hearts indicates reorganization and re-growth, as confirmed by a histologic examination.

Several important problems related to the validation of the generalized polar decomposition for the quantification and interpretation of all the intrinsic polarization parameters of a complex turbid medium have been analyzed in Ghosh et al. (2008, 2009b, 2010) and Wood et al. (2009). The gist of this issue is that the generalized polar decomposition, Eq. (52), represents the initial Mueller matrix as an ordered product of the Mueller matrices of the constituent polarization effects, thus implying that they act in succession. However, in actual tissues, all the polarization effects are exhibited simultaneously and not in an ordered sequential manner. To realize the potential of the polar decomposition approach for real tis-

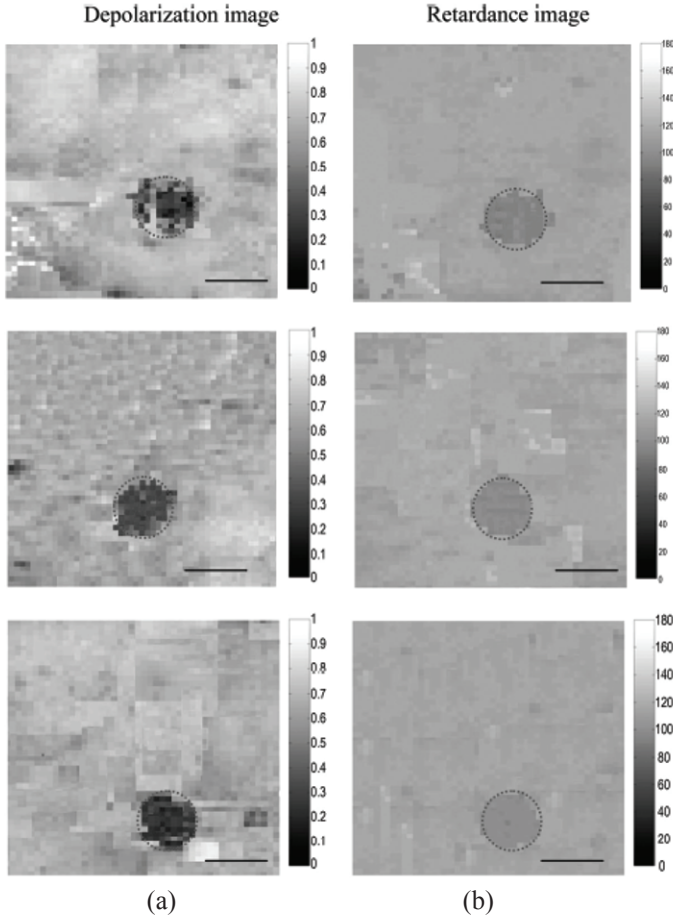


Fig. 2. As in Fig. 1, but for precancerous tissues.

sues, Ghosh et al. (2008, 2009b) have tested the validity of the decomposition process in the case of multiply-scattering media exhibiting simultaneous linear birefringence, optical activity, and depolarization for the forward and backward directions. The simultaneous effects of linear birefringence and optical activity have been accounted for by using the Jones N-matrix formalism (Jones 1948; Wood et al. 2007) in a Monte Carlo simulation of light propagation between successive scattering events. The simulation results for the forward direction were found to corroborate the experimental findings and demonstrate that the Mueller matrix decomposition methodology can successfully extract the individual polarization characteristics of a medium that exhibits simultaneous linear birefringence, optical activity, and multiple-scattering effects. In the backward geometry, it was shown that the simultaneous determination of the intrinsic values of linear and circular bi-

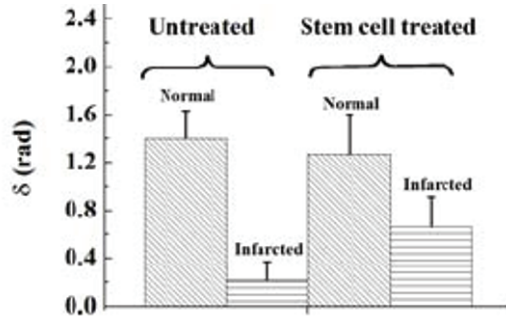


Fig. 3. The results of linear birefringence measurements from control and stem-cell-treated groups of infarcted tissues (after Ghosh et al. 2008).

refringence can be accomplished by decomposing the Mueller matrix measured at a distance from the point of illumination greater than the transport mean free path.

In Wood et al. (2009), the validity of the polar decomposition was demonstrated *in vivo* by carrying out intravital measurements in a dorsal skin window chamber mouse model using collagenase to induce changes in tissue structure and birefringence. In the window chamber model, the skin layer of the mouse was removed from a 10 mm diameter region on the dorsal surface, and a titanium saddle was sutured in place to hold the skin flap vertically. The treatment and measurements were carried out under general anesthesia. The values for linear and circular birefringence, depolarization, and diattenuation were extracted from the experimentally derived Mueller matrices at successive moments in time. The properties of the referenced and treated tissues were alternatively controlled by the histological examination.

In Fig. 4, the values of the birefringence δ and the net depolarization coefficient Δ , Eq. (57), in treated and control regions of the tissue are plotted as functions of time following the collagenase injection. The birefringence ranges from 1.2 to 0.3 rad, in contrast to the control region where the values remain essentially constant at 1 rad. The decrease in birefringence is likely due to the denaturation of the collagen fibers which reduces the structural anisotropy. This reduction in depolarization (Δ decreases from 0.63 to 0.45) is also due to the destruction of the collagen fibers, since these represent one of the primary scattering structures in the tissue. These results were confirmed by histology.

It is quite evident that due to the non-commuting nature of the matrix multiplication, the multiplication order in Eq. (52) is ambiguous. This results in the possibility of six different decomposition orders of multiplication which can be grouped into two families, depending upon the position of the depolarizer and the diattenuator matrices (Lu and Chipman 1996; Morio and Goudail 2004; Ossikovski et al. 2007). The three decompositions with the depolarizer set after the

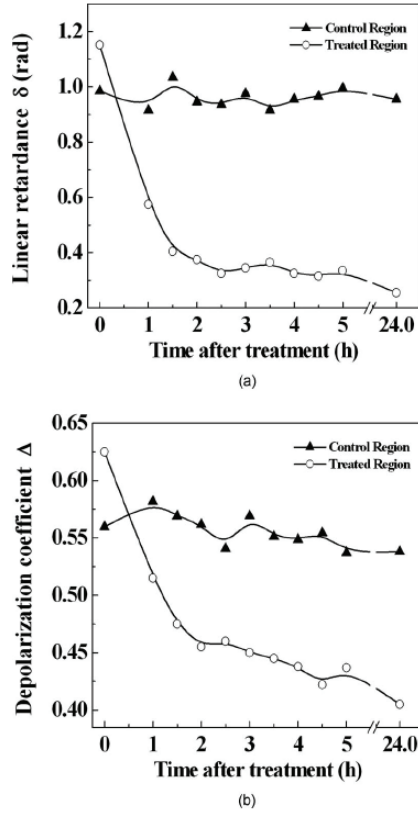


Fig. 4. (a) Birefringence and (b) depolarization as functions of time after the collagenase injection in the treated and control regions (after Wood et al. 2009).

diattenuator form the first family, while the three decompositions with the depolarizer preceding the diattenuator constitute the second family:

$$\begin{array}{cc}
 \text{First family} & \text{Second family} \\
 \left\{ \begin{array}{l} \mathbf{M}_\Delta \mathbf{M}_R \mathbf{M}_P \\ \mathbf{M}_\Delta \mathbf{M}_P \mathbf{M}_R \\ \mathbf{M}_R \mathbf{M}_\Delta \mathbf{M}_P \end{array} \right. & \left\{ \begin{array}{l} \mathbf{M}_P \mathbf{M}_R \mathbf{M}_\Delta \\ \mathbf{M}_R \mathbf{M}_P \mathbf{M}_\Delta \\ \mathbf{M}_P \mathbf{M}_\Delta \mathbf{M}_R \end{array} \right. \quad (58)
 \end{array}$$

In Ghosh et al. (2010), the influence of the multiplication order for two cases from both families represented by the first rows in Eq. (58) was studied.

The experimental Mueller matrices were measured for the dermal tissue of an athymic nude mouse (NCRNU-M, Taconic), *in vivo* from a dorsal skinfold window chamber mouse model, using a high-sensitivity turbid-polarimetry system (Wood et al. 2009). Table 1 shows the experimental Mueller matrix (in transmission), the corresponding decomposed constituent matrices, and the determined

Table 1. The experimental Mueller matrix and the decomposed constituent matrices from dermal tissue (top) and the values of the polarization parameters extracted from the decomposed matrices (bottom).

M

$$\begin{pmatrix} 1 & 0.0707 & 0.0348 & -0.0060 \\ 0.0480 & 0.4099 & 0.0077 & 0.0650 \\ 0.0162 & -0.0184 & 0.2243 & -0.3580 \\ 0.0021 & -0.0465 & 0.3571 & 0.1783 \end{pmatrix}$$

M_Δ

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0.0193 & 0.4006 & 0 & 0 \\ 0.0076 & 0 & 0.4596 & 0 \\ -0.0060 & 0 & 0 & 0.3768 \end{pmatrix}$$

M_R

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.9935 & 0.0697 & 0.0897 \\ 0 & 0.0435 & 0.4960 & -0.8673 \\ 0 & -0.1049 & 0.8655 & 0.4897 \end{pmatrix}$$

M_P

$$\begin{pmatrix} 1 & 0.0707 & 0.0348 & -0.0060 \\ 0.0707 & 0.9994 & 0.0012 & 0 \\ 0.0348 & 0.0012 & 0.9975 & 0 \\ -0.0060 & 0 & 0 & 0.9969 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0.0516 & 0.0301 & -0.0037 \\ 0 & 0.3994 & 0 & 0 \\ 0 & 0 & 0.4588 & 0 \\ 0 & 0 & 0 & 0.3758 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.9935 & 0.0701 & 0.0898 \\ 0 & 0.0436 & 0.4960 & -0.8672 \\ 0 & -0.1047 & 0.8655 & 0.4898 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0.0480 & 0.0162 & 0.0021 \\ 0.0480 & 0.9999 & 0.0004 & 0 \\ 0.0162 & 0.0004 & 0.9988 & 0 \\ 0.0021 & 0 & 0 & 0.9987 \end{pmatrix}$$

Parameters	Estimated values using the order $\mathbf{M}_\Delta \mathbf{M}_R \mathbf{M}_P$	Estimated values using the order $\mathbf{M}_P \mathbf{M}_R \mathbf{M}_\Delta$
d	0.079	0.051
Δ	0.58	0.59
ψ (deg)	0.51	0.50
δ (rad)	1.06	1.058

values of the polarization parameters. The constituent matrices obtained via the decomposition according to the first rows of Eq. (58) ($\mathbf{M}_\Delta \mathbf{M}_R \mathbf{M}_P$ and $\mathbf{M}_P \mathbf{M}_R \mathbf{M}_\Delta$) are shown in the second and third rows, respectively. The results summarized in the table show that the polar decompositions using the two selected multiplication orders of the constituent matrices give quite similar values for the retrieved polarization parameters. An insignificant difference is observed in the linear dichroism value only, the other parameters being almost identical. We can thus conclude that the individual polarization parameters of the tissue could be decoupled and quantified despite their simultaneous occurrence, even in the presence of multiple scattering.

It should be noted that although almost all existing polarimetric matrix models assume successive multiplication of the constituent Mueller matrices, the physical nature of the interrelation between this successive multiplication and the actual simultaneity of the polarization effects in real anisotropic media has not been clarified completely. Therefore, the results of Ghosh et al. (2008, 2009b, 2010) and Wood et al. (2009) have the utmost practical importance.

References

- Abhyankar, K. D., and A. L. Fymat, 1969: Relations between the elements of the phase matrix for scattering. *J. Math. Phys.* **10**, 1935–1938.
- Anderson, D. G. M., and R. Barakat, 1994: Necessary and sufficient conditions for a Mueller matrix to be derivable from a Jones matrix. *J. Opt. Soc. Am. A.* **11**, 2305–2319.
- Antonelli, M.-R., A. Pierangelo, T. Novikova, et al., 2010: Mueller matrix imaging of human colon tissue for cancer diagnostics: how Monte Carlo modeling can help in the interpretation of experimental data. *Opt. Express* **18**, 10200–10208.
- Applequist, J., 1987: Optical activity: Biot's bequest. *Am. Sci.* **75**, 59–67.
- Azzam, R. M., and N. M. Bashara, 1987: *Ellipsometry and Polarized Light* (North-Holland, New York).
- Barakat, R., 1963: Theory of the coherency matrix for light of arbitrary spectral bandwidth. *J. Opt. Soc. Am.* **53**, 317–322.
- Barakat, R., 1987: Conditions for the physical realizability of polarization matrices characterizing passive systems. *J. Mod. Opt.* **34**, 1535–1544.
- Barron, L. D., 1982: *Molecular Light Scattering and Optical Activity* (Cambridge University Press, Cambridge, UK).
- Benoit, A.-M., O. K. Naoun, V. Louis-Dorr, et al., 2001: Linear dichroism of the retinal nerve fiber layer expressed with Mueller matrices. *Appl. Opt.* **40**, 565–569.
- Berry, M. V., and M. R. Dennis, 2003: The optical singularities of birefringent dichroic chiral crystals. *Proc. R. Soc. London, Ser. A.* **459**, 1261–1292.
- Bickel, W. S., and M. E. Stafford, 1980: Biological particles as irregularly shaped particles. In D. Schuerman, Ed., *Light Scattering by Irregularly Shaped Particles* (Plenum, New York), pp. 299–305.
- Bickel, W. S., J. F. Davidson, D. R. Huffman, and R. Kilkson, 1976: Application of polarization effects in light scattering: a new biophysical tool. *Proc. Natl. Acad. Sci. USA* **73**, 486–490.
- Bicout, D., Ch. Brosseau, A. S. Martinez, and J. M. Schmitt, 1994: Depolarization of multiply scattered waves by spherical diffusers: influence of the size parameter. *Phys. Rev. E.* **49**, 1767–1770.
- Boerner, W.-M., 1992: *Direct and Inverse Methods in Radar Polarimetry* (Kluwer, Dordrecht).
- Bohren, C. F., and E. R. Huffman, 1983: *Absorption and Scattering of Light by Small Particles* (Wiley, New York).
- Bronk, B. V., S. D. Druger, J. Czege, and W. P. van de Merwe, 1995: Measuring diameters of rod-shaped bacteria *in vivo* with polarized light scattering. *Biophys. J.* **69**, 1170–1177.
- Brosseau, Ch., 1990: Analysis of experimental data for Mueller polarization matrices. *Optik* **85**, 83–86.
- Brosseau, Ch., 1998: *Fundamentals of Polarized Light: A Statistical Optics Approach* (Wiley, New York).
- Brosseau, Ch., C. R. Givens, and A. B. Kostinski, 1993: Generalized trace condition on the Mueller–Jones polarization matrix. *J. Opt. Soc. Am. A.* **10**, 2248–1151.
- Bueno, J. M., 2001: Indices of linear polarization for an optical system. *J. Opt. A: Pure Appl. Opt.* **3**, 470–476.
- Bustamante, C., M. F. Maestre, D. Keller, and I. Tinoco, Jr., 1984: Differential scattering (CIDS) of circularly polarized light by dense particles. *J. Chem. Phys.* **80**, 4817–4823.

- Chipman, R. A., 1995: Polarimetry. In *Handbook of Optics*, Vol. II (McGraw Hill, New York).
- Chipman, R. A., 1999: Depolarization. *Proc. SPIE* **3754**, 14–20.
- Chipman, R. A., 2005: Depolarization index and the average degree of polarization. *Appl. Opt.* **44**, 2490–2495.
- Chung, J., W. Jung, M. J. Hammer-Wilson, et al., 2007: Use of polar decomposition for the diagnosis of oral precancer. *Appl. Opt.* **46**, 3038–3045.
- Cloude, S. R., 1986: Group theory and polarization algebra. *Optik* **7**, 26–36.
- Cloude, S. R., and E. Pottier, 1995: Concept of polarization entropy in optical scattering. *Opt. Eng.* **34**, 1599–1610.
- Collett, E., 1993: *Polarized Light: Fundamentals and Applications* (Marcel Dekker, New York).
- de Boer, J. F., and T. E. Milner, 2002: Review of polarization sensitive optical coherence tomography and Stokes vector determination. *J. Biomed. Opt.* **7**, 359–371.
- de Boer, J. F., T. E. Milner, M. J. C. van Gemert, and J. S. Nelson, 1997: Two dimensional birefringence imaging in biological tissue by polarization-sensitive optical coherence tomography. *Opt. Lett.* **22**, 934–936.
- Deibler, L. L., and M. H. Smith, 2001: Measurement of the complex refractive index of isotropic materials with Mueller matrix polarimetry. *Appl. Opt.* **40**, 3659–3667.
- Dorman, B. P., and M. F. Maestre, 1973: Experimental differential light scattering corrections to the circular dichroism of T2 bacteriophage. *Proc. Natl. Acad. Sci. USA* **70**, 255–259.
- Espinosa-Luna, R., and E. Bernabeu, 2007: On the Q(M) depolarization metric. *Opt. Commun.* **277**, 256–258.
- Everett, M. J., K. Schoenerberger, B. W. Colston, Jr., and L. B. Da Silva, 1998: Birefringence characterization of biological tissue by use of optical coherence tomography. *Opt. Lett.* **23**, 228–230.
- Fanjul-Vélez, F., and J. L. Arce-Diego, 2010: Polarimetry of birefringent biological tissues with arbitrary fibril orientation and variable incidence angle. *Opt. Lett.* **35**, 1163–1165.
- Fry, E. S., and G. W. Kattawar, 1981: Relationships between elements of the Stokes matrix. *Appl. Opt.* **20**, 2811–2814.
- Gerrard, A., and J. M. Burch, 1975: *Introduction to Matrix Methods in Optics* (Wiley, New York).
- Ghosh, N., H. S. Patel, and P. K. Gupta, 2003: Depolarization of light in tissue phantoms—effect of a distribution in the size of scatterers. *Opt. Express* **11**, 2198–2205.
- Ghosh, N., M. F. G. Wood, and I. A. Vitkin, 2008: Mueller matrix decomposition for extraction of individual polarization parameters from complex turbid media exhibiting multiple scattering, optical activity, and linear birefringence. *J. Biomed. Opt.* **13**, 044036.
- Ghosh, N., M. F. G. Wood, Sh.-h. Li, et al., 2009a: Mueller matrix decomposition for polarized light assessment of biological tissues. *J. Biophoton.* **2**, 145–156.
- Ghosh, N., M. F. G. Wood, and I. A. Vitkin, 2009b: Polarimetry in turbid, birefringent, optically active media: a Monte Carlo study of Mueller matrix decomposition in the back-scattering geometry. *J. Appl. Phys.* **105**, 102023.
- Ghosh, N., M. F. G. Wood, and I. A. Vitkin, 2010: Influence of the order of the constituent basis matrices on the Mueller matrix decomposition-derived polarization parameters in complex turbid media such as biological tissues. *Opt. Commun.* **283**, 1200–1208.

- Gil, J. J., 2000: Characteristic properties of Mueller matrices. *J. Opt. Soc. Am. A* **17**, 328–334.
- Gil, J. J., 2007: Polarimetric characterization of light and media. *Eur. Phys. J. Appl. Phys.* **40**, 1–47.
- Gil, J. J., and E. Bernabeu, 1985: A depolarization criterion in Mueller matrices. *Opt. Acta* **32**, 259–261.
- Gil, J. J., and E. Bernabeu, 1986: Depolarization and polarization indexes of an optical system. *Opt. Acta* **33**, 185–189.
- Gil, J. J., and E. Bernabeu, 1987: Obtainment of the polarizing and retardation parameters of a non-depolarizing optical system from the polar decomposition of its Mueller matrix. *Optik* **76**, 67–71.
- Givens, C. R., and A. B. Kostinski, 1993: A simple necessary and sufficient condition on physically realizable Mueller matrices. *J. Mod. Opt.* **40**, 471–481.
- Gopala Rao, A. V., K. S. Mallesh, and Sudha, 1998a: On the algebraic characterization of a Mueller matrix in polarization optics. I. Identifying a Mueller matrix from its N matrix. *J. Mod. Opt.* **45**, 955–987.
- Gopala Rao, A. V., K. S. Mallesh, and Sudha, 1998b: On the algebraic characterization of a Mueller matrix in polarization optics. II. Necessary and sufficient conditions for Jones-derived Mueller matrices. *J. Mod. Opt.* **45**, 989–999.
- Gross, C. T., H. Salamon, A. J. Hunt, et al., 1991: Hemoglobin polymerization in sickle-cells studied by circular polarized light scattering. *Biochim. Biophys. Acta* **1079**, 152–160.
- Hadley, K. C., and I. A. Vitkin, 2002: Optical rotation and linear and circular depolarization rates in diffusively scattered light from chiral, racemic, and achiral turbid media. *J. Biomed. Opt.* **7**, 291–299.
- He, H., N. Zeng, R. Liao, et al., 2010: Application of sphere–cylinder scattering model to skeletal muscle. *Opt. Express* **18**, 15104–15112.
- Hielscher, A. H., A. A. Eick, J. R. Mourant, et al., 1997: Diffuse backscattering Mueller matrices of highly scattering media. *Opt. Express* **1**, 441–453.
- Hoekstra, A. G., and P. M. A. Sloot, 1993: Dipolar unit size in coupled-dipole calculations of the scattering matrix elements. *Appl. Opt.* **18**, 1211–1213.
- Hoekstra, A. G., and P. M. A. Sloot, 2000: Biophysical and biomedical applications of non-spherical scattering. In M. I. Mishchenko, J. W. Hovenier, and L. D. Travis, Eds., *Light Scattering by Nonspherical Particles: Theory, Measurements, and Applications* (Academic Press, San Diego), pp. 585–602.
- Hovenier, J. W., 1969: Symmetry relations for scattering of polarized light in a slab of randomly oriented particles. *J. Atmos. Sci.* **26**, 488–499.
- Hovenier, J. W., 1970: Principles of symmetry for polarization studies of planets. *Astron. Astrophys.* **7**, 86–90.
- Hovenier, J. W., 1994: Structure of a general pure Mueller matrix. *Appl. Opt.* **33**, 8318–8324.
- Hovenier, J. W., and D. W. Mackowski, 1998: Symmetry relations for forward and backward scattering by randomly oriented particles. *J. Quant. Spectrosc. Radiat. Transfer* **60**, 483–492.
- Hovenier, J. W., and C. V. M. van der Mee, 2000: Basic relationships for matrices describing scattering by small particles. In M. I. Mishchenko, J. W. Hovenier, and L. D. Travis, Eds., *Light Scattering by Nonspherical Particles: Theory, Measurements, and Applications* (Academic Press, San Diego), pp. 61–85.

- Hovenier, J. W., H. C. van de Hulst, and C. V. M. van der Mee, 1986: Conditions for the elements of the scattering matrix. *Astron. Astrophys.* **157**, 301–310.
- Hu, Ch.-R., G. W. Kattawar, M. E. Parkin, and P. Herb, 1987: Symmetry theorems on the forward and backward scattering Mueller matrices for light scattering from a non-spherical dielectric scatterer. *Appl. Opt.* **26**, 4159–4173.
- Hunter, D. G., J. C. Sandruck, S. Sau, et al, 1999: Mathematical modeling of retinal birefringence scanning. *J. Opt. Soc. Am. A.* **16**, 2103–2111.
- Hurwitz, H., and C. R. Jones, 1941: A new calculus for the treatment of optical systems. II. Proof of three general equivalence theorems. *J. Opt. Soc. Am.* **31**, 493–499.
- Johnston, R. G., S. B. Singham, and G. C. Salzman, 1988: Polarized light scattering. *Comments Mol. Cell. Biophys.* **5**, 171–192.
- Jones, R. C., 1941: A new calculus for the treatment of optical systems. I. Description and discussion of the calculus. *J. Opt. Soc. Am.* **31**, 488–493.
- Jones, R. C., 1942: A new calculus for the treatment of optical systems. IV. *J. Opt. Soc. Am.* **32**, 486–493.
- Jones, R. C., 1947: A new calculus for the treatment of optical systems. V. A more general formulation and description of another calculus. *J. Opt. Soc. Am.* **37**, 107–110.
- Jones, R. C., 1948: A new calculus for the treatment of optical systems. VII. Properties of the N-matrices. *J. Opt. Soc. Am.* **38**, 671–685.
- Jones, R.C., 1956: A new calculus for the treatment of optical systems. VIII. Electromagnetic theory. *J. Opt. Soc. Am.* **46**, 126–131.
- Kaplan, B., and B. Drevillon, 2002: Mueller matrix measurements of small spherical particles deposited on a c-Si wafer. *Appl. Opt.* **41**, 3911–3918.
- Kaplan, B., G. Ledanois, and B. Drevillon, 2001: Mueller matrix of dense polystyrene latex sphere suspensions: measurements and Monte Carlo simulation. *Appl. Opt.* **40**, 2769–2777.
- Kim, K., L. Mandel, and E. Wolf, 1987: Relationship between Jones and Mueller matrices for random media. *J. Opt. Soc. Am. A.* **4**, 433–437.
- Kim, Y. L., P. Pradhan, M. H. Kim, and V. Backman, 2006: Circular polarization memory effect in low-coherence enhanced backscattering of light. *Opt. Lett.* **31**, 2744–2746.
- Kliger, D. S., J. W. Lewis, and C. E. Randall, 1990: *Polarized Light in Optics and Spectroscopy* (Academic Press, New York).
- Kokhanovsky, A. A., 2001: *Light Scattering Media Optics: Problems and Solutions* (Praxis, Chichester, UK).
- Kokhanovsky, A. A., 2003a: Parameterization of the Mueller matrix of oceanic waters. *J. Geophys. Res.* **108**, 3175.
- Kokhanovsky, A. A., 2003b: *Polarization Optics of Random Media* (Praxis, Chichester, UK).
- Kokhanovsky, A. A., 2003c: Optical properties of irregularly shaped particles. *J. Phys. D: Appl. Phys.* **36**, 915–923.
- Kostinski, A. B., 1992: Depolarization criterion for incoherent scattering. *Appl. Opt.* **31**, 3506–3508.
- Kostinski, A. B., C. R. Givens, and J. M. Kwiatkowski, 1993: Constraints on Mueller matrices of polarization optics. *Appl. Opt.* **32**, 1646–1651.
- Landau, L. D., E. M. Lifshitz, and L. P. Pitaevskii, 1984: *Electrodynamics of Continuous Media* (Pergamon Press, Oxford).
- Lankaster, P., and M. Tismenetsky, 1985: *The Theory of Matrices* (Academic Press, San Diego).

- Lawless, R., Y. Xie, P. Yang, et al., 2006: Polarization and effective Mueller matrix for multiple scattering of light by nonspherical ice crystals. *Opt. Express* **14**, 6381–6393.
- Li, X., and G. Yao, 2009: Mueller matrix decomposition of diffuse reflectance imaging in skeletal muscle. *Appl. Opt.* **48**, 2625–2631.
- Li, X., J. C. Ranasinghesagara, and G. Yao, 2008: Polarization-sensitive reflectance imaging in skeletal muscle. *Opt. Express* **16**, 9927–9935.
- Lofftus, K. D., M. S. Quinby-Hunt, A. J. Hunt, et al., 1992: Light scattering by Proccentrum micans: a new method and results. *Appl. Opt.* **31**, 2924–2931.
- Lopatin, V. N., A. V. Priezzhev, A. D. Apanasenko, et al., *Methods of Light Scattering in the Analysis of Dispersed Biological Media* (Fizmatlit, Moscow) (in Russian).
- Louis-Dorr, V., K. Naoun, P. Allé, et al., 2004: Linear dichroism of the cornea. *Appl. Opt.* **43**, 1515–1521.
- Lu, S.-Y., and R. A. Chipman, 1994: Homogeneous and inhomogeneous Jones matrices. *J. Opt. Soc. Am. A*, **11**, 766–773.
- Lu, S.-Y., and R. A. Chipman, 1996: Interpretation of Mueller matrices based on polar decomposition. *J. Opt. Soc. Am. A*, **13**, 1106–1113.
- Maestre, M. F., C. Bustamante, T. L. Hayes, et al., 1982: Differential scattering of circularly polarized light by the helical sperm head from the octopus *Eledone cirrhosa*. *Nature* **298**, 773–774.
- Maitland, D. J., and J. T. Walsh, 1997: Quantitative measurements of linear birefringence during heating of native collagen. *Laser Surg. Med.* **20**, 310–318.
- Maksimova, I. L., S. N. Tatarintsev, and L. P. Shubochkin, 1992: Multiple scattering effects in laser diagnostics of bioobjects. *Opt. Spectrosc.* **72**, 1171–1177.
- Manhas, S., M. K. Swami, P. Buddhiwant, et al., 2006: Mueller matrix approach for determination of optical rotation in chiral turbid media in backscattering geometry. *Opt. Express* **14**, 190–202.
- Manickavasagam, S., and M. P. Mengüç, 1998: Scattering-matrix elements of coated infinite-length cylinders. *Appl. Opt.* **37**, 2473–2482.
- Mar'enko, V. V., and S. N. Savenkov, 1994: Representation of arbitrary Mueller matrix in the basis of matrices of circular and linear anisotropy. *Opt. Spectrosc.* **76**, 94–96.
- Mishchenko, M. I., and J. W. Hovenier, 1995: Depolarization of light backscattered by randomly oriented nonspherical particles. *Opt. Lett.* **20**, 1356–1358.
- Mishchenko, M. I., and L. D. Travis, 2000: Polarization and depolarization of light. In F. Moreno, and F. González, Eds., *Light Scattering from Microstructures* (Springer, Berlin), pp. 159–175.
- Mishchenko, M. I., J. W. Hovenier, and L. D. Travis, Eds., 2000: *Light Scattering by Nonspherical Particles: Theory, Measurements, and Applications* (Academic Press, San Diego).
- Mishchenko, M. I., L. D. Travis, and A. A. Lacis, 2002: *Scattering, Absorption, and Emission of Light by Small Particles* (Cambridge University Press, Cambridge, UK).
- Mishchenko, M. I., L. D. Travis, and A. A. Lacis, 2006: *Multiple Scattering of Light by Particles. Radiative Transfer and Coherent Backscattering* (Cambridge University Press, Cambridge, UK).
- Morio, J., and F. Goudail, 2004: Influence of the order of diattenuator, retarder, and polarizer in polar decomposition of Mueller matrices. *Opt. Lett.* **29**, 2234–2236.
- Mueller, H., 1948: The foundation of optics. *J. Opt. Soc. Am.* **38**, 661.
- Muttiah, R. S., Ed., 2002: *From Laboratory Spectroscopy to Remotely Sensed Spectra of Terrestrial Ecosystems* (Kluwer, Dordrecht).

