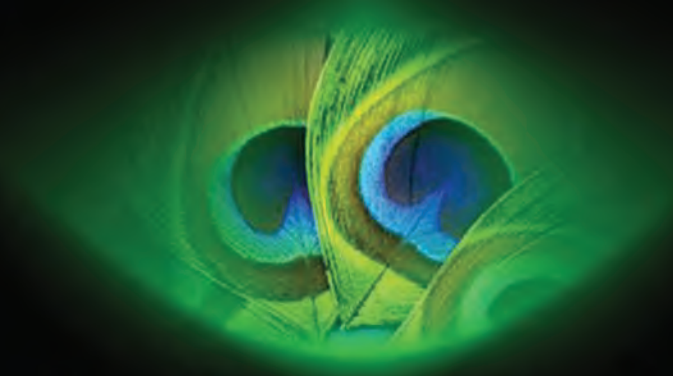




Bernard C. Kress · Patrick Meyrueis

# APPLIED DIGITAL OPTICS

From Micro-optics to Nanophotonics



 WILEY





# **APPLIED DIGITAL OPTICS**





# APPLIED DIGITAL OPTICS

## FROM MICRO-OPTICS TO NANOPHOTONICS

**Bernard C. Kress**

*Photonics Systems Laboratory, Université de Strasbourg, France*

**Patrick Meyrueis**

*Photonics Systems Laboratory, Université de Strasbourg, France*



A John Wiley and Sons, Ltd, Publication

This edition first published 2009  
© 2009 John Wiley & Sons, Ltd

*Registered office*

John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at [www.wiley.com](http://www.wiley.com).

The right of the author to be identified as the author of this work has been asserted in accordance with the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs and Patents Act 1988, without the prior permission of the publisher.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book. This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

***Library of Congress Cataloguing-in-Publication Data***

Kress, B.

Applied digital optics : from micro-optics to nanophotonics / Bernard C. Kress, Patrick Meyrueis.  
p. cm.

Includes bibliographical references and index.

ISBN 978-0-470-02263-4 (cloth)

1. Optical MEMS. 2. Nanophotonics. 3. Integrated optics. 4. Signal processing--Digital techniques.  
5. Diffraction gratings. I. Meyrueis, Patrick. II. Title.  
TK8360.O68.K74 2009  
621.36--dc22

2009004108

A catalogue record for this book is available from the British Library.

ISBN: 978-0-470-02263-4

Set in 9/11pt, Times by Thomson Digital, Noida, India.  
Printed in Great Britain by CPI Antony Rowe, Chippenham, Wiltshire.

*To my lovely wife Mei-Mei, whose unconditional love and support made this book possible. I even learned to appreciate her constant nagging, which drove me up the wall but helped me finish this project.*

**Bernard**

*I would like to dedicate this book to all my university colleagues, students, Photonics Systems Laboratory staff, my assistant Anne and members of institutions and companies all over the world that allowed us, by contributing to or supporting our microphotronics and nanophotonics activities in research and education, to gather the information that made this book possible.*

**Patrick**



# Contents

|                                                                |             |
|----------------------------------------------------------------|-------------|
| <b>About the Authors</b>                                       | <b>xi</b>   |
| <b>Foreword by Professor Joseph Goodman</b>                    | <b>xiii</b> |
| <b>Foreword by Professor Trevor Hall</b>                       | <b>xv</b>   |
| <b>Acknowledgments</b>                                         | <b>xvii</b> |
| <b>Acronyms</b>                                                | <b>xix</b>  |
| <b>Introduction</b>                                            | <b>1</b>    |
| Why a Book on Digital Optics?                                  | 1           |
| Digital versus Analog                                          | 2           |
| What are Digital Optics?                                       | 2           |
| The Realm of Digital Optics                                    | 3           |
| Supplementary Material                                         | 4           |
| <b>1 From Refraction to Diffraction</b>                        | <b>5</b>    |
| 1.1 Refraction and Diffraction Phenomena                       | 5           |
| 1.2 Understanding the Diffraction Phenomenon                   | 5           |
| 1.3 No More Parasitic Effects                                  | 8           |
| 1.4 From Refractive Optics to Diffractive Optics               | 9           |
| 1.5 From Diffractive Optics to Digital Optics                  | 11          |
| 1.6 Are Diffractives and Refractives Interchangeable Elements? | 13          |
| <b>2 Classification of Digital Optics</b>                      | <b>15</b>   |
| 2.1 Early Digital Optics                                       | 15          |
| 2.2 Guided-wave Digital Optics                                 | 16          |
| 2.3 Free-space Digital Optics                                  | 17          |
| 2.4 Hybrid Digital Optics                                      | 19          |
| <b>3 Guided-wave Digital Optics</b>                            | <b>21</b>   |
| 3.1 From Optical Fibers to Planar Lightwave Circuits (PLCs)    | 21          |
| 3.2 Light Propagation in Waveguides                            | 22          |
| 3.3 The Optical Fiber                                          | 25          |
| 3.4 The Dielectric Slab Waveguide                              | 27          |
| 3.5 Channel Waveguides                                         | 28          |
| 3.6 PLC In- and Out-coupling                                   | 30          |
| 3.7 Functionality Integration                                  | 36          |

|           |                                                                         |            |
|-----------|-------------------------------------------------------------------------|------------|
| <b>4</b>  | <b>Refractive Micro-optics</b>                                          | <b>47</b>  |
| 4.1       | Micro-optics in Nature                                                  | 47         |
| 4.2       | GRIN Lenses                                                             | 49         |
| 4.3       | Surface-relief Micro-optics                                             | 55         |
| 4.4       | Micro-optics Arrays                                                     | 58         |
| <b>5</b>  | <b>Digital Diffractive Optics: Analytic Type</b>                        | <b>71</b>  |
| 5.1       | Analytic and Numeric Digital Diffractives                               | 73         |
| 5.2       | The Notion of Diffraction Orders                                        | 73         |
| 5.3       | Diffraction Gratings                                                    | 76         |
| 5.4       | Diffractive Optical Elements                                            | 90         |
| 5.5       | Diffractive Interferogram Lenses                                        | 106        |
| <b>6</b>  | <b>Digital Diffractive Optics: Numeric Type</b>                         | <b>111</b> |
| 6.1       | Computer-generated Holograms                                            | 111        |
| 6.2       | Designing CGHs                                                          | 115        |
| 6.3       | Multiplexing CGHs                                                       | 149        |
| 6.4       | Various CGH Functionality Implementations                               | 151        |
| <b>7</b>  | <b>Hybrid Digital Optics</b>                                            | <b>157</b> |
| 7.1       | Why Combine Different Optical Elements?                                 | 157        |
| 7.2       | Analysis of Lens Aberrations                                            | 157        |
| 7.3       | Improvement of Optical Functionality                                    | 163        |
| 7.4       | The Generation of Novel Optical Functionality                           | 166        |
| 7.5       | Waveguide-based Hybrid Optics                                           | 169        |
| 7.6       | Reducing Weight, Size and Cost                                          | 171        |
| 7.7       | Specifying Hybrid Optics in Optical CAD/CAM                             | 173        |
| 7.8       | A Parametric Design Example of Hybrid Optics via Ray-tracing Techniques | 175        |
| <b>8</b>  | <b>Digital Holographic Optics</b>                                       | <b>181</b> |
| 8.1       | Conventional Holography                                                 | 181        |
| 8.2       | Different Types of Holograms                                            | 185        |
| 8.3       | Unique Features of Holograms                                            | 188        |
| 8.4       | Modeling the Behavior of Volume Holograms                               | 192        |
| 8.5       | HOE Lenses                                                              | 199        |
| 8.6       | HOE Design Tools                                                        | 203        |
| 8.7       | Holographic Origination Techniques                                      | 203        |
| 8.8       | Holographic Materials for HOEs                                          | 207        |
| 8.9       | Other Holographic Techniques                                            | 212        |
| <b>9</b>  | <b>Dynamic Digital Optics</b>                                           | <b>217</b> |
| 9.1       | An Introduction to Dynamic Digital Optics                               | 217        |
| 9.2       | Switchable Digital Optics                                               | 223        |
| 9.3       | Tunable Digital Optics                                                  | 235        |
| 9.4       | Reconfigurable Digital Optics                                           | 244        |
| 9.5       | Digital Software Lenses: Wavefront Coding                               | 250        |
| <b>10</b> | <b>Digital Nano-optics</b>                                              | <b>253</b> |
| 10.1      | The Concept of 'Nano' in Optics                                         | 253        |
| 10.2      | Sub-wavelength Gratings                                                 | 253        |

|                                                               |            |
|---------------------------------------------------------------|------------|
| 10.3 Modeling Sub-wavelength Gratings                         | 255        |
| 10.4 Engineering Effective Medium Optical Elements            | 267        |
| 10.5 Form Birefringence Materials                             | 272        |
| 10.6 Guided Mode Resonance Gratings                           | 275        |
| 10.7 Surface Plasmonics                                       | 277        |
| 10.8 Photonic Crystals                                        | 279        |
| 10.9 Optical Metamaterials                                    | 288        |
| <b>11 Digital Optics Modeling Techniques</b>                  | <b>295</b> |
| 11.1 Tools Based on Ray Tracing                               | 295        |
| 11.2 Scalar Diffraction Based Propagators                     | 298        |
| 11.3 Beam Propagation Modeling (BPM) Methods                  | 321        |
| 11.4 Nonparaxial Diffraction Regime Issues                    | 323        |
| 11.5 Rigorous Electromagnetic Modeling Techniques             | 326        |
| 11.6 Digital Optics Design and Modeling Tools Available Today | 327        |
| 11.7 Practical Paraxial Numeric Modeling Examples             | 330        |
| <b>12 Digital Optics Fabrication Techniques</b>               | <b>339</b> |
| 12.1 Holographic Origination                                  | 340        |
| 12.2 Diamond Tool Machining                                   | 342        |
| 12.3 Photo-reduction                                          | 346        |
| 12.4 Microlithographic Fabrication of Digital Optics          | 347        |
| 12.5 Micro-refractive Element Fabrication Techniques          | 385        |
| 12.6 Direct Writing Techniques                                | 388        |
| 12.7 Gray-scale Optical Lithography                           | 394        |
| 12.8 Front/Back Side Wafer Alignments and Wafer Stacks        | 406        |
| 12.9 A Summary of Fabrication Techniques                      | 408        |
| <b>13 Design for Manufacturing</b>                            | <b>413</b> |
| 13.1 The Lithographic Challenge                               | 413        |
| 13.2 Software Solutions: Reticle Enhancement Techniques       | 418        |
| 13.3 Hardware Solutions                                       | 445        |
| 13.4 Process Solutions                                        | 449        |
| <b>14 Replication Techniques for Digital Optics</b>           | <b>453</b> |
| 14.1 The LIGA Process                                         | 453        |
| 14.2 Mold Generation Techniques                               | 455        |
| 14.3 Embossing Techniques                                     | 459        |
| 14.4 The UV Casting Process                                   | 464        |
| 14.5 Injection Molding Techniques                             | 464        |
| 14.6 The Sol-Gel Process                                      | 471        |
| 14.7 The Nano-replication Process                             | 472        |
| 14.8 A Summary of Replication Technologies                    | 475        |
| <b>15 Specifying and Testing Digital Optics</b>               | <b>479</b> |
| 15.1 Fabless Lithographic Fabrication Management              | 479        |
| 15.2 Specifying the Fabrication Process                       | 480        |
| 15.3 Fabrication Evaluation                                   | 494        |
| 15.4 Optical Functionality Evaluation                         | 510        |

|                                                                    |            |
|--------------------------------------------------------------------|------------|
| <b>16 Digital Optics Application Pools</b>                         | <b>521</b> |
| 16.1 Heavy Industry                                                | 522        |
| 16.2 Defense, Security and Space                                   | 532        |
| 16.3 Clean Energy                                                  | 539        |
| 16.4 Factory Automation                                            | 541        |
| 16.5 Optical Telecoms                                              | 544        |
| 16.6 Biomedical Applications                                       | 548        |
| 16.7 Entertainment and Marketing                                   | 553        |
| 16.8 Consumer Electronics                                          | 554        |
| 16.9 Summary                                                       | 574        |
| 16.10 The Future of Digital Optics                                 | 574        |
| <b>Conclusion</b>                                                  | <b>581</b> |
| <b>Appendix A: Rigorous Theory of Diffraction</b>                  | <b>583</b> |
| A.1 Maxwell's Equations                                            | 583        |
| A.2 Wave Propagation and the Wave Equation                         | 583        |
| A.3 Towards a Scalar Field Representation                          | 584        |
| <b>Appendix B: The Scalar Theory of Diffraction</b>                | <b>587</b> |
| B.1 Full Scalar Theory                                             | 587        |
| B.2 Scalar Diffraction Models for Digital Optics                   | 594        |
| B.3 Extended Scalar Models                                         | 595        |
| <b>Appendix C: FFTs and DFTs in Optics</b>                         | <b>597</b> |
| C.1 The Fourier Transform in Optics Today                          | 597        |
| C.2 Conditions for the Existence of the Fourier Transform          | 600        |
| C.3 The Complex Fourier Transform                                  | 600        |
| C.4 The Discrete Fourier Transform                                 | 601        |
| C.5 The Properties of the Fourier Transform and Examples in Optics | 604        |
| C.6 Other Transforms                                               | 606        |
| <b>Index</b>                                                       | <b>611</b> |



# About the Authors

Bernard Kress has been involved in the field of digital optics since the late 1980s. He is an associate professor at the University of Strasbourg, France, teaching digital optics. For the last 15 years Dr Kress has been developing technologies and products related to digital optics. He has been working with established industries around the world and with start-ups in the Silicon Valley, California, with applications ranging from optical data storage, optical telecom, military and homeland security applications, LED and laser displays, industrial and medical sensors, biotechnology systems, optical security devices, high power laser material processing, to consumer electronics. He is on the advisory boards of various photonics companies in the US and has also been advising venture capital firms in the Silicon Valley for *due diligence* reviews in photonics, especially in micro- and nano-optics.

He holds more than 25 patents based on digital optics technology and applications, and is the author of more than 100 papers on this subject. He has taught several short courses given at SPIE conferences. His first book on digital optics, *Digital Diffractive Optics* (2000), was published by John Wiley & Sons, Ltd and has been translated into Japanese in 2005 (published by Wiley-Maruzen). He is also the author of a chapter in the best seller *Optical System Design* (2007), edited by R. Fisher and published by McGraw-Hill. Bernard Kress can be contacted at [bernard@applieddigitaloptics.com](mailto:bernard@applieddigitaloptics.com).

Patrick Meyrueis is full professor at the University of Strasbourg since 1986 (formerly Louis Pasteur University). He is the founder of the Photonics Systems Laboratory which is now one of the most advanced labs in the field of planar digital optics. He is the author of more than 200 publications and was the chairman of more than 20 international conferences in photonics. He was the representative of the Rhenaphotonics cluster and one of the founders of the CNOP in 2001 (national French committee of optics and photonics). He is now acting as the scientific director of the Photonics Systems Lab and the head of the PhD and undergraduate program in the ENSPS National School of Physics in Strasbourg.



# Foreword by Professor Joseph Goodman

The field of digital optics is relatively new, especially when compared with the centuries-long life of the more general field of optics. While it would perhaps have been possible to imagine this field a century or more ago, the concept would not have been of great interest, due to the lack of suitable sources, computing power and fabrication tools. But digital optics has now come of age, aided by the extraordinary advances in lasers, processor speed and the remarkable development of tools for fabricating such optics, driven in part by the tools of the semiconductor industry.

It was perhaps in the seminal work of Lohmann on computer-generated holograms that interest in the field of digital optics was launched. Lohmann based his experimental work on the use of binary plotters and photo-reduction, but today the plotting tools have reached a level of sophistication not even imagined at the time of Lohmann's invention, allowing elements with even sub-wavelength structure to be directly fabricated on a broad range of materials.

*Applied Digital Optics* is a remarkable compendium of concepts, techniques and applications of digital optics. The book includes in-depth discussions of guided-wave optics, refractive optics, diffractive optics and hybrid (diffractive/refractive) optics. Also included is the important area of 'dynamic optics', which covers devices with diffractive properties that can be changed at will. The optics of sub-wavelength structures is also covered, adding an especially timely subject to the book.

Most interesting to me is the extremely detailed discussion of fabrication and replication techniques, which are of great importance in bringing diffractive optics to the commercial marketplace. Finally, the wide-ranging discussion of applications of digital optics is almost breathtaking in its range and coverage.

Professors Kress and Meyrueis provide therefore a comprehensive overview of the current state of research in the field of digital optics, as well as an excellent analysis of how this technology is implemented today in industry, and how it might evolve in the next decade, especially in consumer electronics applications.

In summary, this book will surely set the standard for a complete treatment of the subject of digital optics, and will hopefully inspire even more innovation and progress in this important field.

*Professor Joseph W. Goodman  
William Ayer Professor, Emeritus  
Department of Electrical Engineering, Stanford University  
Stanford, CA, USA*



# Foreword by Professor Trevor Hall

It was my privilege to host Bernard Kress at an early stage in his career. I was very impressed by his creativity, determination and tireless energy. I knew then that he would become a champion in his field of diffractive optics.

*Applied Digital Optics* is the second book written by Bernard and Professor Patrick Meyrueis from the Photonics Systems Laboratory (LSP) at Université de Strasbourg (UdS) in France. While their first book, *Digital Diffractive Optics*, was solely dedicated to diffractive optics, this one covers a much wider range of fields associated with digital optics, namely: waveguide optics, refractive micro-optics, hybrid optics, optical MEMS and switchable optics, holographic and diffractive optics, photonic crystals, plasmonics and metamaterials. Thus, the book's subtitle, *From Micro-optics to Nanophotonics*, is indeed a faithful description of its broad contents. After reviewing these optical elements throughout the first chapters, emphasis is set on the numerical modeling techniques used in industry and research to design and model such elements. The last chapters describe in detail the state of the art in micro-fabrication techniques and technologies, and review an impressive list of applications using such optics in industry today.

Professors Kress and Meyrueis have been investigating the field of digital optics at LSP since the late 1980s, when photonics was still struggling to become a fully recognized field, like electronics or mechanics. The LSP has been very active since its creation, not only by promoting education in photonics but also by promoting national and international university/industry relations, which has yielded a number of impressive results: publications, patents, books, industrial applications and products as well as university spin-offs both in Europe and the USA. This experience fueled also several European projects, such as the Eureka FOTA project (Flat Optical Technologies and Applications), which coordinated 27 industrial and academic partners, or more recently the European NEMO network (Network in Excellence in Micro-Optics).

The LSP has thus become today one of the premier laboratories in photonics and digital optics, through education, research and product development, and this book serves as a testimonial to this continuous endeavor.

*Professor Trevor Hall*  
*Director, Centre for Research in Photonics*  
*University of Ottawa, School of Information Technology and Engineering*  
*Ottawa, Canada*



# Acknowledgments

We wish to acknowledge and express many thanks to the following individuals who helped directly or indirectly in the production of the material presented within this book:

Prof. Pierre Ambs (ESSAIM, Mulhouse, France)  
Prof. Stephan Bernet (Innsbruck Medical University, Austria)  
Mr Ken Caple (HTA Enterprises Inc., San Jose, USA)  
Dr Chris Chang (Arcus Technology Inc., Livermore, USA)  
Prof. Pierre Chavel (IOTA, Paris, France)  
Mrs Rosie (Conners Photonics Corp., Milpitas, USA)  
Mr Tom Credelle (Holox Inc., Belmont, USA)  
Dr Walter Daschner (Philips Lumileds, San Jose, USA)  
Mr Gilbert Dudkiewicz (Telmat Industrie S.A., Soultz, France)  
Mrs Judy Erkanat (Tessera Corp. San Jose, USA)  
Dr Robert Fisher (Optics 1 Corp., Los Angeles, USA)  
Prof. Joël Fontaine (INSA, Strasbourg, France)  
Prof. Joseph Ford (UCSD, San Jose, USA)  
Dr Keiji Fuse (SEI Ltd, Osaka, Japan)  
Prof. Joseph Goodman (Stanford University, Stanford, USA)  
Prof. Michel Grossman (UdS, Strasbourg, France)  
Prof. Trevor J. Hall (University of Ottawa, Canada)  
Mrs Kiomi Hamada (Photosciences Inc., Torrance, USA)  
Dr Phil Harvey (Wavefront Technologies Inc., Long Beach, USA)  
Mr Vic Hejmadi (USI Inc., San Jose, USA)  
Dr Martin Hermatschweiler (Nanoscribe GmbH, Germany)  
Dr Alex Kazemi (Boeing Corp., Pasadena, USA)  
Prof. Ernst-Bernhart Kley (FSU, Jena, Germany)  
Prof. Sing H. Lee (UCSD, San Diego, USA)  
Mr Ken Mahdi (Rokwell Collins Inc., Santa Clara, USA)  
Prof. Jan Masajada (Wroclaw Institute of Technology, Wroclaw, Poland)  
Dr Nicolas Mauduit (Vision intégrée, Paris, France)  
Prof. Juergen Mohr (Forschungszentrum Karlsruhe, Germany)  
Mr Paul Moran (American Precision Dicing Inc., San Jose, USA)  
Prof. Guy Ourisson (ULP, Strasbourg, France)  
Prof. Olivier Parriaux (Université St. Etienne, France)  
Prof. Pierre Pfeiffer (UdS, Strasbourg, France)  
Dr Milan Popovitch (SBG Labs Inc., Sunnyvale, USA)  
Dr Steve Sagan (BAE Corp., Boston, USA)

Prof. Pierre Saint-Hilaire (Optical Science Center, University of Arizona, USA)  
Dr Edouard Schmidlin (JPL/NASA, Pasadena, USA)  
Mr Michael Sears (Flextronics Inc., San Jose, USA)  
Prof. Bruno Serio (UdS, Strasbourg, France)  
Dr Michel Sirieix (Sagem SA, Paris, France)  
Dr Ron Smith (Digilens Inc., Sunnyvale, USA)  
Dr Suning Tang (Crystal Research Inc., Fremont, USA)  
Dr Tony Telesca (New York, USA)  
Prof. Hugo Thiepont (Vrije Universiteit Brussel, Belgium)  
Dr Jim Thomas (UCSD, San Diego, USA)  
Prof. Patrice Twardowsky (UdS, Strasbourg, France)  
Dr Jonathan Waldern (SBG Labs Inc., Sunnyvale, USA)  
Dr Paul Wehrenberg (Apple Corp., Cupertino, USA)  
Prof. Ming Wu (UCLA, Los Angeles, USA)  
Prof. Frank Wyrowsky (LightTrans GmbH, Jena, Germany)  
Dr Zhou Zhou (UCSD, San Diego, USA)

We also wish to express our gratitude to all our friends and family, who contributed to the completion of the book (Janelle, Sandy, Erik, Kevin, Dan, Hélène, Sabine, Christine, Claire, etc.), and a special thank you to Geoff Palmer, who did a terrific job in copy editing this book.



# Acronyms

## Optical Design Acronyms

|       |                                       |
|-------|---------------------------------------|
| BPM   | Beam Propagation Method               |
| CGH   | Computer-Generated Hologram           |
| DBS   | Direct Binary Search                  |
| DFT   | Discrete Fourier Transform            |
| DOE   | Diffraction Optical Element           |
| DOF   | Depth Of Focus                        |
| EMT   | Effective Medium Theory               |
| FDTD  | Finite Difference Time Domain         |
| FFT   | Fast Fourier Transform                |
| FZP   | Fresnel Zone Plate                    |
| HOE   | Holographic Optical Element           |
| IFTA  | Iterative Fourier Transform Algorithm |
| M-DOE | Moiré DOE                             |
| MTF   | Modulation Transfer Function          |
| NA    | Numeric Aperture                      |
| PSF   | Point Spread Function                 |
| RCWA  | Rigorous Coupled Wave Analysis        |
| SBWP  | Space Bandwidth Product               |

## Computer Design Acronyms

|         |                                                    |
|---------|----------------------------------------------------|
| CAD/CAM | Computer-Aided Design/Computer-Aided Manufacturing |
| CIF     | Caltech Intermediate Format                        |
| DFM     | Design For Manufacturing                           |
| DRC     | Design Rule Check                                  |
| EDA     | Electronic Design Automation                       |
| EPE     | E-beam Proximity Effect                            |
| GDSII   | Graphical Data Structure Interface                 |
| OPC     | Optical Proximity Correction                       |
| OPE     | Optical Proximity Effect                           |
| RET     | Reticle Enhancement Techniques                     |

## **Fabrication-related Acronyms**

|        |                                              |
|--------|----------------------------------------------|
| AFM    | Atomic Force Microscope                      |
| AOM    | Acousto-Optical Modulator                    |
| ARS    | Anti-Reflection Surface                      |
| CAIBE  | Chemically Aided Ion-Beam Etching            |
| DCG    | DiChromated Gelatin                          |
| GRIN   | GRaded INdex                                 |
| HEBS   | High-Energy Beam-Sensitive Glass             |
| H-PDLC | Holographic-Polymer Dispersed Liquid Crystal |
| HTPS   | High-Temperature PolySilicon                 |
| IC     | Integrated Circuit                           |
| LBW    | Laser Beam Writer                            |
| LC     | Liquid Crystal                               |
| LCD    | Liquid Crystal Display                       |
| LCoS   | Liquid Crystal on Silicon                    |
| LIGA   | Lithography/GAlvanoforming                   |
| MEMS   | Micro-Electro-Mechanical System              |
| MOEMS  | Micro-Opto-Electro-Mechanical System         |
| OCT    | Optical Coherence Tomography                 |
| OE     | Opto-Electronic                              |
| PLC    | Planar Lightwave Circuit                     |
| PSM    | Phase Shift Mask                             |
| RIBE   | Reactive Ion-Beam Etching                    |
| SLM    | Spatial Light Modulator                      |
| VLSI   | Very Large Scale Integration                 |

## **Application-related Acronyms**

|       |                                         |
|-------|-----------------------------------------|
| BD    | Blu-ray Disk                            |
| CATV  | CABLe TV                                |
| CD    | Compact Disk                            |
| CWDM  | Coarse Wavelength Division Multiplexing |
| DVD   | Digital Versatile Disk                  |
| DWDM  | Dense Wavelength Division Multiplexing  |
| HMD   | Helmet-Mounted Display                  |
| HUD   | Head-Up Display                         |
| LED   | Light-Emitting Diode                    |
| MCM   | Multi-Chip Module                       |
| OPU   | Optical Pick-up Unit                    |
| OVID  | Optically Variable Imaging Device       |
| VCSEL | Vertical Cavity Surface-Emitting Laser  |
| VIPA  | Virtual Image Plane Array (grating)     |
| VOA   | Variable Optical Attenuator             |

# Introduction

## Why a Book on Digital Optics?

When a new technology is integrated into consumer electronic devices and sold worldwide in supermarkets and consumer electronic stores, it is usually understood that this technology has then entered the realm of mainstream technology.

However, such progress does not come cheaply, and has a double-edge sword effect: first, it becomes widely available and thus massively developed in various applications, but then it also becomes a commodity, and thus there is tremendous pressure to minimize the production and integration costs while not sacrificing any aspects of performance.

The field of digital optics is about to enter such a stage, which is why this book provides a timely insight into this technology, for the following prospective groups of readers:

- for the research world (academia, government agencies and R&D centers) to have a broad but condensed overview of the state of the art;
- for foundries (optical design houses, optical foundries and final product integrators) to have a broad knowledge of the various design and production tools used today;
- for prospective industries – ‘How can I use digital optics in my products to make them smaller, better and cheaper?’; and
- for the mainstream public – ‘Where are they used, and how do they work?’

This book is articulated around four main topics:

1. The state of the art and a classification of the different physical implementations of digital optics (ranging from waveguide optics to diffractive optics, holographics, switchable optics, photonic crystals and metamaterials).
2. The modeling tools used to design digital optics.
3. The fabrication and replication tools used to produce digital optics.
4. A review of the main applications, including digital optics in industry today.

This introductory chapter will define what the term *digital optics* means today in industry, before we start to review the various digital optics implementation schemes in the early chapters.