- Nagirner, D. I., 1993: Constrains on matrices transforming Stokes vectors. *Astron. Astrophys.* **275**, 318–324.
- Naoun, O. K., V. Louis-Dorr, P. Allé, et al., 2005: Exploration of the retinal nerve fiber layer thickness by measurement of the linear dichroism. *Appl. Opt.* **44**, 7074–7082.
- Ossikovski, R., 2008: Interpretation of nondepolarizing Mueller matrices based on singular-value decomposition. *J. Opt. Soc. Am. A.* **25**, 473–482.
- Ossikovski, R., 2009: Analysis of depolarizing Mueller matrices through a symmetric decomposition. *J. Opt. Soc. Am. A.* **26**, 1109–1118.
- Ossikovski, R., A. De Martino, and S. Guyot, 2007: Forward and reverse product decompositions of depolarizing Mueller matrices. *Opt. Lett.* **32**, 689–691.
- Parke III, N. G., 1948: Matrix optics. PhD thesis (Massachusetts Institute of Technology, Cambridge, MA).
- Parke III, N. G., 1949: Optical algebra. *J. Math. Phys.* **28**, 131–139.
- Perrin, F., 1942: Polarization of light scattering by isotropic opalescent media. *J. Chem. Phys.* **10**, 415–427.
- Potton, R. J., 2004: Reciprocity in optics. *Rep. Prog. Phys.* **67**, 717–754.
- Priezzhev, A. V., V. V. Tuchin, and L. P. Shubochkin, 1989: *Laser Diagnostics in Biology and Medicine* (Nauka, Moscow) (in Russian).
- Quinby-Hunt, M. S., P. G. Hull, and A. J. Hunt, 2000: Polarized light scattering in the marine environment. In M. I. Mishchenko, J. W. Hovenier, and L. D. Travis, Eds., *Light Scattering by Nonspherical Particles: Theory, Measurements, and Applications* (Academic Press, San Diego), pp. 525–554.
- Ranasinghesagara, J. C., and G. Yao, 2007: Imaging 2D optical diffuse reflectance in skeletal muscle. *Opt. Express* **15**, 3998–4007.
- Rojas-Ochoa, L. F., D. Lacoste, R. Lenke, et al., 2004: Depolarization of backscattered linearly polarized light. *J. Opt. Soc. Am. A.* **21**, 1799–1804.
- Salzmann, G. C., S. B. Singham, R. G. Johnston, and C. F. Bohren, 1990: Light scattering and cytometry. In M. R. Melamed, T. Lindmo, and M. L. Mendelsohn, Eds., *Flow Cytometry and Sorting* (Wiley, New York), pp. 81–107.
- Sankaran, V., J. T. Walsh, Jr., and D. J. Maitland, 2002: Comparative study of polarized light propagation in biologic tissues. *J. Biomed. Opt.* **7**, 300–306.
- Savenkov, S. N., and K. E. Yushtin, 2000: New classes of objects in polarimetry including the isotropic depolarization. *Radiotekhnika* **116**, 3–11 (in Russian).
- Savenkov, S. N., O. I. Sydoruk, and R. S. Muttiah, 2005: The conditions for polarization elements to be dichroic and birefringent. *J. Opt. Soc. Am. A.* **22**, 1447–1452.
- Savenkov, S. N., V. V. Marienko, E. A. Oberemok, and O. I. Sydoruk, 2006: Generalized matrix equivalence theorem for polarization theory. *Phys. Rev. E.* **74**, 056607.
- Savenkov, S. N., R. S. Muttiah, K. E. Yushtin, and S. A. Volchkov, 2007a: Mueller-matrix model of an inhomogeneous, linear, birefringent medium: single scattering case. *J. Quant. Spectrosc. Radiat. Transfer* **106**, 475–486.
- Savenkov, S. N., O. I. Sydoruk, and R. S. Muttiah, 2007b: Eigenanalysis of dichroic, birefringent, and degenerate polarization elements: a Jones-calculus study. *Appl. Opt.* **46**, 6700–6709.
- Saxon, D. S., 1955: Tensor scattering matrix for electromagnetic fields. *Phys. Rev.* **100**, 1771–1775.
- Schmitt, J. M., and G. Kumar, 1996: Turbulent nature of refractive-index variations in biological tissue. *Opt. Lett.* **21**, 1310–1312.

- Sekera, Z., 1966: Scattering matrices and reciprocity relationships for various representations of the state of polarization. *J. Opt. Soc. Am.* **56**, 1732–1740.
- Shapiro, D. B., P. G. Hull, A. J. Hunt, and J. E. Hearst, 1994a: Calculations of the Mueller scattering matrix for a DNA plectonemic helix. *J. Chem. Phys.* **101**, 4214–4221.
- Shapiro, D. B., M. F. Maestre, W. M. McClain, et al., 1994b: Determination of the average orientation of DNA in the octopus sperm *Eledone cirrhosa* through polarized light scattering. *Appl. Opt.* **33**, 5733–5744.
- Shindo, Y., 1995: Applications of polarized modulator technique in polymer science. *Opt. Eng.* **34**, 3369–3384.
- Shurcliff, W. A., 1962: *Polarized Light: Production and Use* (Harvard University Press, Harvard, MA).
- Simon, R., 1982: The connection between Mueller and Jones matrices of polarization optics. *Opt. Commun.* **42**, 293–297.
- Simon, R., 1987: Mueller matrices and depolarization criteria. *J. Mod. Opt.* **34**, 569–575.
- Simonenko, G. V., T. P. Denisova, N. A. Lakodina, and V. V. Tuchin, 2000: Measurement of an optical anisotropy of biotissues. *Proc. SPIE.* **3915**, 152–157.
- Smith, M. H., 2001: Interpreting Mueller matrix images of tissues. *Proc. SPIE.* **4257**, 82–89.
- Solleillet, P., 1929: Sur les paramètres caractérisant la polarisation partielle de la lumière dans les phénomènes de fluorescence. *Ann. Phys. Biol. Med.* **12**, 23–97.
- Sridhar, R., and R. Simon, 1994: Normal form for Mueller matrices in polarization optics. *J. Mod. Opt.* **41**, 1903–1915.
- Studinski, R. C. N., and I. A. Vitkin, 2000: Methodology for examining polarized light interactions with tissues and tissue-like media in the exact backscattering direction. *J. Biomed. Opt.* **5**, 330–337.
- Tuchin, V. V., 1993: Lasers light scattering in biomedical diagnostics and therapy. *J. Laser Appl.* **5**, 43–60.
- Tuchin, V. V., Ed., 1994: *Selected Papers on Tissue Optics Applications in Medical Diagnostics and Therapy* (SPIE Press, Bellingham, WA).
- Tuchin, V. V., Ed., 2002: *Handbook of Optical Biomedical Diagnostics* (SPIE Press, Bellingham, WA).
- Tuchin, V. V., L. V. Wang, and D. A. Zimnyakov, 2006: *Optical Polarization in Biomedical Applications* (Springer, Berlin).
- Twietmeyer, K. M., R. A. Chipman, A. E. Elsner, et al., 2008: Mueller matrix retinal imager with optimized polarization conditions. *Opt. Express* **16**, 21 339–21 354.
- van de Hulst, H. C., 1957: *Light Scattering by Small Particles* (Wiley, New York).
- van de Merwe, W. P., D. R. Huffman, and B. V. Bronk, 1989: Reproducibility and sensitivity of polarized light scattering for identifying bacterial suspensions. *Appl. Opt.* **28**, 5052–5057.
- van de Merwe, W. P., Z.-Z. Li, B. V. Bronk, and J. Czege, 1997: Polarized light scattering for rapid observation of bacterial size changes. *Biophys. J.* **73**, 500–506.
- van der Mee, C. V. M., 1993: An eigenvalue criterion for matrices transforming Stokes parameters. *J. Math. Phys.* **34**, 5072–5088.
- van der Mee, C. V. M., and J. W. Hovenier, 1992: Structure of matrices transforming Stokes parameters. *J. Math. Phys.* **33**, 3574–3584.
- Vansteenkiste, N., P. Nignolo, and A. Aspect, 1993: Optical reversibility theorems for polarization: application to remote control of polarization. *J. Opt. Soc. Am. A.* **10**, 2240–2245.

- Voss, K. J., and E. S. Fry, 1984: Measurement of the Mueller matrix for ocean water. *Appl. Opt.* **23**, 4427–4436.
- Wang, L. V., G. L. Coté, and S. L. Jacques, Eds., 2002: Special Section on Tissue Polarimetry. *J. Biomed. Opt.* **7**, 278–397.
- Whitney, C., 1971: Pauli-algebraic operators in polarization optics. *J. Opt. Soc. Am.* **61**, 1207–1213.
- Wood, M. F. G., X. Guo, and I. A. Vitkin, 2007: Polarized light propagation in multiply scattering media exhibiting both linear birefringence and optical activity: Monte Carlo model and experimental methodology. *J. Biomed. Opt.* **12**, 014029.
- Wood, M. F. G., N. Ghosh, E. H. Moriyama, et al., 2009: Proof-of-principle demonstration of a Mueller matrix decomposition method for polarized light tissue characterization *in vivo*. *J. Biomed. Opt.* **14**, 014029.
- Xing, Z.-F., 1992: On the deterministic and nondeterministic Mueller matrix. *J. Mod. Opt.* **39**, 461–484.
- Xu, L., J. Ding, and A. Y. S. Cheng, 2002: Scattering matrix of infrared radiation by ice finite circular cylinders. *Appl. Opt.* **41**, 2333–2348.
- Zietz, S., A. Belmont, and C. Nicolini, 1983: Differential scattering of circularly polarized light as an unique probe of polynucleosome superstructures. *Cell Biophys.* **5**, 163–187.
- Zubko, E., Y. G. Shkuratov, and G. Videen, 2004: Coherent backscattering effect for non-zero elements of Mueller matrix of discrete media at different illumination-observation geometries. *J. Quant. Spectrosc. Radiat. Transfer* **89**, 443–452.



Picnic at the Main Astronomical Observatory.



From left to right: Philippe Goloub, Oleg Dubovik, and Thomas Ruutz at Kyiv Pechersk Lavra.



Poster session. From left to right: Felix Yanovsky, William Martin, Sergey Savenkov, Vladimir Barun, and Dmitry Petrov.

Plasmonic optical properties and the polarization modulation technique

B. K. Serdega, S. P. Rudenko*, L. S. Maksimenko, and I. E. Matyash

V. Lashkaryov Institute of Semiconductor Physics, National Academy of Sciences of Ukraine, 45 Prospekt Nauky, 03028 Kyiv, Ukraine

Abstract. Surface-plasmon resonance features in nanosized gold films are studied in the Kretschmann geometry using the polarization modulation technique. The reflection coefficients R_s^2 and R_p^2 for the s - and p -polarized light, respectively, as well as their polarization difference $\Delta\rho = R_s^2 - R_p^2$ are measured as functions of the angle of incidence of electromagnetic radiation, wavelength, and film thickness. Manifestations of classical and topological size effects and a qualitative model of the transition between them in nanosized gold films are discussed. It is found that the magnitude and the sign of curvature of the angular dependence of the polarization difference $\Delta\rho(\theta)$ are related to the excitation of surface plasmons by either p -polarized light or both s - and p -polarized light in the case of homogeneous or cluster-structured metal films, respectively. It is demonstrated that nanocomposite films with gold nanoparticles embedded in a dielectric matrix can exhibit anisotropy of dielectric properties not only in the case of oblique incidence of light but also because of their macroscopically heterogeneous structure.

Keywords: polarization modulation, surface plasmon resonance, gold films, size effects

1. Introduction

The development of nanoscience and nanotechnology has made it possible to create new nanosized materials having unique electronic and optical properties quite different from those of the bulk state. Such optical properties of nanostructures of noble metals remain the focus of investigations [1]. Among these are representative variations of color, and changes in transmission and reflection spectra with varying size, shape, and dielectric material surrounding the metal nanoparticles, as well as the distance between them. In addition, special attention has been paid to effects related to the resonance absorption of electromagnetic radiation in metals and manifestations of the *surface plasmon resonance* (SPR) [2].

The origin of this absorption is attributed to the interaction between light and mobile surface charges, typically the conduction electrons in metals. This light-matter interaction leads to surface plasmon–polariton modes having a greater

* Corresponding author. E-mail: rudenko.svetlana@gmail.com

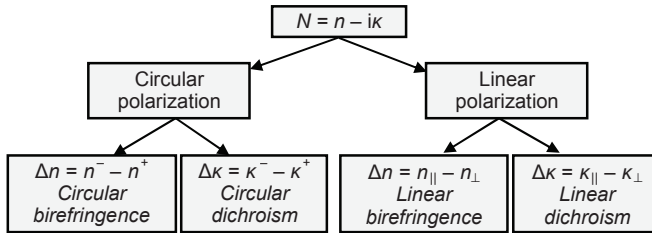


Fig. 1. Classification of optical polarization effects: $n_{\parallel,\perp}$, $\kappa_{\parallel,\perp}$ are the refraction and absorption coefficients parallel and perpendicular to the optical axis; $n^{+,-}$, $\kappa^{+,-}$ are the refraction and absorption coefficients for left- and right-circular polarization state (after [9]).

momentum than light of the same frequency. Consequently, the electromagnetic fields associated with them cannot propagate away from the surface; rather, the fields decay exponentially in strength with distance from the surface. The boundary conditions for electromagnetic fields lead to different conditions for the occurrence of plasmons in the case of bulk material, planar metal–dielectric interfaces, and metal particles. It is, therefore, useful to distinguish these modes by referring to bulk plasmons, surface plasmon–polaritons, or localized plasmons, respectively. Since the time SPRs were discovered [3,4], interest constantly has grown, with the emphasis shifting from investigations of their physical content [5,6] to applications in material science, medicine [7], chemical biology [8], and other fields.

To excite surface plasmon–polaritons with photons, one must use appropriate techniques. Two of the most popular are presented in [3,4]. While differing in details, these two techniques have a common feature in that the dependence of the coefficient of internal reflection from a prism surface on the angle of incidence is measured in the so-called Kretschmann and Otto geometries. The probe radiation should be linearly polarized, and the electric field vector should be parallel to the plane of light incidence (p -polarization). The minimal amplitude value in the above dependence indicates SPR.

The number of publications on SPRs reporting results obtained by various methods of investigation is extremely large. There is no need to list these methods here because they are all used in solid-state physics. When a phenomenon has been thoroughly and deeply investigated, new data can be obtained using a radically new technique only.

2. Polarization modulation technique

Here we consider studying SPRs using a *polarization modulation* (PM) technique. Such an approach has its justification. To illustrate this, we consider the existing classification of polarization effects [9] shown in Fig. 1. The interaction of polarized radiation with anisotropic objects brings about the appearance of optical effects that are characterized with a complex refractive index $N = n - i\kappa$, where $i = (-1)^{1/2}$. According to the polarization state of light (linear or circular), its compo-

nents correspond to those of birefringence and dichroism. These effects are characterized by *amplitude* and *phase anisotropy* of the internal reflection of light in the objects under investigation. One of these characteristics, namely *amplitude anisotropy*, may be exemplified by the effect of dichroism, which depends on the coefficient of absorption (reflection) of linearly polarized radiation on the polarization azimuth, i.e., the angle between the electric field vector and the plane of incidence of the electromagnetic wave, relative to the optical axis of the anisotropic sample [10].

Another reason to apply the PM technique when studying SPRs is its considerable information efficiency. The PM technique used in this study can hardly be considered as a new one. Contrary to all known versions of the modulation spectroscopy method [11], PM as a spatial characteristic of electromagnetic radiation is realized by the two-dimensional (2D) action on an electromagnetic wave. This fact, although insignificant *prima facie*, makes the experimental technique more complicated, and in addition, requires development of novel interpretation techniques to be applied to the results obtained. Nevertheless, if one knows how to interpret these results, one can obtain additional information about the objects under investigation, surrounding media and so on, which were inaccessible with the previously used methods.

PM was implemented with the investigation of the photoelastic effect [12], and subsequently became widespread. At the same time this technique has been fruitful in the study of such polarization-dependent effects in anisotropic media as dichroism in transmission [13], pleochroism in photoconductivity [14], and the photovoltaic effects [15,16].

Thus, the *PM technique* is a highly informative and sensitive method of analysis of polarized electromagnetic radiation. The intensity, frequency, phase, and wave vector remain unchanged under its action. Only the polarization state of light changes.

The most important element of the PM technique is the *polarization modulator* based on the photoelastic effect [12]. It consists of two quartz plates (amorphous and piezocrystalline) that are in acoustic contact. The polarization modulator is a dynamic phase plate that can be operated in two modes by controlling its power supply. In the first mode, it acts as a half-wave plate and changes the azimuth of a linearly polarized wave from parallel to perpendicular to the plane of incidence of light twice during the modulation period. In the second mode, it acts as a quarter-wave plate and transforms the linear polarization into right- and left-circular polarization once during the period.

In this case, a stationary quarter-wave phase plate is installed after the modulator to transform light to make the linear polarization orthogonal once during the period. This version is preferred when using a photodetector with sensitivity reduced at the doubled modulation frequency. In both cases, by rotating the modulator around the optical axis of the measuring system, one chooses the position at which the polarization azimuths of emerging light are successively parallel and perpendicular to the plane of incidence (*p*- and *s*-polarizations, respectively). The

constant-intensity radiation modulated with respect to the polarization state is directed to the sample under investigation.

The essence of this method is that the sample is successively irradiated with beams of s - and p -polarized light of constant intensity. In this case, the measurement recorded is the difference in the s - and p -functions whose arguments may be energy, deformation, electric or magnetic fields, reflection angle of light, etc. An advantage of this technique is that the result obtained is reliable even if the signal difference is below the noise level of the individual s - and p -intensities. This means that small differences in the polarizations can be amplified significantly and detected through cancellation of their mutual dependence.

A SPR is also characterized by its polarization dependence: an angular scan of an internal reflection prism with a resonantly sensitive metal film in p -polarized light results in a minimum at resonance. So this demonstrates the *linear amplitude anisotropy* of the internal reflection. The use of the PM technique leads to registration of the *polarization difference* between the reflectances of the intensities of s - and p -polarized radiation: $\Delta\rho = R_s^2 - R_p^2$.

The polarization difference is nearly instantaneously measured in the experiment rather than obtained through mathematical calculations. This physical quantity is a highly reliable characteristic of the smallest amplitude anisotropy of the dielectric properties. This is possible because the dependence of the polarization difference on a certain parameter (the light wavelength λ , deformation, field strength, etc) is obtained by subtracting functions with different polarizations, thereby eliminating the common features of these functions and enhancing their individual features.

2.1. Formalism

In the simplest case, SPR is observed during the propagation of light through a medium having at least three layers. The first layer (glass) should be denser than air, the second layer is a metal or a heavily doped semiconductor with a complex refractive index $N_1 = n - i\kappa$, and the third layer is, generally, air. To describe the SPR as a version of a more general phenomenon, i.e., reflection, the Fresnel equations are used [17]. In this case, the evolution of the polarization state of light passing through the layers and reflecting from the interfaces is represented in a matrix form. The matrix product makes it possible to describe the polarization state of light in the final form as $\mathbf{M}_j = \mathbf{I}_{j01}\mathbf{L}_1\mathbf{I}_{j12}$. Here, $j = s, p$ indicates the polarization state;

$$\mathbf{I}_{j01} = \begin{bmatrix} 1 & r_{j01} \\ r_{j01} & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{I}_{j12} = \begin{bmatrix} 1 & r_{j12} \\ r_{j12} & 1 \end{bmatrix}$$

are the matrices of the interfaces between media 0–1 and 1–2, respectively;

$$\mathbf{L}_1 = \begin{bmatrix} e^{-i\beta} & 0 \\ 0 & e^{i\beta} \end{bmatrix}$$

is the matrix of the film;

$$r_{p01} = \frac{N_1 \cos \theta_0 - N_0 \cos \theta_1}{N_1 \cos \theta_0 + N_0 \cos \theta_1}, \quad r_{p12} = \frac{N_2 \cos \theta_1 - N_1 \cos \theta_2}{N_2 \cos \theta_1 + N_1 \cos \theta_2},$$

$$r_{s01} = \frac{N_0 \cos \theta_0 - N_1 \cos \theta_1}{N_0 \cos \theta_0 + N_1 \cos \theta_1}, \quad \text{and} \quad r_{s12} = \frac{N_1 \cos \theta_1 - N_2 \cos \theta_2}{N_1 \cos \theta_1 + N_2 \cos \theta_2}$$

are the amplitude reflection coefficients for the s - and p -polarized light, where θ_0 , θ_1 , and θ_2 are the angles of light propagation in the glass, metal film, and air, respectively: $\theta_1 = \arcsin[(N_0 \sin \theta_0)/N_1]$, $\theta_2 = \arcsin[(N_0 \sin \theta_0)/N_2]$, and N_0 , N_1 , and N_2 are the refractive indices of the prism, film, and external medium (air), respectively. The phase factor

$$\beta = \frac{2\pi d}{\lambda} \sqrt{N_1^2 - N_0^2 \sin^2 \theta_0}$$

is common for both polarizations, where d is the layer thickness.

The above expressions make it possible to determine the reflection coefficients for electromagnetic waves of both polarizations:

$$R_p = M_{p(2,1)}/M_{p(1,1)}, \quad R_s = M_{s(2,1)}/M_{s(1,1)}, \quad (1)$$

where $M_{p(2,1)}$, $M_{p(1,1)}$, $M_{s(2,1)}$, and $M_{s(1,1)}$ are the corresponding matrix elements for the s - and p -polarizations. Therefore, the expression for the measured *polarization difference* of the reflection coefficients for the s - and p -polarizations will take the form

$$\Delta\rho = R_s^2 - R_p^2. \quad (2)$$

Firstly, let us refine the terms in order not to confuse the parameter measured with PM, in our case, with one of the four Stokes vector components. Indeed, according to [10], this parameter corresponds to the definition of the Q component (in terms of the notation adopted in [18]), as a difference of the intensities of two in-phase linearly polarized waves whose polarization azimuths are oriented parallel and perpendicular to the optical axis. Note that the intensity of this component, in the general case of elliptical polarization, can be only a part of the total intensity. However, the PM technique in this version operates only with the numerator of the value: $P = (R_{\perp} - R_{\parallel})/(R_{\perp} + R_{\parallel})$, which is referred to as pleochroism and is used to illustrate the contrast of polarization-dependent effects. In this case, the other components of the Stokes vector are, in principle, absent in all situations because of operation only with one wave whose polarization azimuth is modulated. At the same time, as was shown in [14], the singularities in the functional expressions in the numerator and denominator are cancelled out as a result of their division, thus leading to a loss of information. That is why the parameter $\Delta\rho = R_s^2 - R_p^2$ contains information about the difference in the functional dependences of its terms that disappears as a result of the division of the polarization difference by the sum. It should be emphasized that only the sign of the signal is reversed after

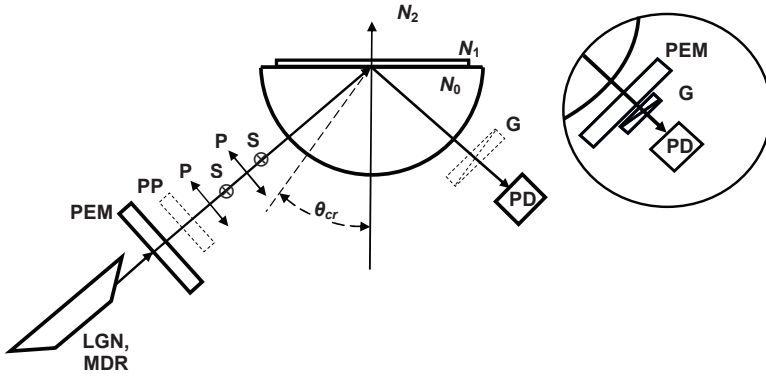


Fig. 2. Optical scheme of the experimental setup for measuring the angular characteristics of the polarization difference $\Delta\rho$ with application of polarization modulation. Notation: LGN – He–Ne laser; MDR – monochromator; PP – quarter-wave phase plate; PEM – photoelastic polarization modulator; p, s – linear polarizations whose azimuths are parallel and perpendicular, respectively, to the plane of incidence; G – Glan prism; PD – photodetector; θ_{cr} – critical angle of total internal reflection; and N_0, N_1, N_2 – refractive indices of glass, metal, and air, respectively (after [19]). The inset shows the part of the experimental setup for measuring the Stokes components [20].

the rearrangement of terms in Eq. (2) for the polarization difference; this is not of importance when using synchronous phase detection.

Thus, such a parameter as *polarization difference* $\Delta\rho$, that is the difference between the reflection coefficients R_s^2 and R_p^2 for s - and p -polarized radiation, respectively, should be considered only as an analog of the Q component of the Stokes vector.

2.2. Laboratory setup

Figure 2 shows the experimental optical scheme incorporating PM according to the technique for measuring both the characteristics of internal reflection and the polarization difference of the reflection coefficients for s - and p -polarized light [19,20].

An LGN-113 He–Ne laser with fixed wavelengths of 0.63 and 1.15 μm and an MDR-4 monochromator with a KGM-150 halogen lamp equipped with a linear polarizer are used as sources of linearly polarized light. The monochromator makes it possible to generate light with a wavelength in the range 0.4–2 μm and a minimum intensity no less than 10^{12} photons per cm^2s at a slit width of 0.1 mm. In both cases, the direction of linearly polarized light makes an angle of 45° with the modulator axes. A photoelastic modulator PEM transforms the wave polarization in such a way that during this period, the electric field azimuth changes its position twice, from parallel to perpendicular relative to the plane of light incidence.

Depending on the experiment, the modulator acts as an *analyzer* or *converter* of the light polarization state. These two components are separated by the ar-

range of the modulator in the optical scheme as follows. If the modulator coupled with a linear polarizer is placed behind the sample, we have an analyzer of the polarization state and can measure the *phase anisotropy* (see inset in Fig. 2). In this case, the performance of the polarization analysis of light becomes more simple and convenient [21]. It is well known that the *polarization analysis* of light is a decomposition procedure for an elliptically polarized electromagnetic wave into the Stokes components. Among them are one circular component (V) which is determined by the phase difference between the orthogonal components of linearly polarized waves and characterizes the magnitude of the phase anisotropy, and two linear components (Q and U) which are differences between the corresponding radiation intensities and characterize the amplitude anisotropy.

To register the Q component, it is easy to measure an analogous parameter such as the polarization difference $\Delta\rho$ [22]. To this end, the modulator should be placed in front of the sample. In this case, we have a converter of the polarization state and can measure the *amplitude anisotropy*. So, after interaction with the prism and resonantly sensitive film on the prism working surface, the light is directed to the photodetector (PD; a silicon or germanium photodiode). The light absorbed by the PD generates a signal having a variable component at the doubled modulation frequency in the first case (half-wave plate) and at the modulation frequency in the second case (quarter-wave plate). This component is proportional to the reflection coefficient difference for the p - and s -polarizations. This latter case has some advantages over the former one, namely, a more acceptable operating mode of the modulator power supply and the possibility to use a more sensitive PD.

Gold layers of different thicknesses are deposited onto the working surface of a total internal reflection half-cylinder. Since the reflection coefficients R_p and R_s are different, the signal is proportional to their difference:

$$\Delta I \approx (R_s^2 - R_p^2) \sin \omega t,$$

where t is time; this signal can be measured from the PD by a selective nanovoltmeter with a synchronous-phase detector at the modulation frequency ω .

In the same version, the arrangement of a linear polarizer G before the PD leads to modulation of the intensity of the polarized light beam with respect to which the polarizer is oriented for transmission. Then, the intensity measurement makes it possible to obtain the SPR characteristic in the conventional form, i.e., as a narrow angular dependence of the reflection coefficient for p -polarization only, $\Delta\rho = R_p^2$, whose amplitude becomes minimal at resonance excitation.

As for the practical significance, the detective ability of the PM technique with respect to the amplitude and phase anisotropy is 10^{-4} and 10^{-6} , respectively, which is very important for any current sensor devices.

2.3. Objects under investigation

Thin gold films (Au 99.999%) 5 to 200 nm thick were deposited on a quartz substrate at room temperature by thermal evaporation from a molybdenum heater

in vacuum (10^{-3} Pa). The deposition rate was $1\text{--}1.5\text{ nm s}^{-1}$ and the film thickness was defined by the deposition time. Quartz plates 1 mm thick served as substrates. The free surface of the substrate is in contact with a half-cylinder segment of the same material on an immersion liquid (glycerin), and they form a half-cylinder. More careful measurement of the film thickness was performed by atomic force microscopy (AFM) NanoScope IIIa (Digital Instrument, USA) when studying sample topology. To this end, we use the mode of periodic contact of the film surface with a silicon needle (rated needlepoint up to 10 nm). The 5-nm-thick film had a granular structure whose grain size correlates with thickness. The films with thickness over 10 nm had a smooth and continuous structure with low roughness. Some of these films were subjected to heat treatment at temperatures of 120°C and 230°C in air for different time intervals.

Thin gold films are also obtained by using pulsed-laser deposition on quartz substrates. The beam of a YAG: Nd³⁺ laser (wavelength of 1.06 μm , pulse energy of 0.2 J, pulse duration of 10 ns, and pulse repetition of 25 Hz) scanned target in a vacuum chamber in an argon atmosphere at a pressure of 10–20 Pa. We obtained two kinds of film. The first one was a thin gold film formed using an Au-target and deposited from a direct high-energy flux of erosion torch particles. The substrate is located normally to the torch axis at a distance of 20 to 25 mm from the target. The density of the irradiation energy is 5 J cm^{-2} . The thicknesses of these films do not exceed 50 nm. The second type of film represents a composite nanostructure including Au nanoparticles in an Al₂O₃ matrix. It is formed from the reverse low-energy flux of the erosion torch particles. The substrate is placed in the target plane at a distance of 5–10 mm from the torch axis. The target consists of gold and aluminum bits. The concentration of gold varies from 20 to 50%. The density of the irradiation energy is changed within $5\text{--}20\text{ J cm}^{-2}$. Some of these films are subjected to heat treatment at a temperature of 400°C for 10 minutes. The surface morphology of the films of the first type is rough and consists of nanoparticles with different lateral dimensions and considerable height dispersion. The films of the second type are much more homogeneous, without clusters on the surface. They are characterized by smaller grain sizes. The films of the first type are single-phase with low porosity; whereas, the porosity in the films of the second type is significant (20%–40%). The films of the second type represent three-phase systems including Au nanoparticles in Al₂O₃ matrices and air.

3. Results and discussion

3.1. Polarization modulation characteristics of nanosized gold films

Figure 3 shows the dependence of the experimental value of polarization difference $\Delta\rho$ on the angle of incidence of light for the Au film of thickness $d = 50$ nm at a wavelength of $\lambda = 632$ nm. The curves of $R_p^2(\theta)$ and $R_s^2(\theta)$ obtained by angular scanning of a prism with a resonantly sensitive film by p - and s -polarized beams are also shown here for comparison. It can be seen that the signal amplitude becomes maximal at the resonance angle θ_{res} , i.e., the angle at which the quantity

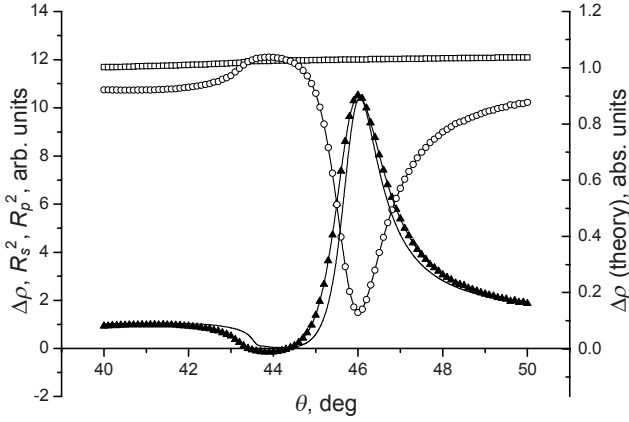


Fig. 3. Dependence of the reflection coefficients of (\square) s - and (\circ) p -polarized light (R_s^2 and R_p^2 , respectively), and (\blacktriangle) polarization difference $\Delta\rho$, and (solid line) the calculated dependence of $\Delta\rho$ on the angle of incidence of light for a gold film with $d = 50$ nm. The calculation was performed using the values $L = 11$ nm, $\lambda = 632$ nm, $N_0 = 1.456$, $N_2 = 1.003$, and Eq. (2) for $N_1 = n_1 - i\kappa_1$ (after [22]).

$R_p^2(\theta)$ generally becomes minimal [23]. It should be emphasized that permutation of terms in Eq. (2) affects only the signal sign and is of little importance in synchronous phase detection. The amplitude of this signal is phase independent because the signal is measured with respect to one of the $R_s^2(\theta)$ or $R_p^2(\theta)$ curves that plays the role of the abscissa axis in this measurement technique. Note that the $R_p^2(\theta)$ and $\Delta\rho(\theta)$ curves are not mirror reflected with respect to some horizontal axis, since the $R_s^2(\theta)$ curve exhibits dispersion near the resonance angle, which becomes especially pronounced with a decrease in the film thickness. This dispersion is related to the sharp kink in the $R_s^2(\theta)$ curve at the critical angle in the absence of a metal film that is smoothed out when the film thickness increases.

Figure 3 demonstrates also a good agreement between the experimental and calculated (from Eq. (2)) characteristics of the polarization difference $\Delta\rho$ for a film of the same 50-nm thickness. We suggest that the optical properties of metal films, such as the refraction and extinction coefficients, decay with thickness exponentially. The following functions were used in this model:

$$n_1 = 0.2 + \exp(-0.2 - d/L), \quad \kappa_1 = 3.6[1 - \exp(-d/L)]. \quad (3)$$

Here the characteristic length L plays the role of a fitting parameter. Its value is determined from the condition of best agreement between the calculated and experimental data at a wavelength $\lambda = 0.632$ μm to be $L = 11.0 \pm 0.5$ nm. The amplitude values of the optical constants n_1 and κ_1 were referenced to vacuum values at zero film thickness; with an increase in the film thickness, they were extrapolated to the bulk values reported in [24]. For the wavelength used in our experiment, they are $n_1 = 0.2$ and $\kappa_1 = 3.6$.

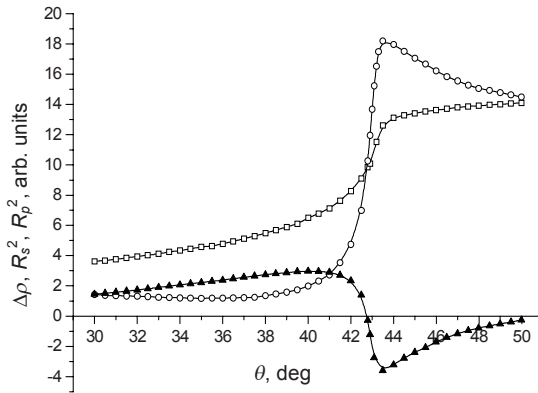


Fig. 4. Dependence of the reflection coefficients for (\square) s - and (\circ) p -polarized light (R_s^2 and R_p^2 , respectively) and (\blacktriangle) polarization difference $\Delta\rho$ on the angle of incidence of light for a Au film with $d = 5$ nm; $\lambda = 632$ nm (after [22]).

To interpret the negative value near the $\theta = 44^\circ$ angle (Fig. 3), we use the angular dependence of all three polarization characteristics, $R_s^2(\theta)$, $R_p^2(\theta)$, and $\Delta\rho(\theta)$ measured for one of the thinnest samples studied. To explain this feature, Fig. 4 shows separately the dependences of the coefficients of reflection of s - and p -polarized light from the prism with a 5-nm-thick Au film. It can be seen that these curves cross each other near the critical angle ($\theta \approx \theta_{cr}$) due to the violation of the conditions of total internal reflection for the metal film. Note that only the s -polarized wave exhibits the anomalous reflection that causes the negative sign in the polarization difference $\Delta\rho(\theta)$ that is also shown in Fig. 4. The range of angles at which the polarization difference becomes negative depends on the film thickness as well. The lower limit of this range corresponds to the critical angle $\theta_{cr} = 43^\circ$, with an extreme amplitude at $d = 5$ nm, while the upper limit is extended to the value corresponding to beam grazing with a decrease in the film thickness. In addition, these negative amplitude values of polarization difference depend on the sample thickness non-monotonically.

The evolution of the SPR angular dependence, which is caused by changes in the thickness or nature of the metal film or changes in the dielectric properties of the external medium, measured by the classical method, has been fairly well studied [23]. However, the shape of the characteristics of the polarization difference differs significantly from that of these dependencies. This fact stimulates new consideration of the influence of sample thickness on the evolution of the PM dependence. To this end, we consider the angular dependence $\Delta\rho(\theta)$ for samples with thicknesses $d = 0, 5, 23, 35, 50, 90, 120,$ and 200 nm. These results (Fig. 5) are more convenient for visual analysis when compared to theoretical quantities that are presented in the form of a set of curves calculated from Eq. (2) and shown in the 3D inset in Fig. 5 for the angular range $\theta = 40^\circ - 50^\circ$ in the vicinity of the plasmon resonance and the range of film thicknesses from 0 to 100 nm.

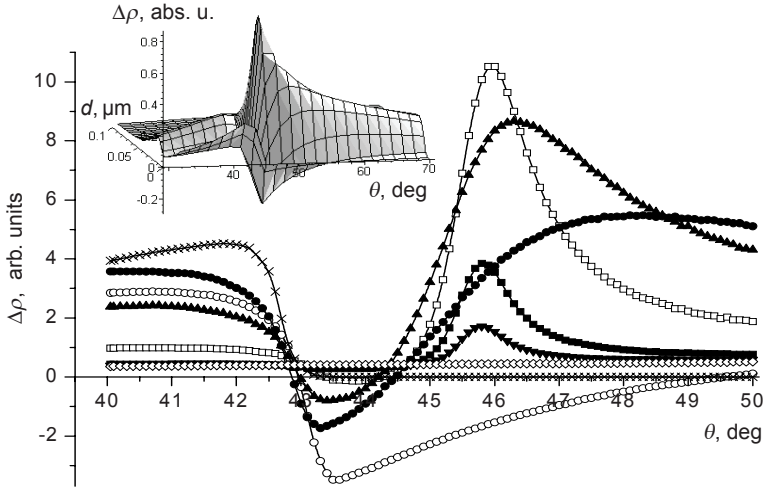


Fig. 5. Dependence of the polarization difference $\Delta\rho$ on the angle of incidence of light for the Au films with thicknesses of (\times) 0, (\circ) 5, (\bullet) 23, (\blacktriangle) 35, (\square) 50, (\blacksquare) 90, (\blacktriangledown) 120, and (\diamond) 200 nm; $\lambda = 632$ nm; the inset shows the three-dimensional dependence of $\Delta\rho$ on the same parameters (θ , d), calculated from Eq. (2) (after [22]).

A comparison of the plots in Fig. 5 demonstrates qualitative agreement between the calculated and experimental data. Concerning quantitative agreement, a theoretical dependence can be put into correspondence with each individual experimental curve. Also this series of angular dependences of $\Delta\rho(\theta)$ for different film thicknesses demonstrates the *classical size effect*. It is known that the abnormal phenomena appear in the optical properties of metal films when the film thickness becomes comparable to the mean free path of electrons.

It is necessary to note some characteristic features of this set. Two of them are related to the positive extrema of the parameter $\Delta\rho$ that are located on each side of the value $\theta_{cr} = 43^\circ$. The set of extrema that are located near $\theta_{res} \cong 46^\circ$ within a narrow angular range illustrates resonant interaction of light with surface plasmon-polaritons. This first component of the angular polarization difference $\Delta\rho(\theta)$ is referred to as the resonance component. The shape of the curves at $\theta < \theta_{cr}$ is determined by the properties of incomplete internal reflection. The second component is revealed when the increase of metal thickness reduces the peculiarities of attenuated internal reflection of light. The third feature of the angular dependences of the polarization difference is that they, along with the SPR component, contain a component due to the conventional reflection from the metal. The two components last mentioned are referred to as non-resonance components of the polarization difference.

An analysis of Eq. (2) that describes the SPR phenomenon shows that both the magnitude of the negative polarization difference $\Delta\rho$ and its distribution in the an-

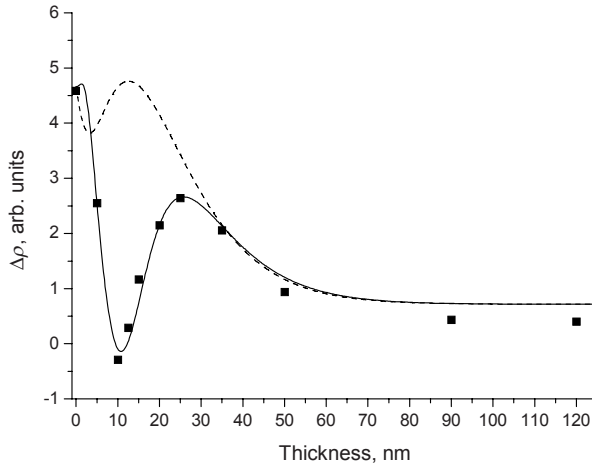


Fig. 6. Dependence of the polarization difference $\Delta\rho$ on the thickness d of a Au film at an angle of incidence of light $\theta = 42^\circ$ and $\lambda = 632$ nm: (■) experimental data (solid) and results of the calculations (dashed), taking into account and disregarding the dependence of the optical constants on the film thickness. The parameters are the same as in Fig. 3 (after [22]).

gle–thickness coordinates depend to a great extent on the dependence relating the permittivity and thickness of the metal film. Figure 6 shows the dependence of the amplitude values of $\Delta\rho(\theta)$ at $\theta = 42^\circ$ (less than θ_{cr}) on the thickness of the Au film on the prism surface. Here the squares are the experimental data, and the solid and dashed curves show, respectively, the results of the calculations performed taking into account and disregarding the dependence of the optical constants on the film thickness. For such a dependence, we use the dependence of the refractive and absorption indices of films on their thickness d from Eq. (3) at $L = 11$ nm. This extrapolation is justified by a more-than-satisfactory agreement between the experimental data and the solid curve. The physical meaning of the parameter L can be interpreted as the film thickness at which an increase results in the bulk properties. In other words, this is an average value in the dependence between the optical constants of gold films and their thickness. This situation can manifest itself, for example, in the fact that the electromagnetic interaction of a wave with plasma, characterized by a set of parameters (one of which is the mean free path of electrons under the field action), becomes independent of the film thickness.

The non-resonance component of light reflection from a metal is most pronounced for relatively thick films in which the resonant part of the angular characteristic $\Delta\rho_{spr}(\theta)$, due to its smallness, does not distort the angular dependence of the metallic reflection $\Delta\rho_{met}(\theta)$. The angular dependence $\Delta\rho_{spr}(\theta)$ obtained for the Au film with $d = 120$ nm at $\lambda = 0.6$ μm is shown in Fig. 7 (circles), in comparison with the dependence $\Delta\rho_{met}(\theta)$ obtained by measuring the external reflection from the same film (squares) [25]. Recall that, in the case of external reflection, SPR

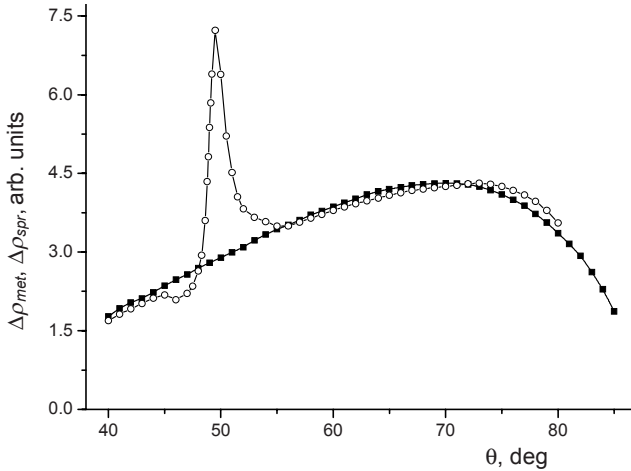


Fig. 7. Angular dependence of the polarization differences $\Delta\rho(\theta)$ of the internal $\Delta\rho_{spr}$ (○) and external $\Delta\rho_{met}$ (■) reflection for a Au film with $d = 120$ nm; $\lambda = 600$ nm (after [25]).

does not manifest itself, and PM reveals only the polarization difference $\Delta\rho$ that decreases with an increase in the reflection coefficients for the s - and p -polarizations. Note that both characteristics are shown in relative units obtained by normalization of the magnitudes to the same value. It is noteworthy that the observed discrepancy between the curves $\Delta\rho_{spr}(\theta)$ and $\Delta\rho_{met}(\theta)$ (Fig. 7) is not due to the measurement error; being highly reliable, this difference has a physical meaning that is as follows.

The relatively simple dependence $\Delta\rho_{met}(\theta)$ is determined by two characteristic parameters: the amplitude at maximum and its angular position. These parameters are related by the corresponding expressions to the refractive and absorption indices of the sample. One can easily make sure of it using the conventional Fresnel equations for external reflection of a material with a complex refractive index. Varying the refractive and absorption indices, one can obtain characteristics without the SPR component that are in satisfactory agreement with the curves in Fig. 7, at least, in their monotonic parts. Their analysis makes it possible to obtain additional data on the relationship between the dielectric function and the metal film thickness, which is a separate problem. The results of measuring the internal reflection parameters reported in the form provided by applying the PM technique are not only convenient for analysis but also help expand the experimental possibilities as well. This is facilitated by the ability to observe the resonant component against the background of a small total signal $\Delta\rho(\theta)$ due to its amplification.

Amplification becomes necessary when the SPR component in the characteristic becomes vanishingly small for a particular reason, like an increase in the metal film thickness or a change in the wavelength. Nevertheless, even against the background of a small polarization difference $\Delta\rho(\theta)$ for reflection from metal, this quantity remains reliably detectable.

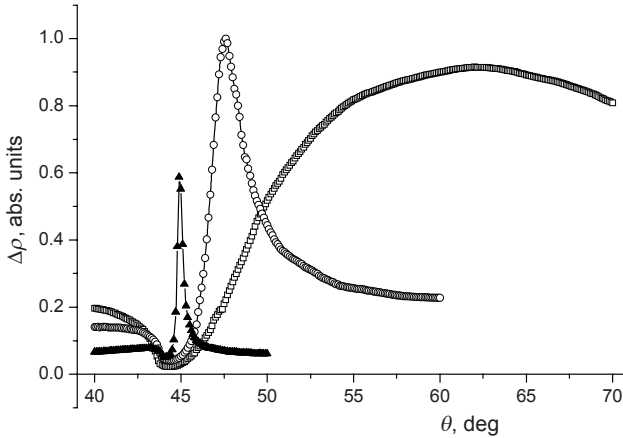


Fig. 8. Angular dependences of polarization difference $\Delta\rho(\theta)$ for a Au film with $d = 50$ nm at the wavelengths of (\square) 500, (\circ) 600, and (\blacktriangle) 900 nm (after [22]).

PM techniques make it possible to investigate the evolution of the angular dependences $\Delta\rho(\theta)$ in the wide wavelength range $\lambda = 0.4 - 2 \mu\text{m}$. The results of the measurements at three wavelengths (Fig. 8) give an idea about the change in the shape and peak-to-peak ratio of the $\Delta\rho(\theta)$ dependence with an increase in the wavelength. The angle θ_{res} corresponding to the maximum amplitude $\Delta\rho$ decreases with wavelength and tends to a constant value at θ_{cr} without exceeding it. This trend allows us to obtain the experimental dispersion SPR characteristic relating the plasmon resonance frequency and wave vector for the Au film with a thickness of 50 nm (Fig. 9). This characteristic for a given sample, its nature and shape, is typical of SPR and is described by a certain set of parameters, like the ranges of variation in the function and argument.

3.2. Size-dependent effects of internal reflection in gold-cluster films

Most nanosized metal films do not have a uniform layer of metal [26], but rather have an inhomogeneous cluster structure. The optical properties of such films are determined by the nature of the light scattering from individual clusters or nanoparticles that may be entirely separated from or slightly touching each other [27]. The interaction of electromagnetic radiation with these particles has specific features due to the relation of their transverse dimensions to the electron mean free path. In other words, the relation of the relaxation time of a system of free electrons to the reciprocal frequency of a wave electric field determines the degree of response of the cluster electronic system to the external radiation. Hence, if the metal film thickness is comparable with the electron mean free path or de Broglie wavelength then the anomalous optical properties of the films appear; i.e., classical and quantum size effects take place, respectively [28,29]. A

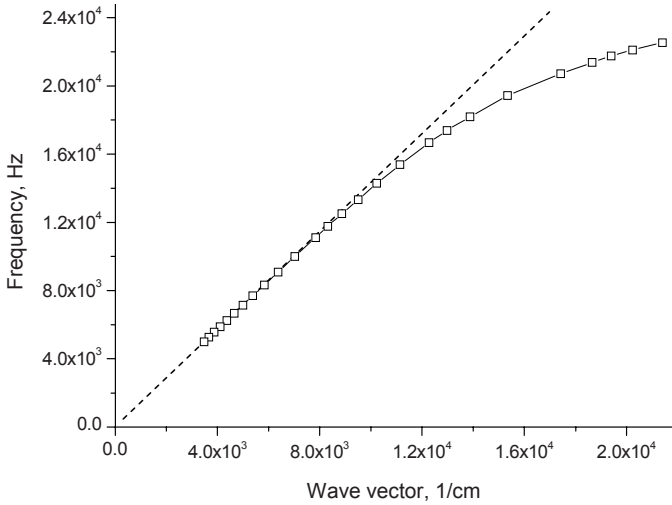


Fig. 9. Dispersion characteristic of SPR for a Au film with $d = 50$ nm: (\square) experimental data and (dashed line) light in vacuum (after [22]).

critical thickness of 11 nm was obtained in [30] from transmission and absorption spectra of gold films. By changing the wavelength of incident radiation and recording the reflection (transmission) of light by a film of a fixed thickness (or by changing the film thickness at a fixed wavelength), one can differentiate the effects of size on these films that may include the *classical size effect* [31] as well as its modification associated with the surface shape or cluster sizes (the *topological size effect* [32]). The specific features of polarization-dependent SPR effects in ultrathin gold films, in particular, in cluster structures are poorly studied. That is why they are of interest in the view of PM technique investigations [33].

The Fresnel formulas that strictly determine the propagation of light through interfaces of homogeneous media cannot be used in the case of ultrathin films. The known approximate models [34] based on the introduction of effective properties to analyze the light-matter interaction may be useful but cannot adequately describe this process and, hence, are not sufficient to determine the dielectric properties and topology of an object. The problem is also complicated by the fact that, as will be shown below, the dielectric properties and SPR in clustered metal films are determined by their topological characteristics rather than thickness.

Figure 10 shows the angular dependences $R_s^2(\theta)$, $R_p^2(\theta)$, and $\Delta\rho(\theta)$, at a wavelength of $\lambda = 1.15 \mu\text{m}$ for an Au film with a thickness $d = 5$ nm. For comparison, a typical dependence $R_p^2(\theta)$ for a sample with a thickness of 50 nm that is optimal for SPR [4] is also shown. By comparing the $R_p^2(\theta)$ curves for these samples, we can see that these curves have important properties that were not identified previously. First, their positive extrema near the angle θ_{cr} are almost the same despite the fact that the film thicknesses differ by an order of magnitude. Second, the angular positions of these maxima coincide. And third, both curves tend to decrease

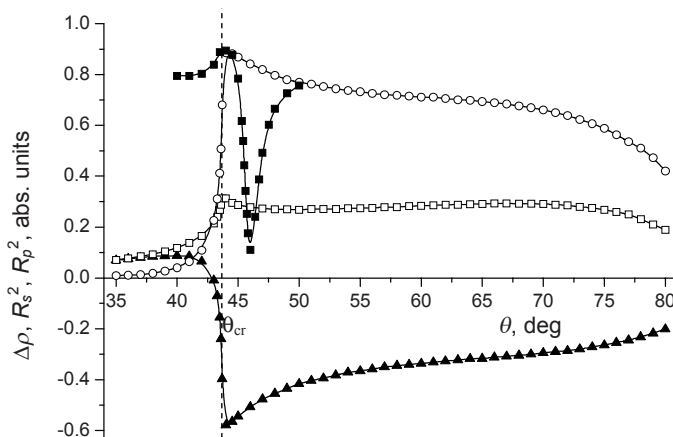


Fig. 10. Dependence of the reflection coefficients for (\square) s - and (\circ) p -polarized light (R_s^2 and R_p^2 , respectively) and (\blacktriangle) polarization difference $\Delta\rho$ for a Au film with $d = 5$ nm on the angle of incidence of light; (\blacksquare) Similar dependence of $R_p^2(\theta)$ for a Au film with $d = 50$ nm; $\lambda = 1150$ nm (after [33]).

with increasing angle. This tendency is typical of the resonance interaction of a wave with an electron gas. Although the degree of decay and the band width of these dependences are different, each of them reflects the same condition of the interaction of light with the electron system.

From the condition following from the equality of wave vectors of the incident wave and the surface plasmon [5], we have for the case of a metal film with a geometrically plane surface:

$$\sqrt{\varepsilon_g} \sin \theta_{res} = \sqrt{\varepsilon'(\omega)\varepsilon_a / [\varepsilon_a + \varepsilon'(\omega)]}.$$

Here ε' is the real part of the complex permittivity of gold, $\varepsilon = \varepsilon' + i\varepsilon''$, ε_g and ε_a are the permittivities of the prism material and ambience, respectively; and the angle θ_{res} corresponds to the minimum of the $R_p^2(\theta)$ curve. However, when passing from a film that is homogeneous throughout its thickness to a film consisting of clusters, this relation, as well as the Fresnel equations, are inadequate. Nevertheless, external radiation can also excite an SPR in a cluster (island) film; the only difference is that the excited surface plasmon in this case is localized in a region of the particle surface such that the normal makes an angle θ_{res} with the incident wave vector. It is easy to imagine that if clusters were hemispherical, then, in some region of incidence angles, it would be possible to find a plane surface on this hemisphere that satisfies the dispersion condition. In this case, the narrow resonance band of the $R_p^2(\theta)$ function (black squares in Fig. 10) spreads over a wide angular range and transforms into the curve with open circles. Because of this, the amplitude measured from the maximum of the total internal reflection coefficient, $R_p^2(\theta) = 1$, decreases.



Fig. 11. Angular dependences of polarization difference $\Delta\rho(\theta)$ for a Au film with $d = 5$ nm at the wavelengths of (\times) 450, (\square) 500, (\blacktriangle) 600, (\blacksquare) 800, (\diamond) 900, (\bullet) 1000, (\circ) 1200 nm. The inset shows the surface topology of this film (after [33]).

Note that this situation also agrees with the Heisenberg uncertainty relation $\Delta x \Delta p \geq h/2$ applicable to quantum particles, including surface plasmons. In particular, the localization of a surface plasmon on a decreasing area of the cluster, corresponding to the approximately linear part of the hemispheric surface, leads to an uncertainty in the plasmon wave vector and, hence, to a spread in the direction and magnitude of the excitation wave vector. This, in turn, implies that the resonance condition for the interaction of light with the film electronic system is satisfied in a wide range of incidence angles.

The hemispherical approximation of the cluster shape also explains the resonant nature of the interaction of the *s*-polarized light with the film (see Fig. 10). This is evidenced by the fact that the angular dependence of the reflection coefficients $R_p^2(\theta)$ and $R_s^2(\theta)$ differ only quantitatively, which is indicated by the polarization difference $\Delta\rho$. Therefore, from the angular dependence of the polarization difference $\Delta\rho(\theta)$, one can evaluate the surface plasmon excitations by both *s*- and *p*-polarized electromagnetic radiation. In this case, it seems reasonable to consider the evolution in the behavior of the polarization difference $\Delta\rho$ caused by a change in the film topology, first of all, by a change in the size of its clusters. In practice, it is more convenient to compare the $\Delta\rho(\theta)$ dependence obtained at a fixed sample topology and different wavelengths. From the viewpoint of the relation between the cluster size and the electron mean-free path, which is determined by the action of the wave field at a distance corresponding to one wave period, or wavelength, these two methods are equivalent to a first approximation.

The angular dependences of the polarization difference $\Delta\rho(\theta)$ were measured on a sample of Au film with a thickness $d = 5$ nm over a wide wavelength range,

from 450 to 1200 nm (Fig. 11). The investigation of the film topology showed (inset in Fig. 11) that the film consisted of separate cylindrical clusters with diameters of 30–80 nm and heights of 5–25 nm. In the $1 \times 1\text{-}\mu\text{m}$ area, the height drop was 30.24 nm and the root-mean-square (RMS) roughness was 4.95 nm.

Initially, we obtained a set of crossing curves, corresponding to different wavelengths of incident light that differed in both amplitude and shape. For greater convenience, all curves were normalized to a constant value of extrema. The observed small shift of the extremum to larger angles with increasing wavelength agrees with the dispersion of the refractive index of the half-cylinder material. However, the most important fact is that the curve corresponding to the polarization difference for the wavelength $\lambda = 800$ nm is, unlike the rest of the curves, linear in a wide region. The other curves can be divided into two groups that fall on both sides of the linear dependence. Taking into account such behavior, we have a system that can be described by opposite signs of curvature. As is shown by our analysis, the sign of curvature reflects the resonance nature of the interaction of differently polarized light with the electron system. Thus, the positive curvature of $\Delta\rho(\theta)$ characterizes the resonance mechanism of simultaneous interaction of s - and p -polarized light with clusters. The $\Delta\rho(\theta)$ dependence with the negative curvature is related to the resonance solely with the p -polarized light. Finally this series of angular dependencies of $\Delta\rho$ in a wide wavelength range demonstrates the *topological size effect*.

To substantiate and systematize our results, we measured all three polarization characteristics of an Au film with $d = 10$ nm in the initial state and after thermal annealing. It is known that annealing of a thin film results in aggregation of the metal in balls [35]. Owing to this fact, the film turns, partially or completely, into an island having a constant mass thickness. This results in violation of not only the total internal reflection but also the reflection at angles smaller than the critical one.

The corresponding film topologies in the initial state and after annealing are shown in Fig. 12a,b. The initial Au film was solid, with a surface consisting of grains 20–45 nm in diameter and 0.5–2 nm in height, with fuzzy boundaries. The film was very smooth, the height drop was 6.78 nm, and the RMS was 0.91 nm. Annealing at a temperature of 230°C in air radically changed the film structure. Similarly to [33], the film was structured into separate clusters of irregular shape, with increased diameters (30–50 nm) and heights (6–30 nm). The roughness characteristics were close to those of the initial 5-nm film; the height drop was 35.89 nm and RMS roughness was 5.69 nm.

The results of initial film measurements (Fig. 13a) show clear evidence of surface plasmon–polariton resonance interaction for p -polarized radiation only (opened circles) and almost complete lack of reflection of s -polarized radiation (opened squares). An increase in the reflection coefficient of s -polarized radiation after annealing (Fig. 13b) is indicative of the typical SPR interaction of light with the cluster electron system. Moreover, the sharpness of the extremum of $R_p^2(\theta)$ that characterizes the resonance degree is reduced after annealing. This change,

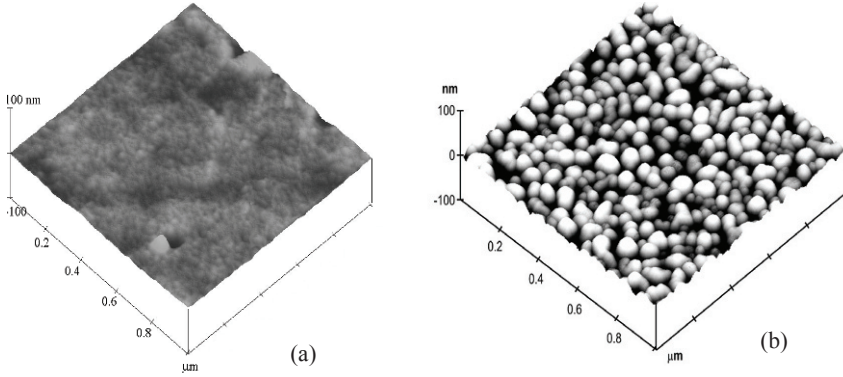


Fig. 12. Surface topology of a Au film with $d = 10$ nm (a) in the initial state and (b) after annealing (after [33]).

along with the unchanged curve shape, is explained by a decrease in the effective film surface due to the hemispherical surfaces that do not satisfy the above relation. The negative sign of curvature of the $\Delta\rho(\theta)$ curve in the range of angles exceeding the critical one is also changed to positive, which is more evidence for SPR excited by s -polarized light. Also the polarization characteristics and the AFM images of the film with $d = 5$ nm in the initial state and the film with $d = 10$ nm after annealing are similar. This is shown by coincidence of the curve with diamond symbols of the angular polarization difference $\Delta\rho(\theta)$ for the sample with $d = 10$ nm and the same curve for the film with $d = 5$ nm.

For a more in-depth study of the transition process from classical to topological size effect, we investigated the gradual influence on the polarization characteristics of the surface structure using gradual annealing. Three Au films with the

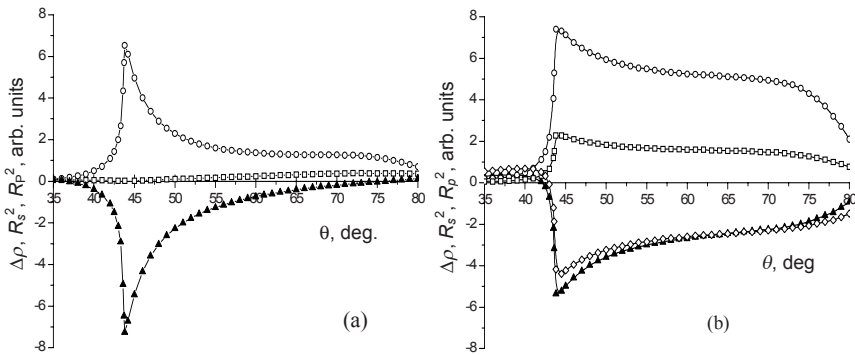


Fig. 13. Dependence of reflection coefficients for (\square) s - and (\circ) p -polarized light (R_s^2 and R_p^2 , respectively) and (\blacktriangle) polarization difference $\Delta\rho$ for a Au film with $d = 10$ nm on the angle of incidence of light: (a) in the initial state and (b) after annealing. A similar dependence $\Delta\rho(\theta)$ is seen for a Au film with $d = 5$ nm at $\lambda = 1150$ nm (\diamond) (after [33]).

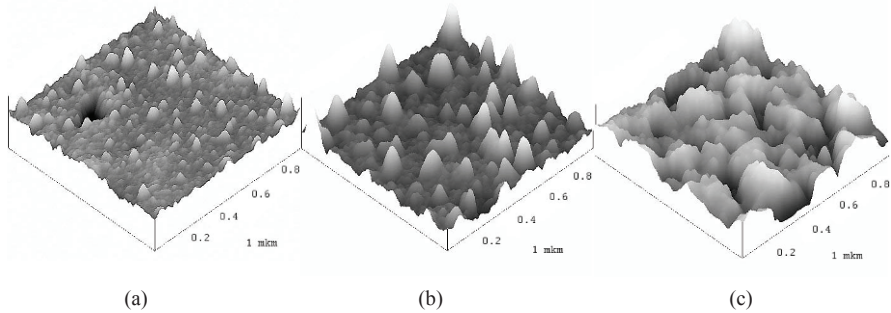


Fig. 14. Surface topology of a Au film with $d = 20$ nm (a) in the initial state, (b) after annealing at $T = 130^\circ\text{C}$, and (c) after annealing at $T = 230^\circ\text{C}$.

same thickness, $d = 20$ nm in the initial state, had similar angular characteristics of $\Delta\rho(\theta)$. The surface structure of the first film (sample 1) remained unchanged. Restructuring of the second and third surfaces was performed by annealing at temperatures of 130°C (sample 2) and 230°C (sample 3). The corresponding topologies are shown in Fig. 14. Sample 1 was solid and homogeneous, with a surface RMS of 1.91 nm. The film was very smooth (Fig. 14a). Annealing of sample 2 led to increasing roughness and, as a result, its surface consisted of grains 35–80 nm in diameter and 5–13 nm in height. Moreover the height drop was 7.41 nm and the RMS was 2.33 nm (Fig. 14b). Further annealing of sample 3 led to the formation of a porous structure and separate clusters 80–180 nm in diameter and 18–25 nm in height. The height drop was 21.73 nm and the RMS was 7.24 nm (Fig. 14c).

In further consideration, it is necessary to mark another peculiarity in the behavior of the polarization difference $\Delta\rho(\theta)$ [36]. Figure 15 shows three angular characteristics of $\Delta\rho(\theta)$ for samples 1, 2 and 3. There are two intersection points of $\Delta\rho$ with the abscissa axis at some angles that lay on each side of the critical angle θ_{cr} . At these points the polarization difference is equal to zero and reflection does not depend on the polarization state of the wave. Therefore, the *isotropic reflection* condition for s - and p -polarized radiation takes place, with $R_s^2 = R_p^2$. This peculiarity cannot be described by the Fresnel formulas, either its position in the angular dependence or especially the number of intersections. One of them (located at $\theta < \theta_{cr}$) is caused by the presence of an absorbing medium on the surface of the totally reflecting prism. Its origin is as follows. Action of the electric field of the s -polarized wave that is not restricted by the surface results in stronger absorption of the energy of the s -polarized wave in the absorbing medium as compared to that of the p -polarized wave. The decrease of the reflection coefficient is the reason for the anomalous reflection. Another isotropic point in the angular dependence of the polarization difference (located at $\theta > \theta_{cr}$) is due to the resonance interaction of radiation with the electron system of the cluster. Its angular position depends on the resonance condition of the interaction between the cluster and ra-

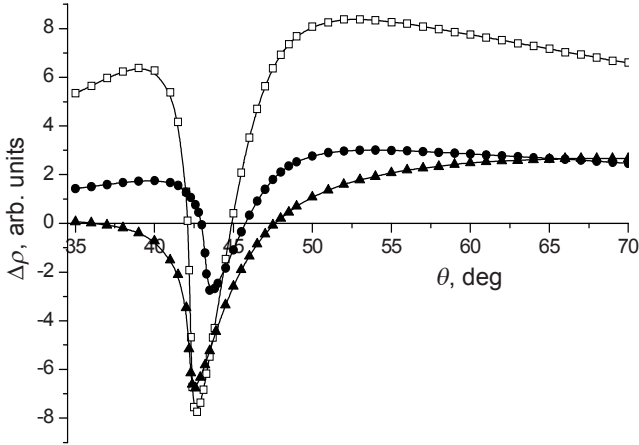


Fig. 15. Angular dependence of the polarization differences $\Delta\rho(\theta)$ for a Au film with $d = 20$ nm (\square) in the initial state, (\bullet) after annealing at $T = 130^\circ\text{C}$, and (\blacktriangle) after annealing at $T = 230^\circ\text{C}$; $\lambda = 630$ nm.

diation with p -polarization or both s - and p -polarizations. Also this point is very sensitive to any changes, for example, to those of wavelength and structural features of films.