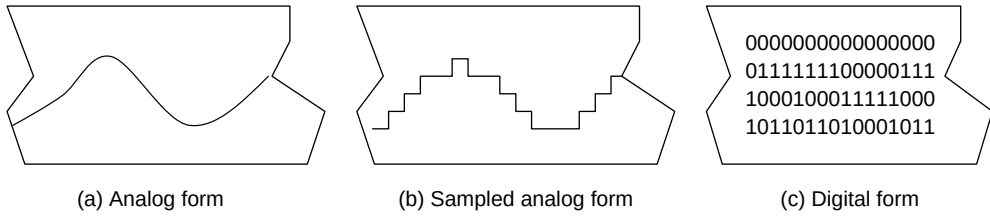


**Figure 1** Analog systems versus digital systems

## Digital versus Analog

In attempting to define the term ‘digital’ as introduced in the title of this book, one has to consider its counterpart term ‘analog’. The ‘digital’ versus ‘analog’ concept can also be understood when considering the term ‘continuous’ versus ‘discrete’ (see Figure 1).

History has proved that the move from analog systems to digital systems in technology (especially in electronics) has brought about a large number of improvements, for example:

- added flexibility (easy to program) and faster, more precise, computers;
- new functionalities (built-in error detection and correction algorithms etc.);
- ease of miniaturization (very large scale integration, VLSI); and
- ease of mass replication (microlithographic fabrication techniques).

## What are Digital Optics?

As far as optics are concerned, the move from analog (conventional lenses, mirrors and fiber optics) to digital (planar optical elements composed of microscopic structures) has been mainly focused on the last two points: miniaturization and mass replication. This said, new or improved optical functionalities have also been discovered and investigated, especially through the introduction of digital diffractive optics and digital waveguide optics, and their hybrid combination, as will be discussed in detail in the chapters to come.

Miniaturization and mass-production have begun to lead the optical industry toward the same trend as in the micro-electronics industry in the 1970s, namely to the integration of densely packed planar systems in various fields of application (optical telecoms, optical data storage, optical information processing, sensors, biophotonics, displays and consumer electronics).

At first sight, the term ‘digital optics’ could lead one to think that such elements might be either digital in their functionality (in much the same way that digital electronics provide digital signal processing) or digital in their form (much like digital – or binary – microscopic shapes rather than smooth shapes). Well, it actually takes none of these forms.

The adjective ‘digital’ in ‘digital optics’ refers much more simply to the way they are designed and fabricated (both in a digital – or binary – way). The design tool is usually a digital computer and the fabrication tool is usually a digital (or binary) technology (e.g. by using binary microlithographic fabrication techniques borrowed from the Integrated Circuit, or IC, manufacturing industry).

Figure 2 details the similarities between the electronic and optic realms, in both analog and digital versions. In the 1970s, digital fabrication technology (binary microlithography) helped electronics move from single-element fabrication to mass production in a planar way through very large scale integration (VLSI). Similarly, identical microlithographic techniques would prove effective in helping the optics industry to move from single-element fabrication (standard lenses or mirrors) down to planar integration

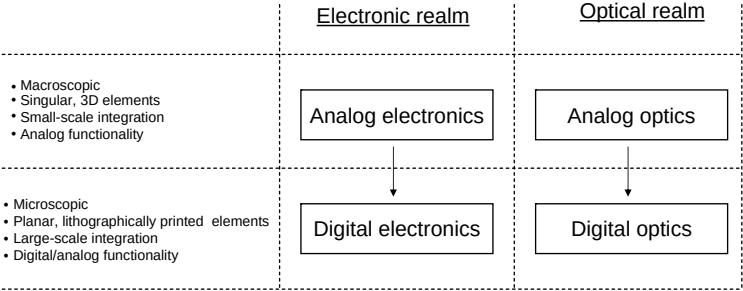


**Figure 2** Analogies between the electronics and optics realms

with similar VSLI features. The door to planar optics mass production has thus been opened, exactly as it was for the IC industry 30 years earlier, with the noticeable difference that there was no need to invent a new fabrication technology, since this had already been developed for digital electronics.

However, it is important to understand that although the fabrication technique used may be a binary microfabrication process, the resulting elements are not necessarily binary in their shape or nature, but can have quasi-analog surface reliefs, analog index modulations, gray-scale shades or even a combination thereof.

Also, their final functionality might not be digital – or binary – as a digital IC chip would be, but could instead have parallel and/or analog processing capabilities (information processing or wavefront processing). This is especially true for free-space digital optics, and not so much for guided-wave digital optics.

It is therefore inaccurate to draw a quick comparison between analog electronics versus digital electronics and analog (refractive) optics versus digital (diffractive or integrated) optics, since both optical elements (analog or digital) can yield analog or digital physical shapes and/or processing capabilities.

**The Realm of Digital Optics**

Now that we have defined the term ‘digital optics’ in the previous section, the various types of digital optical elements will be described.

The realm of digital optics (also referred to as ‘micro-optics’ or ‘binary optics’) comprises two main groups, the first relying on free-space wave propagation and the second relying on guided-wave propagation (see Figure 3).

The various optical elements defining these two groups (free-space and guided-wave digital optics) are designed by a computer and fabricated by means similar to those found in IC foundries (microlithography).

Figure 3 shows, on the free-space optics side, three main subdivisions, which are, in chronological order of appearance, refractive micro-optical elements, diffractive and holographic optical elements, and nano-optics (photonic crystals). On the guided-wave optics side, there are also three main subdivisions, which are, again in chronological order of appearance, fiber optics, integrated waveguide optics and nano-optics. It is worth noting that nano-optics (or photonic crystals) can actually be considered as guided-wave optics or free-space optics, depending on how they are implemented (as 1D, 2D or 3D structures).

This book focuses on the analysis of free-space digital optics rather than on guided-wave optics. Guided-wave micro-optics, or integrated optics, are well described in numerous books, published over

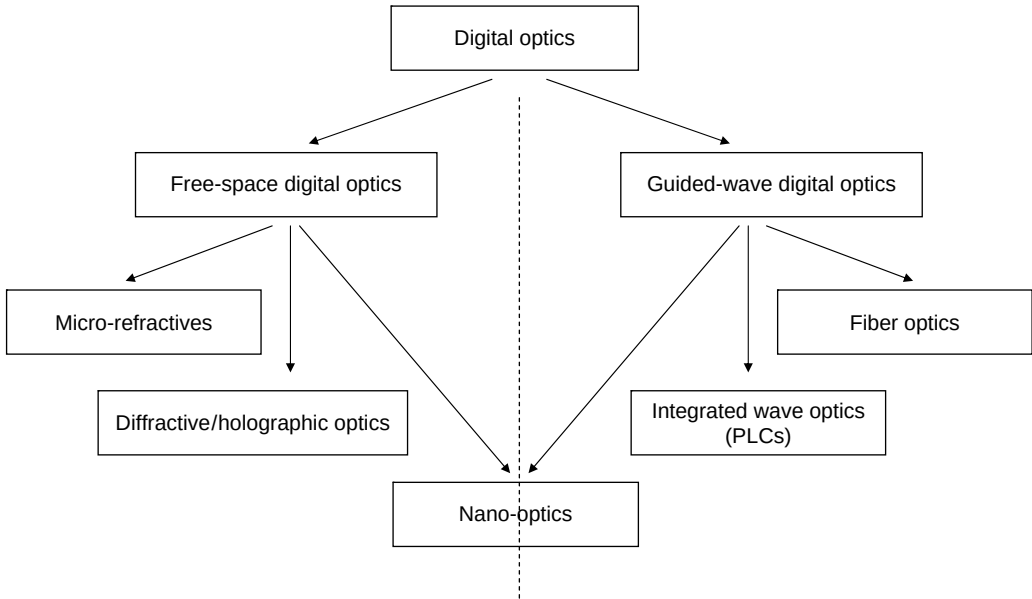


**Figure 3** The realm of digital optics

more than three decades, and dedicated books on ‘guided-wave’ photonic crystals have been available for more than five years now.

However, the combination of free-space digital optics and guided-wave digital optics is a very important and growing field, sometimes also referred to as ‘planar optics’, and that is what will be described in this book.

## Supplementary Material

Supplementary book material is available at [www.applieddigitaloptics.com](http://www.applieddigitaloptics.com) including information about workshops and short courses provided by the authors. The design and modeling programs used in the book can be downloaded from the website.

# 1

## From Refraction to Diffraction

### 1.1 Refraction and Diffraction Phenomena

In order to predict the behavior of light as it is affected when it propagates through digital optics, we have to consider the various phenomena that can take place (refraction, reflection, diffraction and diffusion). Thus, we have to introduce the dual nature of light, which can be understood and studied as a corpuscle and/or an electromagnetic wave [1].

The corpuscular nature of light, materialized by the photon, is the basis of ray tracing and the classical optical design of lenses and mirrors. The wave nature of light, considered as an electromagnetic wave, is the basis of physical optics used to model diffractive optics and other micro- or nano-optical elements, such as integrated waveguides, and photonic crystals (see Chapters 3–10).

In the simple knife-edge example presented in Figure 1.1, the corpuscular nature of light (through ray tracing) accounts for the geometrical optics, whereas the wave nature of light (physical optics) accounts not only for the light present in the optical path, but also for the light appearing inside the geometrical shadow (the Gibbs phenomenon). According to geometrical optics, no light should appear in the geometrical shadow. However, physical optics can predict accurately where light will appear within the geometrical shadow region, and how much light will fall in particular locations.

In this case, the laws of reflection and refraction are inadequate to describe the propagation of light; diffraction theory has to be introduced.

### 1.2 Understanding the Diffraction Phenomenon

Diffraction comes from the limitation of the lateral extent of a wave. Put in simple terms, diffraction arises when a wave of a certain wavelength collides with obstacles (amplitude or phase obstacles) that are either singular or abrupt (the knife-edge test, Young's holes experiment) smooth but repetitive (the sinusoidal grating), or even abrupt and repetitive (binary gratings). The smaller the obstacles are, the larger the diffraction effects become (and also the larger the diffraction angles become).

Today, when harnessing diffraction to be used in industrial applications, the obstacles are usually designed and fabricated as pure phase obstacles, either in reflection or in transmission [2–4]. Fine-tuning of the obstacle's parameters through adequate modeling of the diffraction phenomenon can yield very specific diffraction effects with a maximum intensity (or diffraction efficiency).

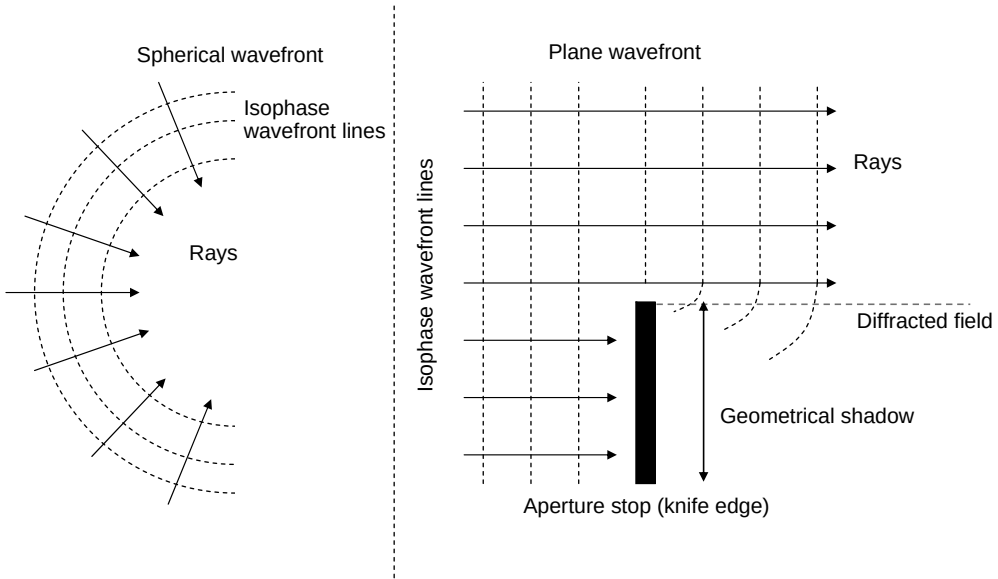


**Figure 1.1** The dual nature of light: geometrical and physical optics

### 1.2.1 Chronological Stages in Understanding Diffraction Phenomena

The diffraction phenomenon was demonstrated for the first time by Leonardo da Vinci (1452–1519) in a very rudimentary way. The first accurate description of diffraction was introduced by Francesco Maria Grimaldi (1618–1663) in his book published in 1665, two years after his death. In those times, corpuscular theory, which was widely believed accurately to describe the propagation of light, had failed to explain the diffraction phenomenon. In 1678, Christian Huygens (1629–1695) proposed a wave theory for the propagation of light that described diffraction as a source of secondary spherical disturbance (see Appendix B). Sir Isaac Newton (1642–1727) had been a strong advocate of the corpuscular theory since 1704. His strong influence over contemporary scientists had halted progress in understanding diffraction during the 18th century. In 1804, Thomas Young (1773–1829) introduced the concept of interference, which directly proceeds from the wave nature of light. Augustin Jean Fresnel (1788–1827) brought together the ideas of Huygens and Young in his famous memoir. In 1860, James Clerk Maxwell (1831–1879) identified light as an electromagnetic wave (see Appendix A). Gustav Kirchhoff (1824–1887) gave a more mathematical form to Fresnel's expression of diffraction. His work basically relied on two assumptions concerning the field at the diffraction aperture. Although those assumptions were quite empirical, his formulation provided a good approximation of the real diffracted field. In 1884, Arnold J.W. Sommerfeld (1868–1951) refined Kirchhoff's theory. Thanks to Green's theorem, he suppressed one of the two assumptions that Kirchhoff had made earlier, to derive the so-called Rayleigh–Sommerfeld diffraction theory.

Table 1.1 summarizes, in a chronological way, the understanding of optics as both a corpuscular phenomenon and an electromagnetic field.

When studying the propagation of light in a homogeneous or nonhomogeneous medium – such as a lens, a waveguide, a hologram or a diffractive element (through refraction, diffraction or diffusion) – the refractive index is one of the most important parameters. Light travels through a transparent medium (transparent to its specific wavelength) of index  $n$  at a speed  $v_n$  that is lower than its speed  $c$  in a vacuum. The index of refraction,  $n$ , in a transparent medium is defined as the ratio between the speed of light in a



Table 1.1 Chronological events in the understanding of optics

|  |      |                                                                                                                            |
|--|------|----------------------------------------------------------------------------------------------------------------------------|
|  |      | ...                                                                                                                        |
|  |      | 130 Claudius Ptolemaeus tabulates angles of refraction for several media                                                   |
|  |      | 1305 Dietrich von Freiberg uses water filled flasks to study the reflection/refraction in raindrops that leads to rainbows |
|  |      | 1604 Johannes Kepler describes how the eye focuses light                                                                   |
|  |      | 1611 Marko Dominis discusses the rainbow in <i>De Radiis Visus et Lucis</i>                                                |
|  | 1621 | <b>1611 Johannes Kepler discovers total internal reflection, a small-angle refraction law and thin lens optics</b>         |
|  | 1678 | <b>1621 Willebrord Snell states his law of refraction</b>                                                                  |
|  |      | 1637 René Descartes quantitatively derives the angles at which rainbows are seen with respect to the the Sun's elevation   |
|  |      | <b>1678 Christian Huygens states his principle of wavefront sources</b>                                                    |
|  |      | 1704 Isaac Newton publishes <i>Opticks</i>                                                                                 |
|  |      | 1728 James Bradley discovers the aberration of starlight and uses it to determine the speed of light                       |
|  |      | 1752 Benjamin Franklin shows that lightning is electricity                                                                 |
|  |      | 1785 Charles Coulomb introduces the inverse-square law of electrostatics                                                   |
|  |      | 1800 William Herschel discovers infrared radiation from the Sun                                                            |
|  |      | 1801 Johann Ritter discovers ultraviolet radiation from the Sun                                                            |
|  | 1801 | <b>1801 Thomas Young demonstrates the wave nature of light and the principle of interference</b>                           |
|  |      | 1809 Etienne Malus publishes the law of Malus, which predicts the light intensity transmitted by two polarizing sheets     |
|  |      | 1811 François Arago discovers that some quartz crystals will continuously rotate the electric vector of light              |
|  |      | 1816 David Brewster discovers stress birefringence                                                                         |
|  |      | 1818 Siméon Poisson predicts the Poisson bright spot at the center of the shadow of a circular opaque obstacle             |
|  |      | 1818 François Arago verifies the existence of the Poisson bright spot                                                      |
|  |      | 1825 Augustin Fresnel phenomenologically explains optical activity by introducing circular birefringence                   |
|  |      | 1831 Michael Faraday states his law of induction                                                                           |
|  |      | 1845 Michael Faraday discovers that light propagation in a material can be influenced by external magnetic fields          |
|  |      | 1849 Armand Fizeau and Jean-Bernard Foucault measure the speed of light to be about 298 000 km/s                           |
|  |      | 1852 George Stokes defines the Stokes parameters of polarization                                                           |
|  |      | 1864 James Clerk Maxwell publishes his papers on a dynamical theory of the electromagnetic field                           |
|  |      | 1871 Lord Rayleigh discusses the blue sky law and sunsets                                                                  |
|  | 1873 | <b>1873 James Clerk Maxwell states that light is an electromagnetic phenomenon</b>                                         |
|  |      | 1875 John Kerr discovers the electrically induced birefringence of some liquids                                            |
|  |      | 1895 Wilhelm Röntgen discovers X-rays                                                                                      |
|  |      | 1896 Arnold Sommerfeld solves the half-plane diffraction problem                                                           |
|  |      | ...                                                                                                                        |

**Table 1.2** Refractive indices for conventional (natural) and nonconventional materials

| Media                            | Refractive index                                 | Type                                                    | Examples                                                |
|----------------------------------|--------------------------------------------------|---------------------------------------------------------|---------------------------------------------------------|
| <b>Conventional materials</b>    |                                                  |                                                         |                                                         |
| Vacuum                           | 1 exactly                                        | Natural                                                 | —                                                       |
| Air (actual)                     | 1.0003                                           | Natural                                                 | —                                                       |
| Air (accepted)                   | 1.00                                             | —                                                       | —                                                       |
| Ice                              | 1.309                                            | Natural                                                 | —                                                       |
| Water                            | 1.33                                             | Natural                                                 | Liquid lenses                                           |
| Oil                              | 1.46                                             | Natural/Synthetic                                       | Immersion lithography                                   |
| Glass (typical)                  | 1.50                                             | Natural                                                 | BK7 lenses                                              |
| Polystyrene plastic              | 1.59                                             | Natural/Synthetic                                       | Molded lenses                                           |
| Diamond                          | 2.42                                             | Natural                                                 | TIR in jewelry                                          |
| Silicon                          | 3.50                                             | Natural                                                 | Photonic crystals                                       |
| Germanium (IR)                   | 4.10                                             | Natural                                                 | IR lenses                                               |
| Media                            | Refractive index                                 | Type                                                    | Examples                                                |
| <b>Nonconventional materials</b> |                                                  |                                                         |                                                         |
| Metamaterials                    | Negative indices                                 | Synthetic, active materials (plasmon)                   | High-resolution lens, Harry Potter’s invisibility cloak |
| Bose–Einstein condensate         | $n \gg 1$ , validated at $n > 1\,000\,000\,000!$ | Synthetic, $T = 0^\circ\text{K}$ ( $v < 1\text{ mph}$ ) | Low-consumption chips, telecom                          |
| ?                                | $0 < n < 1.0$                                    | Improbable ( $v > c$ )                                  | Telecom, time machine,...                               |

vacuum ( $c$ ) and the speed of light in the medium. This index can also be defined as the square root of the product of the permittivity and permeability of the material considered for the specific wavelength of interest (for most media,  $\mu = 1$ ):

$$\begin{cases} n = \frac{c}{v_n} \\ n = \sqrt{\epsilon \cdot \mu} \end{cases} \tag{1.1}$$

At this point, one could ask whether there would be a medium with indices that are positive but lower than 1 (which would mean that light would travel faster than the speed of light in a vacuum). This is largely improbable: however, there are media in which the phase velocity of light is greater than  $c$ , but cannot be used to send energy or signals at a speed in excess of  $c$ .

It is worth noting that the range of refractive indices in nature is much higher than one would imagine (from air = 1.0 to glass = 1.5). For example, silicon (Si) has a quite high index of 3.5 for infrared (IR) wavelengths, which enables the fabrication of photonic crystals in which the index change has to be the highest possible in order to achieve full photonic band gaps (see Chapter 10). Table 1.2 lists the refractive indices for some common materials. Interestingly, the range of refractive indices found in nature can be extrapolated by the fabrication of synthetic materials known as metamaterials (see also Chapter 10), and even materials with negative indices can be produced.

### 1.3 No More Parasitic Effects

History shows us that optical engineering has usually considered diffraction effects to be negative and parasitic. These effects usually manifest when the imaging resolution limit is approached. They are

considered detrimental to the proper operation of optical instruments. It is only recently that such effects have been considered as being advantageous, and have been included in the optical engineer's standard design toolbox. The catalyst has been mainly new microfabrication techniques, and their availability to optical engineers as borrowed from the IC industry (see Chapters 12–14).

Many diffractive elements have counterparts in the classical realm of optical elements. However, the similarity is only superficial, since their behavior under various operating configurations can be very different.

Furthermore, we will see in Chapter 7 that, in many cases, diffractives are actually best used in addition to refractive or reflective optics, in order to provide new and/or extended optical functionality (such as hybrid achromat or athermal singlets). So, a negative effect has been transformed into an advantage thanks to recent developments in both modeling and fabrication techniques and technologies [5].

## 1.4 From Refractive Optics to Diffractive Optics

Let us consider a simple example. At first sight, a linear grating and a prism may seem to bend an incoming laser beam in the same direction, but the similarity stops there, because different effects appear quickly as soon as one deviates from this particular operating configuration (e.g. by using an incoherent light source, varying the depth of the grooves, launching light at different angles, changing the wavelength or the polarization etc.). The same thing happens when one attempts to compare a refractive lens and its counterpart, the diffractive Fresnel lens, where chromatic dispersion appears in opposite directions, since the signs of the lenses' respective Abbe  $V$  numbers are opposite, even though their focusing power and phase profiles might be exactly the same. Figure 1.2 shows both Snell's law (from geometrical optics) and the grating equation (from physical optics), which accounts for the amount of light bending, for a small prism and a linear blazed grating.

As one gets closer to a blazed grating structure, one can consider the various periods of this grating as many individual refractive micro-prisms, and therefore not only apply the grating equation to the entire blazed grating (array of micro-prisms) [6], but also apply Snell's law of refraction to each individual micro-prism, as depicted in Figure 1.3.

It is interesting to note that the light bending angle  $\alpha$  predicted by refraction through the local micro-prism structures and the light bending angle  $\beta$  predicted by diffraction through the blazed grating are not necessarily the same. In effect, they are equal only in one very specific case: when the geometry of the

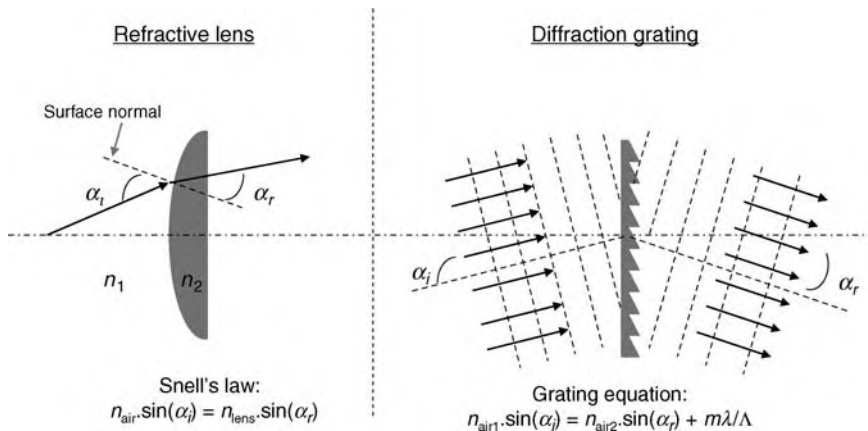


**Figure 1.2** Snell's law of refraction and the grating equation, both of which rule the amount of bending of light

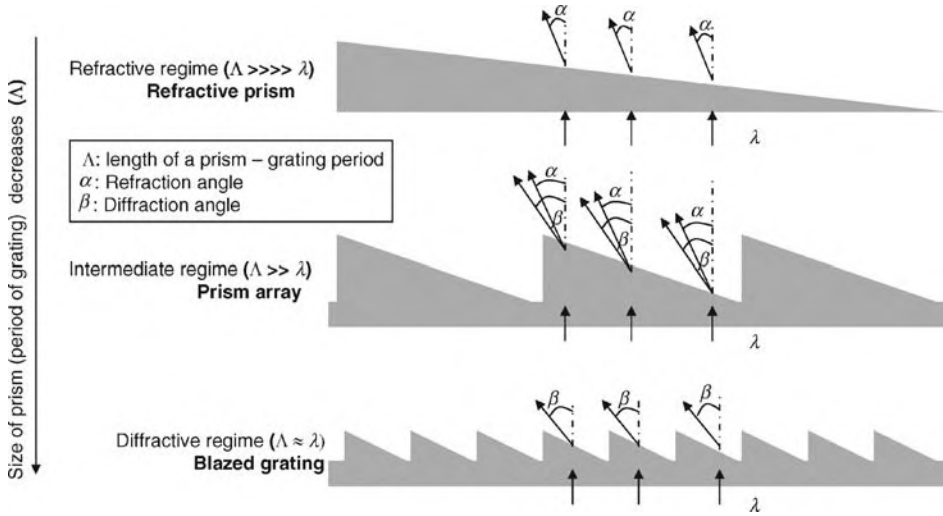


**Figure 1.3** The blazed grating and its micro-prism array structure

micro-prism is carefully chosen (its height, length and refractive index carefully optimized), as shown in Figure 1.4. Maximum diffraction efficiency is then reached for the blazed grating (which can theoretically reach 100% efficiency when both previously described effects are yielding the same bending angle).

Snell's law predicts the amount of refraction (bending of light) at a given optical interface between a medium of refractive index  $n_1$  and a medium of refractive index  $n_2$ , and thus also gives the expression for the angle of the refracted beam through each of the micro-prisms [1]:

$$n_1 \sin(\alpha_1) = n_2 \sin(\alpha_2) \Leftrightarrow \sin(\alpha + \gamma) = n_1 \sin(\alpha) \quad (1.2)$$

Physical optics, or the grating equation, predicts the diffraction angle of an electromagnetic wave at a similar interface (refractive indices  $n_1$  and  $n_2$ ), but this time constituted by a linear array of

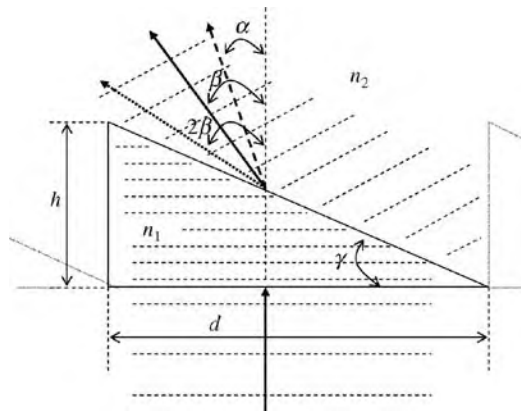


**Figure 1.4** The local micro-prism effect and the global grating effect

micro-prisms [1]:

$$\sin(\beta) = m \frac{\lambda}{\Lambda} \quad (1.3)$$

Intuitively, the maximum efficiency will thus occur when  $\alpha = \beta$ :

$$\alpha = \beta \Rightarrow h = \frac{\lambda}{n_1 - \sqrt{1 - \frac{\lambda}{\Lambda}}} \Rightarrow h = \frac{\lambda}{n_1 - 1} \quad (1.4)$$

Therefore, by using the same concepts to increase the light bending efficiency ( $\alpha = \beta$ ), and by carefully shaping the overall grating geometry (the grating period, groove height, groove angles and refractive index), one can design any type of diffractive grating or diffractive lens to yield a specific optical functionality (aspheric lenses, circular gratings, etc. – see Chapters 5 and 6).