If the topology of a film surface changes into a cluster structure, the angle of isotropic reflection becomes sensitive to the degree of film clustering owing to changes in the ratio between the intensities of polarizations. Since it is difficult to give the definition of the degree of clustering, we will analyze the dispersion angle of isotropic reflection $\theta_{|\Delta\rho=0} = f(\lambda)$ that is located to the right of the critical angle ($\theta_{|\Delta\rho=0} > \theta_{cr}$).

Figure 16 shows the dependence of the angle of isotropic reflection on wavelength for Au films 1, 2 and 3 with the same thickness of 20 nm. One can see that the angle corresponding to the condition of equality of the reflection coefficients depends non-monotonically on the radiation wavelength. The same behavior of the curves in the wavelength range $\lambda = 400 - 700$ nm for the initial (film 1) and annealed at $T = 130^\circ\text{C}$ (film 2) samples characterizes their morphology as solid. Both curves also have a break, but at different wavelengths. For these dispersion curves $\theta_{|\Delta\rho=0}(\lambda)$ tends to the critical angle θ_{cr} with increasing wavelength in the range of $\lambda > 500$ nm. Therefore, the classical size effect dominates in gold films 1 and 2. For gold film 3, its dispersion characteristic of angle of isotropic reflection has a point of inflection at a wavelength $\lambda = 500$ nm, which can be treated as an indication of porous film structure. The common feature for all dispersion curves in Fig. 16 is their location on a scale of ordinates in the range of angles close to the critical angle θ_{cr} , with the common resonant mechanism of surface plasmon-polaritons excitation by p -polarized radiation only.

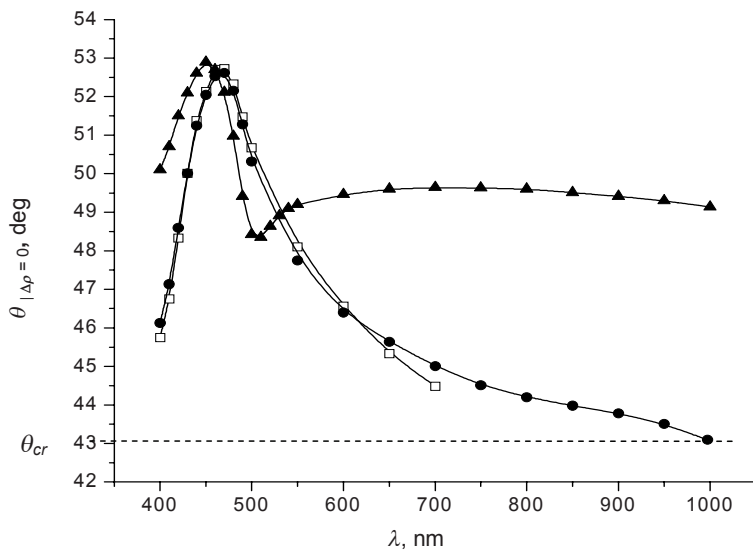


Fig. 16. Spectral dependence of the angle of isotropic reflection of light $\theta_{|\Delta\rho|=0}(\lambda)$ for a Au film with $d = 20$ nm (\square) in the initial state, (\bullet) after annealing at $T = 130^\circ\text{C}$, and (\blacktriangle) after annealing at $T = 230^\circ\text{C}$.

Similar research was performed for film thicknesses $d = 5, 10,$ and 20 nm before and after annealing at a temperature $T = 230^\circ\text{C}$. The results obtained on $\theta_{|\Delta\rho|=0}(\lambda)$ for these Au films are presented in Fig. 17. With the help of additional investigations of the angular characteristics of polarization difference, it is possi-

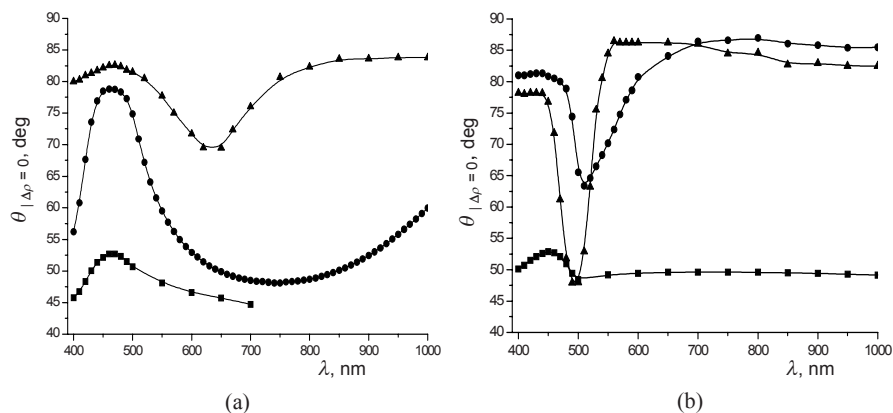


Fig. 17. Spectral dependence of the angle of isotropic reflection of light $\theta_{|\Delta\rho|=0}(\lambda)$ for Au films with thickness of (\blacktriangle) 5, (\bullet) 10, (\blacksquare) 20 nm: (a) in the initial state, (b) after annealing at $T = 230^\circ\text{C}$ (after [36]).

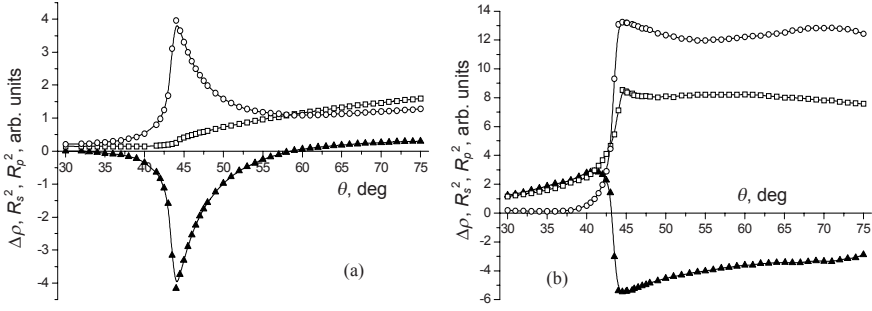


Fig. 18. Dependence of reflection coefficients for (\square) s - and (\circ) p -polarized light (R_s^2 and R_p^2 , respectively) and (\blacktriangle) the polarization difference $\Delta\rho$ on the angle of incidence of light for a Au film with $d = 20$ nm obtained by pulsed laser deposition (a) in the initial state and (b) after annealing; $\lambda = 630$ nm (after [37]).

ble to interpret the behavior of the curves. One can see that the curves corresponding to smaller thicknesses are arranged higher than those for the previously considered samples with thickness $d = 20$ nm. This peculiarity is associated with different thickness of Au films. The angular arrangement of these curves at $\lambda > 500$ nm can be treated as follows. On the one hand, if a curve tends to a glancing angle, then the excitation of surface plasmons will take place by both p - and s -polarized radiation. On the other hand, if the curve tends to the critical angle then the excitation of surface plasmons by p -polarized radiation only will occur. Moreover, at wavelengths $\lambda > 500$ nm, points of inflection take place for all curves.

In conclusion, one can say that in the first region the topological size effect is predominant, with excitation of surface plasmons in gold films having a cluster structure; whereas, in the second region the classical size effect dominates, with excitation of surface plasmon–polaritons in gold films having a homogeneous structure. Figure 17 illustrates the transition from the classical size effect to the topological one. However, in-depth determination of the physical origin of the dispersion of the isotropic reflection point $\theta_{|\Delta\rho=0}(\lambda)$ located to the left of the critical angle θ_{cr} remains beyond the scope of this chapter.

3.3. Mechanism of resonance interaction in gold–dielectric composite films

The above results from single-phase nanosized gold films obtained with the PM technique can provide new information about the surface plasmon excitation in gold–dielectric nanostructures [37]. For performing such experiments, composite nanostructures including gold nanoparticles in the aluminum oxide matrix were used (see Section 2.3).

At first, the characteristics of the angular polarization difference $\Delta\rho(\theta)$ at a wavelength $\lambda = 630$ nm were obtained, both before and after thermal annealing, for an Au film with thickness $d = 20$ nm prepared using the pulsed laser deposition technique (type 1, see Section 2.3). They are presented in Fig. 18. One can see the

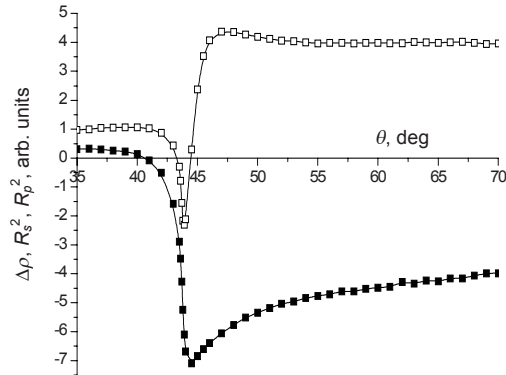


Fig. 19. Angular dependence of the polarization differences $\Delta\rho(\theta)$ for a Au film with $d = 20$ nm obtained by (\square) thermal evaporation and by (\blacksquare) pulsed laser deposition; $\lambda = 1150$ nm (after [37]).

change of sign curvature from negative (corresponding to the initial sample) to positive (corresponding to the annealed sample). After annealing, the dispersion in the curve $R_s^2(\theta)$ becomes more intense, and the magnitude of the resonance extremum in the curve $R_p^2(\theta)$ decreases correspondingly. This indicates that the condition of the surface plasmon excitation is fulfilled over the whole range of angles exceeding the critical one. Moreover, if the Au films obtained by thermal evaporation in vacuum have a mirror surface then these films deposited from the direct beam of the erosive torch are characterized by a rough relief caused by different sizes of separate granules, crystallized drops, and nanoparticles with a shape varying from spherical to ellipsoidal and spherical ones.

Besides, the comparison of the angular characteristics of $\Delta\rho(\theta)$ at a wavelength $\lambda = 1150$ nm for two Au films of the same thickness ($d = 20$ nm) demonstrates the pronounced difference in their structure (Fig. 19). One of these films was considered in Fig. 18 and another one, obtained by thermal evaporation in vacuum, was considered previously in Fig. 15. Such reflectance behavior confirms that the corresponding samples have cluster structure with high roughness and smoother homogeneous structure, respectively. We can consider this cluster film as a nanocomposite film consisting of gold nanoparticles and some dielectric component (for example, air). The penetration depth of the electromagnetic wave in the metal increases because of the presence of the dielectric component. As a result, the dispersion relations are fulfilled over the wider range of incidence angles.

Figure 20 shows all three angular polarization characteristics, $R_s^2(\theta)$, $R_p^2(\theta)$, and $\Delta\rho(\theta)$, at a wavelength $\lambda = 630$ nm for the nanocomposite film including Au nanoparticles in Al_2O_3 matrixes. One can see that the resonance characteristics of the interaction of s - and p -polarized radiation with electrons of Au nanoparticles is more pronounced in this film as compared with pure gold films. It is unusual that

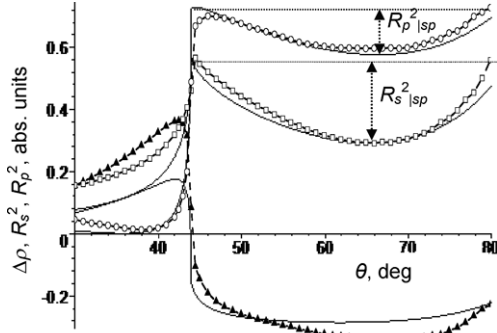


Fig. 20. Dependence of the reflection coefficients for (\square) s - and (\circ) p -polarized light (R_s^2 and R_p^2 , respectively) and (\blacktriangle) polarization difference $\Delta\rho$ on the angle of incidence of light for a Au/Al₂O₃ film obtained by pulsed laser deposition. A similar dependence is calculated using the values $N_0 = 1.445$, $N_2 = 1.003$, $\lambda = 630$ nm, $d = 100$ nm, $N_{1s} = n_{1s} - i\kappa_{1s}$, $N_{1p} = n_{1p} - i\kappa_{1p}$, $n_p = 1.53$, $\kappa_p = 0.065$; $n_s = 1.51$; $\kappa_s = 0.08$ (solid curves) (after [37]).

the intensity of the s -polarized wave is higher than in the case of the p -polarized wave based on their estimation from the ratio of the amplitudes at the critical angle and in the minimum region. This could be due to the anisotropy of the dielectric properties of Au nanoparticles as well as to their preferred orientation. In addition, as was shown in [38], the role of the substrate appears to be different for the two states of the polarized radiation at the expense of the polarization of the substrate by the wave field. In this case, the film under investigation is characterized by the anisotropy of dielectric properties, not only due to the oblique incidence of light, but also due to the macroscopically heterogeneous structural character.

The application of the Fresnel formulas to the description of the experimental results obtained for our media is possible only with regard for the following condition. To the film, we assign a certain indicatrix of dielectric properties based on the complex refractive index separately for s - and p -polarized radiations. Eventually, we describe satisfactorily all the three results of measuring the angular characteristics $R_s^2(\theta)$, $R_p^2(\theta)$, and so $\Delta\rho(\theta)$ by means of the Fresnel formulas with the use of multiparametric fitting of the components of the complex refractive index for two orthogonal polarizations. The fitting procedure appears to be simple as the directions of variation of the curve shape (i.e., the amplitudes of the extrema and their angular positions) differ in the case of independent variations of the refractive and absorption indices. In addition, we do not use the dependence of the optical constants on the film thickness, as was done in [22] according to Eq. (3), because its thickness is much larger than the characteristic parameter of this dependence. The best agreement between the results of calculations and measurements is reached if we assign the only possible combinations of the refractive indices $n_s = 1.51$, $n_p = 1.53$ and absorption indices $\kappa_s = 0.08$, $\kappa_p = 0.065$ to the composite film.

The obtained optical constants allow us to derive a conclusion about the role of Au nanoparticles in the optical parameters of a composite substance. Their presence in the dielectric medium results in a decrease of the effective refractive index of the composite film as compared to that of the Al_2O_3 matrix ($n = 1.765$ at $\lambda = 632$ nm). As to the absorption index, its value is lower than that in a homogeneous gold film ($\kappa = 3.6$ for the same wavelength). In other words, the optical constants of the system take on intermediate values between the components of the film. Their specific magnitudes are somehow related to the solution concentration.

4. Conclusions

For the first time, the use of the polarization modulation technique of electromagnetic radiation for investigations of the SPR effect in gold nanosized films provides the possibility to obtain information on the structural and morphological peculiarities of films with thicknesses up to 5 nm. New features of linear amplitude anisotropy of the internal reflection in the SPR effect were revealed that are related to the size effects in the homogeneous and cluster gold films as well as in gold–dielectric nanocomposite structures.

A new registration principle of the SPR effect is offered by measuring the polarization difference $\Delta\rho$ of the reflection coefficients for s - and p -polarized light analogous to the Q component of the Stokes vector. It is easy to determine the domination of metal or dielectric optical properties in metal films owing to the presence of resonant and non-resonant components in the angular polarization difference. It is found that the gold film thickness value of 11 ± 0.5 nm is the characteristic parameter of the exponential dependence of the refraction and extinction coefficients on the gold film thickness. This characteristic parameter is the transitional value of optical properties from dielectric to metal ones.

The PM technique that operates with spatially dependent azimuths of the electric field allows us to separate the anomalous behavior of the optical properties of homogeneous and cluster gold films associated with the simultaneous manifestations of both the classical and topological size effects. A qualitative model of transition between them in the nanosized gold films is presented. It is revealed that the sign of curvature of the angular dependence of the polarization difference $\Delta\rho(\theta)$ shows two different resonance mechanisms of interaction between radiation and the electron system of nanosized gold films. The surface plasmons are excited by p -polarized light or both s - and p -polarized light if a metal film has homogeneous or cluster structure, respectively. This effect is demonstrated for gold films having different degrees of clustering that is represented in the frequency dispersion characteristic of the angle of isotropic reflection where $\Delta\rho = 0$. These new features in the surface plasmon resonance effect can serve for testing the topological properties of the surface of thin absorbing films as well as for the determination of their optical parameters and thicknesses.

Moreover, for nanocomposite films with gold nanoparticles within a dielectric matrix, it is demonstrated that such films can be characterized by an anisotropy of

their dielectric properties, not only because of oblique incidence of light but also owing to their macroscopically heterogeneous structure.

References

1. Maier, S. A., and H. A. Atwater, 2005: *J. Appl. Phys.* **98**, 011101.
2. Raether, H., 1986: *Surface Plasmons on Smooth and Rough Surfaces and on Gratings* (Springer, Berlin).
3. Otto, A., 1968: *Z. Phys.* **216**, 398.
4. Kretschmann, E., 1971: *Z. Phys.* **241**, 313.
5. Agranovich, V. M., and D. L. Mills, Eds., 1985: *Surface Polaritons. Electromagnetic Waves at Surfaces and Interfaces* (North-Holland, Amsterdam) (also Nauka, Moscow, 1985).
6. Dmitruk, N. L., V. G. Litovchenko, and V. L. Strizhevsky, 1989: *Surface Polaritons in Semiconductors and Dielectrics* (Naukova Dumka, Kyiv) (in Russian).
7. Aoki, K., H. Saito, C. Itzstein, et al., 2006: *J. Clin. Invest.* **116**, 1525.
8. Chau, L.-K., Y.-F. Lin, S.-F. Cheng, and T. Lin, 2006: *Sens. Actuators B* **113**, 100.
9. Mar'enko, V. V., and S. N. Savenkov, 1995: *Opt. Spektrosk.* **78**, 614 (in Russian).
10. Born, M., and E. Wolf, 1999: *Principles of Optics* (Cambridge University Press, Cambridge).
11. Cardona, M., 1972: *Modulation Spectroscopy* (Mir, Moscow) (in Russian).
12. Jaspersen, S. N., and S. E. Schnatterly, 1969: *Rev. Sci. Instr.* **40**, 761.
13. Venger, E. F., I. E. Matyash, and B. K. Serdega, 2003: *Fiz. Tekh. Poluprovodn. (St. Petersburg)* **37**, 1188 (in Russian); *Semiconductors* **37**, 1160.
14. Matyash, I. E., and B. K. Serdega, 2004: *Fiz. Tekh. Poluprovodn. (St. Petersburg)* **38**, 684 (in Russian); *Semiconductors* **38**, 657.
15. Zykov, V. G., and B. K. Serdega, 1991: *Fiz. Tekh. Poluprovodn. (Leningrad)* **25**, 2173 (in Russian); *Sov. Phys. Semicond.* **25**, 1308.
16. Berezhinsky, L. J., E. F. Venger, I. E. Matyash, et al., 2005: *Fiz. Tekh. Poluprovodn. (St. Petersburg)* **39**, 1164 (in Russian); *Semiconductors* **39**, 1122.
17. Azzam, R. M. A., and N. M. I. Bashara, 1989: *Ellipsometry and Polarized Light* (North Holland, New York).
18. Gerrard, A., and J. M. Burch, 1975: *Introduction to Matrix Methods in Optics* (Wiley, London).
19. Berezhinsky, L. J., I. E. Matyash, V. P. Maslov, et al., 2007: UA Patent No. 26 551, Byull. Izobret., No. 15 (September 25).
20. Berezhinsky, L. J., E. F. Venger, V. P. Maslov, et al., 2007: UA Patent No. 80876, Byull. Izobret., No. 18 (November 12).
21. Berezhinsky, L. J., I. E. Matyash, S. P. Rudenko, and B. K. Serdega, 2008: *Optoelektron. Poluprovodn. Tekh.* **43**, 83 (in Russian).
22. Berezhinsky, L. J., L. S. Maksimenko, I. E. Matyash, et al., 2008: *Opt. Spektrosk.* **105**, 281 (in Russian); *Opt. Spectrosc.* **105**, 257.
23. Rengevych, O. V., Yu. M. Shirshov, Yu. V. Ushenin, and A. G. Beketov, 1999: *Semicond. Phys. Quant. Electron. Optoelectron.* **2**, 28.
24. Palik, E. D., Ed., 1998: *Handbook of Optical Constants of Solids* (Academic Press, New York).
25. Berezhinsky, L. J., I. E. Matyash, S. P. Rudenko, and B. K. Serdega, 2008: *Semicond. Phys. Quantum Electron. Optoelectron.* **11**, 63.

26. Shklyarevsky, I. N., A. I. Usoskin, V. P. Lebedev, and E. N. Alekseev, 1974: *Opt. Spektrosk.* **36**, 785 (in Russian).
27. Kreibig, U., and M. Vollmer, 1995: *Optical Properties of Metal Clusters* (Springer-Verlag, Berlin).
28. Wood, D. M., and N. W. Ashcroft, 1982: *Phys. Rev. B* **25**, 6255.
29. Dmitruk, N. L., O. S. Kondratenko, S. A. Kovalenko, and I. B. Mamontova, 2006: *Fiz. Khim. Tverd. Tila* **7**, 39 (in Ukrainian).
30. Dryzek, J., and A. Czaplá, 1987: *Phys. Rev. Lett.* **58**, 721.
31. Chau, L.-K., Y.-F. Lin, S.-F. Cheng, and T. Lin, 2006: *Sens. Actuators B* **113**, 100.
32. Khlebtsov, B. N., A. Bogatyrev, L. A. Dykman, and N. G. Khlebtsov, 2007: *Opt. Spektrosk.* **102**, 269 (in Russian); *Opt. Spectrosc.* **102**, 233.
33. Berezhinsky, L. J., O. S. Litvin, L. S. Maksimenko, et al., 2009: *Opt. Spektrosk.* **107**, 281 (in Russian); *Opt. Spectrosc.* **107**, 264.
34. Bruggeman, D. A. G., 1935: *P. I. Ann. Phys. (Leipzig)* **24**, 636.
35. Hiraia, M., and A. Kumar, 2006: *J. Appl. Phys.* **100**, 014309.
36. Maksimenko, L. S., I. E. Matyash, S. P. Rudenko, and B. K. Serdega, 2009: *Semicond. Phys. Quantum Electron. Optoelectron.* **12**, 129.
37. Kaganovich, E. B., I. M. Kizyak, E. G. Manoilo, et al., 2009: *Ukr. J. Phys.* **54**, 621.
38. Bozhenko, V. V., L. G. Grechko, M. L. Dmitruk, and O. Yu. Semchuk, 2003: *Fiz. Khim. Tverd. Tila.* **4**, 118 (in Ukrainian).



Lena Andriets (left) and Svetlana Rudenko
at the Kyiv Pechersk Lavra.

Inferring microstructure and turbulence properties in rain through observations and simulations of signal spectra measured with Doppler–polarimetric radars

Felix Yanovsky*

National Aviation University, Prospect Komarova 1, 03680 Kyiv, Ukraine

Abstract. Doppler radars are able to measure important parameters of the target velocity. In contrast, polarimetric radars are very sensitive to features of the target shape and orientation relative to the radar beam direction. This chapter describes a novel Doppler–polarimetric approach to radar remote sensing. The combination of the Doppler ability and polarization diversity in the radar technology enables more comprehensive investigations of objects and phenomena in radar coverage. The discussion is adapted to the case of atmospheric remote sensing. A special case of cloud and precipitation observations is considered in greater detail. Mathematical models of signals and spectra of Doppler–polarimetric returns are discussed. It is demonstrated (theoretically, by simulation, and by real data processing) that important parameters of dynamic characteristics and microstructure of meteorological objects can be retrieved from Doppler–polarimetric observations. These results lead to new interesting and important applications like turbulence intensity measurement, drop size distribution estimation, recognition of type of scatterers, detection of hail zones, etc.

Keywords: microwave scattering, polarimetry, phenomenological model, atmospheric remote sensing, spectral differential reflectivity, turbulence intensity retrieval, Doppler radar

1. Introduction

The purpose of remote sensing is to derive information about properties of the outlying objects under observation using special technique without coming into physical contact with the object. Radar meteorology uses remote-sensing techniques to study physics of clouds and precipitation and to obtain meteorological information for different important applications such as aviation, hydrology, climatology, agriculture, and weather manifestation.

The advent of meteorological radars (initially non-coherent) was a very significant benefit for both operational work and research in the atmosphere. Radar

* Corresponding author. E-mail: yanovsky@i.com.ua, felix.yanovsky@ieee.org

reflectivity (Z) measurement was the first stage of quantitative observations in radar meteorology. Rapid advances occurred when Doppler radars were introduced into the practice of radar meteorology. The early years of Doppler radar in meteorology are described by Rogers (1990). However, neither reflectivity nor Doppler measurements can separately solve the problem of obtaining unambiguous information on microstructure and wind dynamics in clouds and precipitation.

Multi-parameter techniques, which involve simultaneous application of a set of informative parameters and/or a set of different facilities promise new possibilities. Probably, radar polarimetry provides the best and, at least, the simplest way to switch from single-parameter to multi-parameter measurements. It does not require an extra frequency contribution. The study of polarization features of non-spherical water drops was actually started in the 1950th (Seliga et al. 1990). The problem of elliptically polarized electromagnetic wave scattering by non-spherical atmospheric particles was solved in the Rayleigh approximation, and the first experimental measurements were performed (Shupiatsky 1959). A period of rapid advance in radar polarimetry occurred since the mid 1970s up to mid 1980s, when the potential of the differential reflectivity (Z_{dr}) was analyzed (Seliga and Bringi 1976) and then much work was done to develop and better understand diverse meteorological applications of radar polarimetry. In addition to the differential reflectivity, other polarimetric parameters such as the linear depolarization ratio (L_{dr}), specific differential phase (K_{dp}), and cross correlation coefficient (ρ_{hv}) were introduced. Since then, the progress in weather radar polarimetry has been gradual but steady.

While the first polarimetric methods were developed for non-coherent radars, currently we can see a growth of activity in radar polarimetry related to Doppler–polarimetric studies. The advent of multi-polarization techniques gave a strong impetus to the development of quantitative operational radar measurements of rainfall. It has also enabled radar meteorologists to study physical processes in precipitation in more detail. The correct use of Doppler–polarimetric information could help answer many questions concerning the microstructure and dynamics of rain. A detailed analysis of the capabilities of multi-polarized Doppler radars for the remote sensing of precipitation was performed by Russchenberg (1992). There are two comprehensive books available to the scientific community and engineers on the Doppler weather radar (Doviak and Zrnić 1993) and polarimetric weather radar (Bringi and Chandrasecar 2001) theory and application.

The combination of polarization diversity with Doppler measurements opens additional possibilities for the study of microphysics and dynamics of scatterers in the atmosphere, and the effect of this combination, i.e., Doppler polarimetry, should be significantly greater than the sum of the effects of the Doppler capability and polarimetry applied separately. Doppler–polarimetric radars enable one to estimate many variables related to different properties of hydrometeors. Nevertheless, the number of variables is always limited, and the fact that the variables are not completely independent is even more important. However, it is obvious that a simple increase of the number of observables is not the best way to maxi-

mize the efficiency of remote sensing.

One of the first coherent radar systems developed for the study of atmospheric objects using different polarizations for both transmission and reception was the Delft Atmospheric Research Radar (DARR) designed and implemented in The Netherlands. It was the 9-cm Frequency Modulated Continuous Wave (FM-CW) Doppler radar installed on the roof of the 21-storey TU-Delft building not far from the Dutch sea shore (Ligthart and Nieuwkerk 1980). The next Doppler–polarimetric product of the TU-Delft was the Transportable Atmospheric RADar (TARA; Yanovsky et al. 1997; Heijnen et al. 2000). Both DARR and TARA were designed as fully polarized systems in the linear orthogonal polarization basis.

The advent of the Doppler–polarimetric technique of observation is accompanied by the rapid growth in the number of measured variables. The means of data interpretation became better, but also much more complicated. It was necessary to examine the relationships between different Doppler and polarimetric parameters and, perhaps, to find new measurands characteristic of and suitable for Doppler polarimetry. This was described in Yanovsky (1998a), where new measurands such as the spectral differential reflectivity $sZdr(v)$ and the spectral linear depolarization ratio $sLdr(v)$, which are functions of the Doppler velocity (or frequency), were introduced and studied, as well as other parameters such as the slope of $sZdr$ (SLP) and differential Doppler velocity (DDV). The DDV was proposed and studied by Wilson et al. (1997) as a parameter for the drop size distribution retrieval. The idea of differential reflectivity and differential phase representation as a Doppler distribution was suggested earlier in Kezys et al. (1993), although not much analysis was done there. Using spectral analysis, the polarimetric parameters can be expressed as functions of Doppler frequency or radial velocity. This approach has been used for different applications (e.g., Yanovsky and Ligthart 2000; Unal et al. 2001; Yanovsky et al. 2005; Bachmann and Zrnić 2006; Wang et al. 2008).

Different meteorological dual-polarization Doppler radars (research-level and operational) have been developed in Great Britain, France, Germany, The Netherlands, and the USA for different frequency bands (S, C, X). For example, SELEX-Gematronik manufactures dual-polarization Doppler radars METEOR 635C and METEOR 1600C. The KOUN radar in the USA, operated by the National Severe Storms Laboratory (NSSL), is a research prototype of the dual-polarization WSR-88D, and the US network of WSR-88D is being upgraded (Zrnić and Boren 2008). Taking into account that a massive introduction of Doppler–polarimetric meteorological radars into operational practice is expected soon, it is very important to develop new efficient approaches and algorithms of signal processing and data interpretation; the existing research-level radars are an important tool for that. It is appropriate to remind the reader that among the numerous other experimental results obtained with the DARR in the 1990s, an unexpected one was obtained during Doppler–polarimetric measurements of overcast rain. It concerned a significant correlation between Ldr and the Doppler spectrum width observed sometimes (Russchenberg 1994). This result was preliminarily interpreted

as being caused by turbulence, but definitely required a more rigorous investigation.

Currently much effort is expended in working up different schemes to relate observed polarimetric signatures with various properties of scattering hydrometeors. The inverse scattering problem is always an incorrectly posed one from the mathematical point of view. The verification of solutions of such problems is difficult. So, how to solve the inverse scattering problem? One approach consists of modeling the scattering characteristics of various hydrometeor populations, computing the expected observables, and comparing them with experimental data. Unfortunately, compatible *in situ* measurements are usually unavailable. However, it could be possible to derive compatible data from indirect measurements related to multi-parameter and multimode applications. For example, drop size distributions derived from Doppler spectral data obtained with vertically looking radars could be compared with polarimetric measurements at grazing incidence, or both Doppler and polarimetric information could be extracted from the signal received at elevation angles between 0° and 90° , and so on.

The problem of data interpretation could be simplified if reliable and physically transparent models of the relationships between Doppler–polarimetric observables and measured properties of the object were available. Several specific models have been developed and described in the literature. For example, a model for the simulation of radar scattering from precipitation is presented in Kwiatkowski et al. (1995). This model takes into account polarization features caused by nonspherical shapes and canting of drops and provides the rms scattering matrix for an ensemble of canting drops with a prescribed two-parameter canting-angle distribution. The model of the effect of turbulence on the radar signal coming from clouds and precipitation (Yanovsky 1996) takes into account the behavior of inertial droplets under the action of turbulence. One can find several important partial models in Russchenberg (1992). However those models use integrated parameters of the drop size distribution as initial data and cannot relate everything simultaneously: the turbulence parameters, drop size distribution, shape and spatial orientation of drops on the one hand, and the Doppler spectrum and polarization observables on the other. A successful attempt to construct a general Doppler–polarimetric spectrum model was made in Yanovsky (1998b), and subsequently this phenomenological model was widely discussed (Yanovsky 2002) and used for turbulence estimation (Yanovsky et al. 2005).

The purpose of this chapter is to describe the concept and the most important details of the complex mathematical model of Doppler–polarimetric returns from precipitation, taking into account the microstructure and dynamic characteristics of the object under observation and the mode of sounding by the polarimetric radar with given performance specifications, as well as to demonstrate (theoretically, by simulation, and by real data analyses) that important parameters of turbulence and microstructure of meteorological objects can be retrieved from Doppler–polarimetric observations.

2. Doppler–polarimetric approach

2.1. Doppler approach

A Doppler radar is able to measure important parameters of the target velocity. Let us consider a droplet as the target. The radial velocity of the droplet relative to the radar depends on a number of factors. The droplet is falling down with the vertical rate of fall V_g , which contributes a radial component V_{gr} . The influence of the wind and turbulence $V_w(t)$ causes a radial component $V_{wr}(t)$. In the case of a moving carrier (with a speed V_c) of the radar, there is an unwanted contribution to the radial velocity V_{cr} which may change in time with changing angle $\alpha(t)$ between the direction of movement and the antenna beam axis. In the case of a ground-based radar, the V_{cr} component is absent. In reality, the situation is much more complicated. Due to a number of moving scatterers in the radar resolution volume, the reflected signal contains the entire spectrum of Doppler frequencies forming the Doppler spectrum. The Doppler spectrum $S(*)$ is the power spectrum of a complex signal expressed as a function of the Doppler frequency $S(f)$ or velocity $S(v)$. The spectrum $S(v)$ is interpreted as the reflectivity-weighted distribution of radial velocities of scatterers in the resolution volume. Thus, $S(v)dv$ is equal to the received power in the velocity interval dv . Based on this definition, $S(v)$ should be normalized according to

$$\int_{-\infty}^{\infty} S(v) dv = \bar{P}_{Rx}. \tag{1}$$

Here, \bar{P}_{Rx} is the mean received power which can be calculated from the radar equation:

$$\bar{P}_{Rx} = \frac{C_{PS}}{R^2} Z |K|^2, \tag{2}$$

where C_{PS} is the dimension factor depending on the performance specifications of the radar, R is the distance between the radar and the object observed (i.e., the resolution volume), and Z is the radar reflectivity factor. The latter is a parameter of the object; in the case of a rain, it can be computed by summing up the sixth powers of the diameters of all the drops contained in a unit volume. Assuming that the drop size distribution $N(D)$ is a continuous function of the drop size D , the reflectivity factor can be written as follows:

$$Z = \int_0^{\infty} D^6 N(D) dD. \tag{3}$$

The reflectivity factor is related to the reflectivity η (i.e., the specific radar cross section of a volume-distributed target) via the wavelength λ according to

$$Z = \frac{\eta \lambda^4}{\pi^5 |K|^2}. \quad (4)$$

Here, $|K|^2$ is the dielectric factor depending on the complex refractive index of the target (≈ 0.93 for water and ≈ 0.19 for ice). The reflectivity η is defined as

$$\eta = \sum_i \sigma_i N_i, \quad (5)$$

where N_i is the number of hydrometeors per unit volume with a radar cross section (RCS) σ_i , and the summation is performed over all the hydrometeors in the unit volume. In accordance with the definition given above, the Doppler spectrum $S(v)$ can be modeled as being proportional to the following integral:

$$S(v) \propto \int_{D_{\min}}^{D_{\max}} p_r(v/D) \sigma(D) N(D) dD. \quad (6)$$

Here, $p_r(v/D)$ is the probability density function (PDF) of the radial velocity of a scatterer of size D , $\sigma(D)$ is the RCS of this scatterer, and $N(D)$ is the drop size distribution; the integration extends from the smallest, D_{\min} , to the largest, D_{\max} , scatterer sizes, e.g., drops in the case of a rain.

On the other hand, the Doppler spectrum estimate $\hat{S}_v(v)$ can be obtained by taking a Fourier transform over the received signal reflected from the object. Three parameters of the Doppler spectrum are most important in the Doppler approach:

$$\int_{-\infty}^{\infty} S(v) dv = Z \text{ -- the radar reflectivity factor,} \quad (7)$$

$$\frac{1}{Z} \int_{-\infty}^{\infty} v S(v) dv = \bar{v} \text{ -- the mean Doppler velocity,} \quad (8)$$

$$\frac{1}{Z} \int_{-\infty}^{\infty} (v - \bar{v})^2 S(v) dv = \sigma_v^2 \text{ -- the Doppler velocity variance.} \quad (9)$$

The parameters (7), (8), and (9) determined via processing of the reflected signal are widely used in the practice of meteorological observations with Doppler radars (Doviak and Zrnić 1993). For correct data interpretation, certain problems of the Doppler approach must be addressed or at least accounted for, such as: the ambiguity of velocity measurements owing to a modulated sounding waveform; spectrum broadening owing to the limited beam width; the influence of the sounding waveform and antenna pattern; the influence of wind; the inertia of scatterers when measuring turbulence; and the effect of the carrier velocity. This technique is not sensitive to the shape of scatterers.

2.2. Polarimetric approach

If a radar enables one to control the polarization of the transmitted waveform and to measure the polarization of the received signal, we can speak of a polarimetric approach to remote sensing. Polarimetric radars are very sensitive to features of the target shape and non-symmetric target orientations with respect to the radar beam.

When an arbitrary linearly polarized wave is incident on a target, the back-scattered field is given by

$$\begin{bmatrix} E_H^s \\ E_V^s \end{bmatrix} = [S] \begin{bmatrix} E_H^i \\ E_V^i \end{bmatrix} = \begin{bmatrix} s_{HH} & s_{HV} \\ s_{VH} & s_{VV} \end{bmatrix} \begin{bmatrix} E_H^i \\ E_V^i \end{bmatrix}, \quad (10)$$

where the superscripts i and s denote the incident and scattered fields, respectively, the quantities s_{mn} are, in general, complex, the subscripts $m, n = H, V$ represent any combination of orthogonal polarizations for transmission and reception, and $[S]$ is the amplitude scattering matrix. The backscattered RCS matrix $[\sigma]$ is related to the scattering matrix components via the following relation:

$$[\sigma] = \begin{bmatrix} \sigma_{HH} & \sigma_{HV} \\ \sigma_{VH} & \sigma_{VV} \end{bmatrix} = 4\pi R^2 \begin{bmatrix} |s_{HH}|^2 & |s_{HV}|^2 \\ |s_{VH}|^2 & |s_{VV}|^2 \end{bmatrix}. \quad (11)$$

It thus follows that once the scattering matrix is specified, the target backscattered RCS matrix can be computed for any combination of transmitted and received polarizations.

In atmospheric measurements, the temporal behavior of an element of the scattering matrix can be described as a random process. Radar meteorologists often use various second-order moments grouped into a three-by-three covariance matrix (Doviak and Zrnić 1993):

$$C = \begin{bmatrix} \langle |s_{HH}|^2 \rangle & \langle s_{HV} s_{HH}^* \rangle & \langle s_{VV} s_{HH}^* \rangle \\ \langle s_{HH} s_{HV}^* \rangle & \langle |s_{HV}|^2 \rangle & \langle s_{VV} s_{HV}^* \rangle \\ \langle s_{HH} s_{VV}^* \rangle & \langle s_{HV} s_{VV}^* \rangle & \langle |s_{VV}|^2 \rangle \end{bmatrix}. \quad (12)$$

The elements of the covariance matrix (12) are the correlation and cross-correlation functions R_{HH} , R_{HV} , and R_{VV} of the random processes s_{HH} , s_{HV} , and s_{VV} .

With regard to rain, a polarization radar can sense the shape of raindrops. Since small droplets are almost spherical while the bigger drops are more oblate, the differential reflectivity defined as

$$Zdr = 10 \log(Z_{HH}/Z_{VV}) = 10 \log(\sigma_{HH}/\sigma_{VV}) = 10 \log(|s_{HH}|^2/|s_{VV}|^2) \quad (13)$$

is a measure of the mean shape of raindrops in the resolution volume.

Certain types of hydrometeors deviate from a sphere even more than large water drops, for example, ice crystals, hailstones, and snowflakes. Many other polarimetric observables sensitive to the shape and orientation of scatterers are

widely used in atmospheric radar research (Bringi and Chandrasekar 2001). Among them are the following:

$$Ldr = 10 \log \frac{|s_{HV}|^2}{|s_{VV}|^2} = 10 \log \frac{|s_{VH}|^2}{|s_{VV}|^2} - \text{the linear depolarization ratio,} \quad (14)$$

$$\rho_{hv}(0) = \frac{\langle s_{VV} s_{HH}^* \rangle}{\sqrt{\langle |s_{HH}|^2 \rangle \langle |s_{VV}|^2 \rangle}} - \text{the cross-correlation coefficient,} \quad (15)$$

$$K_{DP} = \frac{\phi_{DP}(R_1) - \phi_{DP}(R_2)}{2(R_2 - R_1)} - \text{the specific differential phase,} \quad (16)$$

where $\phi_{DP}(R) = \phi_H(R) - \phi_V(R)$ is the phase difference between the horizontally- and vertically-polarized pulses at a given point R in the propagation path.

The benefits of the polarimetric technique are the following: (i) polarimetry normally deals with relative quantities, which allows one to bypass difficult absolute measurements; (ii) polarimetry leads to multi-parametric systems; (iii) polarimetric parameters are very sensitive to the shape and orientation of scatterers; and (iv) the scattering matrix actually provides a “signature” of a target, and so polarimetric parameters can be used for target recognition.

The practical implementation of the polarimetric approach requires dealing with the following problems: (i) the antenna design must ensure the identity of antenna patterns for different polarizations, a reliable isolation between the co-polar and cross-polar components, and a controllable-polarization capability; and (ii) one needs to develop complicated signal processing algorithms. In addition, the polarimetric approach provides no information about the target velocity. Moreover, the scattering matrix elements must be measured simultaneously. In practice, the measurements of the co-polar elements of the scattering matrices s_{HH} and s_{VV} are separated by a finite time interval Δt , which must be smaller than the decorrelation time of the signal reflected from the hydrometeors. A procedure correcting for the non-simultaneity of HH and VV polarimetric measurements was proposed by Unal and Moisseev (2004).

2.3. Doppler polarimetry

A combination of the spectral analysis and polarization diversity in radar technology enables a more comprehensive characterization of objects and phenomena. According to Kozlov et al. (2002), the Doppler polarimetry is a methodology for the determination of both the Doppler velocity (radial component) and the polarization signature of a moving scatterer. When the scatterer is moving, the phase of the received scattered signal is determined by the polarization-dependent properties of the scatterer and its radial velocity. However this definition is too general to be useful.

More specifically, by combining the Doppler and polarimetric information, one attempts to measure and interpret different polarimetric parameters per a

Doppler velocity bin (Yanovsky et al. 2001). Let us explain the transition from the traditional polarimetric approach to the Doppler polarimetry by considering the differential reflectivity Zdr defined by Eq. (13). From Eqs. (12) and (13) it is seen that Zdr can be determined from the correlation functions at zero time lag $R_{HH}(0)$ and $R_{VV}(0)$ (Unal et al. 2001):

$$Zdr(n) = 10 \log \frac{R_{HH}(0, n)}{R_{VV}(0, n)}, \quad (17)$$

where

$$R_{xx}(0, n) = \langle s_{xx}(t, n) s_{xx}^*(t, n) \rangle \quad (18)$$

are covariance matrix elements. In Eqs. (17) and (18), the integer n represents the range bin considered, t is time, the subscript xx stands for HH or VV , and the angular brackets indicate averaging over time.

The covariance matrix elements (18) and the differential reflectivity (17) derived from two of these elements represent the average polarimetric properties of the range bin considered. In this case, the differential reflectivity describes an average particle shape. The value 0 dB indicates, for example, the predominantly spherical shape ($R_{HH} = R_{VV}$). One can attempt to retrieve more specific information about polarimetric properties of the scatterers inside the resolution volume by using the spectral analysis.

When the random processes are stationary, the second-moment spectral analysis can be performed (Ryzhkov 2001). The dynamic properties of the targets are then also considered. Using the time series of the scattering matrices, the Fourier transform of the correlation and cross-correlation functions leads to the power spectra HH , HV , and VV as well as the cross spectra (HV, HH) , (VV, HH) , and (VV, HV) . They represent the elements of the spectral covariance matrix which is then defined for different ranges and Doppler frequencies. This results in a complete target description combining polarimetric and dynamic properties of the radar target. We call this description the ‘‘radar Doppler polarimetry’’; the spectral covariance matrix is a Doppler–polarimetric result (Unal et al. 2001). Unlike Eq. (12), the target spectral covariance matrix can be defined for each range bin and each Doppler bin l . It can be expressed as follows (Unal and Moisseev 2004):

$$[\hat{C}(l\omega_D)] = \frac{1}{k} \sum_{i=1}^k \begin{bmatrix} \hat{s}_{HH}(i, l\omega_D) \hat{s}_{HH}^*(i, l\omega_D) & \hat{s}_{HH}(i, l\omega_D) \hat{s}_{HV}^*(i, l\omega_D) & \hat{s}_{HH}(i, l\omega_D) \hat{s}_{VV}^*(i, l\omega_D) \\ \hat{s}_{HV}(i, l\omega_D) \hat{s}_{HH}^*(i, l\omega_D) & \hat{s}_{HV}(i, l\omega_D) \hat{s}_{HV}^*(i, l\omega_D) & \hat{s}_{HV}(i, l\omega_D) \hat{s}_{VV}^*(i, l\omega_D) \\ \hat{s}_{VV}(i, l\omega_D) \hat{s}_{HH}^*(i, l\omega_D) & \hat{s}_{VV}(i, l\omega_D) \hat{s}_{HV}^*(i, l\omega_D) & \hat{s}_{VV}(i, l\omega_D) \hat{s}_{VV}^*(i, l\omega_D) \end{bmatrix}. \quad (19)$$

Here, the caret indicates that the matrix or parameter is expressed in the frequency domain, meaning that $\hat{s}_x(l\omega_D)$ is defined as a discrete time Fourier transform of $s_x(nT_m)$, where mT_m is the time lag during the correlation function determina-

tion, n is the sample size, ω_D is the Doppler frequency resolution, and k is the number of averages of the Doppler spectra.

Using this approach, all polarimetric parameters (13)–(16) can be expressed as functions of the Doppler velocity or frequency, that is, for each velocity bin the spectral differential reflectivity $sZdr(v)$ and other useful measurands (functions and parameters) can be introduced.

3. Theory and models for rain

In this section, we consider a complex phenomenological model which yields polarization properties of the radar signal coming from raindrops as functions of their radial velocity. The main objective is to relate Doppler–polarimetric observables with weather object parameters for further data interpretation. This is important for meteorological target detection and recognition, but also is useful in cases when meteorological objects are in the state of clutter.

3.1. Phenomenological model

Hereinafter, our phenomenological model is understood as a mathematical model describing a body of knowledge of phenomena and processes under study (objects) and taking into account the results of empirical observations of certain objects as well as interrelations between different elements of these objects. This complex model may not follow entirely from the fundamental theory, but is consistent with the theory and can involve other theoretical and empirical models as its components. Our phenomenological model considers interrelations between the various elements of an object as well as modes of their existence. Therefore, it enables one to investigate certain features of a phenomenon or an object in a wide range of conditions, which is often impossible to do via natural experiments, especially if the parameters and the structure of the object in question cannot be controlled.

3.2. Concept of modeling

We assume that the radar resolution volume is filled with particles which may differ in size, concentration, shape, orientation, velocity, and permittivity.

The concept of mathematical modeling is presented here in accordance with Yanovsky (2002) and Yanovsky et al. (2001). The model yields the Doppler spectra $S_{mn}(v)$ for different combinations of polarization for transmission (second index) and reception (first index) of waves with $m = x, y$ and $n = x, y$; x and y represent the linear orthogonal polarization basis. In the special case of the horizontal–vertical polarization basis, $x = h$ (horizontal) and $y = v$ (vertical), the model yields three Doppler–polarimetric spectra: $S_{HH}(v)$, $S_{VV}(v)$, and $S_{HV}(v)$. Based on these spectra, polarization observables such as the spectral differential reflectivity $sZdr(v)$ and spectral linear depolarization ratio $sLdr(v)$ are calculated. They are defined as follows:

$$sZdr(v) = 10 \log[S_{HH}(v)/S_{VV}(v)], \quad (20)$$

$$sLdr(v) = 10 \log[S_{HV}(v)/S_{VH}(v)]. \quad (21)$$

The model calculates three intermediate kinds of Doppler spectra. The first one is calculated without turbulence and is caused by gravity only; the second one is caused by turbulence of a given intensity but without taking into account the fall velocity of particles; the third one accounts for both turbulence and gravity.

The model is initialized with:

- parameters of the drop size distribution;
- parameters of atmospheric turbulence;
- radar-system parameters (wavelength, range resolution, and antenna beam width).

The main source of experimental data used for model verification is the TARA system (Heijnen et al. 2000) which uses a linear polarization basis. Therefore, the model is developed for the same polarization basis. However, other orthogonal bases can provide the same information.

In the model, the following main stages can be identified:

- a parameterization of the rain drop size distribution is assumed as well as a relationship between the raindrop size and the “stagnant air” fall velocity;
- the velocity distribution of raindrops caused by turbulence is derived. It accounts for the inertia of the drops and the turbulence scale;
- the combined velocity distribution caused by both turbulence and the terminal velocity that the raindrops would have in stagnant air is derived;
- the shape of the raindrops is taken into account for each interval of the drop size distribution; the water drops are modeled as spheroids;
- the RCSs for co-polar and cross-polar radar signals are calculated as functions of the equivalent drop diameter, shape, and orientation;
- the Doppler spectrum of the radar signal scattered from an ensemble of particles is calculated for different polarizations using statistical distributions of particle sizes and shapes;
- the polarization variables are calculated as functions of the Doppler velocity; and finally
- the Doppler–polarimetric spectra and polarization variables are related to the parameters of turbulence and the microstructure of rain in different conditions.

3.3. Initial models

Initial models should be chosen and substantiated for the object, type of radar and its performance specifications, mode of sounding, wavelength, coordinates and polarization basis, and conditions of wave propagation.

3.3.1. Rain microphysics

The microstructure of rain is described by statistical distributions of size, shape, fall velocity, orientation, and number concentration.

Drop fall velocity in stagnant air. According to the Stokes law, the fall velocity of raindrops in stagnant air is related to their size. The relationship between the drop diameter D and the drop fall velocity v_f is approximated in Atlas et al. (1973) as follows:

$$v_f(D) = (9.65 - 10.3) \times e^{-0.6D}, \quad D \geq 0.109 \quad (D \text{ in mm; } v_f \text{ in ms}^{-1}). \quad (22)$$

This relationship was derived from data taken at the sea level and should be corrected when other altitudes are considered (Foote and Toit 1969).

Particle shape. The shape of a falling raindrop is not exactly spherical but rather is flattened at the base. To describe the shape of a raindrop mathematically, it is modeled as a spheroid. The size of a non-spherical particle is characterized by the equivolumetric diameter, which is defined as the diameter of the sphere with the same volume as the spheroid. The shape of the spheroid is characterized by the axial ratio ρ , i.e., the ratio of the shortest to the longest particle dimensions.

Theoretical calculations and measurements of the shape of raindrops have been performed by Shupiatsky (1959), Pruppacher and Beard (1970), Pruppacher and Pitter (1971), Beard and Chuang (1987), Kubesh and Beard (1993), Tokay and Beard (1996) and others. They reported different quantitative relationships between the degree of oblateness and the particle size. These results were combined by Yanovsky (1998a) into the following relationship between the axial ratio ρ and the equivolumetric diameter D :

$$\rho(D) = 0.5[\exp(-D^2/27) + 0.5] \quad (D \text{ in mm}). \quad (23)$$

Equation (23) can be used in practice for $D \geq 0$, whereas the traditional approximation by the fourth-order polynomial (Pruppacher and Klett 1997) gives realistic results only for $D > 2.12$ mm.

Spatial particle orientation. Wind variations may force the raindrop to cant. The canting angle is defined as that between the axis of rotational symmetry corresponding to the shortest dimension of the spheroid and the vertical direction. Brussaard (1976) related the canting of an individual raindrop to the vertical wind shear described by the altitude gradient of the horizontal wind velocity and found that usually this gradient is small at altitudes above 100 m and, consequently, the canting angle is small ($< 3^\circ$). Based on this result, we assume a zero mean canting angle. Furthermore, we assume that turbulence causes a random distribution of canting angles. The relationship between turbulence and canting is not known, but is often assumed to result in a Gaussian distribution of canting angles $p(\delta)$, $\bar{\delta}$ being the mean canting angle and σ_δ being the rms (Russchenberg 1992).

Drop size distribution. The average raindrop size distribution is often described by the Marshall–Palmer formula:

$$N(D) = N_0 e^{-\Lambda_d D}, \tag{24}$$

where $N(D)$ is the number of particles with equivalent diameters between D and $D + dD$ per unit volume, $N_0 = 8000 \text{ mm}^{-1} \text{ m}^{-3}$, and the factor Λ_d depends on the rain intensity R (in mmh^{-1}):

$$\Lambda_d = 4.1R^{-0.21} [\text{mm}^{-1}] \text{ and } \Lambda_d = 3.67/D_0, \tag{25}$$

D_0 being the diameter of the median drop volume. Although Eq. (24) is widely used, many experiments have shown that it is not universally applicable: it exaggerates the number of small drops even for the average data. To overcome this problem, the following gamma distribution is used:

$$N(D) = N_0 D^\mu \exp\left(-\frac{3.67 + \mu}{D_0} D\right). \tag{26}$$

It includes the Marshall–Palmer model as a special case for $\mu = 0$. For $\mu \neq 0$, N_0 can be derived from the Marshall–Palmer distribution by keeping the rainwater content constant for a given D_0 . This yields

$$N_0 \approx \frac{264.59(3.67 + \mu)^{\mu+4}}{D_0^\mu (\mu + 3)!}. \tag{27}$$

The normalized raindrop diameter distribution $n(D)$, which can be used as a PDF function, is derived from Eq. (26) by integrating over D from 0 to ∞ :

$$n(D) = D^\mu \exp\left(-\frac{3.67 + \mu}{D_0} D\right) \Bigg/ \left[\left(\frac{D_0}{3.67 + \mu}\right)^{\mu+1} \mu! \right]. \tag{28}$$

Figure 1 shows $n(D)$ calculated for different μ values and $D_0 = 2 \text{ mm}$. The drop size distribution plays a very important role in the development of the general model because it affects both Doppler and polarization characteristics.

Inertia of drops. The inertia of raindrops in a turbulent environment was estimated in Gorelik and Chernikov (1964). In the case of homogeneous and isotropic turbulence, the correlation function of turbulent wind pulsations was considered. Then the correlation function of drop velocities was derived from the solution of a linear equation for the component of drop velocity. The comparison of these two correlation functions yields the condition of obtaining an undistorted spectrum of turbulent pulsations from the drop velocities (Doppler velocities), assuming that the interaction of the drops with the medium is defined by the Stokes law. This consideration allowed Gorelik and Chernikov to derive the relaxation time T of a droplet with a given effective size D . The difficulty with this approach is the tran-

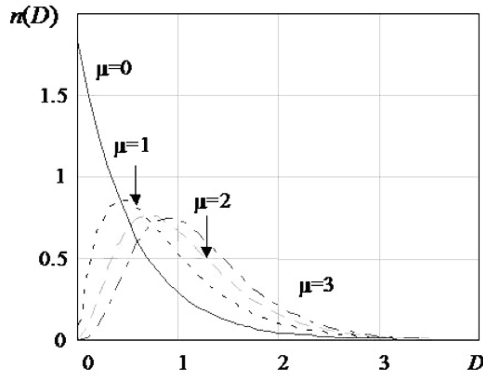


Fig. 1. Normalized gamma distribution of raindrops with $D_0 = 2$ mm and $\mu = 0, 1, 2,$ and 3 . D is given in mm.

Table 1. Relationship between raindrop diameter and relaxation time

D , mm	6	5	4	3	2	1	0.5	0.1
T , s	19.8	19.5	18.8	17	14	8.65	4.38	0.538

sition from the Euler to the Lagrange scale. The relaxation time multiplied by the dimension factor V_w gives the appropriate spatial scale of turbulence. In Gorelik and Chernikov (1964), the value $V_w = 10 \text{ ms}^{-1}$ was used as a typical parameter for the calculations. In the phenomenological model described below, the V_w can be changed during the adaptation. The relationship between the raindrop diameter D and its relaxation time T is summarized in Table 1 taken from Gorelik and Chernikov (1964).

3.3.2. Turbulence

Scales of turbulence. The instantaneous velocity of the turbulent flow can be considered to result from the superposition of three-dimensional fluctuations and the average air motion. The turbulent velocity components follow a normal distribution with a zero mean (Dobrolensky 1969). Turbulence has an eddy nature with a wide spectrum of spatial scales L : from a minimum (inner) scale L_{inner} up to a large (outer) scale L_{outer} which may be comparable to the scale of the airflow as a whole. However, in our model, only the inertial subrange (Frisch 1995) of turbulence is taken into consideration. It includes all scales existing in the free atmosphere, from the smallest ones (several mm) up to about L_{outer} (set to 1500 m). This scale range encompasses the characteristic size of the radar spatial resolution as well as the scales of turbulence dangerous for aircraft.

Energy spectrum of turbulence. The turbulence energy spectrum $S(\Omega)$ is the decomposition of the kinetic energy of turbulence in a Fourier series over wave

numbers $\Omega = 2\pi/L$ (the spatial frequency). In the inertial subrange, where the conditions of homogeneity and local isotropy of turbulence are satisfied, the analytical expression for the spectrum is as follows:

$$S(\Omega) = C \varepsilon^{2/3} \Omega^{-5/3}, \quad (29)$$

where C is a dimensionless constant and ε is the eddy dissipation rate. The Ω is defined as $\Omega = |\mathbf{\Omega}| = 2\pi/L$, where $\mathbf{\Omega}$ is the three-dimensional turbulence wave vector.

The dimensionless constant C depends on the direction of the velocity vector component. The one-dimensional spectra S_u , S_v , and S_w for the components u (along the basic flow) as well as v and w (across the basic flow) are also described by Eq. (29), but with different values of C . The longitudinal, C_u , and transversal, C_v and C_w , constants are related by $C_v = C_w = 4C_u/3$. According to experimental data (Vinnichenko et al. 1968), C_u approximately equals 0.50 with a 20% uncertainty. Taking into account that $C_u \approx 0.327C$ (Vinnichenko et al. 1968), we can derive $C_u \approx 0.40 - 0.60$, $C_v = C_w \approx 0.53 - 0.80$, and $C \approx 1.22 - 1.83$. One can find different estimates of these constants in the literature, but they all have the same order of magnitude.

Variance of turbulence velocity. Estimating the spatial spectrum $S(\Omega)$ experimentally is very difficult. That is why simpler statistical parameters such as the velocity variance σ_v^2 are often used. The velocity variance due to turbulence in a given range of scales can be calculated from the energy spectrum $S(\Omega)$ (Vinnichenko et al. 1968):

$$\sigma_v^2 = \int_{\Omega_{\min}}^{\Omega_{\max}} S(\Omega) d\Omega, \quad (30)$$

where $\Omega_{\min} = 2\pi/L_{ul}$ and $\Omega_{\max} = 2\pi/L_{ll}$ correspond to the upper and lower limits of the turbulence scales considered. Substituting Eq. (29) into Eq. (30), integrating, and taking into account that the effect of L_{ll} , which is much weaker than that of L_{ul} , can be neglected, yields:

$$\sigma_v \approx C_0^{1/2} \varepsilon^{1/3} L_{ul}^{1/3}. \quad (31)$$

In our model, the upper scale L_{ul} does not have to be identical to the outer scale of turbulence. Instead, it depends on the spatial resolution of the radar. Turbulence at scales much larger than the radar resolution does not affect the velocity variance, but changes the observed mean velocity.

Turbulence intensity. The kinetic energy of turbulence is passed sequentially from larger scales to smaller ones and then dissipates at a scale $L \approx L_{inner}$. The latter process is quantified by the eddy dissipation rate ε , which is a fundamental parameter of turbulence characterizing its intensity. It does not depend on the scale of turbulence within the inertial subrange, which makes ε a convenient initial

Table 2. Turbulence classification based on the eddy dissipation rate

$\varepsilon, \text{ cm}^2\text{s}^{-3}$	< 0.2	$0.2 - 3.4$	$3.4 - 42.9$	$42.9 - 550$	> 550
Intensity scale	negligible	light	moderate	heavy	severe

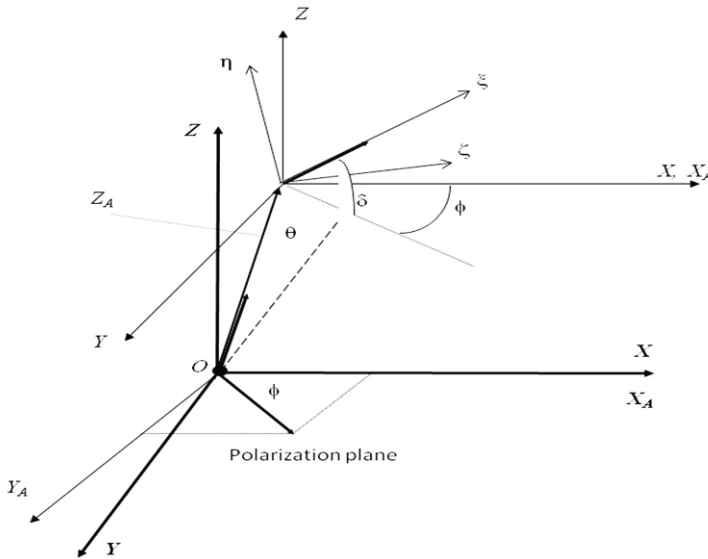


Fig. 2. Coordinates systems. X, Y, Z – main (radar) coordinates; X_A, Y_A, Z_A – antenna coordinates; η, ζ, ξ – particle coordinates; O – radar; Z_A – antenna beam; θ – elevation; δ – particle canting angle; ϕ – polarization angle; and ξ – the symmetry axis of a spheroid.

modeling parameter. The classification of turbulence from the standpoint of aircraft safety in terms of ε is shown in Table 2 (MacCready 1964). The value of ε in cumulonimbus clouds can reach $1000 \text{ cm}^2\text{s}^{-3}$.

The eddy dissipation rate ε and the spatial scale L are the only important parameters of turbulence in our model. They are used to calculate the effect of turbulence on the behavior of the particles in the radar volume.

3.3.3. Coordinates

The coordinates of the mutual locations of the radar system, the moving antenna beam, and the scatterer are shown schematically in Fig. 2, in which:

- O is the point where the radar system is located;
- $X, Y,$ and Z are the basic (radar) coordinates;
- Z_A is the direction of the antenna beam;
- $X_A, Y_A,$ and Z_A are coordinates associated with the antenna beam;
- θ is the antenna elevation (the angle between the OZ_A and XOY planes);

η , ζ , and ξ are the particle coordinates;
 δ is the particle canting angle;
 ϕ is the polarization angle;
 ξ is the axis of symmetry of the spheroid representing the particle.

3.4. Modeling velocity distribution of raindrops

3.4.1. Distribution of the drop fall velocity

In the absence of winds, the drop fall velocity vector is directed straight down; the radar only measures the projection of this vector on the line of sight. Introducing the elevation angle θ in Eq. (22) yields

$$v_f(D) = \alpha - \beta e^{-0.6D}, \quad (32)$$

where $\alpha = 9.65 \sin \theta$ and $\beta = 10.3 \sin \theta$.