## 1.5 From Diffractive Optics to Digital Optics

Opposite to the previous section, an attempt will be made here to move upwards from the diffractive microstructures to the refractive macroscopic structures (see Figure 1.5).

Figure 1.6 shows the similarities between the previous blazed grating and prism, and between the diffractive Fresnel lens and the refractive lens.

One could then argue that diffractives are actually arrays of small refractives, in much the same way that the blazed grating is an array of small refractive prisms.

In order to prove that this is wrong, Figure 1.7 shows optical similarities between an amplitude diffractive grating and a prism, and between an amplitude Fresnel lens (Fresnel Zone Plate) and a refractive lens. There are no similarities in the shape and form of the two elements, but the optical functionalities are the same (or at least the geometrical considerations – the energetic considerations are vastly different).

A grating can diffract light in a given direction, just as a prism would, but it can actually take any physical configuration and form (as long as efficiency is of no concern). Such a grating can be built using any periodical perturbation (phase, amplitude or a combination thereof). The diffraction angle, as the opposite case to the prism, is not a function of the index of the materials, but only a function of the perturbation period. In order to increase the efficiency, and push more light in the desired direction (which

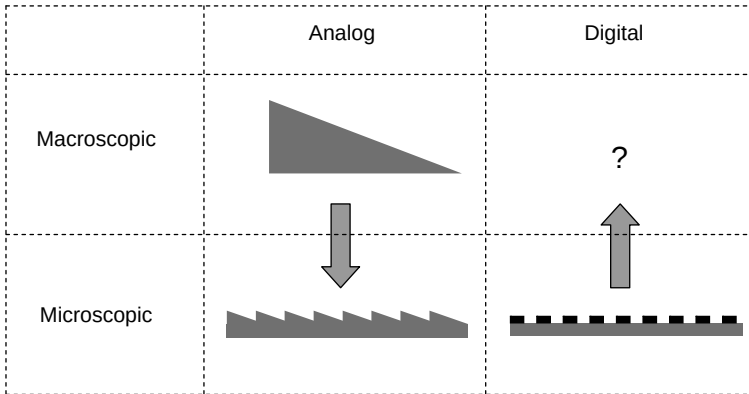


**Figure 1.5** The macroscopic and microscopic realms for analog and digital optics

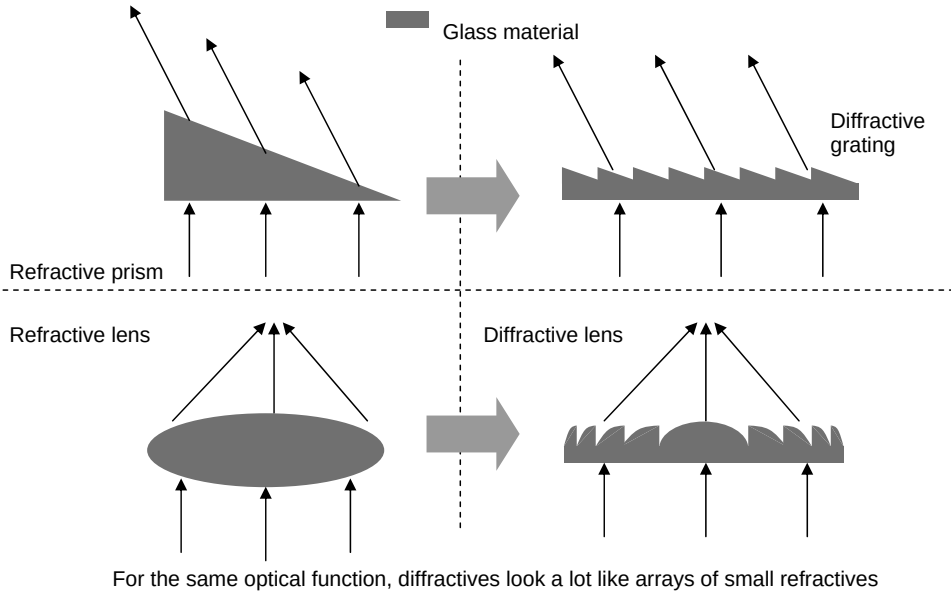


**Figure 1.6** The similarities between the blazed grating and prism and between the diffractive Fresnel lens and the refractive lens

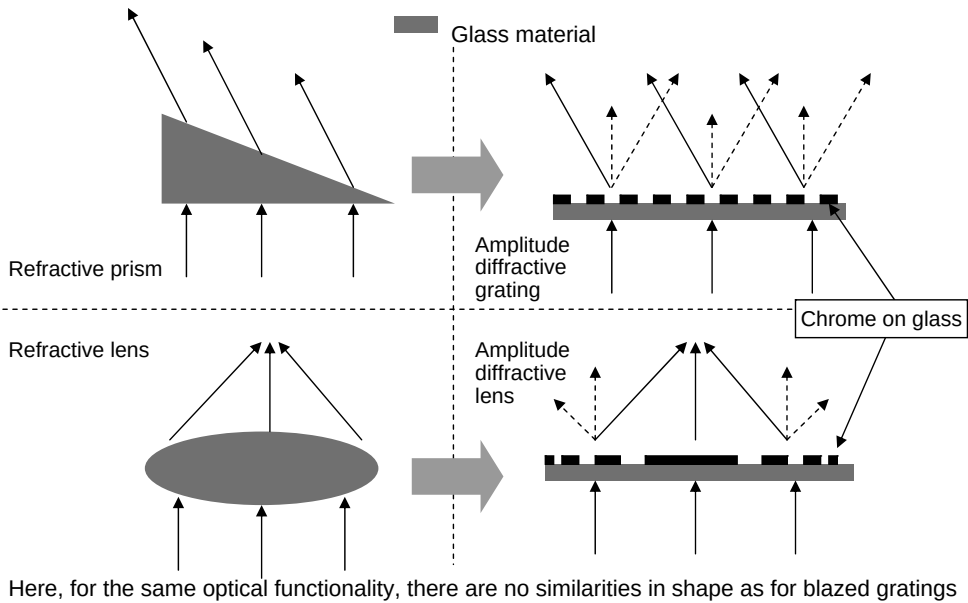


**Figure 1.7** The differences between an amplitude grating and prism, and a Fresnel Zone Plate and a refractive lens

we will call the ‘diffraction order’ in the chapter dedicated to digital diffractive optics), we will move rather to phase perturbations, and carefully optimize the phase perturbation so as to have the maximum efficiency (e.g. by applying the technique described in Equation (1.3)).

## 1.6 Are Diffractives and Refractives Interchangeable Elements?

We have seen in the previous section that it is theoretically possible (if a little more challenging in practice) to move smoothly from refraction through refractive optics (the prism) down to diffraction through diffractive optics (the blazed grating as a micro-prism array). This is, however, not the case when we consider structural binary or digital optics, and attempt to move in the opposite direction, towards the equivalent refractive optics.

Digital optics have no direct refractive counterparts. The replacement of conventional refractive optics (i.e. a refractive lens) by similar digital diffractive optics (i.e. a planar diffractive Fresnel lens), which provides obvious gains in terms of the footprint, weight and perhaps price, usually results – after some difficulties – in a return to reflectives, since the job was tailored for reflectives in the first place. In a general way, if an optical system has been designed for refractive optics, it is very difficult simply to replace reflectives by diffractive counterparts without dramatically altering the functionality of the system. However, if the optical system has been designed to be used with diffractives, the final system can potentially be smaller, lighter and cheaper, and can integrate more complex functionalities than a system that had been designed with refractive optics constraints in mind. This is a typical error that many optical engineers (and especially savvy marketing managers in high tech optical firms) tend to make.

Although diffractives can be a very useful addition to refractive/reflective optics, the simple swap between reflectives and their diffractive counterparts does not work in most cases, owing to the very different nature of diffractives, which are mainly as follows:

- many orders of diffraction can be produced;
- a zero order may be present;
- efficiency may vary strongly with wavelength;
- efficiency may vary with the incoming angle;
- strong spectral dispersion may appear (and with an opposite sign from reflectives); and
- strong thermal effects may appear (and with an opposite sign from reflectives).

Many of the effects and specifications listed here should not be considered as negative, and can actually be used efficiently to refine the functionality of a train of refractive optics (for a simple example, again see Figure 1.7), or to implement novel optical functionalities that cannot be implemented by refractive/reflective optics. However, these effects might be considered as negative effects if one attempts to simply swap elements in the hope of rapid gains in price or size/weight.

## References

- [1] M. Born and E. Wolf, *‘Principles of Optics’*, 6th edn, Pergamon Press, London, 1980.
- [2] K. Miyamoto, ‘The phase Fresnel lens’, *Journal of the Optical Society of America*, **17**, 1961, 17–21.
- [3] K. Iga, Y. Kokubun and M. Oikawa, *‘Fundamentals of Micro-optics’*, Academic Press, Tokyo, 1984.
- [4] H. Nishihara and T. Suhara, ‘Micro Fresnel lenses’, in *‘Progress in Optics, XXIV’*, E. Wolf (ed.), North Holland, Amsterdam, 1987, 3–37.
- [5] H.-P. Herzig, *‘Micro-optics: Elements, Systems and Application’*, Taylor and Francis, London, 1997.
- [6] S. Sinzinger and M. Testorf, ‘Transition between diffractive and refractive micro-optical components’, *Applied Optics*, **34**(26), 1995, 5970–5976.





# 2

## Classification of Digital Optics

As defined in Chapter 1, the adjective ‘digital’ in digital optics does not refer, as in the case of digital electronics, to the digital functionality of the element (digital signal processing) but, rather, to the digital way in which the optics are designed (by a digital computer) and fabricated (by a digital or binary microlithography technology – i.e. successive binary photomasks or reticles).

Chapter 1 has shown that the realm of digital optics can be split into two distinctive groups, free-space digital optics and guided-wave digital optics.

The emphasis in this book will be on free-space digital optics, since guided-wave digital optics are a special subdivision of digital optics, and are better described in numerous books dedicated to fiber optics and integrated waveguide optics. However, Chapter 3 will briefly describe guided-wave digital optics and related technologies.

### 2.1 Early Digital Optics

Mother Nature had developed an infinite variety of high-end refractives, diffractives, waveguides and hybrid optics, and even photonic crystals, long before the first humans were able to carve out the stones with which to smash the skulls of animals that looked reasonably edible. Such examples range from micro-optics to nanophotonics. Figure 2.1 shows a replica of one of the first diffraction gratings (the feather) on the left and one of the first sub-wavelength gratings (or photonic crystals) on the right (morpho-butterfly wings).

While the feather diffracts sunlight into a faint color spectrum (owing to the large periods – which are many times the wavelength – and the amplitude nature of the grating), the butterfly wing produces many more colorful effects, owing to the very small (sub-wavelength) periods and the reflective nature of the gratings (phase gratings).

One of the first supposed attempts by man to produce a diffractive element was to scratch small, densely packed lines onto a shiny material such as metal or glassy lava rocks. Such early diffractives can be called ‘scratch-o-grams’, and are still today an efficient and playful way to teach children (and adults) the beauty of diffractive optics and holograms. ‘Scratch-o-grams’, or hand-drawn holograms, were first popularized by William J. Beaty in 1995. The Alsatian engraver Eugene Lacaque (1914–2005) has actually been recorded in the *Guinness Book of Records* in 1999 for having etched 78 lines per millimeter by hand. Note that such curved gratings with various chirps are actually intermediate elements between gratings and diffractive lenses. Figure 2.2 shows such a ‘scratch-o-gram’ and the method for producing them.

Basically, a ‘scratch-o-gram’ is a spatial multiplexing of several circular scratches with the same radius of curvature, and various center locations and orientations. A circular fringe can be scratched on a planar



**Figure 2.1** The feather and butterfly wings as early micro- and nanostructured optics

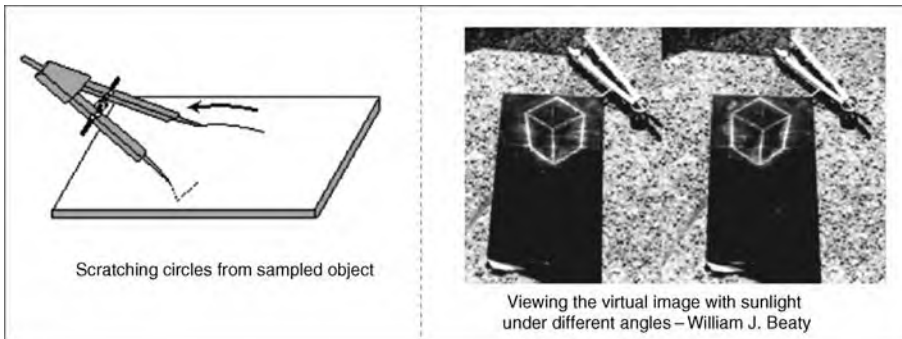


**Figure 2.2** 'Scratch-o-grams'

or curved surface by fixing the center of a compass at one point of a real object drawn on the surface, and scratching a circular fringe. Repeating this process for all the points constituting the (sampled) object generates a spatial multiplexing of various circular fringes on the same surface.

When looking at such 'scratch-o-grams' in sunlight (provided that the surface is clear, free of dust and debris, and makes the sunlight 'shine' through the clean grooves), one can see either the orthoscopic image of the object floating behind the surface or the pseudoscopic image floating in front of the surface.

People have certainly been intrigued by the beauty of natural diffractives such as butterfly wings, and have tried to reproduce such beauty with various painting techniques in caves and later on paper. People have also been intrigued by the mystical nature of such optical phenomena, especially the reproduction of selective angular effects, as well as the generation of a three-dimensional (3D) representation out of a planar surface. There have been reports that such 'scratch-o-grams' have been discovered in ancient Buddhist temples, with pictures of the Buddha floating in space behind carefully scratched glassy stones.

## 2.2 Guided-wave Digital Optics

Guided-wave optics is an important field of digital optics, and is gaining increasing attention from applications pools such as telecommunications, biotechnology and other sensor-related industries. Guided-wave optics rely on both refractive and diffractive effects within a microstructured optical material, which is usually fabricated by microlithographic techniques. Here, the electromagnetic wave is

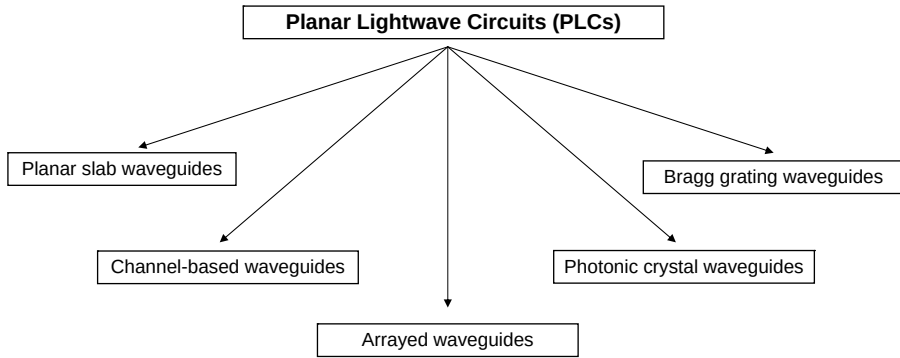


**Figure 2.3** Planar Lightwave Circuits

completely guided within the material by Total Internal Reflection (TIR). However, in many cases, the guided wave can be leaking light into free space, or can be coupled into free space and then re-coupled into guided waves, as in many Planar Lightwave Circuits (or PLCs).

Similarly to free-space digital optics, guided-wave optics can be implemented on various technological platforms; however, they are mostly used in a phase medium, where the index of refraction or the surface relief of the phase material is modulated (as in buried, ridge or diffused waveguides). One main difference is that free-space digital optics can be implemented as an amplitude element, which is not the case for guided-wave optics.

As we will see in Chapter 3, guided-wave optics can be used in numerous applications, ranging from optical telecommunications to integrated sensors, with a potentially high level of functionality in the form of PLCs.

Figure 2.3 shows the different types of guided-wave digital optical elements (or PLCs) available today.

In many cases, guided-wave digital optics are best used in conjunction with free-space digital optics, as we will see in Chapters 3 and 7.

## 2.3 Free-space Digital Optics

Free-space digital optics can be implemented in a number of different physical ways, and can be based on a number of physical phenomena. Throughout this book, we will review refractive micro-optics, diffractive optics, holographic optics and other sub-wavelength elements operating in free space. Two of the most important aspects of digital optics are diffractive optical elements and holographic optical elements.

For an overwhelming number of people (including optical engineers), the field of holography is defined and limited to:

- 3D imaging (display holograms);
- diffractive security tags (used on credit cards, bank notes, compact disks, clothes, etc.); and
- Christmas wrapping paper (roll-embossed holographic gratings).

Similarly, for an overwhelming number of people (including optical engineers), the field of diffractive optics is defined and limited to:

- spectroscopic gratings (used in spectrometers, wavelength demultiplexers, etc.); and
- laser pointer pattern generators.

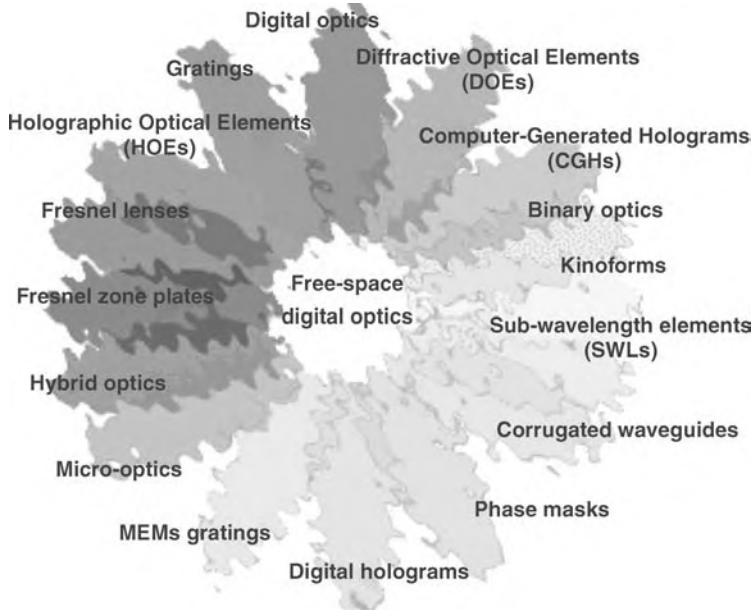


**Figure 2.4** The many names for digital free-space optics

One of the aims of this book is to broaden these established lists.

Marketing engineers and sales managers, venture capitalists, engineers and academics, as well as technical writers, have started to give numerous names to digital optics. Some of the names most commonly used to refer to these elements are binary optics [1], Diffractive Optical Elements (DOEs), Computer Generated Holograms (CGHs), kinoforms [2], zone plates and so on. Figure 2.4 shows a compilation of these various names.

There are roughly five different groups of digital free-space optics that have been reported in the literature since 1967 (when Professor Adolph Lohmann first introduced the concept of the ‘Synthetic Hologram’ [3]), which are categorized not so much according to their optical functionalities but, rather, with reference to the design techniques and the physical implementations used to manufacture them [1]. From that time onwards, Fourier optics has become an important part of modern optics technology, mainly due to the early works of Professor Joseph Goodman at Stanford [4].

Many different techniques can be used to design a diffractive element that produces the same optical functionality. Figure 2.5 summarizes these various types of free-space digital optics.

- Type 1 – Holographic Optical Elements (HOEs) – refers to the traditional optical holographic recording of volume-phase holograms (in phase modulation materials) or surface-relief holograms (in photoresist materials). These elements can be either thin or thick holograms [5], and are described in detail in Chapter 8.
- Type 2 – analytic-type diffractives – refers mostly to elements that can be designed or optimized by means of analytic methods [6] such as ray tracing (as is done in most optical CAD tools), or by solving an analytic equation (as is done for Fresnel lenses or gratings). These are the most common diffractives.
- Type 3 – numeric-type diffractives – refers mostly to elements that cannot be designed or optimized by analytic methods [6–9], and that require stochastic iterative optimization procedures and algorithms.

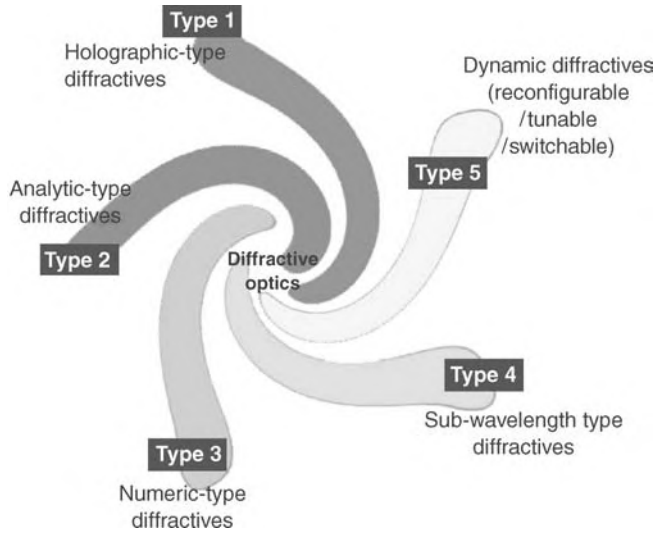


**Figure 2.5** The main different types of free-space digital optical elements

These elements can implement more complex optical functions than analytic-type diffractives, but have their limitations (the amount of CPU power required, the need to rasterize the element in the design process, etc.). They are increasingly used in industry.

- Type 4 – sub-wavelength diffractive elements (or Sub-Wavelength Gratings, SWGs) – refers to elements the basic structures of which are smaller than the reconstruction wavelength: they are thus highly polarization sensitive and they act very differently from the previous two diffractive types. Nano-optical or photonic crystals (photonic lattices) are included in these types, and are described in Chapter 10 as digital nano-optics.
- Type 5 – dynamic diffractives – refers to all the technologies used to implement reconfigurable, tunable or switchable optical functionalities. Note that these elements can actually incorporate any of the four previous elements. This last type of diffractive has recently gained much attention in the emerging optical market and applications (especially in telecom and laser displays – see Chapter 16).

Chapter 5 will discuss these various types of elements in great detail (especially Types 2, 3, 4 and 5). Two additional chapters in this book are dedicated to Holographic Optical Elements (Type 1 – see Chapter 8) and Digital Nano-optics (part of Type 4 – see Chapter 10), since these elements differ in many ways from traditional Digital Diffractive elements.

## 2.4 Hybrid Digital Optics

Hybrid optics consists of mixing different types of optical elements in a single system, in order to improve existing optical systems or to generate new optical functionalities. Hybrid optics are used where a single type of optical element cannot address the optical functionality under a set of constraints, which can include the following:

- footprint, weight and packaging issues;
- efficiency issues;

- budgeting issues; and
- mass-replication issues.

For example, achromatic doublets are well-known pure refractive elements, but hybrid achromatic singlets are smaller, simpler and cheaper to mass-produce (see Chapter 7). Hybrid waveguide gratings are also good examples of hybrid optics, where new functionalities are generated that could not be integrated solely by waveguide optics (see, e.g., the AWG-based PLCs in Chapter 3).

The chapters to follow will review the various elements discussed in this chapter in more detail.

## References

- [1] W.B. Weldkamp and T.J. McHugh, 'Binary optics', *Scientific American*, **266**(5), 1992, 50–55.
- [2] L.B. Lesem, P.M. Hirsch and J. Jordan, 'The kinoform: a new wavefront reconstruction device', *IBM Journal of Research and Development*, **13**, 1969, 150–155.
- [3] J.W. Goodman, '*Introduction to Fourier Optics*', McGraw-Hill, New York, 1968.
- [4] A.W. Lohman and D.P. Paris, 'Binary Fraunhofer holograms generated by computer', *Applied Optics*, **6**, 1967, 1739–1748.
- [5] T.K. Gaylor and M.G. Moharam, 'Thin and thick gratings: terminology clarification', *Applied Optics*, **20**, 1981, 3271–3273.
- [6] H.-P. Herzig, '*Micro-optics: Elements, Systems and Application*', Taylor and Francis, London, 1997.
- [7] B. Kress and P. Meyrueis, '*Digital Diffractive Optics*', John Wiley & Sons, Ltd, Chichester, 1999.
- [8] B. Kress, 'Diffractive Optics Technology for Product Development In Transportation, Display, Security, Telecom, Laser Machining and Biomedical Markets', Short course, SPIE SC787, 2008.
- [9] S. Sinzinger and J. Jahns, '*MicroOptics*', VCH, Weinheim, 1999.

# 3

## Guided-wave Digital Optics

As pointed out in the previous chapter, guided-wave optics is an important field of digital optics. Guided-wave digital optics refers more to integrated waveguides rather than fiber optics (systems that are fabricated through digital lithographic means rather than the conventional fiber perform drawing process).

However, integrated waveguide technology is closely related to fiber optics technology, for applications such as sensors, telecom modulators, DWDM devices and so on (especially to provide input and output interfaces for the digital waveguide device). Such systems are also known in industry as Planar Lightwave Circuits (PLCs) or planar integrated optics.

As it is an intrinsic part of the realm of digital optics, this chapter will introduce the concept of guided-wave optics, define the various modes that can in optical waveguides, and explain the fundamentals of optical couplers and optical modulators. It will also show how free-space planar optics can be used in PLCs in order to integrate novel and complex optical functions.

### 3.1 From Optical Fibers to Planar Lightwave Circuits (PLCs)

As early as 1870, John Tyndall in the United Kingdom demonstrated light guiding in a thin water jet. Ten years later, Alexander Graham introduced for the first time the notion of an optical waveguide, and in the early 1930s the first patents on 'optical tubing' appeared. As early as 1950, a patent for a two-layer glass waveguide (two different indices of refraction) had been applied, and in 1960 the laser was used for the first time as a waveguide light source.

Industry had to wait until 1965 [1] to find out how to take advantage of the low spectral absorption regions (see Figure 3.1), which would lead in the 1980s to the first optical fiber technology backbone of long-distance telephone networks [2,3].

Figure 3.1 shows that there are two low absorption levels around 1.3  $\mu\text{m}$  and 1.5  $\mu\text{m}$ . Today, industry standard values for propagation losses in telecom-grade single-mode optical fibers are 0.4 dB/km @ 1.3  $\mu\text{m}$  and 0.2 dB/km @ 1.55  $\mu\text{m}$ , which are the two most used wavelength regions (for CATV and DWDM/CWDM applications, respectively) [4]. The 10 Gb/s Ethernet application uses a shorter wavelength of 850 nm (VCSEL lasers or laser arrays), but over very small distances, since the absorption level is much greater in that wavelength range. For the 10 Gb/s Ethernet, multimode graded-index or even plastic fibers can be used.

It is worth noting that Erbium Doped Fiber Amplifiers (EDFAs) are actually best suited to work over the DWDM -C and/or -L bands (1.53–1.65  $\mu\text{m}$ ). The C band is mostly used in the United States and Europe, and the L band in Japan for historical reasons. These spectral regions are located outside the water absorption peak, which is a good natural match for the telecom industry. Therefore, wavelengths within

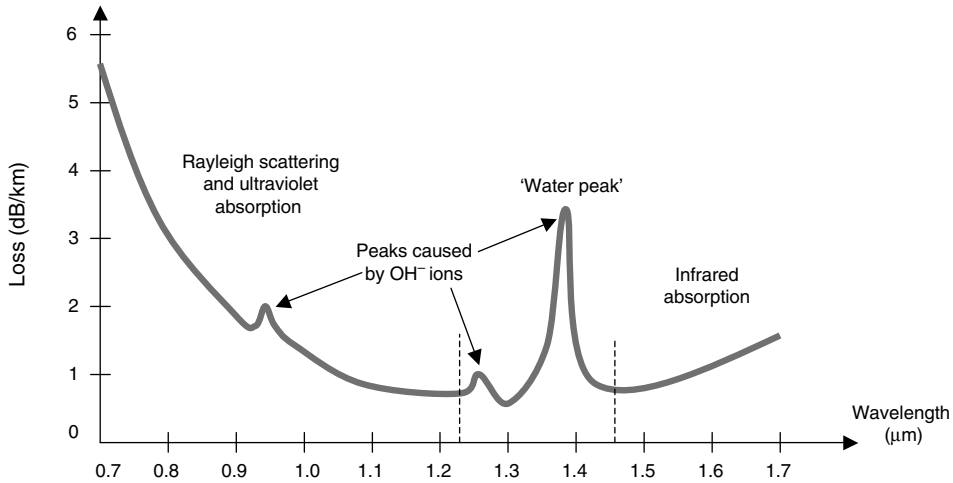


**Figure 3.1** Propagation losses occurring in glass optical fibers as a function of wavelength

the C band (1525–1565 nm) are commonly used today in conjunction with such EDFA amplifiers for submarine backbone networks or very long haul networks. It is worth noting that EDFA amplifiers are also integrated within PLC substrates.