The drop fall velocity v_f is assumed to be a function of the random parameter D , which obeys the known PDF as given by Eq. (28). According to Venttsel' (1998), dividing Eq. (26) by the derivative $|dv_f/dD|$ and substituting $D(v_f)$ yields the analytical expression of the drop fall velocity distribution (Yanovsky et al. 2001):

$$N_f(v_f) = \frac{5N_0}{3(\alpha - v_f)} \left(\frac{5}{3} \ln \frac{\beta}{\alpha - v_f} \right)^\mu \exp \left[\frac{5(3.67 + \mu)}{3D_0} \ln \frac{\alpha - v_f}{\beta} \right], \quad v_f \geq 0. \quad (33)$$

The PDF of the fall velocity can be obtained by normalization:

$$n_f(v_f) = N_f(v_f) / \int N_f(v_f) dv_f. \quad (34)$$

The integration of Eq. (34) over v_f can be done analytically (Yanovsky 1998a), but the resulting expressions are rather bulky and are not given here. The values of Eq. (34) for $D_0 = 1.5$ and $\theta = 45^\circ$ are shown in Fig. 3. The calculations were done for different values of the spread parameter of the gamma size distribution: $\mu = 1, 2, 5, 7$. A significant effect of μ is obvious from these plots. The most probable fall velocity shifts to greater values when μ increases; however, the maximum velocity remains almost the same.

3.4.2. Drop turbulent velocity distribution

The detection of turbulence in clouds and precipitation by a Doppler radar requires scatterers to respond instantly to the turbulent motion. However, scatterers such as raindrops may not respond perfectly due to inertia. In this section the velocity distribution of raindrops due to turbulence is calculated.

Concept of a threshold turbulence scale. To make a raindrop move, a turbulent eddy must have enough energy; the larger the raindrop the more energy is needed.

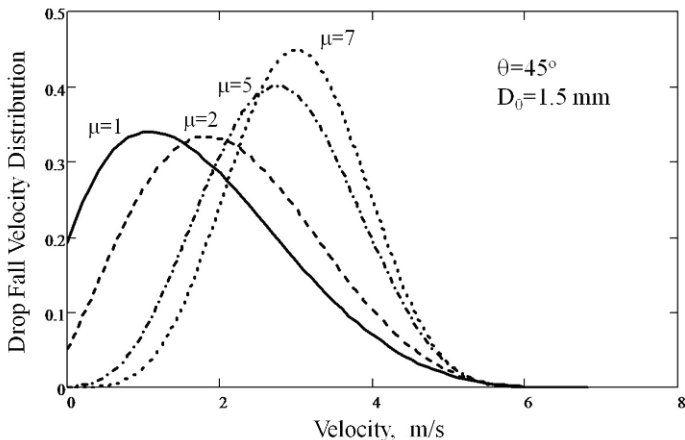


Fig. 3. Radial drop fall velocity distribution for different drop size distributions and antenna elevations under the condition of constant volume of water in m^3 (approximately the same rain rate); v_f is given in ms^{-1} .

In Yanovsky (1996), a new approach was adopted to model the interaction of turbulence and raindrops with the purpose of accounting for the inertia in the simulation of the reflected signal. It was based on an earlier work by Fishman and Yanovsky (1983). In this approach, the key is the assumption that a threshold turbulent scale L_{thres} exists and is specific for a given mass (and size) of drops. According to this approach:

- below a threshold scale, turbulent eddies do not affect raindrops of a certain size; this threshold scale L_{thres} corresponds to a raindrop size D_{thres} ;
- turbulence at scales larger than the threshold scale L_{thres} affects all raindrops with sizes $D < D_{thres}$;
- turbulence at scales smaller than the threshold scale L_{thres} does not affect the raindrops with $D \geq D_{thres}$;
- once raindrops are set into motion they act as perfect tracers of turbulence, i.e., the inertia does not play a role anymore.

Thus each drop diameter D_{thres} corresponds to a unique value of the threshold spatial scale L_{thres} , i.e., L_{thres} and D_{thres} are functionally related values: $L_{thres} = f(D_{thres})$.

General expression. Based on this concept, the drop turbulent velocity distribution for drops with a diameter D can be obtained as follows:

$$p_T(v_T, D, \varepsilon) = \int_{L_{thres}(D)}^{L_{nl}} w_T(v_T/L, \varepsilon) w(L) dL, \tag{35}$$

where $w_T(v_T/L, \varepsilon)$ is the conditional probability density of turbulence velocity and $w(L)$ is the probability density of spatial scales for given turbulence parameters. As follows from Section 3.2.2, any component of the random turbulent velocity field follows the normal distribution law with zero mean and the variance depending on the eddy dissipation rate ε and spatial scales L of turbulence according to Eq. (31):

$$w_T(v_T/L, \varepsilon) = \frac{1}{\sqrt{2\pi C_0 \varepsilon^{1/3} L^{1/3}}} \exp\left(-\frac{v_T^2}{2\pi C_0 \varepsilon^{2/3} L^{2/3}}\right). \quad (36)$$

Turbulence scale distribution. Since L and ε are functionally related random variables (Venttsel' 1998), we derive the function $w(L)$ from

$$w(L) = \frac{S[\Omega(L)]}{|dL/d\Omega|},$$

where the PDF of the variable Ω is defined by Eq. (29) after appropriate normalization. Finally, the normalized density function $w(L)$ can be written as follows (Yanovsky 1996):

$$w(L) = \frac{2}{3} L^{-1/3} L_{ul}^{-2/3}. \quad (37)$$

L_{ul} appears as the upper limit of integration over L and can be interpreted as the largest spatial scale of turbulence taken into account, the lower limit being set to zero.

Analytical solution. Substitution of Eqs. (36) and (37) into Eq. (35) yields:

$$p_T[v_T/L_{thres}(D)] = \frac{\sqrt{2} \varepsilon^{-1/3} L_{ul}^{-2/3}}{3\sqrt{\pi C_0}} \int_{L_{thres}(D)}^{L_{ul}} L^{-2/3} \exp\left(-\frac{v^2 \varepsilon^{-2/3} L^{-2/3}}{2\pi C_0}\right) dL. \quad (38)$$

The integral in Eq. (38) can be expressed analytically in terms of the error function $\text{erf}(x)$ (Venttsel' 1998). Then the equation for $p_T[v_T/L_{thres}(D)]$ is as follows (Yanovsky et al. 2001):

$$p_T[v_T/L_{thres}(D)] = \frac{[*]}{\pi^{1/2} C_0 L_{outer}^{2/3} \varepsilon^{2/3}}, \quad (39)$$

where

$$[*] = 2^{1/2} \varepsilon^{1/3} L_{ul}^{1/3} C_0^{1/2} \exp\left(-\frac{v^2}{2\varepsilon^{2/3} L_{ul}^{2/3} C_0}\right) - 2^{1/2} \varepsilon^{1/3} L_{thres}^{1/3} C_0^{1/2} \exp\left(-\frac{v^2}{2\varepsilon^{2/3} L_{thres}^{2/3} C_0}\right) + \pi^{1/2} v \text{erf}\left(\frac{v}{2^{1/2} \varepsilon^{1/3} L_{ul}^{1/3} C_0^{1/2}}\right) - \pi^{1/2} v \text{erf}\left(\frac{v}{2^{1/2} \varepsilon^{1/3} L_{thres}^{1/3} C_0^{1/2}}\right).$$

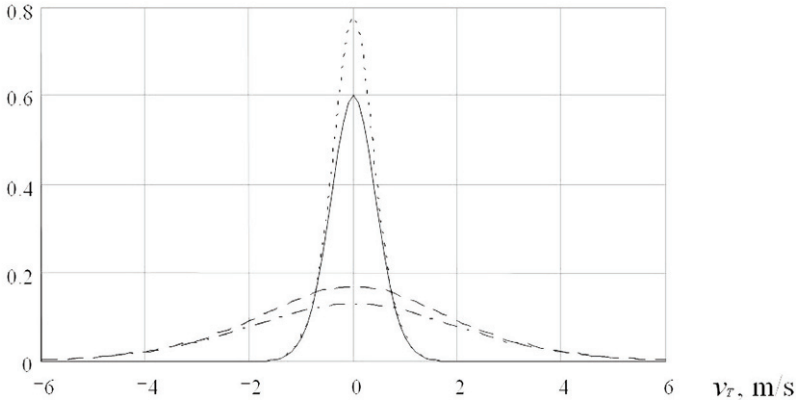
$p_T(v_T/\varepsilon, D)$


Fig. 4. Partial distribution for different ε and D . Dotted curve: $\varepsilon = 1 \text{ cm}^2\text{s}^{-3}$ and $D = 1 \text{ mm}$; solid curve: $\varepsilon = 1 \text{ cm}^2\text{s}^{-3}$ and $D = 4 \text{ mm}$; dashed curve: $\varepsilon = 100 \text{ cm}^2/\text{s}^3$ and $D = 1 \text{ mm}$; dot-dashed curve: $\varepsilon = 100 \text{ cm}^2\text{s}^{-3}$ and $D = 4 \text{ mm}$.

Limits of integration. The lower integration limit in Eq. (35), $L_{thres}(D)$, is the minimal spatial scale affecting all raindrops with equivolumetric diameters $\leq D$. The numerical relationship between the diameter D and the relaxation time presented in Table 1 is used to determine $L_{thres}(D)$. It can be approximated as follows:

$$L_{thres}(D) = 21.17(1 - e^{-0.527D})V_w, \quad (40)$$

where V_w is a constant having the dimension of velocity (ms^{-1}); it relates the droplet relaxation time with the scale of turbulence. Note that D and $L_{thres}(D)$ in Eq. (40) are expressed in millimeters and meters, respectively.

The upper integration limit L_{ul} follows from the maximum spatial scale of turbulence contributing to random motion of scatterers in the radar volume (Yanovsky 1998b). It can be defined as the scale of turbulence that affects individual particles in a single resolution volume differently. Scales larger than L_{ul} only influence the mean particle velocity. In fact, it is the largest characteristic size of a single resolution volume in radial or tangential direction: $L_{ul} = \max(R_{rad}, R_{tan})$.

Calculations. Equation (39) allows one to calculate the partial distribution of turbulence-induced velocity for raindrops of a given diameter D for the turbulence parameters C_0 , L_{ul} , and ε . Figure 4 shows the results for $C_0 = 1.5$, $V_w = 10 \text{ ms}^{-1}$, and $L_{ul} = 1000 \text{ m}$ (Yanovsky et al. 2001). Two pairs of distributions can be seen. The pair of narrower distributions corresponds to light turbulence ($\varepsilon = 1 \text{ cm}^2\text{s}^{-3}$), while the pair of broader ones corresponds to heavy turbulence ($\varepsilon = 100 \text{ cm}^2\text{s}^{-3}$).

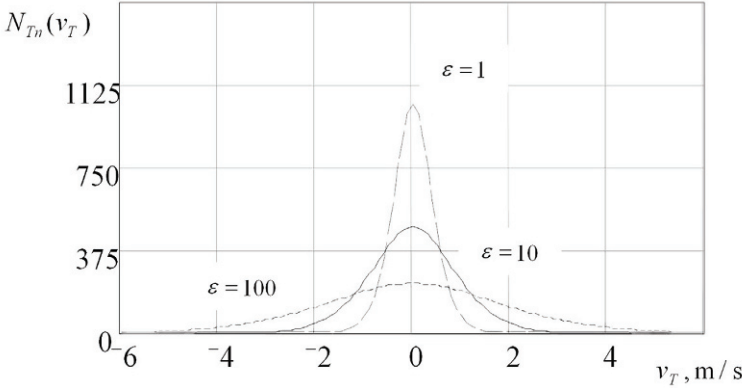


Fig. 5. Distribution of turbulent velocities of the drop ensemble.

Each pair depicts the results for two drop diameters: the upper curve corresponds to small drops ($D = 1$ mm), while the lower one corresponds to large drops ($D = 4$ mm).

The integration of Eq. (39) over D yields the drop turbulent velocity distribution accounting for all the drops within the integration limits:

$$N_T(v_T) = \int_{D_{\min}}^{D_{\max}} p_T(v_T, D) N(D) dD, \tag{41}$$

where $p_T(v_T, D)$ is defined by Eq. (38) or (39); $N(D)$ is defined by Eq. (26); D_{\min} and D_{\max} are the bounding drop-diameter values. Figure 5 shows $N_T(v_T)$ for the median drop size $D_0 = 1.6$ mm, dispersion factor $\mu = 1$, and three values of the eddy dissipation rate ε : 1, 10, and $100 \text{ cm}^2 \text{ s}^{-3}$. The results demonstrate that the greater the eddy dissipation rate ε the broader the distribution caused by turbulence. $N_T(v_T) dv_T$ is the number of raindrops in a unit volume with turbulent velocity component values between v_T and $v_T + dv_T$ provided that $p_T(v_T, D)$ is normalized according to the following condition:

$$\int_{-\infty}^{\infty} p_T(v_T, D) dv_T = 1.$$

3.4.3. Drop velocity distribution caused by both turbulence and gravity

The separate velocity distributions $N_f(v_f)$ and $N_T(v_T)$ due to gravity and turbulence, respectively, were derived above. Expressing the total drop velocity as $v = v_f + v_T$ and neglecting the correlation between the two components, the combined PDF can be determined by convolution. However, taking into account that the terminal fall velocity of a drop is unambiguously related to the drop diameter

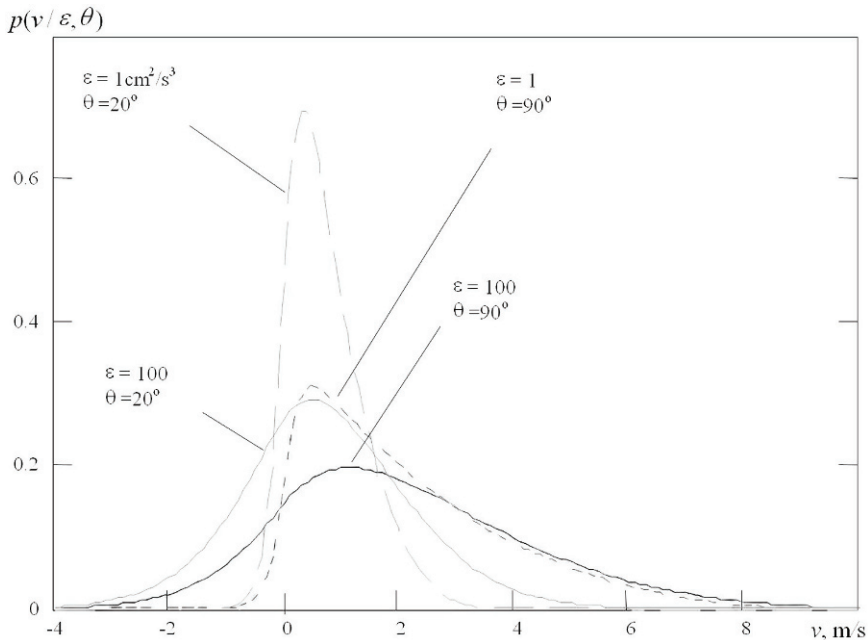


Fig. 6. Radial drop velocity distribution in the Marshall–Palmer case for two values of ε and two antenna elevations θ .

according to Eq. (22), we can write the partial velocity distribution $p_p(v/D)$ for a given drop by substituting $v_T = v - v_f$ in Eq. (39) (Yanovsky et al. 2005). In this case the partial distribution for a given drop diameter D written in the form of a conditional distribution is as follows:

$$p_p(v/D) = p_T[(v - v_f)/D], \tag{42}$$

where p_T is given by Eq. (38) or (39). Integrating over all droplet diameters yields the distribution of radial velocities of the ensemble:

$$p_\Sigma(v) = \int_{D_{\min}}^{D_{\max}} p_T\{[v - v_f(D, \theta)]/D\} N(D) dD. \tag{43}$$

After normalization, one obtains the PDF model for Doppler velocities of the drops in the resolution volume:

$$p(v) = p_\Sigma(v) / \int_{-\infty}^{\infty} p_\Sigma(v) dv. \tag{44}$$

Figure 6 shows $p(v)$ in the Marshall–Palmer case ($\mu = 0$) for light (dashed curves) and heavy (solid curves) turbulence and two modes of sounding: (i) an-

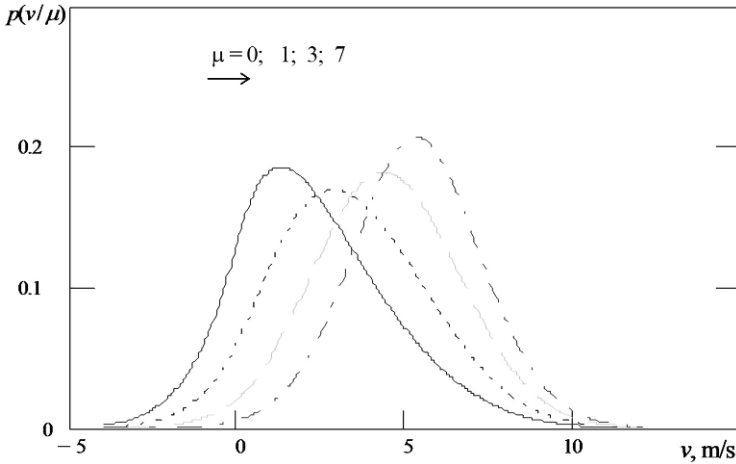


Fig. 7. Radial drop velocity distribution for different μ ($D_0 = 2$ mm, $\varepsilon = 50$ cm²s⁻³, $\theta = 90^\circ$, $L_{ul} = 1000$ m, and $V_w = 3.5$ ms⁻¹).

tenna is pointed towards zenith ($\theta = 90^\circ$) and (ii) the antenna elevation angle is 20° . The rest of the parameters are fixed: $D_0 = 2$ mm, $L_{ul} = 350$ m, and $V_m = 3.5$ ms⁻¹. It is seen that the $\theta = 90^\circ$ curves are shifted to the right of the $\theta = 20^\circ$ curves because the drop radial fall velocity is maximal when the antenna is pointed towards zenith. The maxima corresponding to heavy turbulence are significantly broader than those for light turbulence. The degree of broadening due to turbulence is more apparent at small elevation angles ($\theta = 20^\circ$) than at large elevation angles ($\theta = 90^\circ$). More positive velocities (relative to the radar) are seen in the case of zenith sounding, while more negative velocities occur in heavy turbulence. This figure demonstrates clearly that in the case of sufficiently strong turbulence, negative velocities appear in the convoluted velocity spectrum. These results will serve us as the basis for Doppler spectra calculations

Figure 7 shows the behavior of the drop velocity distribution for different values of the dispersion factor μ of the drop size distribution. The curves become more symmetric when the parameter μ is increased.

Increasing the turbulence eddy dissipation rate ε enhances the spread of the velocity distribution, as shown in **Fig. 8** generated for $\mu = 3$, $D_0 = 1.6$ mm, $\theta = 30^\circ$, $L_{ul} = 1000$ m, and $V_w = 3.5$ ms⁻¹. A similar effect occurs when L_{ul} is increased, e.g., by enlarging the radar resolution cell. Finally, the parameter V_w has a rather weak effect on the resulting drop velocity distribution. **Figure 9** shows the radial drop velocity PDF for different values of the antenna elevation θ while keeping other parameters constant ($\varepsilon = 50$ cm²s⁻³, $\mu = 3$, $D_0 = 1.6$ mm, $L_{ul} = 1000$ m, and $V_w = 3.5$ ms⁻¹). In the case of near-horizontal sounding (the left-hand curve), turbulence mainly contributes to the drop velocity distribution; the role of the gravitational fall velocity becomes more pronounced when the antenna points closer to the zenith.

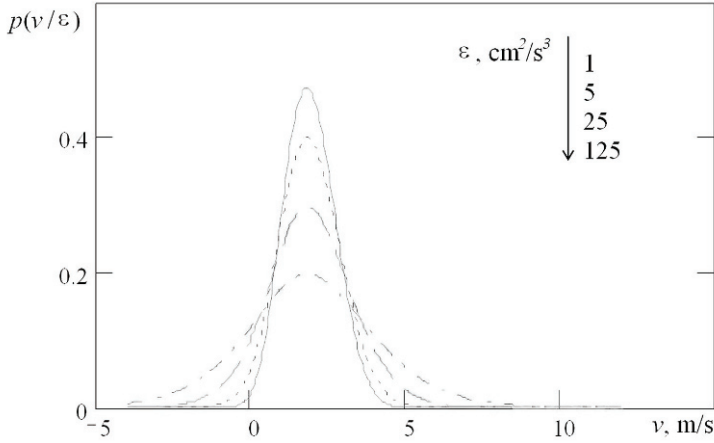


Fig. 8. Radial drop velocity distribution for different ε .

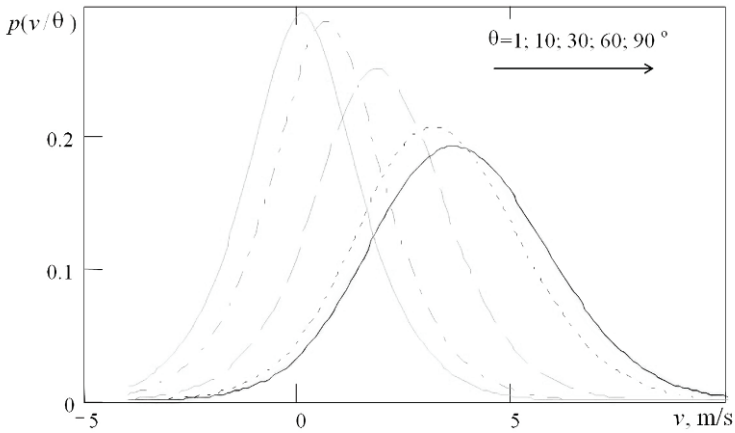


Fig. 9. Radial drop velocity distribution for different θ .

3.5. Polarization parameter models

The models developed in the previous section do not involve explicit assumptions regarding the shape of raindrops. In this section, the non-sphericity, or more specifically, the near-spheroidal shape of the particles is taken into account, thereby enabling one to model polarization parameters of the reflected signals.

3.5.1. Drop shape and orientation parameters affecting polarization

In Section 3.2.1, the axial ratio ρ was introduced as a function of the axes $a_1 = a_2$ and a_3 of a spheroid. Let us assume that the volume of the spheroid is

equal to that of a sphere with a diameter D . Then $a_1(D) = 0.5D[\rho(D)]^{-1/3}$ and $a_3(D) = a_1\rho(D)$. According to De Wolf et al. (1990), the depolarization factors λ_1 , λ_2 , and λ_3 of the spheroid can be calculated via the following formulas:

$$\lambda_3(D) = \frac{1 - [e(D)]^2}{[e(D)]^2} \left[-1 + \frac{1}{2e(D)} \ln \frac{1+e(D)}{1-e(D)} \right], \quad e^2 = 1 - \rho^{-2}, \quad 0 < \rho^{-1} < 1, \quad (45)$$

$$\lambda_3(D) = \frac{1 + [f(D)]^2}{[f(D)]^2} \left[1 - \frac{1}{f(D)} \arctan f(D) \right], \quad f^2 = \rho^{-2} - 1, \quad \rho^{-1} > 1, \quad (46)$$

$$\lambda_1(D) = \frac{1 - \lambda_3(D)}{2}, \quad \lambda_2(D) = \lambda_1(D). \quad (47)$$

Taking into account the relative permittivity of water ε_r , the shape factors of the spheroidal raindrop are as follows:

$$\Lambda_i(D) = [1 + \lambda_i(D)(\varepsilon_r - 1)]^{-1}, \quad i = 1, 2, 3. \quad (48)$$

Thus, in the framework of the accepted model, the drop shape can be calculated for any drop equivolumetric diameter D .

Let us now consider the orientation parameters affecting the polarization of the reflected signal. The particle azimuth α and canting δ angles, the antenna elevation angle θ , and the polarization angle ϕ can be considered parameters characterizing the mutual orientation of the particle and the sounding wave. Simple yet rather bulky formulas for the polarimetric orientation parameters Φ_{HH} , Φ_{VV} , and Φ_{HV} for co-polar (HH , VV) and cross-polar (HV) signals were derived by Ruschenberg (1992):

$$\begin{aligned} \Phi_{HH} = & (\sin \delta \cos \alpha \sin \phi \sin \theta)^2 + (\sin \delta \sin \alpha \cos \phi)^2 \\ & + (\cos \delta \sin \phi \cos \theta)^2 - \frac{1}{2} \sin 2\delta \cos \alpha \sin^2 \phi \sin 2\theta \\ & - \frac{1}{2} \sin 2\delta \sin \alpha \sin 2\phi \cos \theta + \frac{1}{2} \sin^2 \delta \sin 2\alpha \sin 2\phi \sin \theta, \end{aligned} \quad (49)$$

$$\begin{aligned} \Phi_{VV} = & (\sin \delta \cos \alpha \cos \phi \sin \theta)^2 + (\sin \delta \sin \alpha \sin \phi)^2 \\ & + (\cos \delta \cos \phi \cos \theta)^2 - \frac{1}{2} \sin 2\delta \cos \alpha \cos^2 \phi \sin 2\theta \\ & + \frac{1}{2} \sin 2\delta \sin \alpha \sin 2\phi \cos \theta - \frac{1}{2} \sin^2 \delta \sin 2\alpha \sin 2\phi \sin \theta, \end{aligned} \quad (50)$$

$$\begin{aligned} \Phi_{HV} = & \frac{1}{2} \sin 2\phi [(\sin \delta \sin \alpha)^2 - (\sin \delta \cos \alpha \sin \theta)^2 \\ & - \cos^2 \delta + \frac{1}{2} \sin 2\delta \cos \alpha \sin 2\theta] \\ & - \frac{1}{2} \cos 2\phi (\sin 2\alpha \sin^2 \delta \sin \theta + \frac{1}{2} \sin 2\delta \sin \alpha \cos \theta). \end{aligned} \quad (51)$$

According to De Wolf et al. (1990) and Ruschenberg (1992), the combined parameters taking into account both the shape and the orientation of a drop are as follows:

$$Q_{HH}(D, \delta, \alpha, \theta) = \{A_1(D) + [A_3(D) - A_1(D)]\Phi_{HH}(\delta, \alpha, \phi, \theta)\}^2, \quad (52)$$

$$Q_{VV}[D, \delta, \alpha, \theta] = \{A_1(D) + [A_3(D) - A_1(D)]\Phi_{VV}(\delta, \alpha, \phi, \theta)\}^2, \quad (53)$$

$$Q_{HV}(D, \delta, \alpha, \theta) = \{[A_3(D) - A_1(D)]\Phi_{HV}(\delta, \alpha, \phi, \theta)\}^2. \quad (54)$$

3.5.2. Radar cross section of spheroidal drops

The models of RCS calculation for raindrops with an equivolumetric diameter D in the Rayleigh approximation were considered by De Wolf et al. (1990). Using the notation introduced in Section 3.2, the RCS of a spheroidal drop with a relative permittivity ε_r at a wavelength $\lambda \gg D$ is given by

$$\sigma_{xy}(D) = \frac{\pi^5 D^6}{9\lambda^4} |\varepsilon_r - 1|^2 Q_{xy}, \quad x, y = h, v. \quad (55)$$

Here, the complex parameter

$$Q_{xy} = F_{xy}(\Lambda)\Phi_{xy}(\delta, \alpha, \theta) \quad (56)$$

is responsible for the polarization characteristics of the RCS. In Eq. (56), F_{xy} represents the particle shape (parameterized with a vector of shape parameters Λ) according to Eqs. (56)–(58) and (45)–(47), while Φ_{xy} takes into account the particle orientation (parameterized by the canting angle δ and azimuth angle α) and the radar elevation angle θ according to Eqs. (49)–(51). In the special case of a spherical particle, $Q_{HH} = Q_{VV} = 9|2 + \varepsilon_r|^2$ and $Q_{HV} = Q_{VH} = 0$, and so $\sigma_{HH} = \sigma_{VV} = \pi^5 D^6 \lambda^{-4} |\varepsilon_r - 1|^2 |\varepsilon_r + 2|^2$ and $\sigma_{HV} = \sigma_{VH} = 0$, that is, Eq. (55) is reduced to the well-known result of Rayleigh scattering.

3.5.3. Polarization parameters of an ensemble of drops

In order to calculate the conventional polarization parameters of the signal reflected from an ensemble of drops located inside the resolution volume, the integration over all scatterers must be performed. Let us calculate the polarization parameters Zdr and Ldr assuming that $p_\delta(\delta)$ is the PDF of the particle canting angle, the particle azimuth α is a fixed, and $N(D)$ is the drop size distribution:

$$Zdr = 10 \log \frac{\int_{\delta_{\min}}^{\delta_{\max}} \int_{D_{\min}}^{D_{\max}} \sigma_{HH}(D, \delta, \alpha, \theta) N(D, D_0, \mu) p_\delta(\delta) dD d\delta}{\int_{\delta_{\min}}^{\delta_{\max}} \int_{D_{\min}}^{D_{\max}} \sigma_{VV}(D, \delta, \alpha, \theta) N(D, D_0, \mu) p_\delta(\delta) dD d\delta}, \quad (57)$$

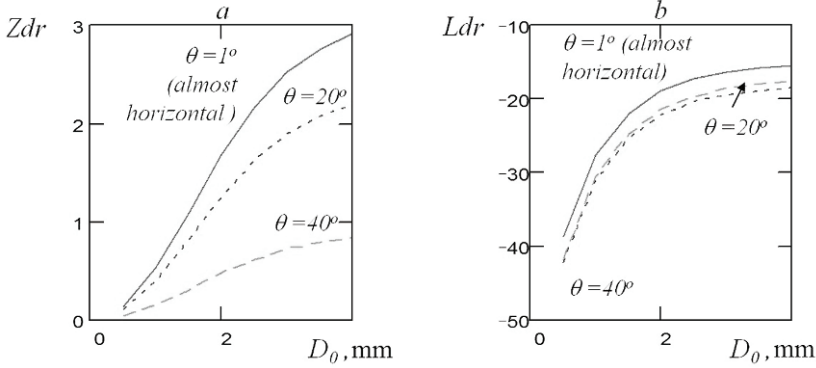


Fig. 10. Zdr and Ldr [dB] as functions of D_0 for different antenna elevations θ .

$$Ldr = 10 \log \frac{\int_{\delta_{\min}}^{\delta_{\max}} \int_{D_{\min}}^{D_{\max}} \sigma_{HV}(D, \delta, \alpha, \theta) N(D, D_0, \mu) p_{\delta}(\delta) dD d\delta}{\int_{\delta_{\min}}^{\delta_{\max}} \int_{D_{\min}}^{D_{\max}} \sigma_{HH}(D, \delta, \alpha, \theta) N(D, D_0, \mu) p_{\delta}(\delta) dD d\delta}. \quad (58)$$

An example of Zdr and Ldr calculations as functions of the median equivolumetric drop diameter D_0 using the Gaussian $p_{\delta}(\delta)$ (see Section 3.3.1) for $\sigma_{\delta} = 30^{\circ}$, $\mu = 0$, $D_{\min} = 0.1$, $D_{\max} = 8$, $\delta_{\min} = 0$, $\delta_{\max} = \pi$, and different antenna elevation angles θ is given in Fig. 10. These dependencies are consistent with common sense and actual measurements. However, owing to the integration over D they cannot be used for modeling Doppler–polarimetric spectral functions such as $sZdr(v)$. For the purpose of Doppler polarimetry, it is not sufficient to obtain the polarization parameters as functions of an integral drop size distribution parameter such as D_0 . The general concept of our model (Section 3.2) requires the polarization parameters as functions of the drop diameter taking into account statistical characteristics of particle orientations.

3.5.4. Averaging over particle orientations

General expressions for the orientation-averaged parameters Q_{xy} , $x, y = H, V$, described by Eq. (56) are as follows:

$$\overline{Q_{xy}}(D, \theta) = \int_{\delta} \int_{\alpha} Q_{xy}(D, \delta, \alpha, \theta) p_{\alpha}(\alpha) p_{\delta}(\delta) d\alpha d\delta, \quad (59)$$

where $p_{\alpha}(\alpha)$ is the drop azimuth distribution and $p_{\delta}(\delta)$ is the drop canting distribution. The polarization angle ϕ in the initial expressions of Q_{xy} can be assumed to be zero without loss of generality. The expressions (52)–(54) with non-averaged values Φ_{xy} , $x, y = H, V$, can be considered as sums of squares. There-

fore, the average values of orientation parameters $\overline{\Phi_{HH}}$, $\overline{\Phi_{HH}^2}$, $\overline{\Phi_{VV}}$, $\overline{\Phi_{VV}^2}$, and $\overline{\Phi_{HV}^2}$ are needed; they were obtained analytically by Yanovsky (1998a). For example, some simple final expressions obtained are as follows:

$$\overline{\Phi_{HH}} \approx \frac{1}{4}[1 - \exp(-2\sigma_\delta^2)], \quad \overline{\Phi_{HH}^2} = \frac{3}{64}(e^{-8\sigma_\delta^2} - 4e^{-2\sigma_\delta^2} + 3), \quad (60)$$

$$\overline{\Phi_{VV}} \approx \frac{1}{4}[\cos^2 \theta + 3 \cos^2 \theta \exp(-2\sigma_\delta^2) - \exp(-2\sigma_\delta^2) + 1]. \quad (61)$$

One can see from the expressions (60) that there is no dependence on the antenna elevation in the case of the HH polarization. Physically, this is because the plane of polarization rotates around the H polarization axis when the elevation angle is changed.

3.5.5. Polarization parameters versus equivolumetric diameter in an ensemble of drops

The average shape–orientation parameters can be derived by combining the expressions (59) with Eqs. (52)–(54) as well as the average orientation parameters $\overline{\Phi_{xy}^i}$, $i = 1, 2$:

$$\overline{Q_{HH}}(D, \theta, \sigma_\delta) = \Lambda_1^2 + 2\Lambda_1(\Lambda_3 - \Lambda_1)\overline{\Phi_{HH}} + (\Lambda_3 - \Lambda_1)^2\overline{\Phi_{HH}^2}, \quad (62)$$

$$\overline{Q_{VV}}(D, \theta, \sigma_\delta) = \Lambda_1^2 + 2\Lambda_1(\Lambda_3 - \Lambda_1)\overline{\Phi_{VV}} + (\Lambda_3 - \Lambda_1)^2\overline{\Phi_{VV}^2}, \quad (63)$$

$$\overline{Q_{HV}}(D, \theta, \sigma_\delta) = (\Lambda_3 - \Lambda_1)^2\overline{\Phi_{HV}^2}. \quad (64)$$

Because no averaging over D is performed, it is not necessary to multiply $\overline{Q_{HH}}$, $\overline{Q_{VV}}$, and $\overline{Q_{HV}}$ by the radar cross section $\sigma_{xy}(D)$. Finally, the differential radar reflectivity and linear depolarization ratio can be calculated as functions of D via the following formulas:

$$Zdr = 10 \log \frac{\overline{Q_{HH}}(D, \sigma_\delta)}{\overline{Q_{VV}}(D, \theta, \sigma_\delta)}, \quad Ldr = 10 \log \frac{\overline{Q_{HV}}(D, \theta, \sigma_\delta)}{\overline{Q_{VV}}(D, \theta, \sigma_\delta)}. \quad (65)$$

These expressions yield quite realistic curves of Zdr and Ldr versus D for different θ and σ_δ (Yanovsky 1998a).

3.6. Doppler–polarimetric characteristics

Doppler–polarimetric spectra and Doppler–polarimetric parameters such as the spectral differential reflectivity introduced theoretically in Section 2 are considered here in more detail for the case of rain.

3.6.1. Polarimetric Doppler spectra

In accordance with our main concept (Section 3.2), in the frequency domain the complex model provides Doppler spectra for different combinations of polari-

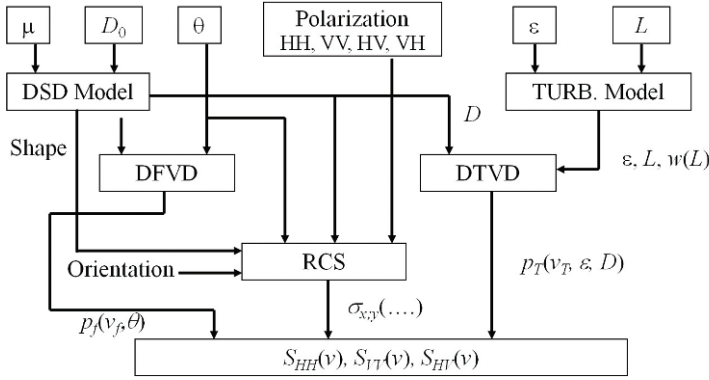


Fig. 11. Generalized structure of the phenomenological model developed for the computation of Doppler spectra under different conditions.

zations for transmitted and received waves. [Figure 11](#) shows the general structure of the model, which uses as input data the parameters of the drop size distribution (DSD; μ and D_0), the antenna elevation (θ), the polarization mode, the parameters of turbulence (ϵ and L), and the radar performance specifications (not shown). Based on the mathematical models described above, the model yields the drop fall velocity distribution (DFVD); drop turbulence velocity distribution (DTVD); and RCS for a given polarization mode, DSD, particle shape, antenna elevation, wavelength, etc. Finally, the requisite spectra and Doppler–polarimetric parameters are computed (Yanovsky 2002).

In the case of the HV linear polarization basis, at least three functions, $S_{HH}(v)$, $S_{VV}(v)$, and $S_{HV}(v)$, are generated for given conditions. They are the models of Doppler energy spectra for different combinations of polarizations of transmitted and received waves, i.e., the Doppler–polarimetric spectra:

$$S_{xy}(v) = \int_{D_{\min}}^{D_{\max}} p(v/D, \epsilon, \theta) \sigma_{xy}(D/\theta, \epsilon) N(D/D_0) dD, \quad x, y = H, V. \quad (66)$$

An example of calculations for the same hypothetic rain event but for different polarizations is shown in [Fig. 12](#) (left-hand panel). The upper curve corresponds to the HH spectrum, while the lower curve represents the VV case. In the case of a slant sounding of rain, the horizontal polarization provides more energy in the reflected signal due to predominantly horizontal orientation of the larger axes of spheroidal drops.

In the absence of turbulence, the Doppler spectrum $S_{xy}^f(v)$ is controlled only by the DFVD and is given by

$$S_{xy}^f(v) dv = N_f(v) \sigma_{xy}(v) dv, \quad (67)$$

while including turbulence yields:

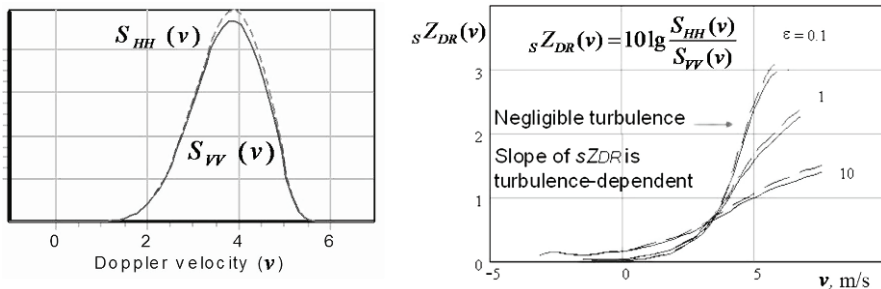


Fig. 12. Generated Doppler spectra for horizontal and vertical polarizations (left-hand panel). Three $sZdr(\nu)$ curves are computed for different turbulence intensities (negligible, light, and moderate) corresponding to eddy dissipation rates $\epsilon = 0.1, 1,$ and $10 \text{ mm}^2 \text{ s}^{-3}$ as well as for two modes of the model (right-hand panel).

$$S_{xy}^{ft}(\nu) = \int_{-\infty}^{\infty} S_{xy}^f(\nu - \nu_T) N_T(\nu_T) d\nu_T, \quad (68)$$

where xy denotes the polarization pair $HH, HV,$ or VV , and $\sigma_{xy}(\nu)$ is the RCS of the particle with velocity ν . Since turbulence produces different wind velocities at different locations in the radar volume, equal-sized raindrops will not appear in the same velocity bin of the Doppler spectrum. This trivial aspect is important when Doppler spectra are calculated for different polarizations, as will be shown later.

3.6.2. Polarization parameters of rain as functions of the Doppler velocity

Following the Doppler–polarimetric approach, one can construct different polarimetric characteristics, for example, those corresponding to Eqs. (13)–(16), but as functions of the Doppler velocity by using the elements of the spectral target covariance matrix (19) instead of the conventional covariance matrix (12). Using the models (66) according to the above discussion, one can calculate the radar observables (20), (21), etc. for different conditions.

An example of calculating the spectral differential reflectivity in rain as a function of the Doppler velocity for different turbulence intensities ϵ and fixed remaining parameters is shown in Fig. 12 (left-hand panel). It is seen that the $sZdr(\nu)$ curve flattens with increasing intensity of turbulence. In the case of negligible turbulence ($\epsilon = 0.1 \text{ cm}^2 \text{ s}^{-3}$), the larger droplets are more oblate and fall faster than the smaller ones. If scatterers become more oblate then $sZdr$ increases. This behavior changes in the case of substantial turbulence. Because of the turbulence-induced random mixing, particles with different shapes and velocities are mixed, resulting in a flattened $sZdr(\nu)$ curve.

In theory, the spectral linear depolarization ratio $sLdr(\nu)$ behaves similarly; however, its values in rain are rather small, typically -30 dB and even much smaller. Obviously, it is difficult to measure such values reliably.

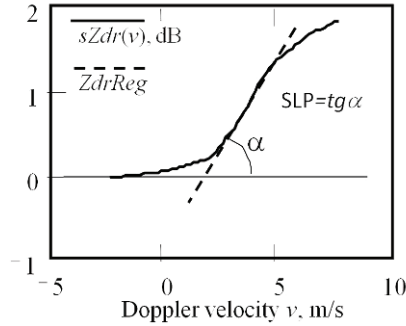


Fig. 13. Definition of the slope of the $sZdr(v)$ curve, SLP, as the tangent of the angle α .

3.6.3. Slope of $sZdr(v)$

Unlike the traditional Zdr parameter (13), the $sZdr(v)$ (20) is a function (generally nonlinear) that in the case of a rain increases monotonously. To characterize a $sZdr(v)$ curve, the slope of $sZdr$ (SLP) was introduced (Yanovsky et al. 2005). The SLP is estimated as the slope ratio of the tangent at the inflection point, as explained in Fig. 13.

The SLP is not the only parameter analyzed previously; another informative parameter is DELTA defined as the difference between the maximum and minimum of the $sZdr(v)$ curve (Yanovsky et al. 2003a). However, the SLP is more sensitive to turbulence intensity.

3.6.4. Differential Doppler velocity

Another Doppler–polarimetric parameter is DDV defined as the difference between the mean Doppler velocities for horizontal and vertical polarizations: $\Delta V = \overline{V_{HH}} - \overline{V_{VV}}$ (Yanovsky et al. 2003b). It can be calculated using modeled Doppler spectra according to

$$\Delta V = \int_{D_{\min}}^{D_{\max}} v S_{HH}(v) dv - \int_{D_{\min}}^{D_{\max}} v S_{VV}(v) dv. \tag{69}$$

The actual DDV varies with elevation angle θ ; this pronounced dependence can be used to retrieve useful information by comparing measurement and model results.

4. Analysis of polarimetric parameters

In this section we consider analysis results based on computing Doppler–polarimetric parameters for different conditions with the help of the phenomenological complex model described above.

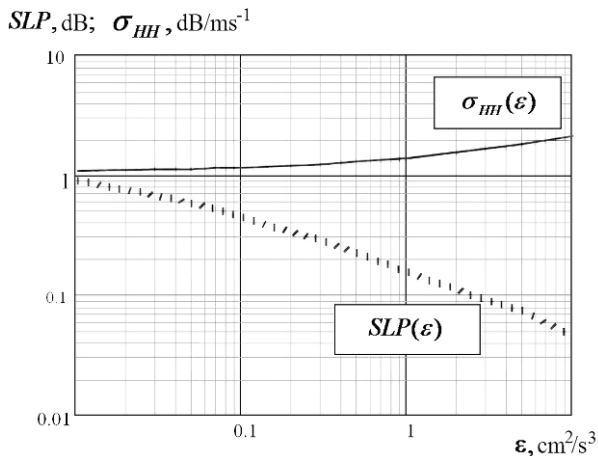


Fig. 14. Relationships between the Doppler-spectrum width σ_v and the $sZdr(v)$ slope, SLP, and the intensity of turbulence in rain defined by the eddy dissipation rate ϵ .

4.1. Relation of SLP and slope of $sZdr(v)$ to turbulence intensity

In the practice of radar meteorology, the main parameter traditionally used to retrieve information on turbulence intensity characterized by the eddy dissipation rate ϵ is the Doppler spectrum width σ_v (Doviak and Zrníć 1993). Figure 14 shows the relationship between $\sigma_v(\epsilon)$ and $SLP(\epsilon)$ computed on the basis of mathematical models and the computer realization of the complex phenomenological model.

Figure 15 confirms the strong influence of turbulence on the behavior of the spectral differential reflectivity (left-hand panel) and the spectral linear depolarization ratio. These plots are calculated for eddy dissipation rates ϵ ranging from 0.1 up to $100 \text{ cm}^2 \text{ s}^{-3}$, which implies a wide range of turbulence intensity from negligible to severe. Figure 16 shows the effect of the drop size distribution on the relationships between SLP and other parameters. More specifically, the left-hand panel depicts SLP as a function of ϵ for different values of the spread parameter μ of the gamma drop size distribution, while the right-hand panel shows the inverse value $1/SLP$ versus the Doppler spectrum width for the same μ .

4.2. Relation of DDV to turbulence intensity and rain microstructure

The first study of DDV as a radar parameter for characterizing the microstructure of weather formation was performed by Wilson et al. (1997). Their paper contains a detailed analysis of the relationship between DDV and the hydrometeor size distributions under the assumption that turbulence does not affect the former.

The effect of turbulence was taken into account by Yanovsky et al. (2003b). Figure 17 shows the relation of DDV for an antenna elevation of 45° to parameters of the drop size distribution for three values of turbulence intensity: $\epsilon = 0.1 \text{ cm}^2 \text{ s}^{-3}$

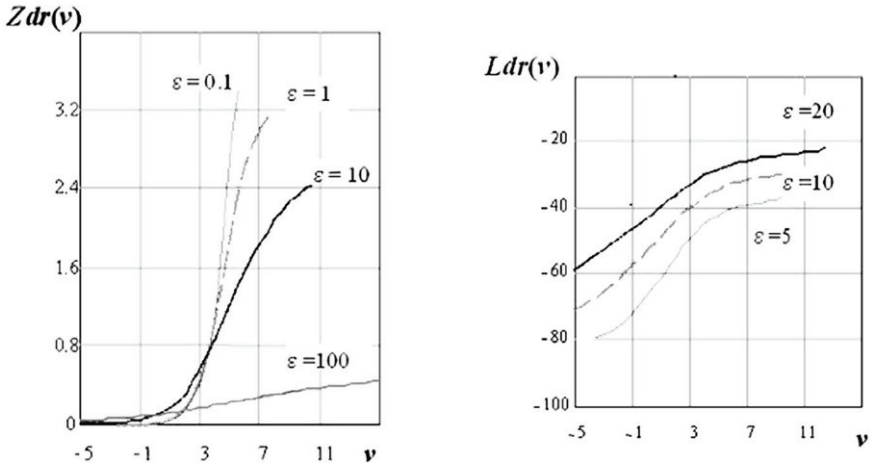


Fig. 15. Spectral differential reflectivity (left-hand panel) and spectral linear depolarization ratio as functions of the Doppler velocity computed for different intensities of turbulence represented by a wide range of eddy dissipation rates ϵ from 0.1 up to $100 \text{ cm}^2 \text{ s}^{-3}$.

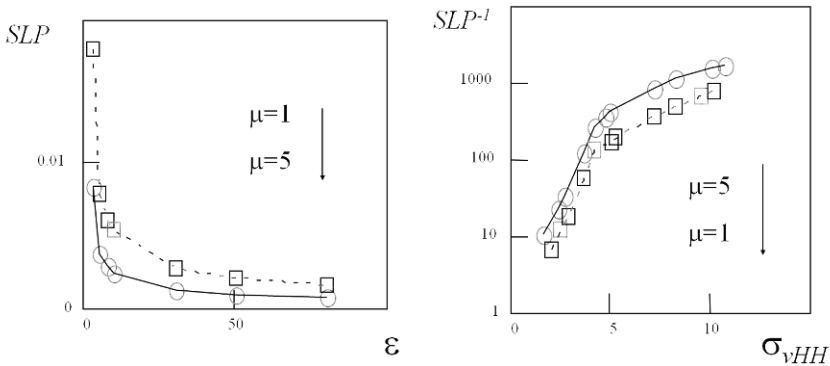


Fig. 16. Relationships between SLP and ϵ (left-hand panel) and between $1/SLP$ and Doppler spectrum width for different spread parameters μ of the gamma drop size distribution.

(solid curve), $\epsilon = 10$ (dotted curve), and $\epsilon = 100$ (dashed curve). These results show a pronounced dependence of DDV on μ and D_0 . It is seen that the solid and dotted curves are close to each other on both panels, which implies that light turbulence does not affect the relationships significantly. However strong turbulence (dashed curves) is rather important.

The direct dependence of DDV on the eddy dissipation rate is illustrated in Fig. 18 corresponding to $\theta = 30^\circ$, $D_0 = 1.5 \text{ mm}$, and $\mu = 5$. One can see that the rate of change of the function $DDV = f(\epsilon)$ is increasing with ϵ and then is almost constant for strong and severe turbulence.

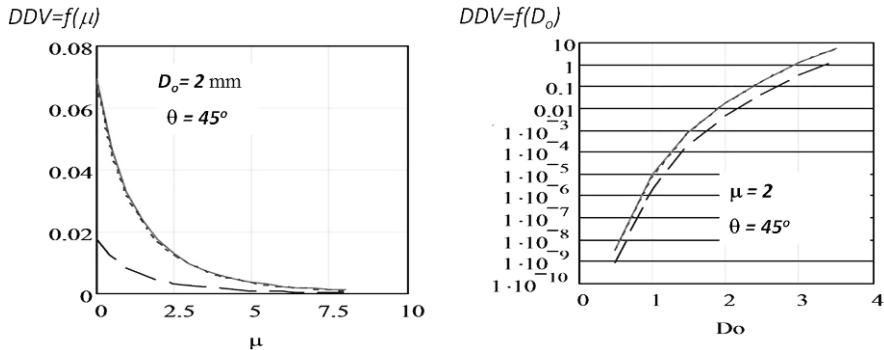


Fig. 17. Dependence of DDV on the parameters of the drop size distribution (μ and D_0) for three values of the turbulence intensity $\varepsilon = 0.1 \text{ cm}^2\text{s}^{-3}$ (solid curve), $\varepsilon = 10$ (dotted curve), and $\varepsilon = 100$ (dashed curve). The antenna elevation is 45° .

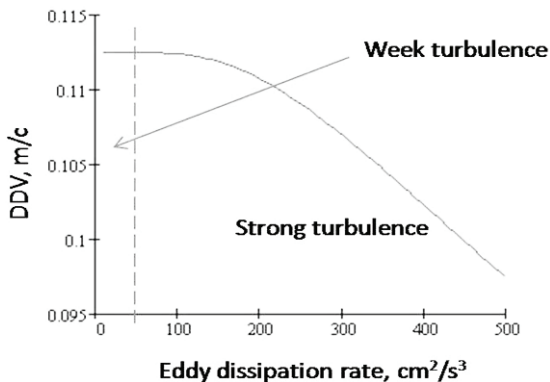


Fig. 18. Dependence of DDV on the eddy dissipation rate for $\theta = 30^\circ$, $D_0 = 1.5 \text{ mm}$, and $\mu = 5$.

Additional analyses of the DDV parameter and its relation to other radar parameters and the object features have been presented by Glushko and Yanovsky (2009, 2010) and Yanovsky and Glushko (2010).

5. Measurements

The Doppler–polarimetric measurements with the radar TARA (Yanovsky et al. 1997; Heijnen et al. 2000) were used for the verification of the model described above. The combination of Doppler and polarimetric measurements is discussed in Unal and Moisseev (2004). Using the measured time series of scattering matrices, we perform the Fourier transforms (or Doppler processing) and calculate the second moments of the resulting spectral scattering matrices, which yields polarimetric spectrograms. Some results were discussed in Yanovsky et al. (2005, 2007).

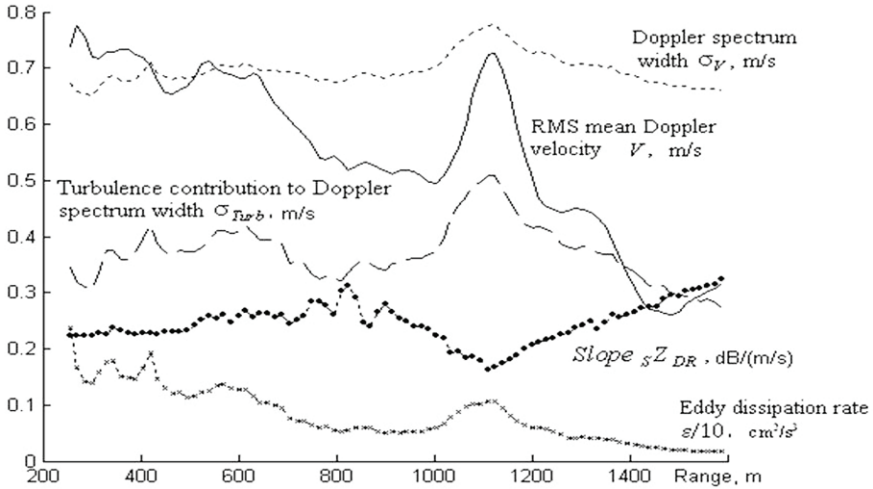


Fig. 19. Comparison of different measures of turbulence.

Measurements were performed with different antenna elevations, including vertical (zenith) sounding, but mostly in the slant sounding mode (45° and 30°). In this chapter we will not discuss the details of signal processing, but it is useful to mention that the received signal was subjected to the procedures of unfolding, clipping, averaging, and smoothing. Moments and other parameters of the Doppler spectra were estimated as well as polarimetric parameters such as $sZdr(v)$, SLP, DDV, etc. The turbulence eddy dissipation rate was retrieved from the Doppler spectrum width using the established methodology (Doviak and Zrnić 1993) and making the correction for the drop fall velocity variance as described in Yanovsky et al. (2005).

5.1. Comparison of different parameters

Comparison of different measures of turbulence is presented in Fig. 19. One can identify the maxima of all parameters, except for SLP ($Slope\ sZdr$), occurring at ~ 1140 m; these maxima are perfectly collocated with the SLP minimum.

Space and time distributions of different Doppler–polarimetric parameters are shown in Fig. 20. The horizontal axis is time and the vertical axis is height, while the value of each parameter is shown by color according to the color bar on the right-hand side of the respective panel. All fields were obtained by processing the same raw data of light overcast rain in The Netherlands (Yanovsky et al. 2003a, 2005). One can clearly see correlation between all these parameters. For example, small (blue) values of SLP (the stripe at the ~ 800 m altitude) correspond to large values of the Doppler spectrum width, the eddy dissipation rate, and the rms Doppler velocity. The behavior of DDV is similar to that of SLP, but it is less sensitive

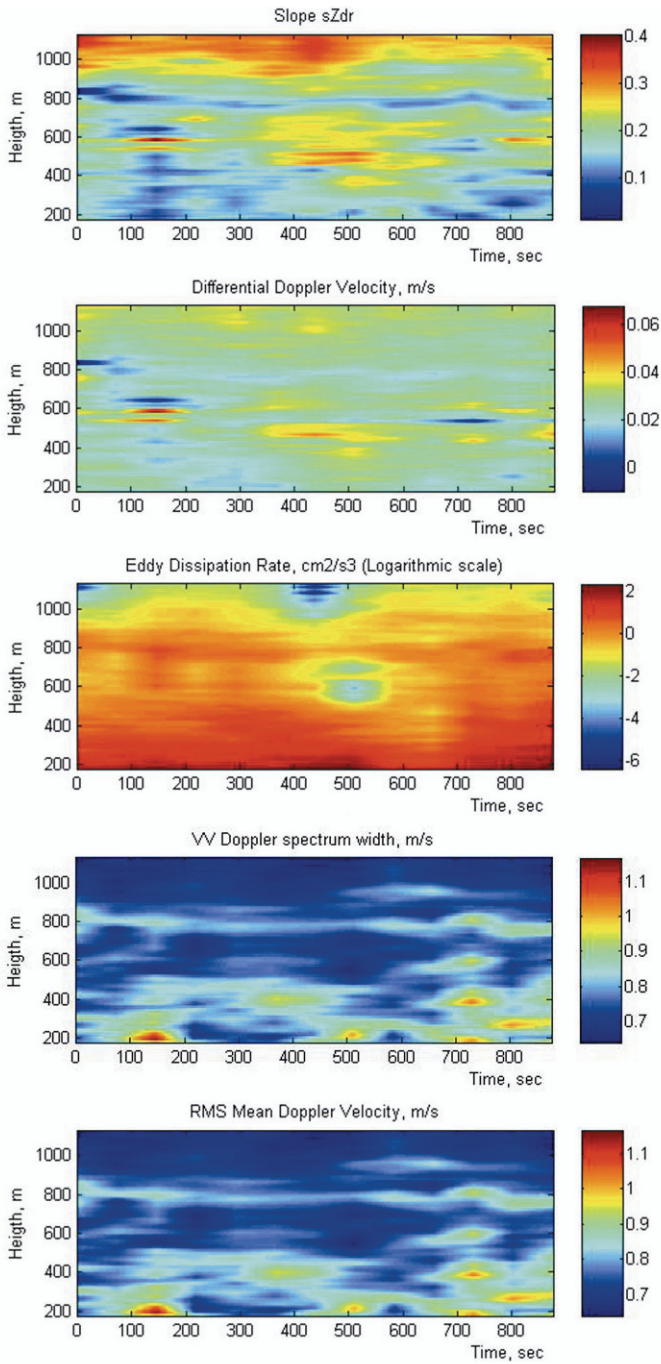


Fig. 20. Comparison of space–time behavior of Doppler–polarimetric parameters.

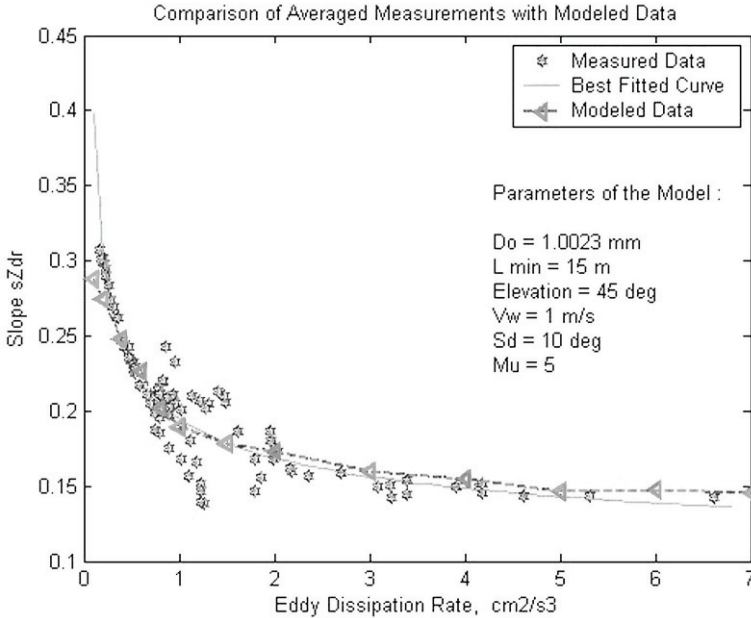


Fig. 21. Comparison of modeled and measured data.

to the turbulence intensity. The measurements show that the above-mentioned correlation is observed not only in the space domain but also in time.

5.2. Comparison between the model and measurements

As an example, in Fig. 21 the model results are shown together with the processed measurements. The former are represented by triangles connected by the dashed curve. The solid gray curve shows the best least-squares fit. We processed a ~15-min measurement dataset accumulated for overcast rain. The key parameters of the model were chosen to maximally correspond to the parameters of the real event.

Specifically, $D_0 = 1.023$ mm is the average median drop diameter retrieved from the reflectivity and rain rate; $L_{ul} = 15$ m is equal to the radar resolution; $\theta = 45^\circ$ is equal to the antenna elevation. The remaining parameters of the model were derived from the best fit to the measured data.

One can see that the model is in a rather good agreement with the measurements and not too far from the best-fit curve.

6. Discussion and applications

So far the model results and the Doppler–polarimetric measurements (reflectivity, Doppler spectrum, spectral differential reflectivity, DDV, SLP) have shown

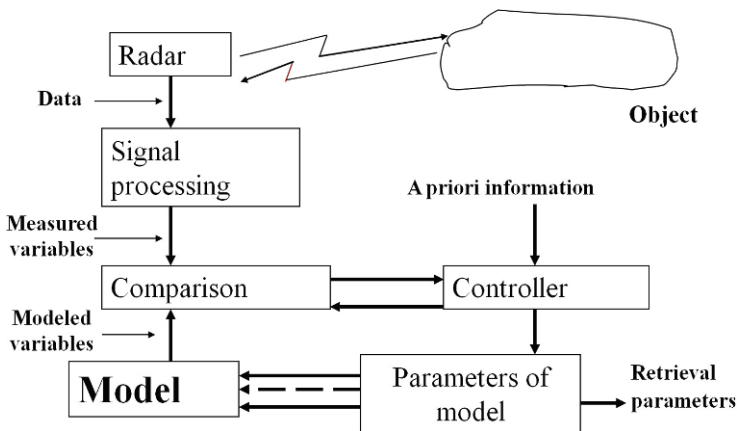


Fig. 22. Combination of modeling and measurements.

consistency, thereby implying that the model can be used to estimate the microphysics of rain events. Several comparisons have been done for light stratiform rains, in which case it is necessary to consider medium or large rain intensity in order to verify the significant differentiation in the $sZdr$ slope with the eddy dissipation rate. For light rain rates (see Fig. 18, where D_0 is ~ 1 mm), the sensitivity of the $sZdr$ curve to the eddy dissipation rate is weak when the turbulence starts to be moderate (eddy dissipation rates $> 3 \text{ cm}^2 \text{ s}^{-3}$). With this example of measurements, we cover only the interval $[0, 3] \text{ ms}^{-1}$ of radial drop velocities in Fig. 15, which is not sufficient to obtain significantly different $sZdr$ slopes. From 5 ms^{-1} , there is a clear differentiation of the $sZdr$ slope when the eddy dissipation rate varies. We need, therefore, moderate to high rain intensities for a further verification and to study whether the retrieval of the eddy dissipation rate from the $sZdr$ slope is feasible.

As follows from Fig. 21, the combination of modeling and measurements makes it possible to obtain much indirect information about the object of observation, specifically, on the microstructure of and turbulence in rain. The mathematical representation of the object described above and the relations between its components and the radar sounding waveform improve our understanding and allow the inference of the microstructure and turbulence properties of rain by comparing modeled and measured data and tuning the model to bring its results in agreement with the object. This is illustrated in Fig. 22.

The approach developed can be used operationally to improve the quality of radar meteorological information. New parameters SLP, DDV, and other polarimetric characteristics are independent of the Doppler spectrum width and other conventional parameters and, therefore, can be used together to obtain more reliable and comprehensive information on the microstructure and turbulence of rain as well as for the retrieval of the hydrometeor type.

7. Conclusions

The Doppler–polarimetric radar is a powerful tool for remote sensing of rain. The potential of Doppler polarimetry exceeds significantly that of the sum of the Doppler technique and polarimetry used separately. Doppler–polarimetric techniques provide promising capabilities for data interpretation. However, the number of variables grows significantly and the mutual interrelations between them are rather complicated. The interpretation of Doppler–polarimetric measurements requires an adequate model capable of relating the Doppler–polarimetric observables to the parameters of precipitation. Such a model has been created and described in this chapter.

The approach presented has facilitated the selection of certain Doppler–polarimetric parameters helping to retrieve turbulence parameters in precipitation as well as the microstructure of rain. The results of actual data processing using different approaches have demonstrated good agreement. The application of independent turbulence-related variables can improve the reliability and accuracy of radar retrievals in rain.

Measurements of weather-related objects by Doppler–polarimetric radars can be very useful in solving the wave propagation problem for the tasks of microwave remote sensing, communications, and radar detection of targets embedded in intensely cluttered backgrounds.

The continued research in this field should result in further improvements of models, the development of algorithms and devices for real-time measurements, more efficient processing, coverage of objects other than rain, optimal combination of different parameters, and adaptive methods of measurements and signal processing.