As we will see, optical fibers have been the first industrial development of optical waveguides. The fabrication of optical fibers does not require any microlithography equipment, although the core dimension might be on the order of several microns, and the core dimension has tolerances well within the realm of sub-micro lithography. Optical fibers are drawn from macroscopic glass preforms, which scale down the size of the core by a very large factor—for example, a typical fused silica preform manufactured by MCVD may be 1 inch in diameter, and the drawn fiber only 250  $\mu\text{m}$  in diameter [5]!

Planar Lightwave Circuits (PLCs) are basically optical fibers fabricated within a planar substrate by lithographic tools [6,7]. PLCs can implement very complex optical functionalities, which are difficult to implement in fibers. Therefore, optical fibers are mainly used to transport lightwave signals over long distances, and are pigtailed to PLCs when it comes to processing the light in complex ways. As we will see below, coupling from fibers to PLCs and back to fibers is not a trivial task.

The main advantages of PLCs over optical fibers are as follows:

- PLCs can implement complex optical functionality (more complex than fibers);
- PLCs can be mass-produced by lithographic methods (plastic embossing or molding);
- hybridization to electronic and mechanic circuits (planar geometry) for planar OE, MEMS or MOEMS devices is easy; and
- PLCs are thin, planar and stackable, and can be built up into 3D systems.

### 3.2 Light Propagation in Waveguides

The basis for electromagnetic wave propagation in a dielectric waveguide relies on Total Internal Reflection (TIR) via rapid or smooth refractive index variations. Usually, the core of an optical fiber (of refractive index  $n_2$ ) or an optical waveguide is composed of a higher-index material than the cladding (of refractive index  $n_1$ ). The cladding is the medium surrounding the core, which enables the basic TIR effect, as depicted in Figure 3.2.



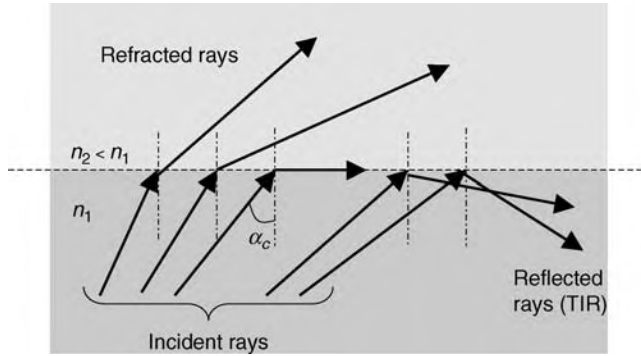


**Figure 3.2** The principle of Total Internal Reflection (TIR)

Total internal reflection occurs when the propagation angle is larger than the Brewster angle  $\alpha_B$ . This angle is also called the critical angle:

$$\alpha_c = \arcsin\left(\frac{n_2}{n_1}\right) \quad (3.1)$$

However, the TIR effect does not occur as a rapid transition between transmission (refraction) and reflection effects at the index interface, but rather as a gradual transition as the incoming angle increases, as is described in the following equation:

$$\begin{cases} T_{||} = \frac{2n_1 \cos(\alpha_i)}{n_2 \cos(\alpha_i) + n_1 \cos(\alpha_t)} U_{||} \\ T_{\perp} = \frac{2n_1 \cos(\alpha_i)}{n_1 \cos(\alpha_i) + n_2 \cos(\alpha_t)} U_{\perp} \end{cases} \quad \begin{cases} R_{||} = \frac{n_2 \cos(\alpha_i) - n_1 \cos(\alpha_t)}{n_2 \cos(\alpha_i) + n_1 \cos(\alpha_t)} U_{||} \\ R_{\perp} = \frac{n_1 \cos(\alpha_i) - n_2 \cos(\alpha_t)}{n_1 \cos(\alpha_i) + n_2 \cos(\alpha_t)} U_{\perp} \end{cases} \quad (3.2)$$

where the subscripts  $||$  and  $\perp$  indicate, respectively, the parallel and orthogonal polarizations of the wave under consideration ( $U$ , incoming;  $T$ , transmitted;  $R$ , reflected), see also Figure 3.3.

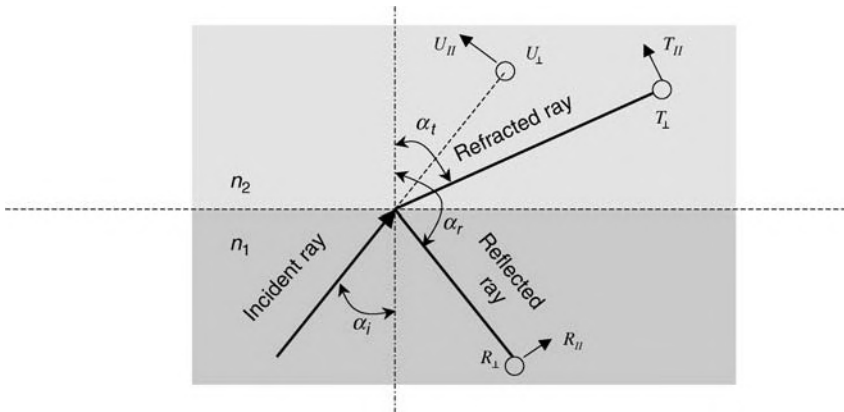


**Figure 3.3** Transmission and reflection at a planar interface

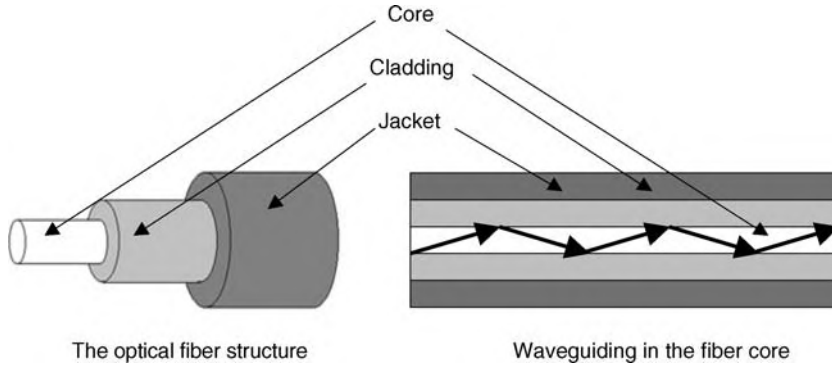


**Figure 3.4** The modern optical fiber structure

Any waveguiding principle is based on the TIR angle for mode confinement in the core. Any waveguide device (an optical fiber, a channel-based waveguide or a planar slab waveguide) is composed of a core, a cladding (and a cladding jacket for the fiber), as depicted in Figure 3.4. However, it is worth noting that the material is not necessarily glass, and can also be plastic, air or even – as will be seen later, in Chapter 10 – a nanostructured Photonic Crystal (PC waveguide) producing an effective refractive index.

The acceptance cone angle for an optical fiber is the largest angle that can be launched in the fiber for which there is still propagation along the core (that is, below the critical angle  $\alpha_c$ ). The numeric aperture (NA) of an optical waveguide is simply the sine of that maximum launch angle. See Figure 3.5, in which a step-index optical waveguide structure is depicted as an example.

The NA of conventional telecom-grade optical fibers (for a single-mode fiber, or SMF) is approximately 0.13 (e.g. the Corning SMF28 fiber). As any TIR is not 100% effective, the common definition of the waveguiding effect is based on a maximum allowed loss of 10 dB at the core/cladding interface.

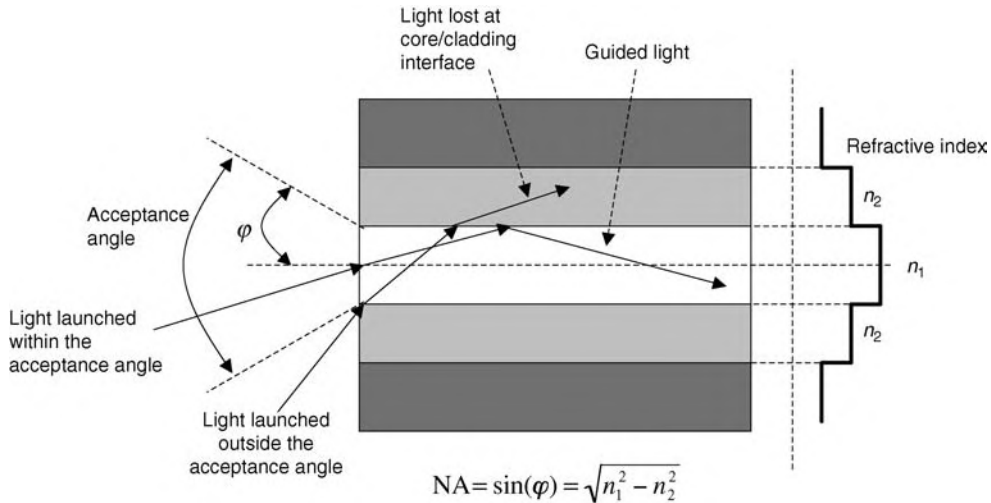


**Figure 3.5** The numeric aperture of an optical waveguide

### 3.3 The Optical Fiber

There are three main optical waveguide structures that are used today in industry:

- the step-index multimode optical waveguide;
- the graded-index multimode optical waveguide; and
- the single-mode optical waveguide.

Figure 3.6 shows the internal structures of the three different waveguide architectures.

In a step-index waveguide (both in multimode and single-mode configuration), the interface between core and cladding is a rapid index step, whereas in graded-index waveguides, the transition from the core index to the cladding index is very smooth and continuous. TIR can occur in both cases. A graded-index fiber is usually a multimode fiber. Refraction through the graded index bends the rays continuously and produces a quasi-sinusoidal ray path.

In some cases, it is interesting to use asymmetric core sections in order to produce polarization-maintaining fibers (in order to lower Polarization-Dependent Loss – PDL – in otherwise high-PDL PLCs by launching only one polarization state). Such sections are described in Figure 3.7.

The multicore optical fiber depicted in Figure 3.7 is not a polarization-maintaining fiber, but can serve many purposes in sensors and telecom applications, by introducing coupling functionalities between each core. Multicore optical fibers with up to 32 cores have been fabricated. Typically, a multicore optical fiber is fabricated by fusing together several preforms on which part of the cladding has been shaved down – ground – in order to have a core that is very close to the surface. The multicore optical fiber is then drawn as a standard fiber.

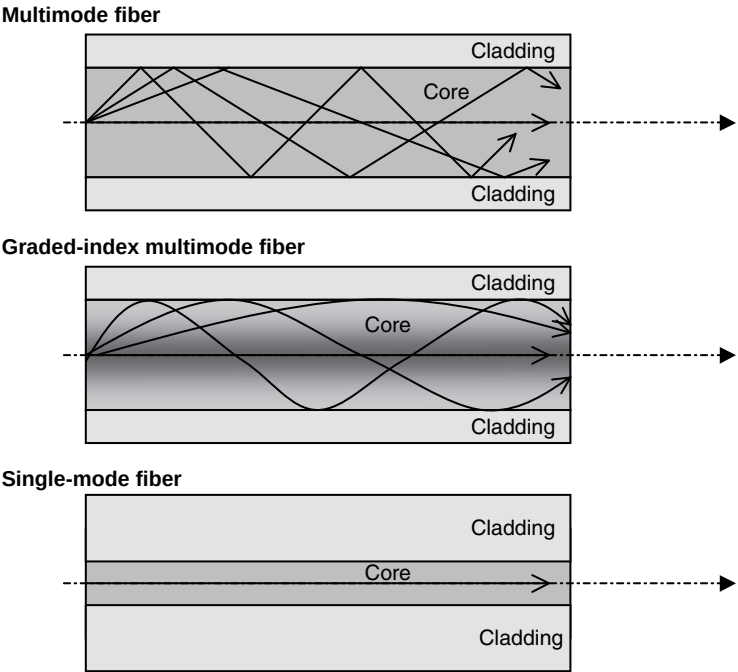


**Figure 3.6** The main optical waveguide structures

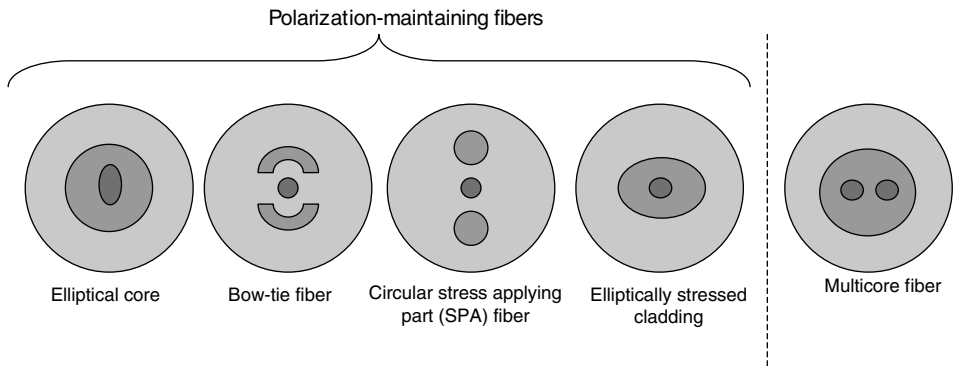


**Figure 3.7** Polarization-maintaining fiber core structures

Table 3.1 summarizes the key parameters of step-index or graded-index optical fibers (where  $a$  is the radius of the core).

The complex amplitude of low- and high-order modes that can travel within an optical fiber is shown in Figure 3.8. The mode size of the fundamental mode is also described.

The higher the mode order, the more energy will be traveling within the cladding. Similarly, the greater the wavelength, the more energy will be propagating into the cladding.

Figure 3.9 shows the two-dimensional mode profiles for some propagating modes in an optical fiber. The light intensity is highest at the center of the fiber.

Depending on the size and geometry of the core, there can be a multitude of modes circulating in the optical fiber (see Table 3.1). In Chapter 16, an example is given of a doughnut mode in a multimode graded-index plastic fiber, which can be matched with a digital diffractive vortex lens in order to minimize coupling losses.

As seen previously (see Figure 3.2), propagation in any waveguide (optical fiber or PLC) has to fulfill the condition of TIR. In Section 3.4, the cut-off frequency, which rules the propagation of the various modes within a planar dielectric slab waveguide, is derived.

**Table 3.1** Key parameters of optical fibers

| Parameter                   | Step-index fiber                                        | Graded-index fiber                                                                     |
|-----------------------------|---------------------------------------------------------|----------------------------------------------------------------------------------------|
| Refractive index, $n$       | $\begin{cases} n_1; r \leq a \\ n_2; r > a \end{cases}$ | $\begin{cases} n_1 \sqrt{1 - 2\Delta(r/a)^\alpha}; r \leq a \\ n_2; r > a \end{cases}$ |
| Numerical aperture, NA      | $\sqrt{n_1^2 - n_2^2}$                                  | $\sqrt{n(r)^2 - n_2^2}; r \leq a$                                                      |
| Normalized frequency, $(V)$ | $\frac{2\pi a}{\lambda} \cdot \text{NA}$                | $\frac{2\pi a}{\lambda} \cdot \text{NA}$                                               |
| Cut-off frequency           | 2.405                                                   | $2.405 \sqrt{1 + 2/\alpha}$                                                            |
| Number of modes             | $V^2/2$                                                 | $\frac{V^2 \alpha}{2(\alpha + 2)}$                                                     |

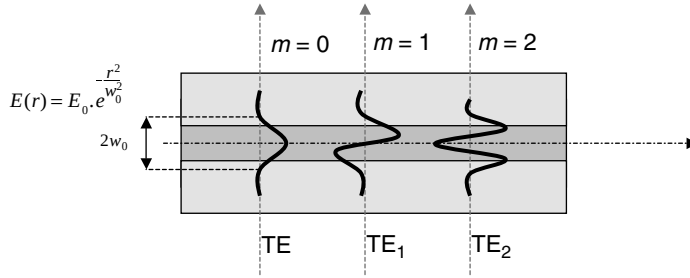


**Figure 3.8** Low and higher modes in an optical fiber

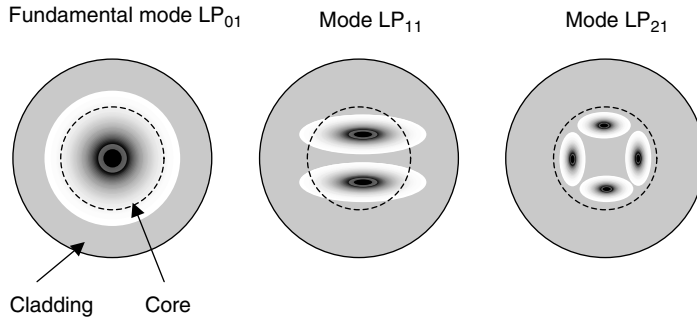


**Figure 3.9** Mode profiles in a circular waveguide

### 3.4 The Dielectric Slab Waveguide

The dielectric slab waveguide is a one-dimensional optical waveguide. The mode confinement is therefore active only in one direction, whereas in the other direction the beam can diverge as it would do in free space. The following section will discuss dielectric channel waveguides that have two-dimensional mode confinement.

In a planar dielectric optical waveguide (slab), the lower cladding index is not necessarily the same as the upper cladding index, as it is for an optical fiber. Here, the upper cladding can even be air, while the lower cladding is usually a low-index glass (see Figure 3.10).

Through TIR, the waves may bounce between the guide walls, as they would for an optical fiber waveguide. Let us consider the scalar wave equation for TE polarization along the  $y$ -axis (Equation (3.3)) (see also Appendices A and B):

$$\nabla_y^2 E_y(x, z) + n_i^2 k_0^2 E_y(x, z) = 0 \quad \text{for } i = 1, 2, 3 \quad (3.3)$$

Appendix B shows that solutions can be written in the form

$$E_y(x, z) = E_i(x) \cdot e^{-j\beta z} \quad \text{for } i = 1, 2, 3 \quad (3.4)$$

where  $\beta$  is the propagation constant, defined as

$$\beta = k_0 \cdot \sin(\phi) \quad (3.5)$$

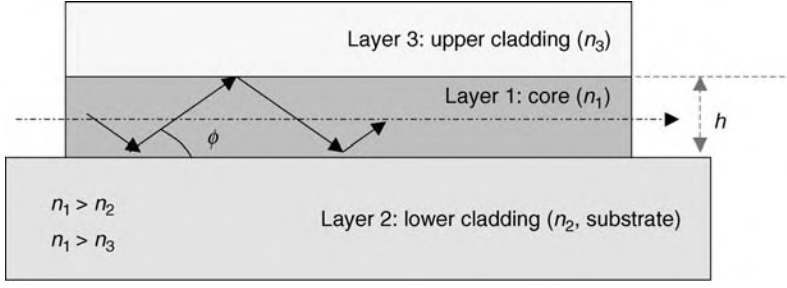


**Figure 3.10** The PLC waveguide structure

For fields that are confined within the waveguide (which are standing waves inside the guiding layer, and evanescent fields outside), there are three solutions (for the three regions):

$$\begin{cases} \text{Layer 1: } E_1 = E \cdot \cos(\kappa x - \phi) \\ \text{Layer 2: } E_2 = E' \cdot e^{\gamma x} \\ \text{Layer 3: } E_3 = E'' \cdot e^{-\delta(x-h)} \end{cases} \quad \text{where} \quad \begin{cases} \kappa = \sqrt{n_1^2 k_0^2 - \beta^2} \\ \gamma = \sqrt{\beta^2 - n_2^2 k_0^2} \\ \delta = \sqrt{\beta^2 - n_3^2 k_0^2} \end{cases} \quad (3.6)$$

After some rearrangement of Equation (3.6), one can derive an eigenvalue equation for the dielectric slab:

$$\tan(\kappa h) = \kappa \frac{(\gamma + \delta)}{(\kappa^2 - \gamma\delta)} \quad (3.7)$$

It can be shown that only certain values of  $\beta$  can satisfy Equation (3.7). This means that the dielectric slab waveguide will only support a finite number of modes. Since Equation (3.7) is very difficult to solve analytically, the eigenvalues  $\beta_n$  must be found numerically.

In order to estimate the number of modes that can travel at a given frequency, we will consider that particular modes are no longer propagating in the guide when their ray angles are close to the critical TIR angle (Equation (3.1)). For symmetric modes, the cut-off condition can be described as follows:

$$\tan\left(\frac{\kappa h}{2}\right) = 0 \quad (3.8)$$

The upper limit on the height  $h$  of the slab waveguide can be expressed as follows:

$$h < \frac{\lambda}{2\sqrt{n_1^2 - n_2^2}} \quad (3.9)$$

### 3.5 Channel Waveguides

Section 3.4 has shown the simple example of a planar slab waveguide. This waveguide was invariant in the  $x$  and  $y$  directions. We will now focus on channel waveguides that confine the guided wave within a small section, much as an optical fiber would do.

However, unlike an optical fiber, the geometry of such channel waveguides can be very carefully tailored and modulated in space in order to implement special optical functionalities. Several channel

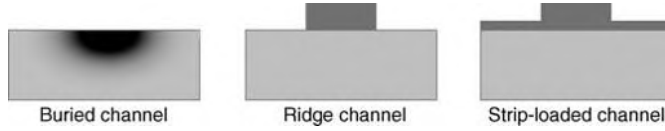


**Figure 3.11** Channel waveguides: physical implementations

waveguides can also be combined on a single substrate; for example, in order to couple energy from one to the other. These are the basis of most telecom and sensor-based PLCs today [8,9].

### 3.5.1 Channel Waveguide Architectures

There are mainly three different physical types of channel waveguide: the buried channel, the ridge channel and the strip-loaded channel (see Figure 3.11).

Diffusion fabrication techniques are often used for the production buried waveguides, similar to the ones described in Chapter 12, or for GRIN lenses (see also Chapter 4). Since it is difficult to produce quasi-circular gradient indices via diffusion techniques, buried waveguides are usually asymmetric and have a weak guidance (the effective core region lies at a very shallow depth below the surface, and light coupling is strong to the cladding or to the air). Therefore, they are best suited to the fabrication of PLCs where additional elements are used on the surface of the guide directly to process the cladding modes that are propagating – as in Bragg gratings on top of the buried core structure (see Chapter 16).

Standard lithography techniques are used to produce either ridge or strip-loaded waveguides. Ridge waveguides are usually used where strong guidance and strong confinement is required. They can also be fabricated to produce complex multichannel configurations (AWGs etc.), as we will see in the following sections. Figure 3.12 shows the typical structure of a double-heterostructure GaAs/GaAlAs planar ridge waveguide. Such structures are heavily used in semiconductor lasers.

Instead of using compositional variations to form a layered guiding structure, waveguides can also be formed by making use of the reduction in the refractive index that follows from an increase in the free carrier concentration. Such index variations can be described as follows (Drude model):

$$n^2 = 1 + \frac{N \cdot q_e^2}{m \cdot \epsilon_0} \cdot \frac{1}{-\omega^2 + \frac{i\omega}{\tau}} \quad (3.10)$$

where  $N$  is the density of free electrons,  $q_e$  the charge of  $e$ ,  $m$  the mass of an electron,  $\epsilon_0$  the dielectric constant,  $\omega$  the excitation frequency and  $\tau$  the relaxation time of one electron.

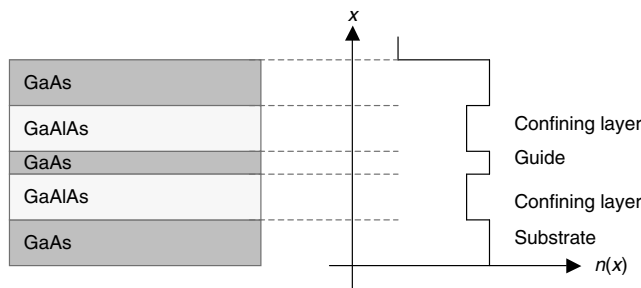


**Figure 3.12** The double-heterostructure GaAs/GaAlAs planar ridge waveguide

### 3.5.2 Low-index Waveguides

Channel waveguides come in two different configurations: low-index and high-index waveguides. Low-index waveguides are usually passive elements such as mode-matching PLCs and so on. The low indices are usually around 1.45–1.55 (glass, SiO<sub>2</sub>, BPSG, polymers etc.).

Such low-index waveguides have the same refractive indices as fibers, therefore reducing the coupling losses (about 0.1 dB per connection). Fabrication is also easy. Furthermore, they exhibit low-temperature sensitivity and polarization-dependent losses (PDLs).

### 3.5.3 High-index Waveguides

High-index waveguides are usually used to produce active PLCs in materials that have refractive indices on the order of 3.25 (indium phosphide – InP, indium gallium arsenide phosphide – InGaAsP, etc.), to produce active devices such as optical amplifiers, detectors, phase shifters and so on (see the final sections of this chapter). As their index is higher, the core section is usually smaller than the core section of low-index waveguides, therefore giving rise to larger mode mismatch between the fibers and the PLCs. Coupling losses can be as high as 3 dB per connection (e.g. half the light is lost). Such high-index PLCs require mostly lensed fibers or mode converter PLCs, as described in the next section. Finally, they have also much higher temperature sensitivity than low-index waveguides.

## 3.6 PLC In- and Out-coupling

Waveguide devices are only useful if one can couple light into these devices, and properly out-couple the processed signal into either a detector or another optical fiber, for further propagation.

In many cases, the in-coupling and out-coupling tasks require alignment efforts, and account for most of the losses in the PLC device. Very precise alignment can be done on the (lithographically patterned) substrate in between several guides and other devices on the same substrate, but when aligning an external single-mode guide to another (a semiconductor laser to a SMF fiber, for example), the task is much more difficult. This is why today laser-to-fiber alignment still remains one of the most challenging engineering tasks for various optical telecom PLCs – namely the packaging and PLC/laser pigtail or free-space coupling.

### 3.6.1 The Mode-matching Condition

In order to use the functionality of a slab or channel waveguide PLC, one needs to inject light into and collect it from the PLC. Usually, PLCs are single mode (especially for telecom applications), and therefore the coupling from a fiber to the PLC remains a difficult task (it is much easier for a slab waveguide, as we will see in Section 3.6.2).

The mode(s) of the PLC has (have) to be matched by the coupling element in order to excite the right mode(s) in the PLC and coupled back from the PLC into the exit fiber.

Figure 3.13 shows the mode-matching diagram, using the wave vector description in both media and the launch angle. The location of all potential propagating vectors forms a circle of radius  $\beta$  (related to its effective index  $n_{\text{eff}}$ ).

If the guide supports more than one bounded mode (say,  $n$  modes), the locations of all possible modes of guided propagation are a set of concentric circles of radii  $\beta_1, \beta_2, \dots, \beta_n$ . In Figure 3.12, we show three possible modes and the six potential mode matches that can occur at that transition (three reflected and three transmitted). One would like to limit the number of reflected modes and have all the modes coupled into transmitted modes in the considered waveguide. If the interface is sufficiently weak, the incoming field will traverse the interface without too much loss, and will couple with the potential modes in the second waveguide structure.