Our approach to the Doppler–polarimetric modeling is promising, deserves further development, and may play an important role in radar remote-sensing data interpretation and signal processing.

Acknowledgments

The project on Doppler polarimetry was supported in part by STW (the Dutch Technology Foundation). It was part of a joint project of the International Research Centre for Telecommunications and Radar (IRCTR) at TU-Delft and the National Aviation University in Kyiv. The author is grateful to H. Russchenberg, C. Unal, L. Ligthart, A. Yarovoy, O. Krasnov, and S. Heijnen for cooperation during collaborative work in TU-Delft.

References

- Atlas, D., R. C. Srivastava, R. S. Sekhon, 1973: Doppler radar characteristics of precipitation at vertical incidence. *Rev. Geophys. Space Phys.* **11**, 1–35.
- Bachmann, S., and D. S. Znić, 2006: Spectral density of polarimetric variables separating biological scatterers in the VAD display. *J. Atmos. Oceanic Technol.* **24**, 1186–1198.

- Beard, K. H., and C. Chuang, 1987: A new model for equilibrium shape of rain drops. *J. Atmos. Sci.* **44**, 1509–1524.
- Bringi, V. N., and V. Chandrasekar, 2001: *Polarimetric Doppler Weather Radar: Principles and Applications* (Cambridge University Press, Cambridge, UK).
- Brussaard, G., 1976: A meteorological model for rain-induced cross polarization. *IEEE Trans. Antennas Propag.* **24**, 5–11.
- De Wolf, D. A., H. W. J. Russchenberg, and L. P. Ligthart, 1990: Effective permittivity of and scattering from wet snow and ice droplets at weather radar wavelengths. *IEEE Trans. Antennas Propag.* **38**, 1317–1325.
- Dobrolensky, Yu. P., 1969: *Dynamics of Flight in Turbulent Atmosphere* (Transport Press, Moscow) (in Russian).
- Doviak, R. J., and D. S. Zrnić, 1993: *Doppler Radar and Weather Observations* (Academic Press, San Diego).
- Fishman, B. E., and F. J. Yanovsky, 1983: To the theory of radar signals' formation reflected from meteorological objects. *Radiotekhnika*, 56–57 (in Russian).
- Frisch, U., 1995: *Turbulence: The Legacy of A. N. Kolmogorov* (Cambridge University Press, Cambridge, UK).
- Foote, G. B., and P. S. du Toit, 1969: Terminal velocity of rain drops aloft. *J. Appl. Meteorol.* **8**, 249–253.
- Glushko, D. N., and F. J. Yanovsky, 2009: DDV – novel Doppler–polarimetric technique for remote sensing of precipitation. In *Proceedings of the 6th European Radar Conference* (Rome), pp. 298–301.
- Glushko, D. N., and F. J. Yanovsky, 2010: Analysis of differential Doppler velocity for remote sensing of clouds and precipitation with dual-polarization S-band radar. *J. Microwave Wireless Technol.* **2**, 391–398.
- Gorelik, A. G., and A. A. Chernikov, 1964: Some results of the radar investigation of the wind structure on heights 50–700 m. *Trudy (Proc.) Central Aerological Observ.* **57**, 3–17 (in Russian).
- Heijnen, S. H., L. P. Ligthart, and H. W. J. Russchenberg, 2000: First measurements with TARA; an S-Band transportable atmospheric radar. *Phys. Chem. Earth. B* **25**, 995–998.
- Kezys, V., E. Torlaschi, and S. Haykin, 1993: Potential capabilities of coherent dual polarization X-band radar. In *Proceedings of the 26th International Conference on Radar Meteorology* (American Meteorological Society, Boston), pp. 106–108.
- Kozlov, A. I., L. P. Ligthart, and A. I. Logvin, 2002: *Mathematical and Physical Modelling of Microwave Scattering and Polarimetric Remote Sensing* (Kluwer, Dordrecht).
- Kubesh, R. J., and K. V. Beard, 1993: Laboratory measurements of spontaneous oscillations for moderate-size raindrops. *J. Atmos. Sci.* **50**, 1089–1098.
- Kwiatkowski, J. M., A. B. Kostinski, and A. R. Jameson, 1995: The use of optimal polarizations for studying the microphysics of precipitation: nonattenuating wavelengths. *J. Atmos. Oceanic Technol.* **12**, 96–114.
- Ligthart, L. P., and L. R. Nieuwkerk, 1980: FM-CW Delft atmospheric research radar. *Proc. IEE F* **127**, 421–426.
- MacCready, P., 1964: Standardization of gustiness values from aircraft. *J. Appl. Meteorol.* **3**, 439–449.
- Pruppacher, H. R., and K. V. Beard, 1970: A wind tunnel investigation of the internal circulation and shape of water drops falling at terminal velocity in air. *Quart. J. R. Meteorol. Soc.* **96**, 247–256.
- Pruppacher, H. R., and J. D. Klett, 1997: *Microphysics of Clouds and Precipitation* (Kluwer, Dordrecht).

- Pruppacher, H. R., and R. L. Pitter, 1971: A semi-empirical determination of the shape of cloud and rain drops. *J. Atmos. Sci.* **28**, 86–94.
- Rogers, R. R., 1990: The early years of Doppler radar in meteorology. In D. Atlas, Ed., *Radar in Meteorology* (American Meteorological Society, Boston), pp. 122–129.
- Russchenberg, H. W. J., 1992: Ground-based remote sensing of precipitation using a multi-polarized FM-CW Doppler radar. PhD Thesis (Delft University, Delft).
- Russchenberg, H. W. J., 1994: Combined radar measurements of the Doppler spectrum and the linear depolarization ratio of rain: a turbulence-canting model. In *Proceedings of the 3rd International Symposium on Tropospheric Profiling* (Hamburg), pp. 456–459.
- Ryzhkov, A. V., 2001: Interpretation of polarimetric radar covariance matrix for meteorological scatterers: theoretical analysis. *J. Atmos. Oceanic Technol.* **18**, 315–328.
- Seliga, T. A., and V. N. Bringi, 1976: Potential use of radar differential reflectivity measurements at orthogonal polarizations for measuring precipitation. *J. Appl. Meteorol.* **15**, 69–76.
- Seliga, T. A., R. G. Humphries, and J. I. Metcalf, 1990: Polarization diversity in radar meteorology: early developments. In D. Atlas, Ed., *Radar in Meteorology* (American Meteorological Society, Boston), pp. 109–114.
- Shupiatsky, A. B., 1959: Radar scattering by non-spherical particles. *Trudy (Proc.) Central Aerological Observ.* **30**, 39–52 (in Russian).
- Tokay, A., and K. V. Beard, 1996: A field study of raindrop oscillations. Part I: Observation of size spectra and evaluation of oscillation causes. *J. Appl. Meteorol.* **35**, 1671–1687.
- Unal, C. M. H., D. N. Moisseev, F. J. Yanovsky, and H. W. J. Russchenberg, 2001: Radar Doppler polarimetry applied to precipitation measurements: introduction of the spectral differential reflectivity. In *Proceedings of the 30th International Conference on Radar Meteorology* (American Meteorological Society, Boston), pp. 316–318.
- Unal, C. M. H., and D. N. Moisseev, 2004: Combined Doppler and polarimetric radar measurements: correction for spectrum aliasing and nonsimultaneous polarimetric measurements. *J. Atmos. Oceanic Technol.* **21**, 443–456.
- Venttsel', E. S., 1998: *Probability Theory* (Vysshaya Shkola, Moscow) (in Russian).
- Vinnichenko, N. K., N. Z. Pinus, C. M. Shmeter, and G. N. Shur, 1968: *Turbulence in Free Atmosphere* (Gidrometeoizdat, Leningrad) (in Russian).
- Wilson, R., A. J. Illingworth, and T. M. Blackman, 1997: Differential Doppler velocity: a radar parameter for characterizing hydrometeor size distributions. *J. Appl. Meteorol.* **36**, 649–663.
- Yadong, W., X. Xiao, and T.-Y. Yu, 2008: Spectral analysis of dual-polarization radar signals in a tornadic supercell storm. In *Proceedings of the 24th Conference on Severe Local Storms* (Savannah, GA), pp. 1–10.
- Yanovsky, F. J., 1996: Simulation study of 10 GHz radar backscattering from clouds and solution of the inverse problem of atmospheric turbulence measurements. *Comput. Electromagn. IEE CP* **420**, 188–193.
- Yanovsky, F. J., 1998a: Model of the relationship between Doppler and polarization parameters of the radar signal from precipitation. *Report No. IRCTR-S-002-98* (Technical University, Delft).
- Yanovsky, F. J., 1998b: Simulation of the processes of interaction between radar signal and turbulized meteorological object. *Visnyk KMUTSA* **1**, 125–136 (in Ukrainian).
- Yanovsky, F. J., 2002: Phenomenological models of Doppler–polarimetric microwave re-

- remote sensing of clouds and precipitation. *IEEE Int. Geosci. Remote Sens. Symp. IGARSS-02*, Vol. 3, pp. 1905–1907.
- Yanovsky, F. J., and D. N. Glushko, 2010: Simulation study of relationships between Doppler polarimetric parameters at microwave remote sensing of precipitation. In *Proceedings of the 7th European Radar Conference* (Paris), pp. 148–151.
- Yanovsky, F. J., and L. P. Ligthart, 2000: Microwave remote sensing of dangerous meteorological phenomena. In *Proceedings of the XIII International Conference on Microwave, Radar and Wireless Communications*, Vol. 3, pp. 70–82.
- Yanovsky, F. J., L. P. Ligthart, and H. W. J. Russchenberg, 1997: Analysis of transportable atmospheric radar (TARA) possibilities for the remote sensing of dangerous meteorological phenomena. In *PIERS* (Cambridge, MA), p. 116.
- Yanovsky, F. J., H. W. J. Russchenberg, and L. P. Ligthart, 2001: Doppler–polarimetric models of microwave remote sensing of rain. In *Proceedings of the 11th Conference on Microwave Technique* (University of Pardubice, The Czech Republic), pp. 47–62.
- Yanovsky, F. J., C. M. H. Unal, and H. W. J. Russchenberg, 2003a: Doppler–polarimetric radar measurements of turbulence in rain. *Report IRCTR-S-006-03* (Technical University, Delft).
- Yanovsky, F. J., C. M. H. Unal, and H. W. J. Russchenberg, 2003b: Relationship between differential Doppler velocity and turbulence intensity in rain: modeling and measurements. In *Proceedings of the 6th International Symposium on Tropospheric Profiling* (Leipzig), pp. 306–308.
- Yanovsky, F. J., H. W. J. Russchenberg, and C. M. H. Unal, 2005: Retrieval of information about turbulence in rain by using Doppler–polarimetric radar. *IEEE Trans. Microwave Theory Tech.* **53**, 444–450.
- Yanovsky, F. J., C. M. H. Unal, H. W. J. Russchenberg, and L. P. Ligthart, 2007: Doppler–polarimetric weather radar: returns from wide spread precipitation. *Telecommun. Radio Eng.* **66**, 715–727.
- Zrnić, D., and D. Boren, 2008: Weather radar – recent developments and trends. In *Proceedings of the Microwaves, Radar and Remote Sensing Symposium* (Kiev), pp. 174–179.



From left to right: Adrian Doicu, Sergey Savenkov, and Felix Yanovsky.

Index

- Absorption**, 45–47, 384, 386, 398–400, 402–406
cross section, 120
symmetry and, 45–47
Rayleigh limit and, 45
- Advanced Very High Resolution Radiometer**, 342
- aerosol**
optical depth, 207, 211–218, 221–225
retrieval of, 313–315, 323, 236, 337
- agglomerate**, 145–146, 150–161, 165, 167, 169, 261–272, 277–290
- albedo**, 268–271, 283, 373, 375–379
of black sky, 373,
- alignment**, 280, 281, 289
- amino acid**, 278, 281
- analog**
asteroidal, 138, 165–166
cometary, 166–170
Titan aerosols (tholins), 139, 146, 156–159 169, 175
- Ångström exponent**, 213, 216, 223, 224
- anisotropy**, 438, 452–455, 561
amplitude of, 453, 456, 475, 479
phase 453, 456, 475, 479
- Ap and Bp (chemically peculiar) stars**, 17
- aspect ratio**, 67
- Asteroid Polarimetric Database**, 241, 254, 415, 416, 419, 421, 431, 434
- asteroid(s)**, 237, 239, 241–243, 250, 253, 254, 259, 262, 265, 266, 270, 409–413, 415–423, 427, 428, 431, 433–436
anomalous characteristics of, 241, 242
brightness opposition effect of, 410, 411, 415–417, 419, 421, 424, 425, 431, 432, 436
opposition effects of, 241–243
phase-angle dependence of, 241–243, 410, 416, 418–421, 426, 428
polarimetry of, 241–243, 251, 253, 254, 409, 421, 433, 434
polarization opposition effect of, 409, 411, 415–417, 419, 421, 423–428, 431, 435
spectral dependence for, 241–243, 409, 423
- astrobiology**, 277, 279, 290
- atmosphere**, 137–139, 163–165
- autocorrelation**, 73
- Backscatter**, 70
- backscattered field**, 507
- backward scattering**, 133
- bacteria**, 277, 287–290
- ballistic model**, 385
- Beta Pictoris**, 266
- bidirectional reflectance distribution function**, 313–321, 323, 327–334, 336–338, 341–343, 366–372, 380
candidate models of, 351
kernel-driven model of, 350, 351
- bidirectional polarization distribution function**, 313–316, 318, 321, 322, 327–333, 335–338
- biomarkers**, 201
- biosignature**, 277, 278, 289, 290
- birefringence**, 437, 438, 453, 454, 456, 458, 459, 461, 462
linear, 438, 453, 454, 458, 460, 461
circular, , 278, 285, 289, 290, 438, 452, 455, 456, 461
- brightness opposition effect**, 241, 243, 249, 256, 257
- Carbon**, 267–269
- Centaurus**, 243, 250, 251, 409, 411, 413, 427–429, 432, 433
polarimetry of, 243, 250, 427–434
- charge shuffling**, 189–190
- chirality**, 454
- circular dichroism**, 278, 282, 285, 289

- cloud effects
 - cloud adjacency effects, 207, 210, 221, 224
 - cloud-induced enhancement, 210, 211, 214, 215
 - subpixel cloud contamination, 207, 224
- clouds, 142–164, 166–170
 - of irregularly shaped particles, 156–165, 166–170
 - of spherical particles, 146, 149–153
- cluster, 261, 267–269, 277, 283, 284, 287
- coated spheres, 149–150, 152–153
- coherent backscattering, 240, 241, 249, 250, 254, 255, 258, 411, 413, 416, 419, 423, 425, 435
- coherent radar, 502, 503
- color, 261, 264, 8, 271, 283
- comet(s), 233, 237, 239–241, 244, 245, 253, 254, 256, 266–270, 277, 279–285, 286, 289, 431, 432, 434
- circular polarization for, 244, 256
- during stellar occultations, 244, 245, 256
- Halley, 262, 266, 268, 269, 279, 281, 282
- polarimetry of, 244, 251, 256, 434
- concentration effects, 383
- confidence region, 349, 353
- continuous medium approximation, 453
- convex sum 440, 442
- cotton effect, 289
- Coulomb screening, 384
- covariance matrix, 507, 509, 530
 - spectral, 509
- cross correlation coefficient, 502, 515
- cross-talk, 10, 12
- crystal waveplates, 187
- Data reduction**, 189
- Database of Comet Polarimetry, 245, 253
- debris disks, 262, 263, 265, 266, 269
- degree of polarization, 440, 444, 446
 - average, 444, 445
- depolarization
 - anisotropic, 446
 - index of, 444, 445
 - isotropic, 446
 - metric, 444, 445
 - ratio, 71, 445
- dichroism, 438, 454
 - circular, 438, 452, 455
 - linear, 438, 454, 463
- dielectric, 118, 124, 126
- differential reflectivity, 502, 503, 507, 509, 510, 528, 530, 532, 533, 537
- spectral, 503, 510, 528, 530, 532, 533, 537
- diffraction pattern, 60
- directional effects, 117, 133
- discrepancy principle
 - generalized, 94
 - linear problems, 93
 - nonlinear problems, 103, 106
- discrete dipole approximation, 45
- DNA 452, 453
- Doppler polarimetry, 508, 509, 527, 539
- Doppler spectrum 503, 504, 505–507, 511, 530, 532, 533, 535, 537, 538
- width of, 503, 532, 533, 535, 538
- Doppler velocity, 503, 506, 508, 509–511, 530, 531, 533, 535
 - differential, 503, 531
 - mean, 506
 - variance of, 506
- double Wollaston, 186
- drop
 - fall velocity, 512, 517, 518, 529, 535
 - size distribution, 503–506, 511, 513, 518, 523, 526, 527, 529, 532–534
 - turbulent velocity, 517, 518, 521, 529
- dual-polarization Doppler radar, 503
- dust, 137–139, 144, 158, 165–170
 - atmospheric, 197–200
 - cometary, 138–139, 166–170, 277, 279–282, 286, 287, 289, 296–299, 306, 307
 - cosmic, 261–272
 - on asteroidal surfaces, 138–139, 154–155, 165, 296
 - on planetary surfaces, 138, 165–166
 - properties of, 138–139, 154–155, 65–170
 - zodiacal cloud, 296, 294, 295
- Eddy dissipation rate**, 515, 516, 519, 521, 523, 530, 532–535, 538

- efficiency factor, 384, 385, 390
 - for absorption, 384
 - for extinction, 384, 385, 390
 - for scattering, 384, 390
- electrodynamic coupling, 383, 386–388, 390, 391, 394, 397, 402, 404
 - lateral, 383, 390, 391, 394, 397, 399, 402
 - near-field, 383
- electric
 - dipole, 121
 - permittivity, 118, 121, 123–124, 126–131
 - resonance, 126–129, 131, 132
 - terms, 126
- electromagnetic interaction, 270, 271
- Euler angles, 63
- Euler scale, 514
- evaporation, thermal, 479
- exoplanet, 194–197, 287, 290
- expected error estimation method
 - linear problems, 92
 - nonlinear problems, 103
- explained variance, 356–358
- extinction, 120–123, 126, 128
 - cross section, 120
 - efficiency, 121–123, 126
- extragalactic objects, 248, 249
- Facet-tracing**, 66
- factor analysis, 353–355
- far zone, 58
- fast cameras, 191
- ferroelectric liquid crystals, 193
- first Jones' equivalence theorem, 457
- Fisher statistics, 346, 349, 353, 360
- forward scattering, 133
- Fourier transform, 71, 72
- Fraunhofer diffraction, 60
- Fresnel coefficients, 55, 66
- Fresnel rhomb, 5
- Fröhlich frequency, 384, 396
 - effective, 396
- Fröhlich resonance, 124, 126
- Galilean satellites**, 240, 241
 - opposition effects for, 240, 241, 253, 256
- Gauss–Newton method, iteratively
 - regularized, 108
- Gaussian surface, 322
- generalized cross-validation
 - linear problems, 96
 - nonlinear problems, 104, 106
- generalized inverse
 - discrete problems, 85
 - regularized, 86
- Hadamard's conditions**, 83
- homochirality, 277, 279, 287, 289
- hydrometeor(s), 502, 504, 506–508, 532, 538
- Ideal depolarizer**, 446
- ill-conditioned index, 343, 358, 380
- ill-posedness, 83
- illumination directions, 155
- imaging polarimetry, 300–302, 307
- imaging technique, 143–144
- information, *a priori* statistical, 341, 360
- informative parameter, 502, 531
- interference, 61
- interplanetary dust, 261, 263, 266, 267, 269, 270, 287, 290
- inverse scattering problem, 504
- Jet-like features in cometary polarization**, 300–302, 304, 307
- Jones matrix, 51, 56, 60, 61, 440–443, 448
- Jones N-matrix formalism, 460
- Julian day, 375
- Kernel-driven model**, 319, 320, 328
- Kimes reflectance data, 374
- Kirchhoff approximation, 322
- Kronecker (tensor) product, 441
- L-curve method**
 - linear problems, 99
 - nonlinear problems, 104, 107
- laboratory simulations (measurements), 261, 262, 265, 277, 279, 287–290
- Lagrange scale, 514
- land cover types, 355
- land surfaces, 313–315, 318, 321, 322, 330

- layers, particulate, 149, 152, 154–155, 157–158, 164–166
 - bare or coated, 152–153
 - of spheres, 149, 154–155
- leaf area index, 350
- least-square deconvolution, 22
- least squares, 347, 348
 - solution, 85, 341, 347, 359,
- left-handed, 118, 121, 126, 127
- Levenberg–Marquardt method,
 - regularizing, 110
- levitation techniques, 141–142
 - gas-lift, 142
 - reduced gravity, 141–142
- Li kernel, 320
- light scattering by dust
 - in comets, 298–302, 304
 - in the zodiacal cloud, 299, 303, 308
 - laboratory measurements of, 304–305, 308
 - numerical simulations of, 305–306, 308
 - on surfaces of asteroids, 297, 298, 303, 304
- linear depolarization ratio, 502, 508, 510, 528, 530, 532, 533
 - spectral, 503, 510, 530, 532, 533
- linear unbiased estimator, 347
- liquid crystal modulators, 192
- Magnetic**
 - dipole, 121
 - field, 279, 281, 289
 - permeability, 118–119, 121, 123, 126–132
 - resonance, 127, 128, 131, 132
 - response, 121
 - terms, 126, 128–132
- metamaterials, 118, 121
- magnetic field
 - in stars, 1–3, 17–25
 - measurement of, 20–21
 - modelling of, 22–25
 - origin of, 2
- marine organisms, 451
- Mars, 277, 281, 415
- maximum likelihood estimation
 - linear problems, 97
 - nonlinear problems, 104, 106
- mean, longitudinal, 20
 - curve of, 22
- Mercury, 415
- metallic nanostructures, 383, 404
 - densely packed, 383–387, 389–393, 395–407
 - gradient, 383, 403–406
 - layered, 383–385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407
- meteorite, 279, 281, 289
- Meteosat Second Generation, 342
- microphysical model, 35–37, 43–44
 - volume integral equation and, 35
 - wavelets and, 37, 43–44
- Mie
 - coefficients, 118–123
 - resonances, 119–131
 - theory, 118
- modified Fresnel model, 321, 322, 329, 331, 332, 335, 336
- Moon, 233, 234, 237, 239, 240, 253, 258, 415, 432
- Mueller–Jones matrix, 448
- Mueller matrix, 50, 59, 61, 70
 - block-diagonal, 448, 449
 - depolarizing, 440, 443, 457
 - pure, 440–443, 453, 454
- multiple correlation coefficient, 349
- multiple scattering, 270, 279, 283
- Nadal–Bréon model**, 318, 321, 328, 329, 331, 332, 335, 336
- National Oceanic and Atmospheric Administration, 342
- near-field pattern, 384, 386–388, 389, 391, 394, 404
- near zone, 51, 54
- nematic liquid crystals, 192
- noise error
 - linear problems, 91
 - nonlinear problems, 102
- normal equation
 - discrete problems, 85
 - regularized, 86
- numerical simulations, 149–151, 153, 169
- numerical weather prediction, 342

- O**blique rotator model, 16, 22
- observations
- aircraft-based, 207–209, 215, 216, 225
 - in situ*, 205, 207, 217–219, 224
 - satellite-based, 206–209, 217, 223–225
 - surface, 215, 216, 218, 221, 224, 225
- optical activity, 177, 282, 284, 285, 287, 289
- optical communications, 117, 131
- optical probe, 299, 300
- organic, 268, 269, 277–282, 289, 290
- P**acking, 263, 286
- density, 154
 - of irregularly shaped particles, 157–158, 164–166
- Parseval theorem, 72
- partially-ordered structure, 384, 391
- particle size, 118, 123, 124, 126, 129–132
- dependence on, 129
- Paschen–Back effect, 17
- phase function (curve), 66, 72, 263–268, 271, 283, 286
- phenomenological model, 504, 510, 514, 529, 532
- photoelastic modulators, 181–182
- photonic confinement, 401, 402, 405
- PlanetPol, 183–184
- planets, 245, 246, 251, 255
- plasmonic array, 386–388
- plasmonic nanoparticle, 386, 387, 391–394, 397, 404
- plasmonic–photonic confinement, 383, 401
- Poincaré sphere, 445
- polar decomposition, 456, 457
- polarimeter, 7, 10, 140–142, 233, 234, 236, 237–239, 251, 253–255, 257, 258, 287, 413, 414, 417, 545, 436
- for clouds of particles, 140–142
 - for particulate layers, 140–141
 - ideal, 7
- polarimetric color, 265–268, 270, 271
- polarimetric dependence on wavelength, 295, 297, 298, 303–306, 308
- polarimetric halo in comets, 300, 301, 304
- polarimetric observation, 504
- polarimetric radar, 501, 502, 504, 507, 539
- polarimetry, 440, 443, 445, 456, 502, 503, 508, 509, 527
- with area detectors, 189
- polarization, 38–39, 41–44, 46, 55
- analysis, 479
 - circular, 41–43, 178, 271, 277–290
 - diagnostics, 179
 - difference, 476–478
 - effects of absorption on, 157–161
 - effects of particle size on, 156–162
 - ellipse, 5
 - ellipsometric interpretation of, 39
 - elliptical, 41–44, 46
 - linear, 38–39, 41–42, 139–140, 178, 261–272
 - measurements of, 140
 - modulation technique, 474–476
 - modulator, 475
 - particle symmetry and, 37–39, 44
 - phase curve, 295, 301, 302
 - ripples, 189
 - sensitivity, accuracy of, 177
- polarization opposition effect, 241, 243, 249, 250, 255, 256
- polarization phase curve
- amplitude of, 146, 151–159, 161–165, 169–170
 - positive branch of, 156–157, 159, 161–165, 169
- POLDER angular geometries, 344
- POLDER BRDF, 343, 345
- POLDER experiment, 343
- porosity, 261, 263, 264, 269, 271, 277, 290
- predictive error, 91
- producing polarized radiation,
- intrinsic, 179
 - secondary mechanisms, 179
- pulsed laser deposition, 480, 495
- $Q(\mathbf{M})$ metric, 445
- quasi-optimality criterion, 98
- quasicrystalline approximation, 388, 391–394, 396, 397, 399, 401, 402, 405
- R**adar cross section, 505, 506, 526, 528
- radar equation, 505

- radar reflectivity factor, 502, 505, 506
radial distribution function, 388
Rahman–Pinty–Verstraete model, 318,
319, 327–331, 333–336
ray tracing, 50, 52, 66
ray trajectory, 54
Rayleigh
 approximation, 120, 121, 123, 128
 particles, 264, 269
 scattering, 45–47
RCS matrix, 507
reference plane, 55
reflectance
 polarized, 208–210, 224, 225, 314, 317,
 318, 321–323, 327, 329, 330, 332,
 333, 335–337
 ratio(s), 211–218
 spectral, 205, 208–212
 total, 317, 319, 323, 324, 326–328, 330,
 333, 337
reflection, 387, 390, 395, 396, 400,
402–404
 coherent, 390
 isotropic, 492–495, 498
reflectivity, 505, 506, 537
refractive index, 12, 15, 17, 19, 118, 119,
121, 123, 126–128
 negative, 126–128
regolith, 296, 297, 304
regularization matrix, 88
relative azimuth angle, 13, 15
relative humidity, 207, 217, 221–225
remote sensing, 261, 262, 267, 271, 277,
284, 287
Research Scanning Polarimeter, 313,
315–318, 9, 323, 328–330, 332, 333,
335, 336
resolution volume, 505, 507, 509, 510,
520, 522, 526
resonance quenching of transmission, 383,
392
resonant position, 118, 123, 124, 126–128,
130, 131
retardation effects, 384
Ricatti–Bessel functions, 122
ridge regression, 341, 346, 359
Ross–Li model, 318–320, 327–331, 334
Ross kernel, 319, 320
Ross–Roujean model, 318–320, 328–331
Roujean kernel, 340
Samples, particulate, 144–145, 147–148
 compact, 147
 fluffy, 148
satellites, 409–411, 414, 423–428, 431,
433–436
 polarimetry of, 434, 436
scattering, 117–123, 133
 amplitude, 72
 backward, 448, 449, 452
 cross section, 120
 efficiency, 120–123
 forward, 448–450
 matrix, 6–7, 284, 504, 507, 508, 509,
 534
 multiple, 444, 446, 447, 450, 451, 543,
 455, 460, 463
 plane, 56
 single, 444, 447
semi-empirical models, 313–319, 321,
323, 329, 330, 336, 337
sensitivity factor, 397–399
SEVIRI/MSG measurements, 343, 374
SEVIRI/MSG angular sampling geometry,
344–346, 360, 366–372
shadow-forming field, 52
shadow function, 73
shadowing, 319, 320, 322
shape function, 54
silicate feature, 269, 272
silicates, 262, 267–270
size effect, 405, 406
 classical, 483, 487, 493, 495
 internal, 405
 topological, 487, 490, 495
size parameter, 119, 121
singular value decomposition, 346
 generalized, 89
 ordinary, 84
skeletal muscle, 451
smoothing error
 linear problems, 90
 nonlinear problems, 102
solar zenith angle, 353, 355, 361, 374
space mission, 261, 262, 269, 272
spatial frequency, 515

- spatial modulation, 179–180, 184
 spatial resolution, 514, 515
 specific differential phase, 502, 508
 spectrum, 281, 287, 289
 spheres, 33–38, 41–42
 internal field of, 33–36
 polarization state and, 41–42
 reflection symmetry of, 35–38
 spherical particles, 445, 447–451
 spheroids, 45–47, 269
 polarization state and, 45–47
 squared sum of residuals, 341, 347, 374
 standard deviation, 374
 stardust, 262, 269
 stars, 246–248, 251, 252, 414, 415, 434
 polarimetry of, 246–248
 statistical inversion method, 346
 statistical regularization, 341, 346
 statistical theory of multiple scattering of
 waves, 383, 388, 407
 Stokes
 parameters, 3–5, 8, 25, 39–40, 50, 265,
 314, 316
 definition of, 3–4
 measurement of, 5
 spherical particles and, 40
 transformation matrix, 440
 vector, 178, 278, 284, 439, 440, 443,
 445, 450, 456, 477, 498
 Stokes–Verdet criterion, 440, 443
 Sun, 246, 256
 surface plasmon resonance, 383, 384, 394,
 396–399, 401–405, 407, 473, 474,
 488, 498
 collective, 396
 symmetry, 34, 37–39, 44–47
 absorption and, 45–47
 breaking of, 44–45
 consequences on scattering, 37–39, 44
 mirror, 442, 449,
 reflection, 34
 rotational, 449, 450

T matrix (method), 198, 249, 267, 277,
 281, 283, 284, 289
T-multivariate statistics, 346
 Taylor expansion, 122
 telescope, 235–238, 241, 245, 248, 253,
 258, 413, 414, 416, 417, 421, 426,
 433, 434
 polarization of, 184
 temporal modulation, 179–184
 thermal emission, 268, 269, 272
 Tikhonov regularization
 linear problems, 86
 nonlinear problems, 101
 tissue
 cancerous, 458
 colon, 452
 connective, 453
 dermal, 462, 463
 infracted, 461
 noncancerous, 458
 precancerous, 458, 460
 Titan, 269, 287
 total error
 linear problems, 90
 nonlinear problems, 102
 transmission, 383, 387, 390–392, 395,
 396, 398, 402, 404, 405
 coherent, 390
 enhanced, 383, 391, 404
 transmittance condition, 441
 trans-Neptunian objects, 243, 241, 243,
 250, 409, 427–434
 polarimetry of, 243, 250, 428, 432–434
 turbulence, 504–507, 511–523, 529–535,
 539–542
 distribution of, 519
 energy spectrum of, 514, 515
 intensity of, 515, 531, 532, 534, 537
 kinetic energy of, 514, 515
 scale of, 511, 514, 515, 517–520
 velocity variance of, 515

Umov law, 240, 242
 unbiased predictive risk estimator method,
 95

Vector spherical harmonics, 118
 vector spherical wave functions, 33–34
 vegetation, 278, 290
 index, 316, 321
 view zenith angle, 353, 355, 371

Wave zone, 60

wavelength effect, 158–160, 165,
167–169
waveplate, 5
 chromatism of, 8
 fringing, 10
Wollaston prism, 6

Z
Zeeman effect, 12–17, 28, 32
Zeeman splitting, 13, 15
Zeeman Doppler imaging, 24
ZIMPOL, 190