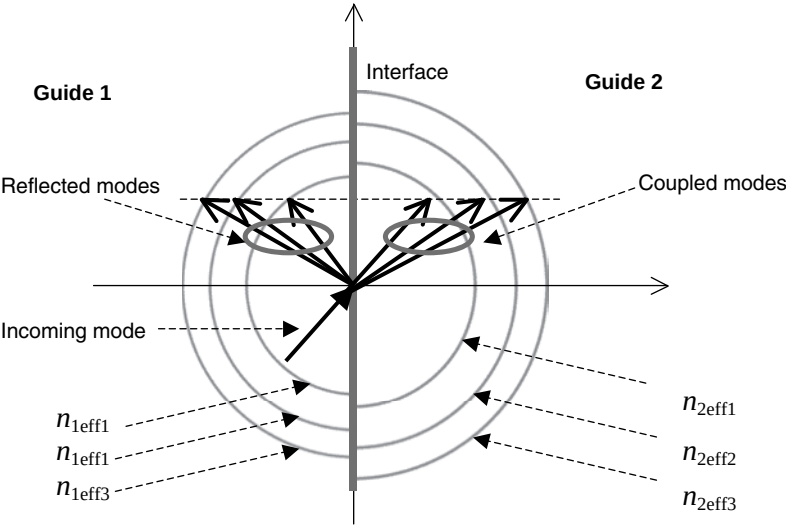


**Figure 3.13** The mode-matching diagram for mode coupling from and into waveguides

A variety of methods can be used to perform mode coupling in slab or channel-based waveguides. These methods include simple prism- or grating-based coupling (for slab waveguides), butt-coupling, ball lenses, tapered fiber ends, diffractive optics, GRIN lenses or tapered mode-matching guides.

3.6.2 Slab Waveguide Coupling

Several methods have been developed in order to couple into a slab waveguide or a channel waveguide. Light coupling into a slab waveguide can be performed via a prism or grating coupler, as shown in Figure 3.14.

When using a prism coupling, the prism is placed close to the surface of the waveguide, with a small gap (usually, the gap is provided by dust or other particles). For optimum coupling, the index of the prism is slightly higher than the index of the substrate. Similarly, a grating can be etched into the substrate, and

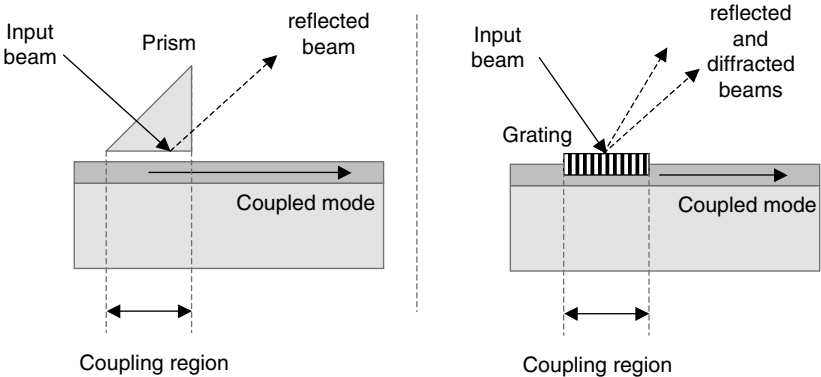


**Figure 3.14** Light coupling into a slab waveguide

can produce the desired mode coupling. Note that in the latter case, the incoming beam could be launched almost normal to the substrate.

In a general way, if one needs to couple a beam from one guide to another, mode matching has to be performed. The orientation of the propagation vector  $\beta$  gives the wave direction and its magnitude describes the effective index of the medium:

$$|\beta| = \frac{2\pi n_{eff}}{\lambda} \quad (3.11)$$

### 3.6.3 Channel Waveguide Coupling

#### 3.6.3.1 Mode Matching and Coupling Losses

The simplest way to match the modes between two waveguides (which have more or less the same geometry) is butt-coupling (that is, placing one in front of the other, as closely as possible, for direct near-field coupling). This technique has the advantage of being simple and (potentially) cheap, but it can lead to severe losses if the cores are misregistered, as can be seen in Figure 3.15.

The losses for the three misalignment geometries depicted in Figure 3.15 are derived as follows:

$$\left\{ \begin{array}{l} \text{offset loss} = \left( \frac{2w_1 w_2}{w_1^2 + w_2^2} \right)^2 \cdot e^{-2 \frac{d^2}{w_1^2 + w_2^2}} \\ \text{tilt loss} = \left( \frac{2w_1 w_2}{w_1^2 + w_2^2} \right)^2 \cdot e^{-2 \frac{(\pi n_2 w_1 w_2 \phi)^2}{\lambda^2 (w_1^2 + w_2^2)}} \\ \text{gap loss} = \frac{4 \left( 4Z^2 + \frac{w_1^2}{w_2^2} \right)}{4Z^2 \left( 1 + \frac{w_2^2}{w_1^2} \right) + \frac{w_1^2 + w_2^2}{w_2^2}} \end{array} \right. \quad \text{where } Z = \frac{g\lambda}{2\pi n_2 w_1 w_2} \quad (3.12)$$

where  $w$  is the mode diameter and  $n_2$  is the refractive index of the cladding: note that this equation only considers mode overlap. Such losses, called coupling losses, added to the standard losses of the fiber or PLC (absorption losses), are called Insertion Losses (IL) and are measured in dB. The task in pigtailling fibers to PLCs is to reduce the IL, especially for optical telecom applications, where a low IL is the name of the game (every photon counts).

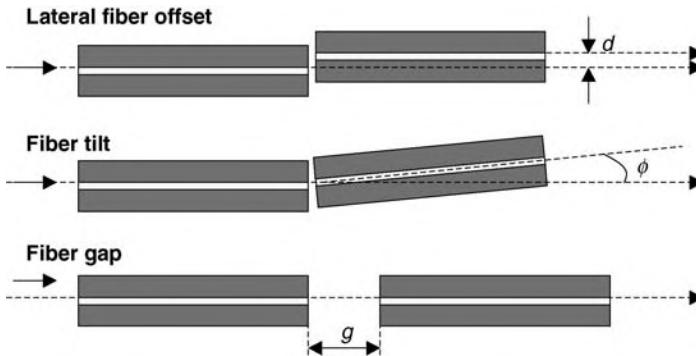


**Figure 3.15** Butt-coupling and losses due to misalignments

Absorption losses can arise from many effects: Rayleigh scattering ( $C_1$ ), fiber imperfection structure ( $C_2$ ), impurities and intrinsic absorption  $A(\lambda)$ :

$$\text{losses}(\lambda) \propto \frac{C_1}{\lambda^4} + C_2 + A(\lambda) \quad (3.13)$$

The Polarization-Dependent Loss (PDL) is another important loss measurement, which varies as a function of the launch polarization, if the PLC is polarization sensitive (unfortunately, in most cases it is). When the pigtailed fiber is not a Polarization Maintaining (PM) fiber, the fiber scrambles the polarization and therefore any polarization state can be injected into the PLC. In telecom applications, it is very desirable to have the lowest PDL, since the PDL can modulate the signal directly and thus reduce the signal-to-noise ratio (SNR) and increase the Bit-Error-Rate (BER). In many cases, the PDL is actually more critical than the IL, since the PDL can create noise and unwanted signals when the polarization changes rapidly, whereas the IL is mostly a fixed value or varies very slowly – for example, changing with temperature or pressure. For Wavelength Division Multiplexing (WDM) applications using PLCs (DWDM or CWDM), typical values for the IL are  $<0.5$  dB and typical PDL values should be  $<0.3$  dB.

The main losses that are important to optical telecom and sensor applications are as follows:

- The *Insertion Loss* (or IL) (in dB): the ratio of the total power at all output ports to the launched power (also sometimes called the excess loss) – this generally includes the coupling losses.
- The *Polarization-Dependent Loss* (or PDL) (in dB): peak-to-peak variations in the IL when the polarization varies – very critical for optical telecom PLCs.
- The *return loss* (in dB): the ratio of the power returned to the input port to the launched power.
- The *extinction ratio* (in dB): the ratio of the residual power in an extinguished polarization state to the transmitted power.

### 3.6.3.2 Injecting Light into and Extracting it from PLCs

In order to decrease the coupling losses (i.e. the IL), it is useful to couple light with an optical element, which can be a ball lens, a GRIN lens or a diffractive lens. One can also make use of the fact that the end tip of the fiber can be shaped into a multitude of forms that can implement lensing (and thus coupling) functions (fiber tapers, chiseled fibers etc.). Figure 3.16 shows such coupling techniques used in industry today. For a more in-depth analysis of GRIN-based fiber couplers, see also Chapter 4.

Apart from simple butt-coupling (Figure 3.15), pigtailling the fibers with half-pitch GRIN lenses, produces the lowest coupling losses (about 0.1–0.2 dB – see A in Figure 3.16).

When it comes to coupling a laser diode into a fiber, the mode size is very different, and coupling techniques include the ball lens (B1), diffractive lenses (B2) or the lensed fiber (B3 – chiseled tapered fiber end tip). Laser diode to fiber coupling is usually a difficult task owing to the size of the laser diode core (about 3  $\mu\text{m}$ ). Moreover, the laser diode waveguide core is not round or square, but rectangular, which produces strong astigmatism. Therefore, laser diode collimators are usually orthogonal cylindrical lenses instead of circular symmetric lenses. Collimating light from fibers is much easier, since the fiber core is round and symmetric.

A desirable feature in laser-to-fiber or fiber-to-detector coupling is the use of arrays; for example, laser diode arrays coupled into fiber arrays in V-grooves for DWDM applications or VCSEL arrays (easier to produce). Figure 3.14 shows two examples of such coupling element arrays, the first one being an array of digital diffractive microlenses (C1) and the second an array of GRIN lenses fabricated in a single planar substrate (for more insights on such GRIN substrates, see Chapters 4 and 11). Chapter 16 also shows an example of an array of 12 diffractive lenses used in a multichannel 10 Gb/s Ethernet network (coupling an array of 12 VCSELs into a 12-fiber bundle with an MC connector).

Note that when using 2D GRIN arrays, due to the fabrication processes used, the core-to-core spacing is usually pretty large (several millimeters – best suited for spacing edge-emitting laser diodes on multiple

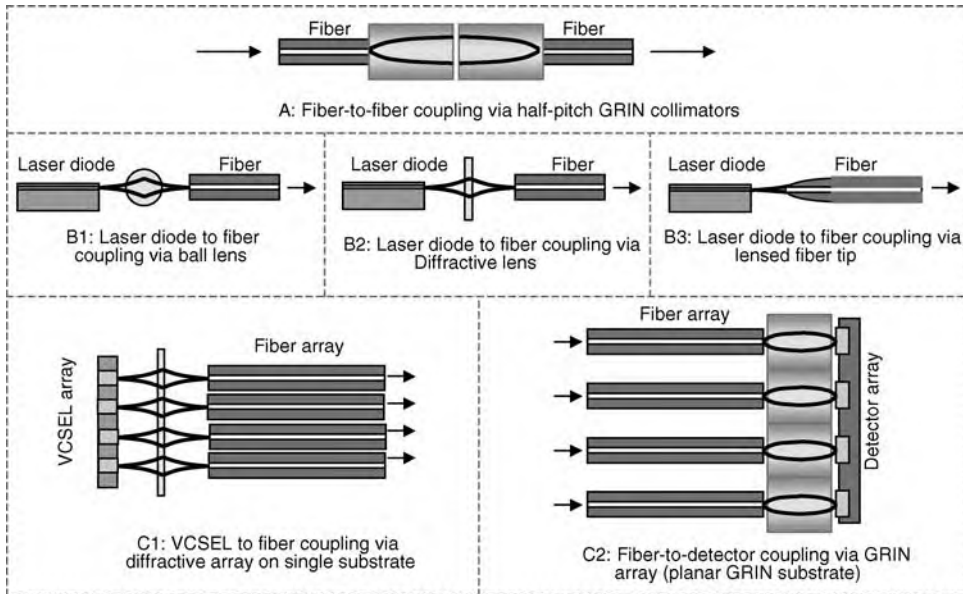


**Figure 3.16** Channel-based waveguide coupling techniques

substrates), whereas with arrays of diffractive lenses the spacing can be very small (tens of microns – best suited for integrated arrays of VCSEL lasers on a single surface-emitting substrate).

### 3.6.3.3 Mode-matching PLCs

The exact mode field diameter  $w$  of the fundamental mode size for a single mode fiber is a function of the cut-off frequency  $V$  (here, for the  $HE_{11}$  mode of a simple step-index SMF fiber, with radius  $a$ ):

$$w = 2a \cdot \left( 0.65 + \frac{1.619}{V^{3/2}} + \frac{2.879}{V^6} \right) \quad (3.14)$$

Usually, PLC waveguide cores have mode sizes or mode configurations that can be very different from those of fibers, therefore emphasizing the need for adequate mode converters. Moreover, multiple-core PLCs and fiber arrays (through silicon V-grooves, for example) usually have different core spacing, therefore emphasizing the need for core spacing adaptors.

Tapered mode-matching arrays are PLC inserts that can be inserted in between different PLCs and/or fiber arrays. Their sole aim is to match modes and to match fiber spacing, in and out. They have no other functionality. Such arrays of tapered waveguides can, for example, convert an array of  $250\ \mu\text{m}$  spaced silicon V-grooves into an irregularly spaced array of spots, which can be located at the focal plane of a free-space grating demultiplexer lens, an *échelette* waveguide grating or an AWG device (see the following sections). Figure 3.17 shows such mode-matching and spacing-matching PLCs.

The PLC inserts described in Figure 3.17 are usually low-index waveguides ( $\text{SiO}_2$ , BPSG, polymer, diffused graded-index waveguides etc.), since they are passive and do not need any active media. High-index PLCs (active devices) will be described in detail in Section 3.7.

Other mode-matching inserts can, for example, convert the narrow mode of a high-index ridge waveguide (e.g.  $4\ \mu\text{m}$  mode size) into the larger mode size of a standard SMF fiber (e.g.  $9\ \mu\text{m}$  mode size).

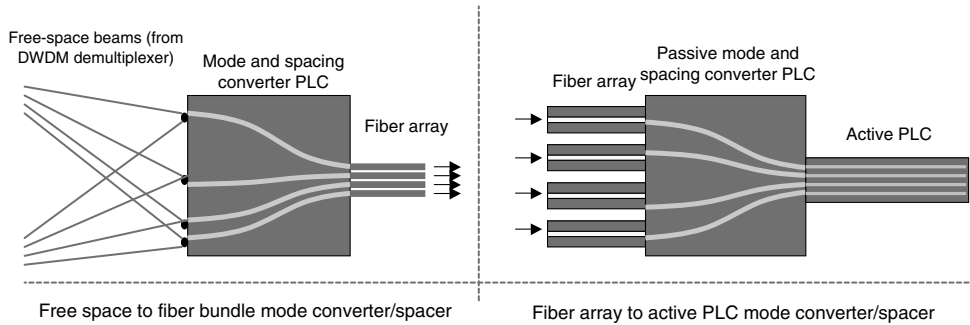


Figure 3.17 Mode-matching PLCs for waveguide arrays

*Note that Chapter 11 shows a simulation of the optimal coupling between a semiconductor laser and a tapered fiber end tip, without any additional element included, performed by a numeric propagation algorithm based on scalar diffraction theory (no ray tracing). Chapter 16 also shows an example of a Gaussian mode conversion into a doughnut mode for graded-index fiber coupling via a diffractive vortex lens.*

3.6.4 Directional Coupling and Y-junctions

Most PLCs (and fiber-based components) used in industry today are based on Y-junctions and directional couplers. Two channel waveguides (or more) are fabricated on a single substrate and separated by a fixed distance. A directional coupler works by coupling together the modes traveling in the same direction. Figure 3.18 shows the basis of a direction coupler.

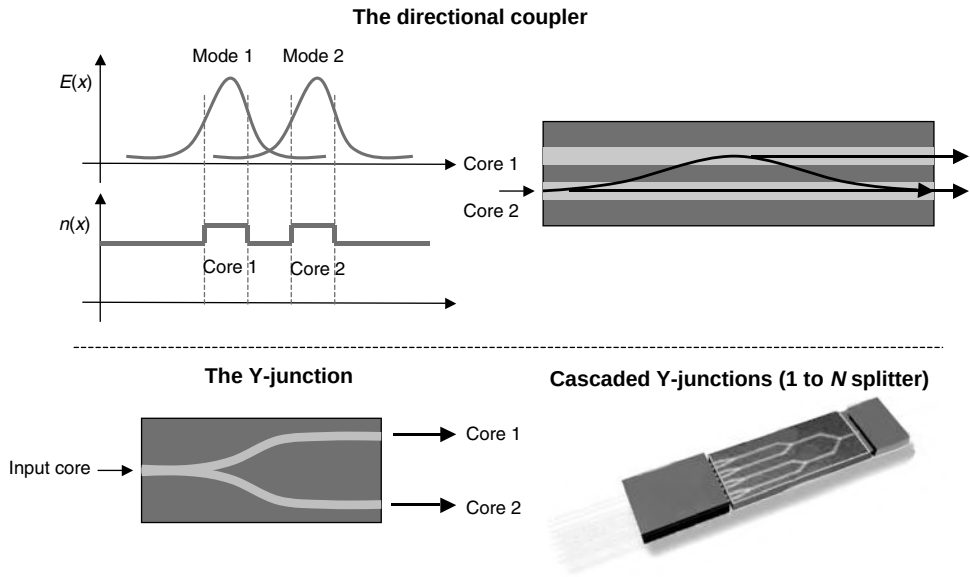


Figure 3.18 The directional coupler and the Y-junction

A Y-junction is basically a power splitter that couples light into one or more cores from a single core. Y-junctions can be used in combination with directional couplers to implement complex functionalities (see Section 3.7). The amount of splitting occurring in a Y-junction is a function of the geometry of each arm.

The most straightforward application example is the power splitter (or 3 dB coupler). However, the directional coupler is the architectural basis for most of the devices that we will describe in Section 3.7. They can be passive (based on low-index materials) or active (based on high-index materials or a combination of low- and high-index materials). Direction couplers can be integrated with numerous channel waveguides – up to several hundred.

Directional coupling from one core to the other is a function of the core spacing, the refractive indices, the size of the cores (the cores can be different) and the lengths of the guides. Note that similar effects can be triggered in multicore optical fibers (see Section 3.1), where the cores can be located in a 2D array, whereas in PLCs the cores are mostly located in a 1D configuration (other than in 3D stacked PLCs).

In order to change the coupling ratio, one can change the wavelength, but more commonly one could change the local index of the core or cladding; for example, by using active high-index materials (see Section 3.7).

We have described in the previous sections the basics of PLC architecture and operation, including slab and channel-based PLCs, directional coupling and mode matching. In Section 3.7, we will derive some interesting functionalities based on these building blocks. These optical functionalities are used today in industry, and especially in optical telecom applications.

## 3.7 Functionality Integration

In previous sections, the basics of optical waveguides, in fiber waveguide form, slab waveguide PLC form or channel waveguide PLC form, have been described. Also discussed were the challenges that coupling between fibers and PLCs offer to the optical engineer, especially in telecom applications. This section summarizes the main applications that use the unique characteristics of PLCs and other digital waveguide devices, especially for the telecom application pool. This section begins with the semiconductor laser waveguide, then shows how such lasers can be modulated via high-index PLCs, then how different wavelength lasers can be combined or split (multiplexed or demultiplexed), and finally how complex optical functionalities can be implemented further down the road in more complex lightwave PLCs.

As most of these applications are based on optical telecoms, it is worth briefly summarizing a typical DWDM network architecture (see Figure 3.19).

Most of the PLC devices that are presented in the next section, as well as the free-space refractive elements (Chapter 4) and the digital diffractive elements presented in Chapters 5 and 6, as well as some of the elements presented in Chapter 16, can find a fit within the previous DWDM architecture.

### 3.7.1 Laser Waveguide Devices

Although the operational details of the semiconductor laser will not be discussed, its typical waveguide structure will be presented (the edge-emitting buried heterostructure). A typical semiconductor laser has a rectangular waveguide cross-section, with a height of about  $1.5\text{ }\mu\text{m}$  and an aspect ratio of about 10:1 (see Figure 3.20). The end facet acts as an emitter, and therefore produces a highly diverging wavefront (a  $10\text{--}15^\circ$  half-angle in the slow axis and a  $30\text{--}40^\circ$  half-angle in the fast axis). Such a wavefront is thus highly astigmatic, since the end facets have very different sizes in the  $x$  and  $y$  directions. While the near-field beam has approximately the same aspect ratio as the waveguide (though elliptical), its far field (angular spectrum) has the opposite elliptical aspect ratio.

In many cases, lasers with lower divergence angles are desired, as is the ability to include several lasers in a linear or even 2D array configuration. While edge-emitting lasers cannot address these issues, Vertical Cavity Surface-Emitting Lasers (VCSELs) can, and thus are used increasingly in applications where arrays are required but wavelength stability is not such an issue (e.g. multichannel 10 Gb/s Ethernet networks @ 850 nm).

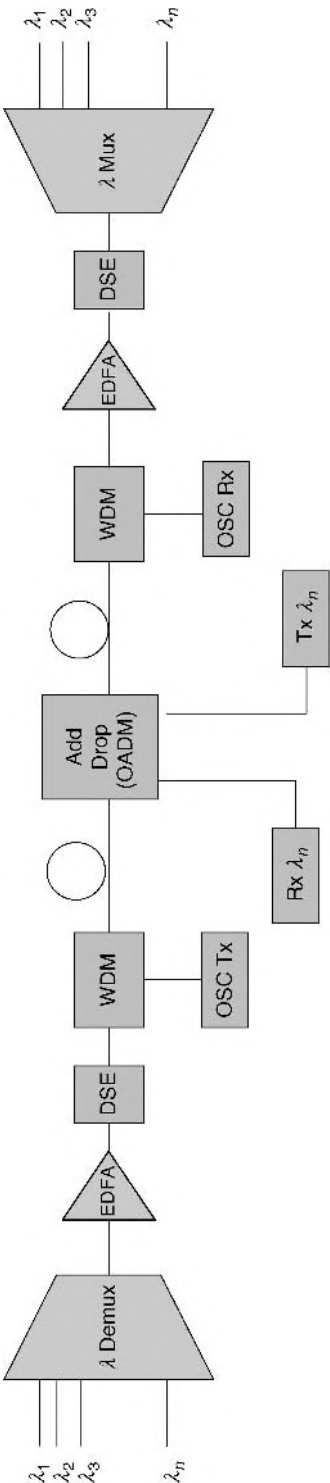


Figure 3.19 A typical DWDM optical network architecture

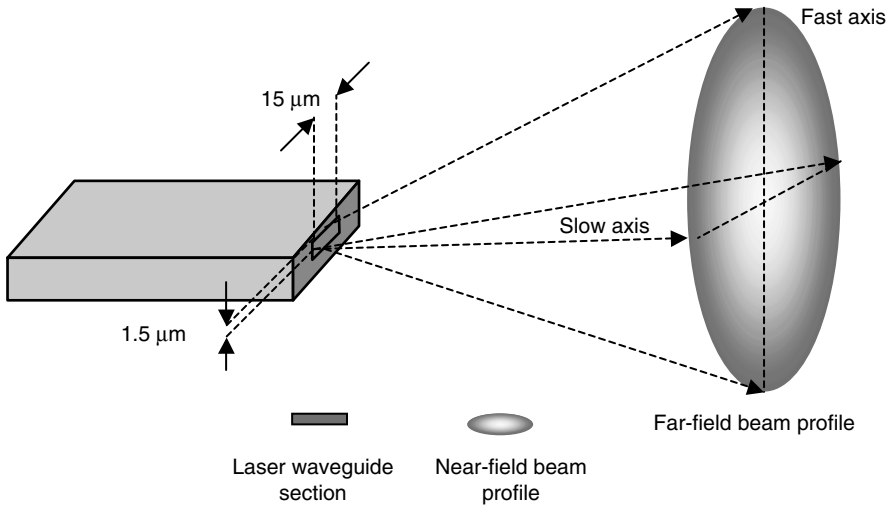


**Figure 3.20** A typical edge-emitting buried heterostructure semiconductor laser architecture

### 3.7.1.1 DBR and DFB Lasers

A typical semiconductor laser as depicted in Fig. 3.20 is basically a Fabry–Perot cavity device. In its simplest form, such a laser produces several spectral lines, and is therefore not optimal for telecom applications where a single spectral line is desired (especially for DWDM applications).

Two solutions have been proposed in order to reduce the spectral bandwidth of such lasers: the Distributed Bragg Reflector (DBR) laser and the Distributed FeedBack (DFB) laser.

In a DBR laser, one or both of the cleaved end mirrors of the laser are replaced by a corrugated reflection grating. Such a laser would provide perfect reflection down the active layer and back only for a perfect Bragg condition involving the wavelength and the end facet grating period and effective index (Equation (3.15)).

$$2 \cdot n_{\text{eff},n} \cdot \Lambda = \lambda_n \quad (3.15)$$

where  $\lambda_n$  is the lasing wavelength,  $\Lambda$  is the grating period and  $n_{\text{eff},0}$  is the effective index in the waveguide for this specific wavelength. When the Bragg condition is not satisfied, the laser does not amplify the corresponding wavelength and this reduces the spectral lines almost to zero.

In a DFB laser, the grating is etched within the active region, and not at each end facet as in a DBR laser (see Figure 3.21). Usually, a quarter-wave phase shift is introduced in the grating periodicity (the quarter-wave-shifted DFB laser).

The Bragg reflections are distributed along the entire active region – hence the name ‘Distributed’. Such DFB lasers produce spectral lines that are similar to those of DBR lasers.

### 3.7.1.2 Multiple-laser Substrates

For many applications, it is desirable to have arrays of lasers that can be packaged with arrays of optics in order to produce planar Opto-Electronic (OE) devices. Such lasers might have similar wavelengths (as in VCSEL laser arrays for 10 Gb/s Ethernet applications) or specific different lasers (as in laser diodes for DWDM applications).



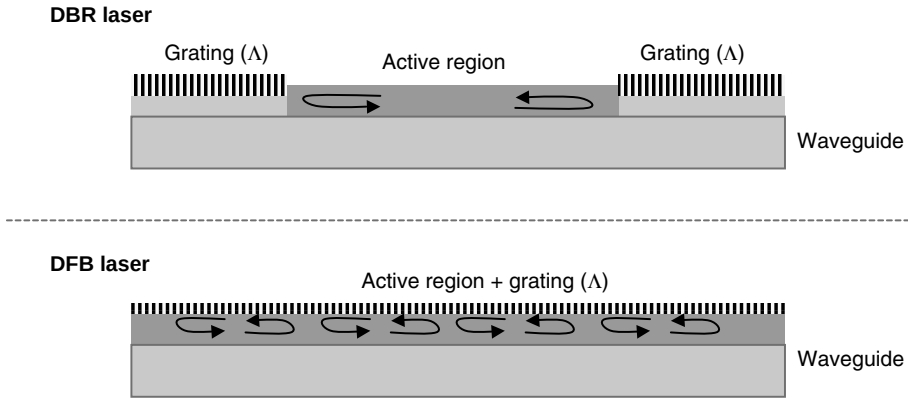


**Figure 3.21** Waveguide structure and Bragg gratings in DBR and DFB lasers

The obvious reason would be to increase the total output power by using multiple lasers instead of single lasers. However, due to the third principle of entropy, it is impossible to combine several laser beams with identical wavelengths and polarizations into a single one with minimal power loss. Such laser beam combiners would have a loss of  $1/N$  for  $N$  combined channels.

The first two examples derived here are laser diode arrays and VCSEL arrays, in which the similar laser beams cannot be combined in a single guide. However, the next two (extra- and intra-cavity laser devices) can be combined, due to the fact that their wavelengths are slightly different; for example, tuned on the DWDM C band.

### **Array Lasers**

Array lasers are simple multichannel ridge waveguides that produce unidirectional or even bidirectional 1D edge-emitting lasers. The powers of such edge-emitting laser strips cannot be combined.

### **VCSEL Laser Arrays**

Most applications that require optical interconnections need two-dimensional arrays of laser sources. This is very difficult when dealing with edge-emitting strips, as seen in the previous section. Such applications can include chip-to-chip interconnections or backplane interconnections in highly parallel optoelectronic (OE) Multi-Chip-Modules (MCMs). More insight on these applications can be found in Chapter 16.

Here, the laser cavity is in the natural dimension for semiconductor processing, which is the normal to substrate direction, and the cavity is also circular. Figure 3.22 shows the internal structure of a VCSEL laser.

The advantages of VCSELs over standard edge-emitting lasers include:

- a vertical cavity, which can be replicated in the  $x$  and  $y$  directions into arrays by lithography;
- the possibility of wafer-scale laser testing (no need to dice the wafer in order to test);
- beam divergence that is lower than the edge emission;
- a circular – rather than elliptical – beam (easier coupling into an optical fiber); and
- lower power consumption due to the thin active layer.

However, even with all these advantages, VCSELs are not (yet) used in DWDM telecom applications today, since they lack wavelength stability due to wafer thickness variations (whereas wavelength stability can be achieved by DBR and DFB lasers). The wavelength range is also an issue (VCSELs are common in the 850 nm regime for 10 Gb/s optical Ethernet, but seldom in the 1.5  $\mu\text{m}$  and visible regions). However,

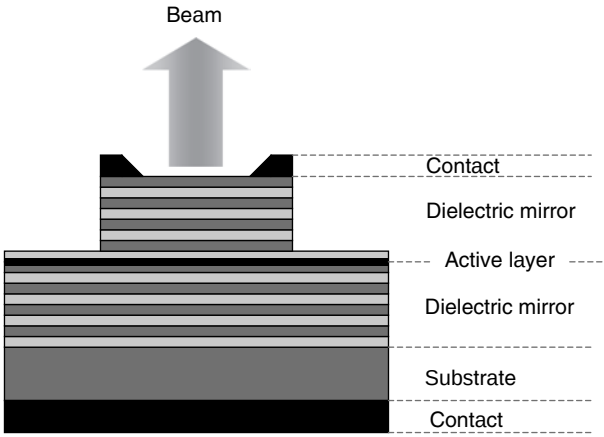


**Figure 3.22** The internal structure of a VCSEL laser

recently, VCSEL manufacturers (Novalux) have demonstrated high-power arrays in the RGB colors for laser display applications (see Chapter 16).

Today, VCSEL laser arrays are fabricated by one of three techniques:

- etched air-post VCSELs;
- ion-implanted VCSELs; or
- selectively oxidized VCSELs (the most commonly used).

**Extra- and Intra-cavity Combined Lasers**

For telecom applications where several different wavelengths are used in a single fiber (DWDM), it is desirable to have a laser module that can directly produce such  $n$  wavelengths in a single output waveguide. This is possible since the wavelengths are different and well defined, and two integration architectures have been proposed for the purpose: the extra-cavity combined laser module and the intra-cavity combined laser module. These two architectures are depicted in Figure 3.23.

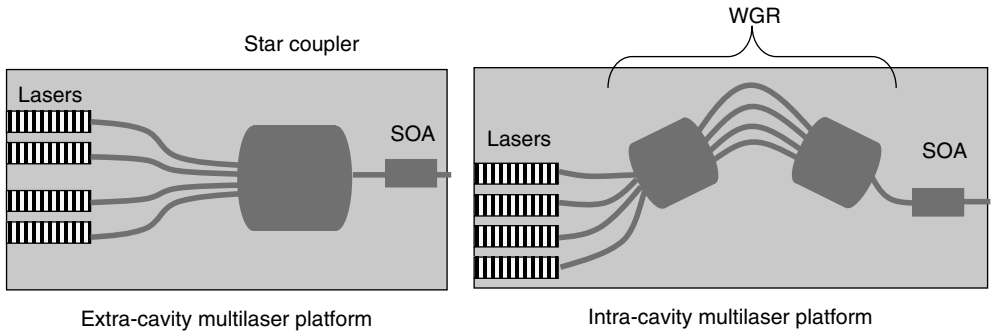


**Figure 3.23** The extra-cavity and intra-cavity multi-laser modules

An extra-cavity multi-laser module uses an array of DFB or DBR lasers linked to a power coupler, which can be an integrated slab star coupler (see the following sections). However, as seen previously, such power couplers have a  $1/N$  loss for  $N$  channels.

An intra-cavity multi-laser module uses an array of DFB or DBR lasers linked this time to a frequency coupler (a shared angular dispersive element). Such a device (e.g. a AWG-based waveguide grating router) is described in Section 3.7.3. The power losses of such an intra-cavity multi-laser module are much lower than for an extra-cavity multi-laser module.

For both architectures, an additional amplifier (a Semiconductor Optical Amplifier, or SOA) can be integrated in the active PLC before out-coupling to the fiber.

### 3.7.2 Phase Modulator Devices

Phase modulators are very important building blocks in many PLC devices. They are based on electrically controllable materials that are inserted into a conventional channel-based single or multiple waveguide structure. Such materials change their effective refractive index as a function of the applied voltage. Among the most common electro-optical materials used in PLCs are  $\text{Ti:LiNbO}_3$  (which has the benefit of being a very simple process) or the more common GaAs/GaAlAs or InP/InGaAsP. Based on the properties of the phase modulation material, the voltage has to be perpendicular or transverse. Such field configurations are easily produced by placing the electrodes right on top on the phase-shifting material (field normal to the substrate) or alongside the active material (field parallel to the substrate).

The PLC architecture is best suited to using these materials, since the waveguide core regions are very small and thus the voltage required to perform the phase change remains very low. If they were to be used with macroscopic optical elements, the associated voltage would be extremely high.

#### 3.7.2.1 Laser Modulators

A simple phase modulator – the Mach–Zehnder interferometer modulator architecture – is depicted in Figure 3.24.

Such a modulator is built upon two Y-junctions and a phase-shifting material in the path of a single core in the interferometer arms. Therefore, by varying the associated voltage, one can extinguish the signal completely or strengthen the output signal to its maximum value.

There are other modulation operations based on the same phase-shifting effect, such as the frequency-shifting modulation, which is especially useful in high-frequency operations. The voltage is controlled in a precise manner at high frequency; for example, along a sawtooth signal. In the linear regions of the sawtooth voltage drive, a more or less linear phase shift is applied to the beam, and the same pattern is applied to the signal. Sinusoidal voltage drives can also be used for specific modulation tasks.

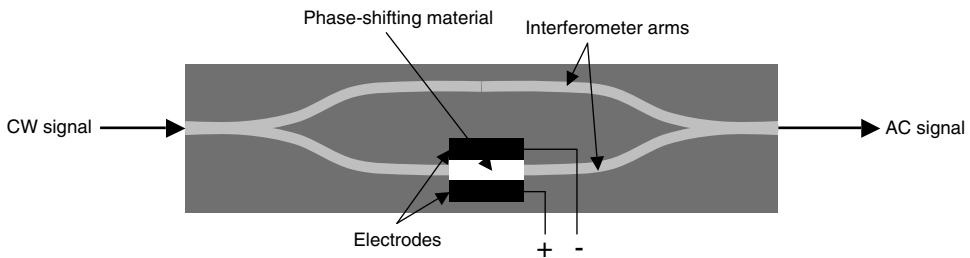


**Figure 3.24** The Mach–Zehnder phase modulator PLC

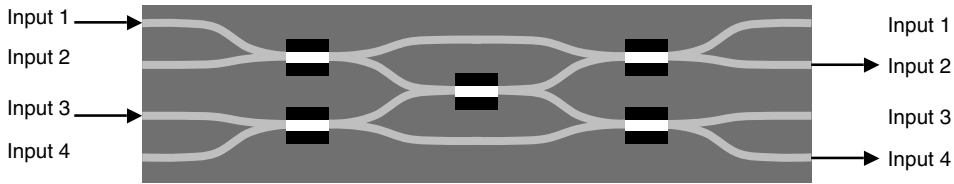


**Figure 3.25** An example of a  $4 \times 4$  optical switch in a PLC architecture

Usually, lasers are not modulated directly at high speed. A continuous laser signal can be produced by the modules described in Section 7.2. A modulator is coupled to the laser to produce the actual signal that will travel along the line. Modulating a laser directly decreases its MTBF, and increases noise and other nonlinear effects that create parasitic signals on the line.

### 3.7.2.2 Variable Optical Attenuators (VOAs)

In a similar way, Variable Optical Attenuators (VOAs) can be integrated in PLCs on a single core in a region of which a phase-shifting material would modulate the local index and therefore modulate the strength of the mode guidance (and thus attenuate the signal).

A VOA can also be implemented as a directional coupler device or Mach–Zehnder modulator (see the previous section), while applying a voltage that would only attenuate the signal, and not cancel it completely.

### 3.7.2.3 Optical Switches

Optical switches are core elements in DWDM optical systems. Complex optical switches can have  $N$  inputs and  $M$  output channels. Figure 3.25 shows a PLC architecture where the basic building blocks, as seen in the previous sections, are used to produce such an optical switch (Y-junctions, directional couplers – or Mach–Zehnder  $2 \times 2$  switches – and phase modulators).

Such integrated optical waveguide switches are competing in the telecom realm with free-space MEMS optical switches, which consist of 1D or 2D arrays of micromirrors that can tilt to a single angle (binary switches, or 2D MEMS switches – similar to DLP microdisplays) or arrays of micromirrors that can tilt to a wide variety of different angles (3D optical MEMS switches). Although such switches operate in free space, the incoming signals are carried by optical fiber arrays, and are often collimated and coupled back into the output optical fibers by arrays of GRIN lenses (see Chapter 4).

## 3.7.3 Integrated Waveguide Dispersion Devices

This section discusses some of the integrated waveguide dispersion PLCs that have been developed in industry, especially for telecom DWDM Mux/Demux applications.

These are, among others, the Mux/Demux devices, grating based and thin film filter [10] (TFF – dichroic based). GRIN lenses combined with TTF layers are described in Chapter 4, and reflective grating based Mux/Demuxes in Chapter 5.

### 3.7.3.1 Arrayed Waveguide Gratings (AWGs)

The AWG, or Waveguide Grating Router (WGR), is an important building block for many Mux/Demux-based PLCs in DWDM devices. One integration application has already been discussed in the previous section (the intra-cavity multi-laser PLC; see Figure 3.21).

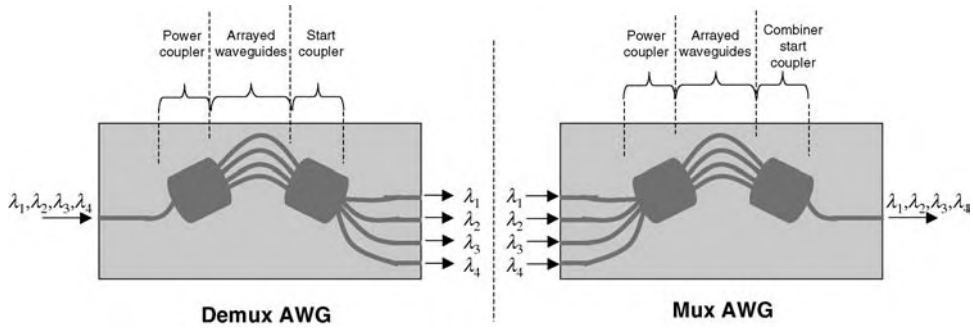


**Figure 3.26** AWGs (or WGRs) in Demux and Mux configurations

Usually, an AWG has one input and many outputs. The input carries many wavelengths, and the outputs carry single wavelengths (i.e. a wavelength demultiplexer – Demux). The AWG can also work in reverse mode, as a wavelength multiplexer (Mux). Figure 3.26 shows AWGs working in the Mux and Demux modes.

The two basic constituent elements in an AWG are star couplers and arrayed waveguides. Arrayed waveguides each have slightly different lengths in between star couplers, in order to imprint on the traveling waves a phase shift that looks like a phase shift introduced by a grating or a prism. Star couplers are basically semifree-space cavities in which diffraction arises as it would in a 3D free-space configuration. The input and end facets of star couplers are slightly curved in order to mimic a lens profile. The star couplers therefore have three functionalities: they out-couple light and couple back light into 1 or  $N$  waveguides, they let the incoming beam(s) diffract freely in free space and they imprint a 1D lensing phase function on the incoming beams.

In a Demux configuration, the incoming signals are dispatched equally on  $N$  channels through a star coupler. The arrayed waveguides imprint on the  $N$  beams a quasi-continuous linear phase shift. All wavelengths are still present in all waveguides. The second star coupler lets these  $N$  beams diffract freely, just as a grating would diffract them, and spectrally disperses the beam constituted by the  $N$  beamlets. The resulting spectral focal plane lies on the exit facet of the star coupler, where  $N$  channels collect the  $N$  individual wavelengths. The  $N$  wavelengths are then out-coupled to  $N$  fibers.

### 3.7.3.2 Integrated Echelette Grating

An integrated échelette grating is very similar to an AWG or a WGR, in the sense that it implements the same functionality in a planar PLC, with free-space regions where natural diffraction occurs. However, in an integrated échelette grating, the grating functionality is performed by a real reflective grating, whereas in the AWG element this grating effect is performed by the arrayed waveguides. Figure 3.27 shows such an integrated échelette grating.

The lensing effect is performed by introducing a curvature on the échelette grating. As the grating is reflective here (in most cases), the input and output ports are on the same side of the PLC, which reduces the coupling and packaging costs.

It is interesting to note that the échelette grating is one of the only elements where the edges of the grating are actually patterned rather than comprising the lateral structure of the grating, as is done for conventional free-space gratings. It is therefore very easy to imprint any curvature onto such a grating, an effect that is very difficult to imprint onto a regular free-space grating (i.e. a grating on a lens).

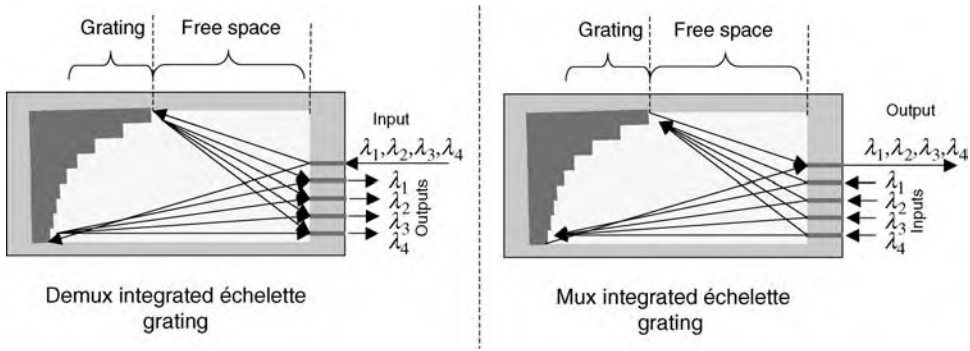


**Figure 3.27** An integrated échelette grating in Mux and Demux configurations

Both AWG and échelette gratings are used extensively in DWDM Demux and Mux applications today.

### 3.7.4 Complex Functionality Integration

Based on the integrated waveguide based Mux and Demux architectures presented in the previous section, numerous complex optical functionality PLCs can be developed for telecom applications and datacom applications. Some of them are described in this section. The functionalities can be extended to more complex applications: however, when cascading elements such as AWGs, one has to be careful about the total power budget of the PLC, since the insertion losses, polarization-dependent losses and return losses are added together.

#### 3.7.4.1 The $N \times N$ Wavelength Router

An AWG-based  $N \times N$  wavelength router PLC module is shown in Figure 3.28. Such a wavelength router routes a set of  $N$  channels into a set of  $N$  output channels. It is a passive device, which means that the routing architecture is set in stone once the router is fabricated.

#### 3.7.4.2 The Add-Drop Module

An AWG-based add-drop PLC module is shown in Figure 3.29. Such an add-drop can add or drop a channel through the use of a set of  $2 \times 2$  switches (for the description of optical switches, see Section 3.7.2.3).

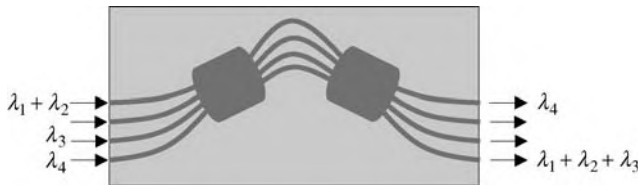


**Figure 3.28** An AWG-based wavelength router PLC module

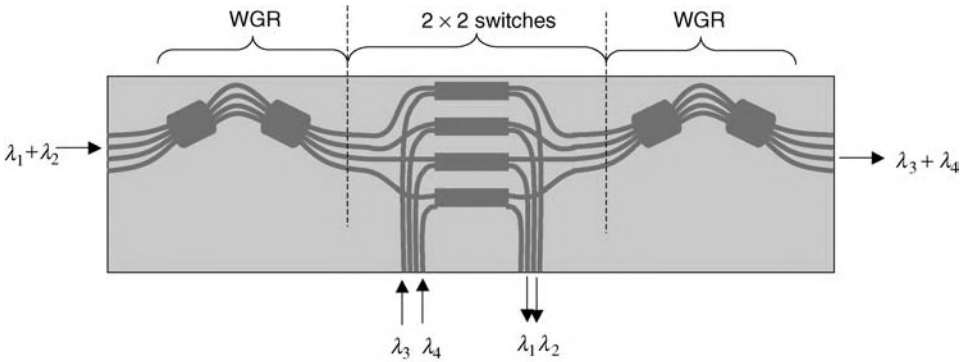


**Figure 3.29** An AWG-based dynamical add-drop PLC module

**3.7.4.3 The Cross-connect Module**

More complex PLCs can be designed, such as the AWG-based cross-connect module PLC shown in Figure 3.30.

Other complex multichannel multi-stage AWG-based PLC-based devices include interleavers, dispersion compensators, more complex cross-connect modules and so on.

However, one has to be very careful about the insertion loss budget that such a complex PLC will create. Typical AWGs have 4–6 dB insertion losses, and these losses add up considerably as one proceeds to cascade them, with optical switches in between.

This chapter has reviewed the fundamentals of digital integrated waveguide technology – or Planar Lightwave Circuits (PLCs) – and has described some of their functional integrations, especially for the telecom industry. In the next few chapters, the focus will be on free-space digital optics.

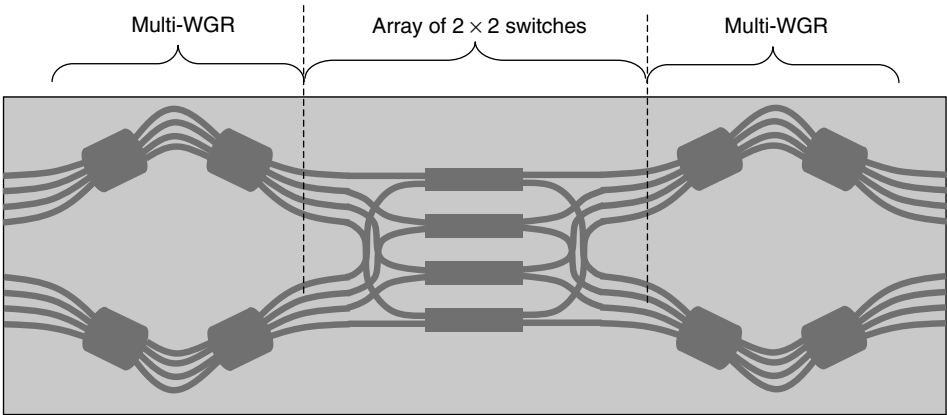


**Figure 3.30** An AWG-based cross-connect PLC module

## References

- [1] S.E. Miller, 'Integrated optics: an introduction', *Bell Systems Technical Journal*, **48** (1969), 2059–2068.
- [2] T. Tamir, '*Integrated Optics*', Springer-Verlag, Berlin, 1982.
- [3] H.-G. Unger, '*Planar Optical Waveguides and Fibers*', The Clarendon Press, Oxford, 1977.
- [4] R.G. Hunsperger, '*Integrated Optics*', Springer-Verlag, Berlin, 1995.
- [5] J.M. Senior, '*Optical Fiber Communication: Principles and Practice*', 2nd edn, Prentice Hall, Cambridge, 1992.
- [6] Ch. Doerr, 'Why waveguides?' In '*Proceedings of OFC 2001*', Technical Digest Series, March 2001, Tutorial Sessions.
- [7] H. Nishihara, M. Haruna and T. Suhara, '*Optical Integrated Circuits*', McGraw-Hill, New York, 1987.
- [8] A.E. Willner, 'Optical communications', in '*Handbook of Photonics*', M.C. Gupta (ed.), CRC Press, Boca Raton, Florida, 1997.
- [9] C. Dragonne, C.H. Henry, I.P. Kaminov and R.C. Kister, 'Efficient multichannel integrated optics start coupler on silicon', *IEEE Photonics Technical Letters*, **1**, 1989, 241–243.
- [10] C.K. Carniglia, 'Design of thin-film interference filters for telecommunications applications', Technical Digest OFC 2001, paper WF4.



# 4

## Refractive Micro-optics

Today, the realm of micro-optics comprises quite an impressive range of optical elements, which will be described in the next six chapters.

This chapter focuses on refractive micro-optics, a sub-field of micro-optics [1–3] that considers refraction/reflection and/or the waveguide phenomenon rather than diffraction/diffusion. Therefore, we consider here structures that are much larger than the wavelength, from about 20 times the operating wavelength up to 1000 times that wavelength. Figure 4.1 shows that refractive micro-optics are actually at the interface between the realms of macro-optics and micro-optics.

Refractive and reflective micro-optics have found many applications in today's industry, and both fields are growing at a rapid pace.

While traditional macroscopic optics are fabricated by conventional polishing and grinding techniques, micro-optics are fabricated by lithographic techniques [4–6]. The physical aspects of such optical elements as well as the three major fabrication techniques are shown in Figure 4.2.

Salt diffusion for GRIN lenses and resist reflow process followed by dry proportional etching [7] are simple tools that provide most of the requirements for refractive micro-optics. Gray-scale lithography is a more complex lithographic technique, which can provide arbitrary surface-relief elements in a single lithography step. For more insight into such fabrication techniques and related processes, including mass replication [8], see the fabrication chapters (Chapters 12–14).

This chapter presents three physical implementations of refractive micro-optics:

- The first is dedicated to GRaded INdex (GRIN) elements and lenses, which are widely used as telecom fiber collimator pigtails (guided-wave to free-space).
- The second is dedicated to individual surface-relief refractive micro-optics, both for imaging and nonimaging applications, with an emphasis on CMOS sensor lenses for integrated imaging applications.
- The third, and also the most used today, is dedicated to various arrays of micro-optics, beginning with the standard microlens arrays, and from there to prism arrays and other retro-reflector arrays.

### 4.1 Micro-optics in Nature

Micro-refractive optical elements have been around for a long time, Mother Nature having developed such elements within the animal kingdom mainly as compound eyes, or arrays of microlenses, combining surface-relief profiles and index modulations. Arrays of micro-refractives will be reviewed later in this chapter. Figure 4.3 shows some of the compound eyes found in animals today.

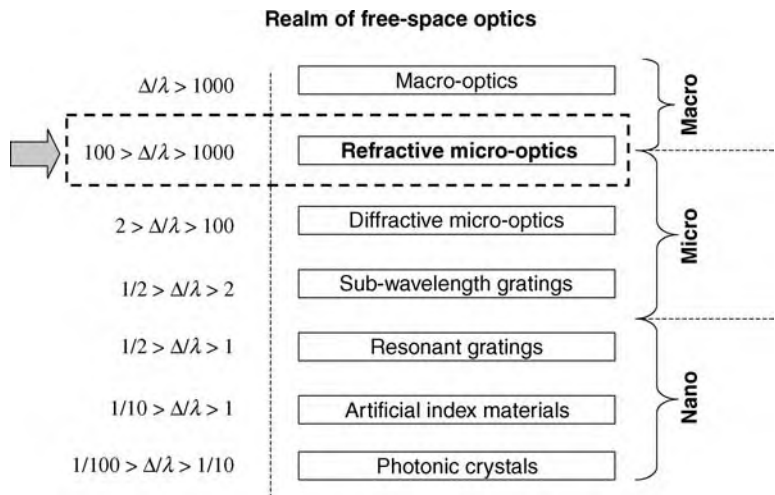


**Figure 4.1** Refractive micro-optics

A compound eye consists of thousands of individual photoreception units. The image perceived through such a compound eye is a combination of inputs from the numerous ommatidia (individual ‘eye units’), which are located on a convex surface and thus point in slightly different directions. Compared with simple eyes, compound eyes have a very large field of view, and can detect fast movement and, in some cases, the polarization of light. Because the individual lenses are so small, the effects of diffraction impose a limit on the possible resolution that can be obtained. This can only be countered by increasing the lens size and f-number – to see with a resolution comparable to that of our simple eyes, us humans would require compound eyes each of which would reach the size of our heads.

Compound eyes fall into two groups: apposition eyes, which form multiple inverted images, and superposition eyes, which form a single erect image. The design and fabrication of similar artificial apposition or superposition microlens arrays will be discussed in this chapter.

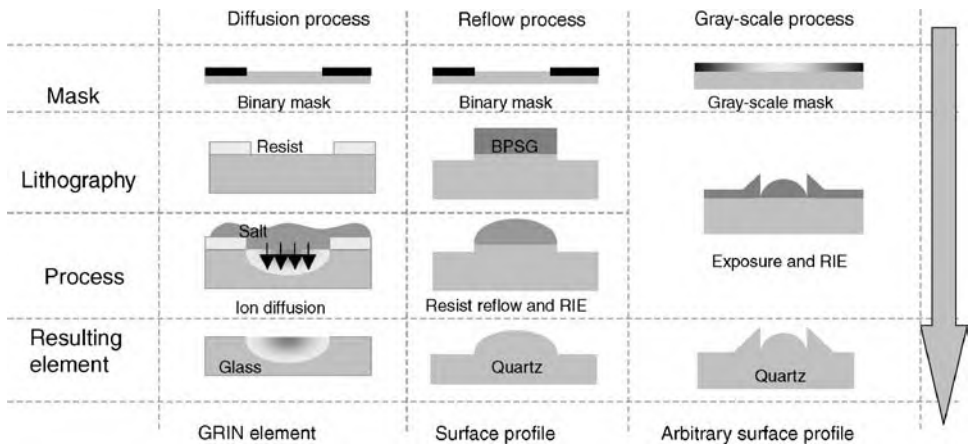


**Figure 4.2** Physical aspects and fabrication techniques for three groups of refractive micro-optics

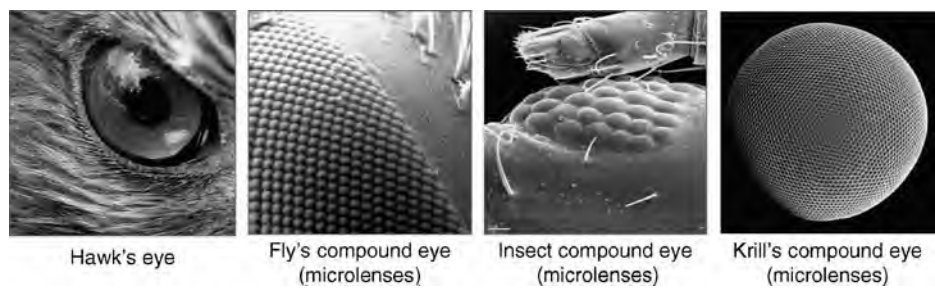


Figure 4.3 Single and compound eye lenses

4.2 GRIN Lenses

At first sight, GRaded INdex (GRIN) elements and lenses do not look like lenses, since most of them do not have any surface profile at which interface refraction or reflection could occur [9]. They look like glass rods or planar substrates. In some cases, however, a surface-relief profile can be added to the GRIN effect to add functionality.

In GRIN elements, the refraction process does not happen at the air/material interface as in standard lenses but, rather, continuously in the material itself, as the index of refraction varies in the inhomogeneous material.

They are therefore very similar to graded-index optical fibers and multilayer Anti-Reflection (AR) surfaces (although these are also based on multiple interferences – as in the Fabry–Perot effect). Figure 4.4 shows the continuous refractive effects found in all three elements (GRIN lenses, graded-index fibers and multilayer AR coatings).

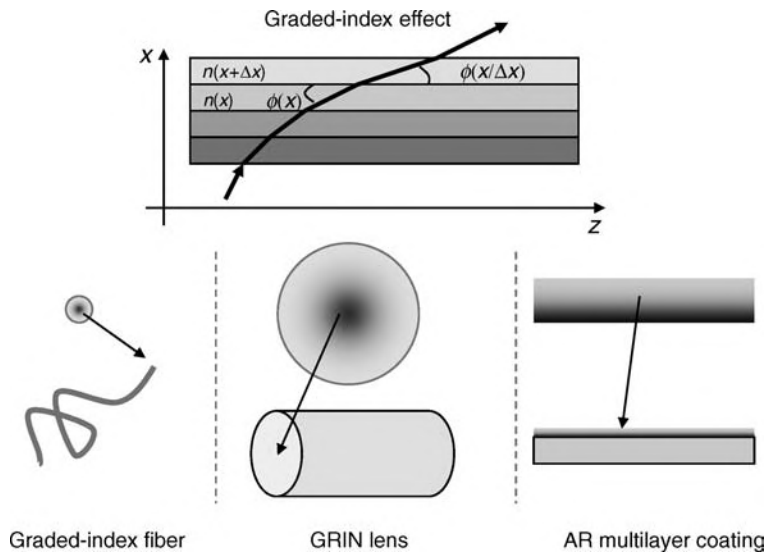


Figure 4.4 Graded-index elements used in industry

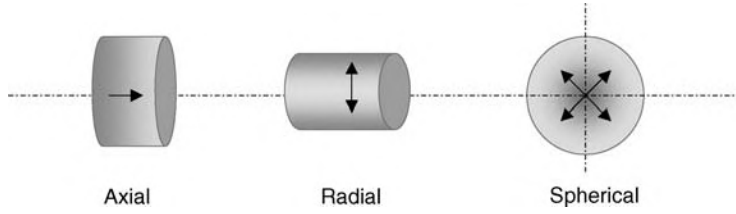


**Figure 4.5** Axial, radial and spherical GRIN lenses

The ray trajectory through GRIN media can thus be described as follows:

$$\frac{d}{ds} \left[ n(\vec{r}) \cdot \frac{d}{ds} (\vec{r}) \right] = \nabla n(\vec{r}) \quad \text{where} \quad \vec{r} = x\vec{i} + y\vec{j} + z\vec{k} \quad (4.1)$$

where  $n(r)$  is the refractive index distribution and  $ds$  is the differential element of the path length along the ray.

There are mainly three different type of GRIN lenses that have been developed in industry [10, 11]: axial, radial and spherical GRIN lenses (see Figure 4.5).

#### 4.2.1 Axial GRINs

In an axial GRIN, the refractive index varies continuously along the optical axis, with constant index planes orthogonal to that direction. The index distribution of an axial GRIN can thus be expressed as

$$\frac{d}{ds} \left( n \cdot \frac{d}{ds} (x) \right) = 0 \quad \text{and} \quad \frac{d}{ds} \left( n \cdot \frac{d}{ds} (y) \right) = 0 \quad (4.2)$$

Such a radial GRIN has practically no spherical aberration [12], whereas a continuous media lens with same dimension could case serious spherical aberrations. It is therefore possible to reduce the spherical aberrations of standard spherical refractive lenses by incorporating a radial index modulation rather than changing its surface profile into an aspheric profile. Such elements are widely used in the telecom industry today.

#### 4.2.2 Radial GRINs

Radial or cylindrical GRIN elements have a radial modulation of their refractive index:

$$n(r) = n_0 \left( 1 - \frac{1}{2} C_0 r^2 \right) \quad (4.3)$$

where  $n_0$  is the index at the center of the lens and  $C_0$  is a positive constant. These elements usually look like rods. When Equation (4.3) is substituted in Equation (4.1), one obtains

$$\frac{d^2}{dz^2} (r) = \frac{1}{n(r)} \frac{d}{dz} (r) = -Cr \Rightarrow r = r_0 \cos(\sqrt{C_0} \cdot z) + \frac{\sin(\sqrt{C_0} \cdot z)}{\sqrt{C_0}} \left( \frac{d}{dz} (r) \right)_{z=0} \quad (4.4)$$

The rays in axial GRIN lenses thus have sinusoidal directions as they propagate in the medium along the optical axis. The period  $L$  of the sinusoidal ray modulation within the radial GRIN lens (see also Figure 4.6) is given by

$$L = \frac{2\pi}{\sqrt{C_0}} \quad (4.5)$$

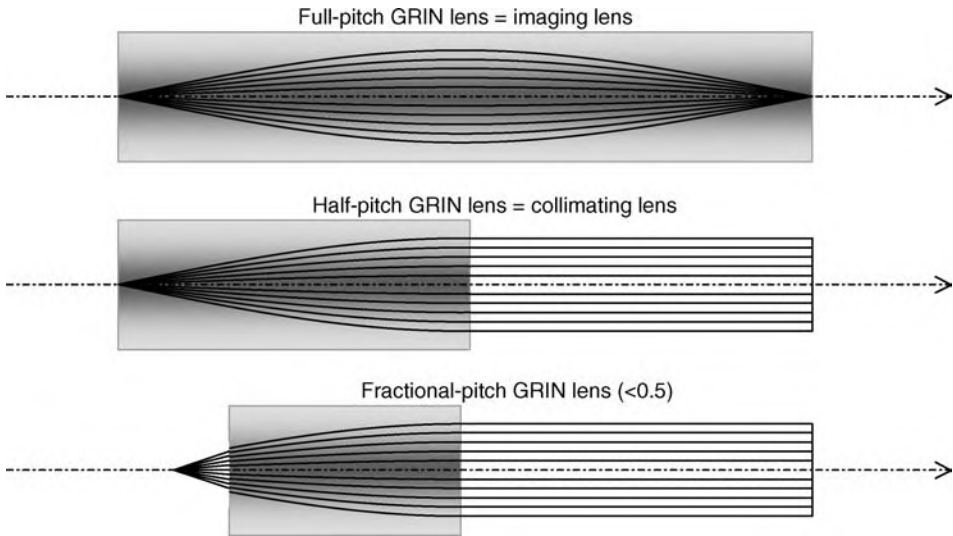


**Figure 4.6** Some of the GRIN lens configuration used in industry

For length  $L/2$  exactly, a radial GRIN lens will produce an inverted image with a magnification factor of unity, and then the initial image will appear again for a length  $L$ .

The telecom industry is a vast user of radial GRIN lenses, pigtailed to end tips of fibers (the half-pitch GRIN; see Figure 4.6) in order to process beams in free space.

The fiber tip is fused to the axis of the half-pitch GRIN lens and produces a quasi-collimated beam. This beam can then be processed (filtered) or redirected by micromirrors into another fiber (2D or 3D optical switching – see Chapter 16). Figure 4.7 shows a filter application in free space between two half-pitch GRIN lenses. Gratings in the free-space region between GRIN lenses can be used to build power splitters; cascaded free-space dichroic filters can be used to build a spectral demultiplexer; and so on (see also Chapter 16).

When moving a half-pitch GRIN away from the fiber end, one can add flexibility to fiber–lens applications. Figure 4.8 shows such a configuration, where a cylindrical glass spacer is used to create a slightly focused output beam in free space for coupling into a lower NA waveguide.

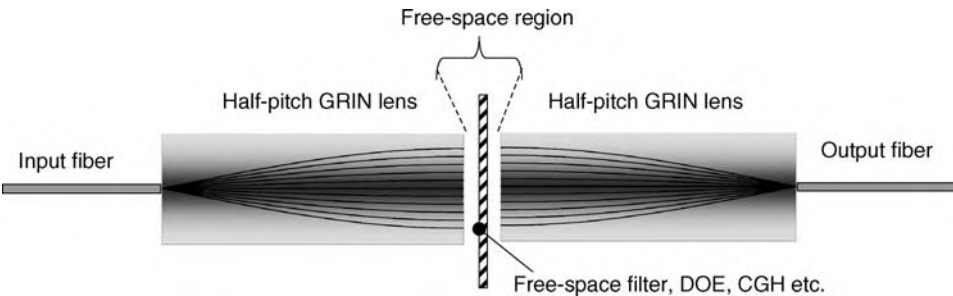


**Figure 4.7** Free-space processing between two half-pitch GRINs

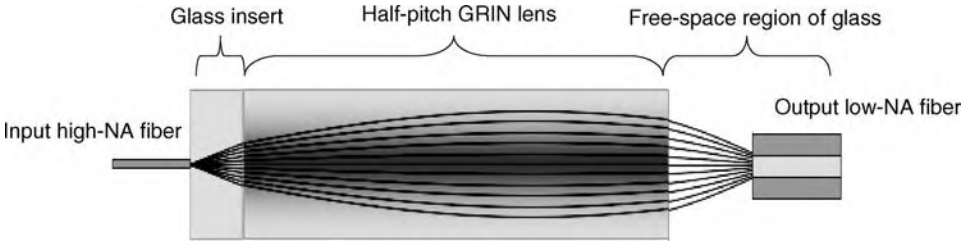


**Figure 4.8** The use of neutral rods in between fibers and GRINs

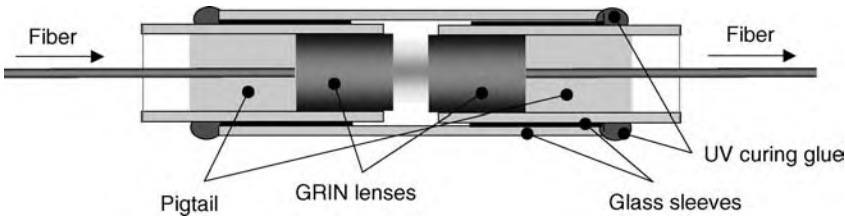


**Figure 4.9** A packaged fiber-to-fiber coupler based on GRIN lenses

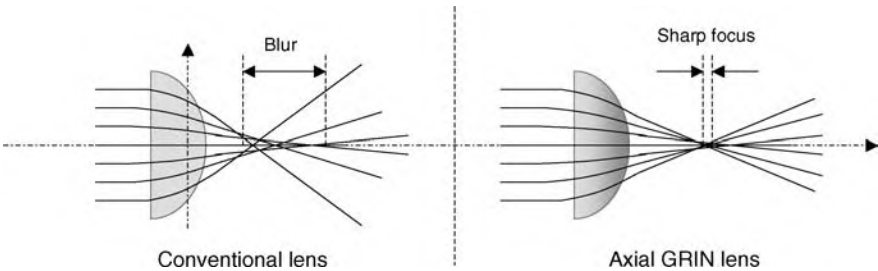


**Figure 4.10** The focusing properties of an axial lens and a conventional lens

Such double-GRIN lens devices can be packaged by using glass sleeves and pigtail rods, as shown in Figure 4.9. Figure 4.10 shows the advantage of using an axial GRIN plano-convex spherical lens over a conventional plano-convex spherical lens.

4.2.3 GRIN Fibers

The other implementation of radial GRINs is the GRIN optical fiber. The preform (before fiber drawing) is actually and exactly a radial GRIN lens as described before. The preform is then heated and drawn into a fiber that has the same index profile but a diameter reduced by a factor of more than 100.

In a GRIN optical fiber, the rays propagate through a sinusoidal movement without ever touching the fiber walls. This is different from traditional step-index fibers, where the light relies on total internal reflection (TIR) (see Chapter 3). However, due to fabrication constraints, such fibers have an index discontinuity at the center of the core, which prevents efficient energy coupling between a source and the

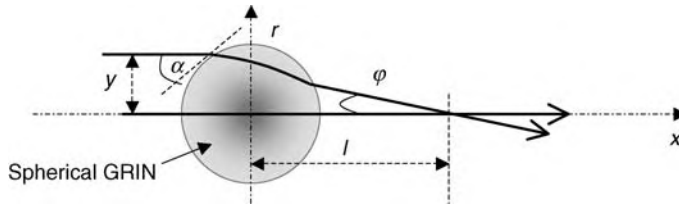


**Figure 4.11** Spherical GRIN lens

core via a traditional coupling lens. Chapter 6, as well as Chapter 16, will show that one can design a special diffractive optical lens (a vortex lens) that can very efficiently couple light into the doughnut mode of such a GRIN fiber without having to suffer from index discontinuity at the optical axis.

#### 4.2.4 Spherical GRINs

Maxwell's fish eye was the first implementation of a spherical or ball GRIN lens. Although being impossible to fabricate at that time, it remained an optical element curiosity where only points on its surface could be imaged sharply.

Ideally, the index of refraction of such a lens gradually varies from the index of the outside medium to a higher index at the center of the ball lens. This is, of course, almost impossible to manufacture if the ball lens is supposed to be functioning in air: it is more likely to function in a liquid (oil, water). Spherical GRIN lenses have been manufactured with more realistic indices, with a step index at the external surface if the ball air lies in air.

The following equation in polar coordinates describes the spherical GRIN lens [12] (see also Figure 4.11):

$$\varphi(r) = \varphi_0 + e \cdot \int_{r_0}^r \left( \frac{1}{r \cdot \sqrt{n^2 r^2 - e^2}} \right) dr, \quad \text{where } e = \pm n \cdot r \cdot \sin(\alpha) \quad (4.6)$$

It is worth noting that the position  $l$  at which the ray intersects the  $x$ -axis is a function of the height of the incoming ray  $y$ . Such a spherical GRIN lens thus has much lower spherical aberrations than a spherical lens made from homogeneous material.

Figure 4.12 illustrates the advantages of spherical GRIN ball lenses over conventional ball lens. GRIN lenses can also be integrated into 2D arrays (see Section 4.1). When integrated into 2D arrays, they can be combined with lithographically patterned diffractive elements.

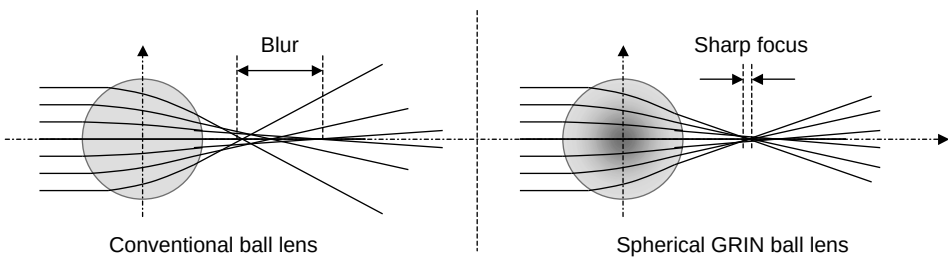


**Figure 4.12** Focusing properties of spherical GRIN and spherical lenses

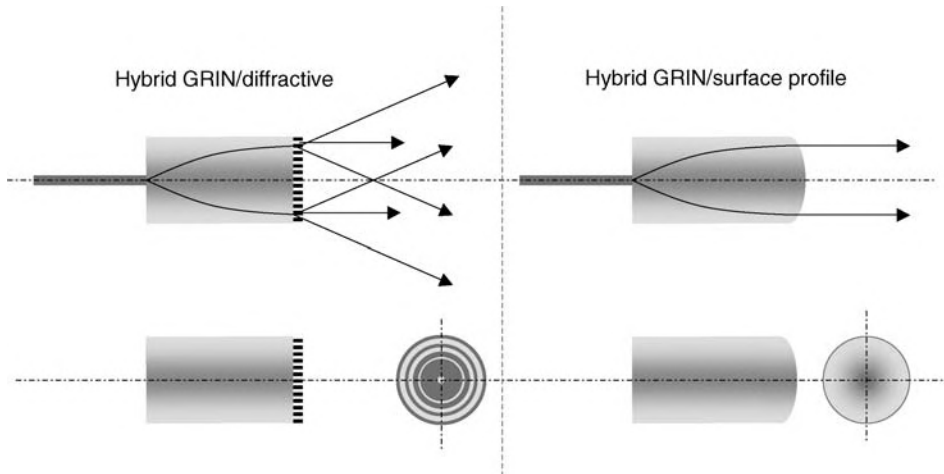


**Figure 4.13** An example of a hybrid GRIN/surface profile and GRIN/diffractive elements

#### 4.2.5 Hybrid GRINs

In order to increase the functionality of GRIN lenses, one can include an axial or radial GRIN on each side of a refractive surface relief (in order to reduce aberrations) and/or etch a diffractive element on one of the surfaces, as depicted in Figure 4.13.

Since GRINs can also be produced in a planar configuration [10, 11], one can lithographically produce sets of diffractive elements on the end surface of these GRINs (either a diffractive Fresnel lens or a beam splitter/beam shaper), or micro-refractive elements can be etched into these substrates via resist reflow and/or RIE (see also Chapter 12).

#### 4.2.6 GRIN Fabrication Techniques

There are mainly two type of GRIN elements used in industry today: the glass GRIN and the polymer GRIN.

##### 4.2.6.1 Glass GRINs

The main fabrication techniques for glass GRINs are diffusion processes, some of which are listed below:

- ion exchange;
- molecular stuffing; and
- the sol-gel process.

For more insight into these fabrication techniques, see Chapter 16. Note that similar fabrication techniques are also used for the production of graded-index waveguides (see Chapter 3).

##### 4.2.6.2 Polymer GRINs

The copolymerization of different monomers (M1, M2, etc.) can build up a whole range of refractive indices that can produce a GRIN Polymer Element [12]. In the copolymerization of monomers M1 and



M2, the refractive index  $n$  of the copolymer is related to the copolymer composition by the Lorentz–Lorenz equation:

$$\frac{n^2 - 1}{n^2 + 2} = \frac{n_1^2 - 1}{n_1^2 + 2}v_1 + \frac{n_2^2 - 1}{n_2^2 + 2}v_2 \quad (4.7)$$

where  $n_1$  and  $n_2$  are the refractive indices of homopolymers M1 and M2, and  $v_1$  and  $v_2$  are the volume fractions of each monomer. Examples of these monomers are methylmethacrylate (MMA) or benzyl methacrylate (BzMA). Often, the following extrapolation of the Lorentz–Lorenz equation is used:

$$n = n_1v_1 + n_2v_2 \quad (4.8)$$

Various fabrication techniques have been proposed for polymer GRIN fabrication, but most of them can be classified into either diffusion processes or processes that make use of monomer reactivity diversity.

### 4.3 Surface-relief Micro-optics

Refractive micro-optics are similar to conventional refractive/reflective optics. The difference, however, is in how they are fabricated. While conventional macroscopic refractive optics are fabricated by the grinding and polishing of glass, digital micro-optics are fabricated by micro-lithography in a wafer-scale fabrication scheme.

#### 4.3.1 Micro-prisms

Prisms are useful refractive (macro- or micro-) elements used to perform numerous tasks, and thus are valuable tools in the optical engineer's toolbox. A few of their functionalities are explained below. Prisms or micro-prisms can:

- Erect or rotate images.
- Change the direction of propagation of a light path:
  - fold a system for compactness; or
  - retro-reflect a signal.
- Disperse the spectrum.
- Control beam parameters:
  - anamorphic telescopes; or
  - by varying the angle, position and path length.
- Divide amplitude or polarization:
  - beam splitters; or
  - beam samplers.

Figure 4.14 shows some of the basic functionalities that folding micro-prisms can implement.

Wedges can also be used for many beam manipulations, such as beam deviation (the Risley prism), beam displacement (the sliding-wedge prism) or focus adjustment via variable path length, through the use of two sliding opposite wedge prisms. Anamorphic micro-prisms can perform spatial beam compressions, as described in Figure 4.15.

Chapter 1 has shown how the basic prism deflection angle can be made equal to the micro-prism array diffraction angle, since in the general case they are not the same. Similarly with spectral dispersion, a micro-prism has one particular spectral dispersion characteristic (see Figure 4.16), while an array of micro-prisms has another spectral dispersion characteristic (usually much larger). Figure 4.16 shows the basic spectral dispersion properties of micro-prisms.

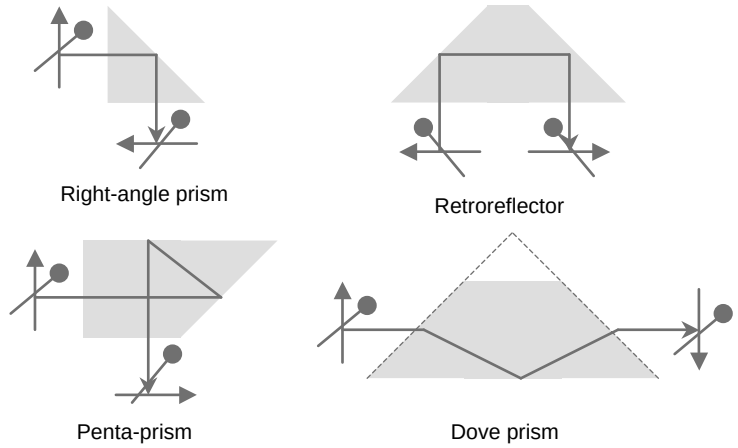


**Figure 4.14** Micro-prism folding functionalities

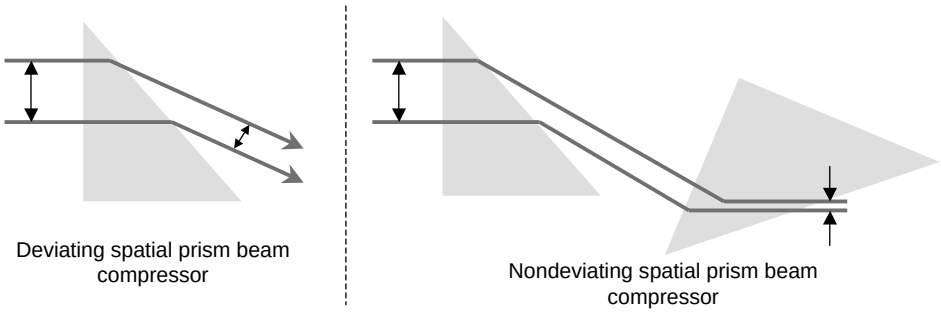


**Figure 4.15** Beam compression through anamorphic prisms

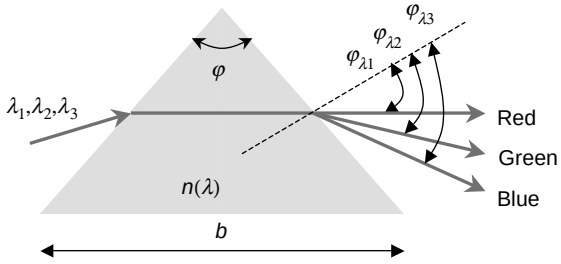


**Figure 4.16** Micro-prism spectral dispersion characteristics

The spectral dispersion of a prism is a function of the dependence of the refractive index on the wavelength,  $n(\lambda)$ , and is not directly related to the geometry of the prism. The resolving power  $R$  for a prism (which is also related to the base prism base length  $b$ , since  $R$  is a spatial spectral resolving criteria) is given by

$$R_{prism} = \frac{\lambda_0}{\delta\lambda} = b \cdot \frac{\partial}{\partial\lambda}(n) \quad (4.9)$$

The resolving power for a grating, which is related to the diffraction order  $m$  considered and the number of grating periods  $N_0$  illuminated, is given by

$$R_{grating} = \frac{\lambda_0}{\delta\lambda} = m \cdot N_0 \quad (4.10)$$

When comparing this spectral resolving power to the spectral dispersion of gratings (see also Section 5.3.5), one can note that in the case of a grating, the larger wavelengths (red) are the most deflected (diffracted), while in the case of a prism, the smallest wavelengths (blue) are the most deflected (refracted). In a prism (or a lens), this is related to the fact that  $n(\lambda)$  increases when  $\lambda$  decreases.

These opposite dispersion effects are the basis of the design of many hybrid refractive/diffractive elements (such as achromatization of hybrid singlets), which will be discussed in detail in Chapter 7. Similar opposite effects occur also when the temperature changes and can give rise to interesting applications (such as athermalization of hybrid singlets – see also Chapter 7).

### 4.3.2 Imaging Microlenses

Consumer electronics has been the main driver in the development of imaging microlenses. Two examples are described here, one from consumer electronics data storage (CD/DVD) and the other from consumer electronics digital cell phone cameras.

#### 4.3.2.1 Optical Pick-up Unit (OPU) Lenses

Optical Pick-up Unit (OPU) microlenses are used to read optical storage media such as CD, DVD and Blu-ray disks. Such lenses have a large NA, and come in numerous different formats.

A simple lens for a single wavelength and a single NA is easy to design and fabricate. A lens that has to be able to read different media with different NAs, different spherical aberrations (due to different disk media thicknesses) and different wavelengths (due to different media dies for RW capabilities) is more difficult to design and fabricate. Such lenses are dual- or triple-focus lenses, and yield different optical functionalities for different wavelengths (see Chapter 16).

#### 4.3.2.2 Wafer-scale Camera Objective Lenses

Today, digital cameras are present in everyday life in the form of camera phones, digital SLR cameras, CCTV surveillance cameras, webcams, automotive safety cameras and so on. Such cameras will become more miniaturized in the future, and increasingly efficient.

Chapter 12 shows how two or more wafers with microlens arrays can be aligned prior to dicing them into individual stacked wafer-scale optical compounds, and then integrated in a miniature camera objective. Chapters 12 and 13 reviewing such wafer-scale fabrication processes, and the various wafer alignment techniques that can be used (microlens alignment, Talbot self-imaging, etching and solder bumps, etc.).

### 4.3.3 Nonimaging Microlenses

Microlenses are now also beginning to be used for nonimaging tasks [13–15], such as beam shaping and beam homogenizing (other than with the use of microlens arrays, as described in the next section). Although this is a task that is best performed by arrays of microlenses (see the next section) and/or diffractive beam shapers/diffusers (see Chapter 6), arbitrary-shaped surface profile microlenses can be designed and fabricated by lithography (either by gray-scale lithography or resist reflow) for quasi-arbitrary beam shaping.

The advantages of using such lenses over diffractive beam shapers are the very high efficiency that one can obtain (100%) and the fact that such lenses are almost achromatic (this is not the case for diffractive beam shapers).

The advantage of using such a single lens over arrays of microlenses, as discussed in the next section, is that such an arbitrary shaped lens can perform beam-shaping tasks that are not limited to perfect paving geometries of lens apertures (square, hexagonal, etc.).

## 4.4 Micro-optics Arrays

The previous sections have reviewed the basic functionalities and characteristics of single GRIN or micro-refractive lenses. We will now describe how they can be used in 1D or 2D arrays and show some applications that use such arrays.

### 4.4.1 GRIN Lens Arrays

GRIN arrays are very efficient elements, since they can now be routinely fabricated in a 1D or 2D array [10], and they provide a perfect planar surface, unlike micro-refractive lenslet arrays and diffractive lens arrays. Such planar arrays can be used in numerous applications, and can also be used as wafers for further processing via microlithography (producing either refractive or diffractive elements on the array).

#### 4.4.1.1 One-dimensional GRIN Arrays

One-dimensional GRIN lens arrays are used in a wide field of planar free-space devices, especially for telecom applications when pigtailed to a fiber. Such devices are very useful in numerous applications to out-couple the guided wave into free space for optical processing and to re-couple that processed wave into the waveguide. Chapter 7 shows such a hybrid system, which is applied to the integration of a free-space  $\lambda$  Mux/Demux with the use of free-space Thin Film Filters (TFFs). An incoming train of wavelengths is collimated by the first GRIN, and as the beam is reflected on the various TFFs within the glass slab, the rejection band of each TFF lets only one wavelength through and reflects the others down the slab.

Note that in some cases, a linearly changing TFF can be used instead of a series of discrete TFFs (fabrication is done through apodizing the filter thicknesses along one axis). In other cases, a single TFF filter is used for all channels, but a precisely wedged glass slab is used so that the angle of each beam inside the slab is slightly different and thus produces another filter response (since the filter response is usually proportional to the incidence angle).

Another common use for 1D GRIN lenses is in free-space MEMS micromirror switch arrays (for a practical example, see Chapter 16). Such micromirror arrays are very sensitive to the degree of collimation of the beams coupled out of the fibers. The waist of the Gaussian beam is usually set precisely at the mid-distance between the input and output GRIN, by adjusting the GRIN length or by using a spacer, as discussed previously. Such optical switches are called 2D switches, since the micromirrors can only tilt to a specific and unique angle. 3D optical switches are much more complex in their architecture and functionality: in 3D optical switches, the micromirrors can switch to a series of pre-set angles or to any angle. However, in this case, there has to be a feedback control of the mirror tile, which is usually performed optically or by torque measurement on the micro-hinges.

#### 4.4.1.2 Two-dimensional GRIN Arrays

The previous two examples used 1D GRIN lens arrays. 2D GRIN arrays are also desirable elements and can be used as substrates (wafers) in microlithography in order to align and pattern additional elements. Figure 4.17 shows a 2D array of GRIN lenses (see, e.g., Nippon Sheet Glass products) with patterned diffractive elements in order to produce highly functional arrays of hybrid microlenses that can compensate chromatic aberrations.

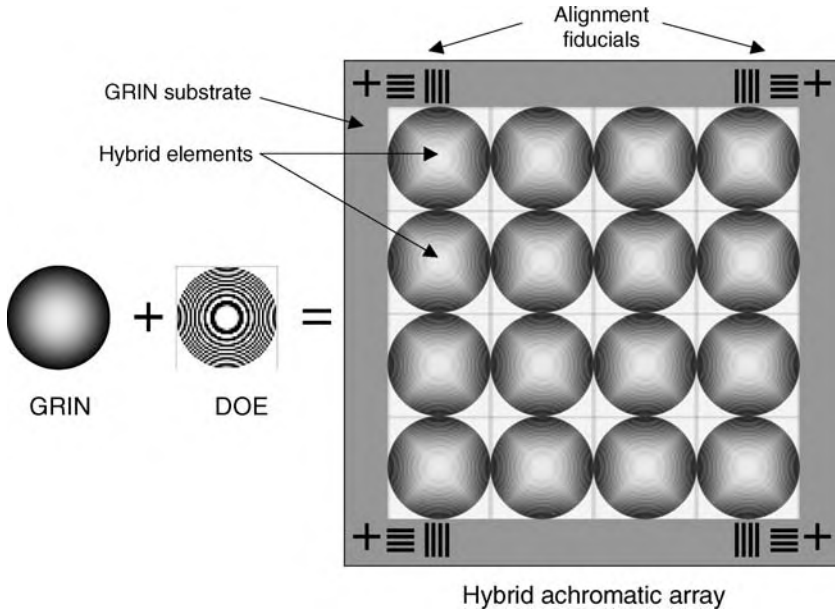


**Figure 4.17** Hybrid GRIN/diffractive arrays of achromatic lenses

Fabricating a hybrid diffractive/refractive (surface profile) achromatic lens arrays is very challenging, since doublet lens arrays are very costly, and hybrid arrays of refractive/diffractive profiles are difficult to produce via lithography since one of the surfaces is not flat. The hybrid planar GRIN/diffractive array thus provides a desirable alternative that is possible to implement with standard lithography, without having to use complex 3D gray-scale lithography.

More complex GRIN hybrids can be fabricated conveniently via standard lithography and Fourier CGHs rather than DOEs, as depicted in Figure 4.17. Examples include arrays of on-axis GRINs with Fourier beam-shaping CGHs (see Chapter 6) or beam-splitting gratings (fan-out gratings). Such hybrid GRIN/diffractive arrays can thus produce arrays of multiple focal length GRINs for special imaging tasks, or arrays of beam-shaping elements with high efficiency, since the GRIN (the lensing function) is almost 100% efficient, and the Fourier CGH, in a symmetric configuration, can have over 80% efficiency in its simplest binary phase profile case.

One major concern when aligning a diffractive structure on top of a GRIN lens array is the lateral registration of the optical axis location of the GRIN array on the substrate. Therefore, alignment marks have to be inserted within the GRIN array prior to its fabrication. This process is usually performed by the GRIN lens manufacturer (see the alignment fiducials in Figure 4.17).

#### 4.4.2 Micro-prism Arrays or Prism Sheets

Micro-prism arrays can be used for many applications. As seen in Chapter 1, a micro-prism array with a fill factor of 100% can actually be a blazed grating if the sizes of the microprisms are small enough, or can produce a cylindrical lens if the sizes of the prisms only vary in one direction. However, refractive micro-prism arrays rely heavily on the sole refractive effect. Diffractive micro-prism arrays are discussed in Chapter 5. Nanoscale prism arrays are discussed in Section 10.4.3. Note that prism arrays can also produce a diffractive Fresnel lens if one attempts to wrap such a chirped array into a circular geometry.

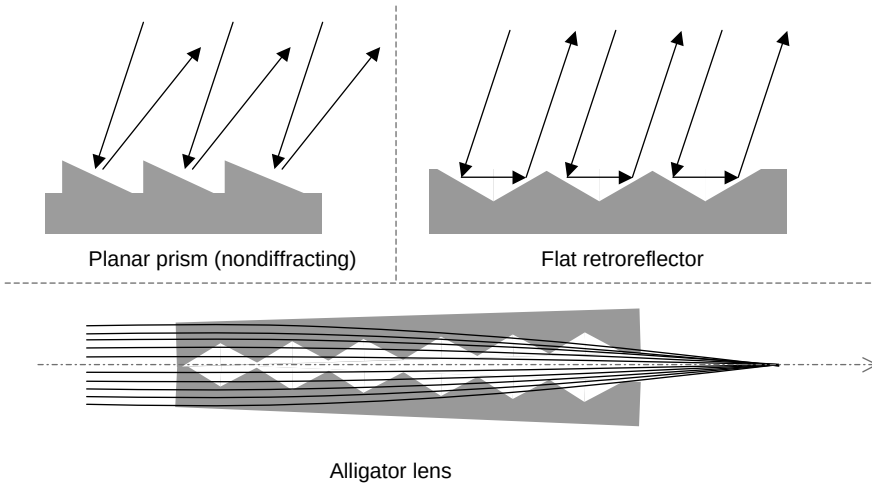


**Figure 4.18** Micro-prism effects

Such elements can be used as retroreflectors (or ‘cat’s eyes’) in applications where a precise and invariant beam reflection angle is required, or beam deflectors, or even tunable lensing effects, as in the alligator lens. Figure 4.18 shows such micro-prism applications [16].

Bear in mind that the arrays of micro-prisms depicted in Figure 4.18 are refractive, not diffractive, which means that they are actually almost achromatic elements, and they do not lose any efficiency when changing the wavelength of light (at least, not as much as diffractives would).

Other applications of refractive prism arrays are extensively used in LCD and OLED display panels, for optimal light extraction via the TIR effect, as described in Figure 4.19.

Such prism arrays are usually called prism sheets or Brightness-Enhancing Films (BEF) (for an overview of various optical BEF and Dual BEF–DBEF film technologies, see also Chapter 16).

Micro-refractive BEF films can be linear prism arrays or bi-dimensional prism arrays (fabricated by diamond turning or diamond ruling). Here, the prism sheets in the BEF film presented in Figure 4.19 are circular symmetric prisms or, rather, partial cones. The almost achromatic nature of the micro-prism is used here as well, although the conical structures can be pretty small, as small as  $10\text{ }\mu\text{m}$  in diameter, which also produces a super-grating effect when the entire array is considered. This parasitic diffraction effect is, however, limited due to the low spatial and temporal coherence of the sources used (LED and OLED).

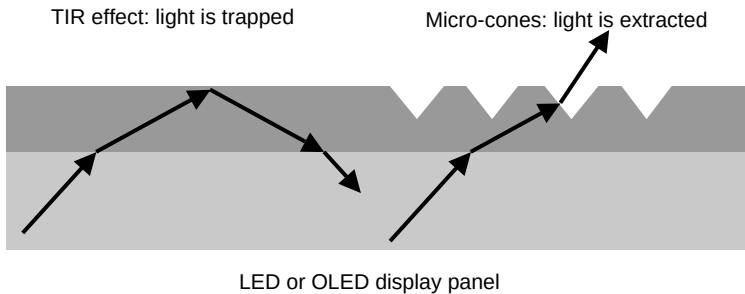


**Figure 4.19** A brightness-enhancing layer via micro-cone arrays and the TIR effect

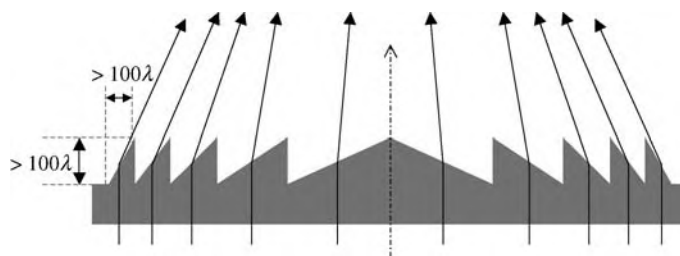


**Figure 4.20** A refractive sawtooth Fresnel lens

#### 4.4.3 Refractive Micro-Fresnel Lenses

Chirped micro-prism arrays can be used to produce numerous micro-optical elements, including cylindrical and circular lenses [17–19]. Such lenses are called micro-Fresnel lenses, and are very often mistakenly referred to as diffractive elements. Figure 4.20 shows such a refractive micro-Fresnel lens, this one having a sawtooth profile generated by simple diamond turning.

Such a micro-Fresnel lens works in the same way as the 19th-century Fresnel lighthouse concentrator. They yield very small diffraction effects (see also Chapter 1), and they work almost entirely in the refractive regime, since the period of the fringes (or micro-prisms) is several hundred times the wavelength, and the height of the structures is several hundred times the wavelength, with a tolerance much larger than the wavelength (and thus perfectly unable to create any local interference, which is the basis of the diffraction effect in diffractive Fresnel lenses – see Chapter 5).

Figure 4.20 shows how ray tracing can be used to model micro-Fresnel lenses through each of its micro-prismatic rings. It is not possible to use such ray tracing in the case of a Fresnel diffractive lens.

#### 4.4.4 Refractive Microlens Arrays

Refractive microlens (or lenslet) arrays are very desirable elements, which are used today in numerous imaging and nonimaging applications in industry [20–23].

Refractive lenslet arrays have numerous advantages over diffractive lenslet arrays, namely:

- they are almost achromatic (no strong dispersion effects, and no drop in efficiency when changing the wavelength or the incoming angle);
- they are very efficient (almost 100%), and they do not produce unwanted orders; and
- they can be easily replicated due to their (relatively) large sizes and smooth profiles.

However, they lack where the diffractives excel, namely in:

- 100% fill factors in any paving geometry [24];
- accurate phase description (precise aspheric, anamorphic, etc.);
- additional functionality (multiple imaging, beam shaping, etc.);
- accurate fabrication via traditional binary lithography (no gray-scale or resist reflow needed); and
- flat substrates.

##### 4.4.4.1 Imaging Lenslet Arrays

Imaging microlens arrays have numerous advantages over their macroscopic counterparts. When the imaging system becomes large, it gets increasingly difficult for a macroscopic lens to correct

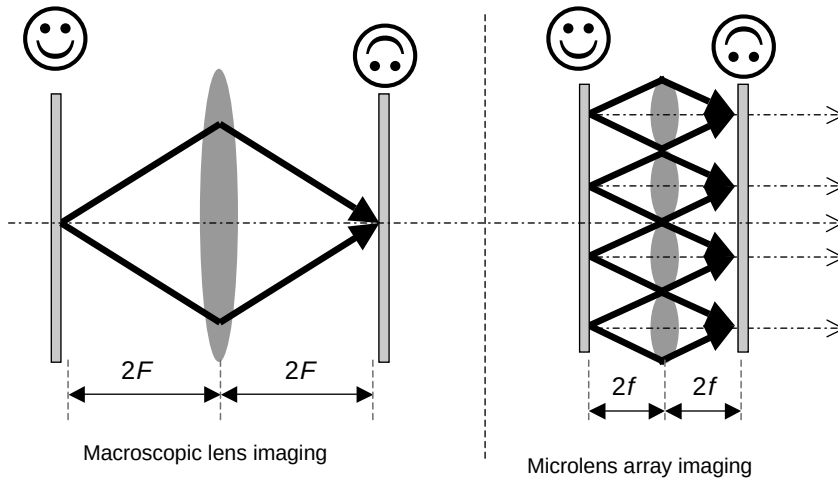


**Figure 4.21** Imaging through a macroscopic lens and a microlens array

all wavefront aberrations, and thus such a lens becomes very expensive to fabricate (highly aspheric surface profile). Simple spherical microlens arrays can yield decent imaging qualities by partitioning the image into a matrix corresponding to the microlens array. Figure 4.21 compares both imaging set-ups.

It is obvious that for the best imaging quality with the highest efficiency, such microlens arrays should have a 100% fill factor (no dead zones in between lenses). Thus, the microlenses should be fabricated over perfect paving geometries such as hexagonal, square or rectangular, and spherical aperture shapes would yield considerable losses. Refractive microlens arrays are easily fabricated over circular apertures, but can be fabricated (less easily) over perfect paving apertures. Diffractive microlens arrays, on the other hand, can be fabricated over any type of aperture with a 100% fill factor. However, diffractive microlens arrays (see Chapter 5) have limited efficiency and are usually not broadband.

Figure 4.21 shows an inverted image resulting from the microlens array  $4f$  configuration. The image can be set upright when using two arrays in telescope configuration, as depicted in Figure 4.22.

However, as the object is emitting light in all spatial directions, like a point source, light from one section of the image hits many lenses in the microlens array, thus producing not just one image but many (composite) images.

For the position of the composite images in the telescope set-up, see Figure 4.22. The brightest image (provided that the source emits most of the light normal to the surface) is the primary image, the others dimming out as one moves away from the optical axis of this particular lens in the array.

### ***Lenticular Arrays For Stereo Imaging***

Lenticular arrays of microlenses are used extensively in stereoscopic and other 2.5D and pseudo-3D screens. Conventional lenticular arrays are cylindrical, and are placed over a sliced and inter-digitated set of images of an object taken at various angles, in order to produce a pseudo-3D viewing effect (see Figure 4.23).

Note that such lenticular arrays can also be two-dimensional (arrays of spherical lenses) in order to produce a vertical parallax on top of the previous horizontal parallax. However, such arrays are very scarce since, for a viewer, the most important parallax is horizontal (especially when the device is a computer screen), since the viewer's head seldom moves up and down.



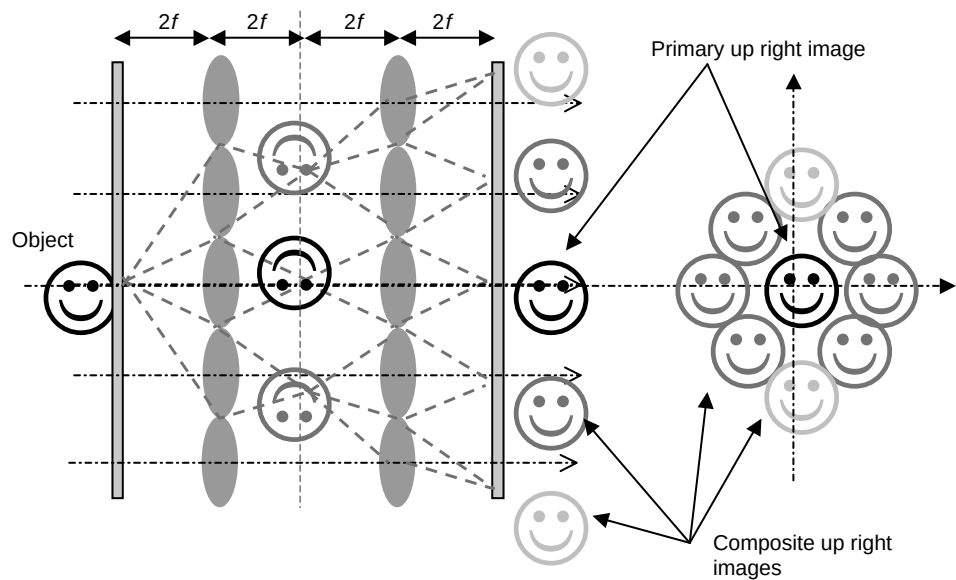


**Figure 4.22** A microlens array in telescope configuration and the location of composite images

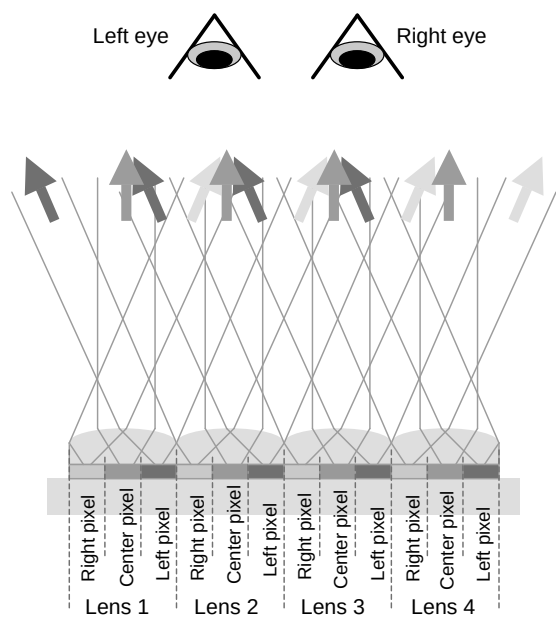


**Figure 4.23** Lenticular arrays used in stereo imaging

### Other Lenslet Array Imaging Tasks

Other imaging applications include lenslet arrays that produce numerous images from a single object, anamorphic lenslet arrays that produce deformed images in one or two dimensions, and so on.

#### 4.4.4.2 Nonimaging Lenslet Arrays

Nonimaging refractive microlens arrays are also elements that are widely used in industry today, such as in collimation arrays and beam steering devices [6, 13, 15].

##### *Laser Collimator Arrays*

Laser collimation arrays are one example: such collimation arrays can be two-dimensional spherical lens arrays for VCSEL array collimation (as in 10 Gb/s Ethernet transceiver stacks – see Chapter 16), or one-dimensional stacked arrays of cylindrical lenses for edge-emitting laser diode arrays (see Figure 4.24). Edge-emitting diode arrays (Chapter 4) produce astigmatic beams that have different diverging angles (slow and fast axis), therefore requiring arrays of cylindrical lenses.

In such applications, it is again difficult to swap refractive arrays with diffractive arrays, not so much because of their spectral dispersion and polychromatic efficiency reduction issues (coherent spatial and temporal beams produced by laser) but, rather, because of the limited angular bandwidth of diffractives, compared to large angular bandwidths of refractives. Besides, for edge-emitting laser diodes, the divergence angle can be such that the required smallest fringe size in an equivalent diffractive lens would be so small that, first, fabrication would be a challenge and, second, the depth and fringe profile would have to be carefully optimized by electromagnetic (EM) tools, since scalar diffraction theory would no longer yield decent predictions of efficiency.

In the case of VCSEL arrays (especially in 2D), the use of diffractive collimator arrays makes much more sense (smaller angles and circular beams).

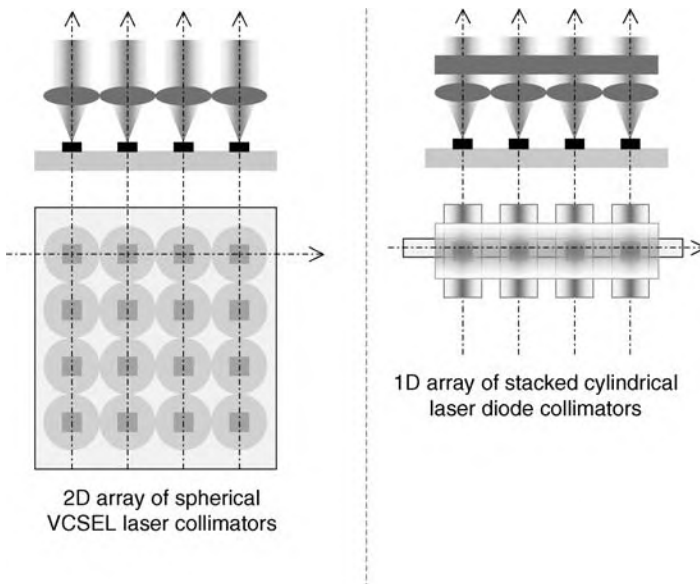


**Figure 4.24** Collimator arrays for VCSEL and edge-emitting laser diodes

Note that such arrays can also be used for laser coupling into fibers if the NA of the lens matches the NA of the optical fiber, and for coupling the light from optical fibers onto detector arrays. Fiber arrays are easily packaged into linear arrays in silicon V-grooves.

### Beam-steering Lenslet Arrays

Another nonimaging task usually implemented with sandwiched refractive microlens arrays is the production of collimated beam-steering devices. Figure 4.25 shows such a collimated beam-steering device, using a first array formed by converging microlenses and a second array formed by diverging lenses in a telescopic arrangement.

An offset between a converging and a diverging lens array in a telescopic arrangement is equivalent to a wedge prism effect. An array of lenslets will thus have the same effect as a micro-refractive prism array (see the previous sections). This prism effect can be tuned by changing the offset; in other words, by sliding one array with regard to the other. The amount of deflection is a function of the vertical offset and is given by

$$\varphi = -\arctan\left(\frac{\Delta x}{d \cdot f^{\#}}\right) \quad (4.11)$$

In a beam-steering device based on microlens arrays, one array moves with regard to the other, in one or two dimensions, to produce a tunable beam steerer in all spatial directions.

As discussed in the first chapter (from refraction to diffraction), the periodic grating gives rise to a diffraction effect. The maximum efficiency is thus achieved when the refractive and diffractive angles are equal.

When considering the use of diffractive lenslet arrays instead of refractive lenslet arrays for beam steerers, other than having the benefit of a 100% fill factor, zero orders (on-axis beams) would appear when using broadband illumination, the steering angle would be a function of the wavelength, and the efficiency would be reduced as the square of the single efficiency value, since two diffractives were being used.

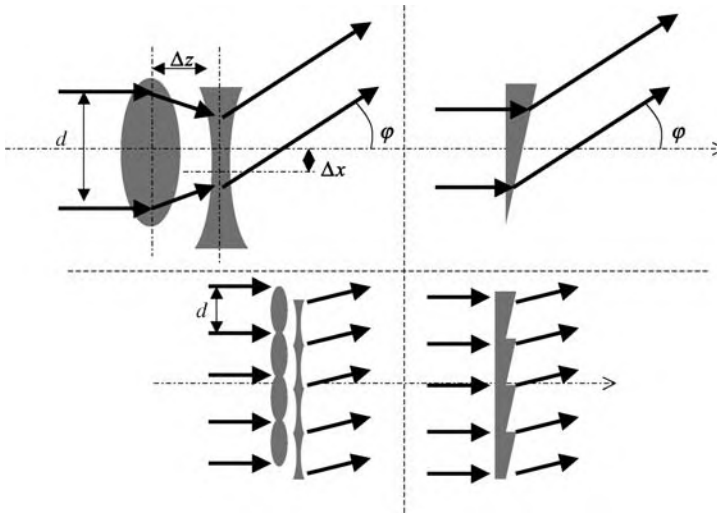


**Figure 4.25** Stacks of collimated beam-steering microlens arrays

### CMOS Sensor Light Collector Arrays

In order to increase the efficiency of CMOS sensors, arrays of microlenses are used to focus more light into the active CMOS (or CCD) cells [25], and to use the light that would otherwise be absorbed by inter-pixel void spaces (for more details on this application, see Chapter 16).

#### 4.4.4.3 Beam Homogenizer Lenslet Arrays

Beam homogenizing is one of the most popular applications of lenslet arrays. Beam homogenization is very similar to beam shaping. Both functionalities can be integrated together in a lenslet. Beam homogenizer arrays are usually called ‘fly’s eye’ arrays, due to their resemblance to lens arrays in the animal kingdom. They are widely used in digital projectors and other imaging devices.

Beam homogenizers can act on any type of beam profile, and they homogenize the beam in the far field, rather than at the focal plane of the lenslet array. Such lenslet beam homogenizers can be used with or without a field lens (see Figure 4.26).

In the case in which such a field lens is used with a Gaussian input beam profile to be converted to a top hat profile, the overlap of the different beamlets is perfect, and thus yields a nice top hat intensity profile. In the case in which one chooses to let the beamlets propagate to a set distance (the far or near field), the overlap is not perfect, thus producing a super-Gaussian beam (in theory, in the far field, these beamlets will also superimpose perfectly).

This top hat has the same section as the individual lens aperture (it produces a square top hat if the lenses are arrays of square apertures, and a circular top hat if the lenses are fabricated on circular apertures).

Note that if one changes the incoming Gaussian beam into a doughnut mode beam, the resulting beam patch will still be square, rectangular or circular, depending on the individual lens aperture geometry.

As efficiency and uniformity are usually of interest here, a fill factor of 100% is highly desired and, thus, a perfect paving geometry is commonly used. The problem with refractive lenslet arrays is that they can be fabricated easily on circular apertures, and then replicated. However, fabricating refractive lenslet arrays on square aperture, hexagonal aperture or other perfect paving geometries pushes the fabrication notch a degree higher. A circular lens aperture is easy to produce, but yield a fill factor below 100%.

#### 4.4.4.4 Beam-shaping Lenslet Arrays

Besides being used as beam homogenizers, as discussed in the previous section, microlens arrays can also be used as beam-shaping elements, by varying the aperture dimension and the shape of the single lenses, and the lens functionality itself.

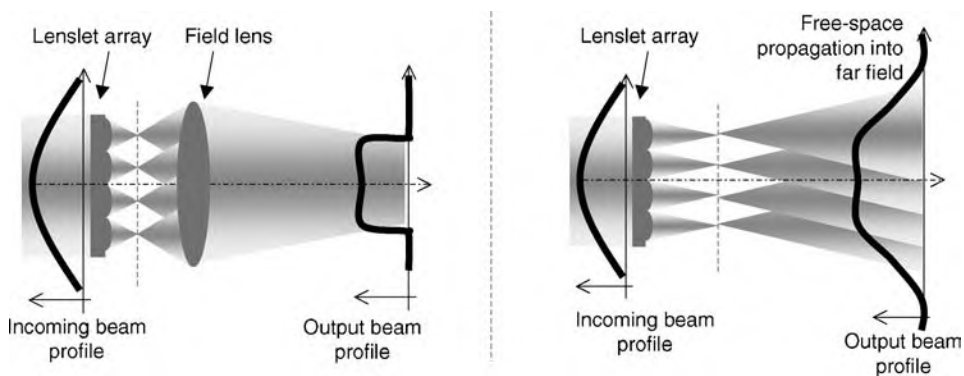


**Figure 4.26** Beam homogenizers based on lenslet arrays

For example, a hexagonal close-packed lens array can shape an incoming beam into a hexagonal output patch, a triangular lens array can shape any incoming beam into a triangular patch, and rectangular arrays can shape any incoming beam into a rectangular patch. All three can be paved with a 100% fill factor, and can therefore be very efficient.

There are two main differences between beam-shaping lenslet arrays and the other beam shapers that are described in this book (refractive and diffractive):

- One main difference between a diffractive or refractive beam shaper is that these lenslet arrays can produce a specific output beam geometry for any incoming beam geometry.
- The other main difference between a diffractive beam shaper and an lenslet beam-shaper array is that the refractive lenslet array has a high efficiency for a wide range of wavelengths (not totally achromatic, though), which means that it can actually work efficiently for white light, whereas a diffractive beam shaper is usually designed and fabricated for a single or narrow wavelength bandwidth, for which it has maximum efficiency.

A diffractive or refractive beam shaper has to have a more or less well-defined incoming beam profile (e.g. a Gaussian to top hat beam shaper). One exception is the Fourier diffractive beam shaper, which allow greater changes in the incoming beam profile (e.g. changing the beam waist, but not changing the beam mode).

Figure 4.27 shows the differences between a circular Gaussian-shaped beam via a refractive hexagonal close-packed array of micro-refractive spherical lenses (left) and the same Gaussian beam shaped via a diffractive beam shaper in the far field.

In the left-hand reconstruction in Figure 4.27, note the hexagonal shape appearing either in the near or far field, with fuzzy edges, mainly due to multiple interferences of all the beamlets produced by the microlens array, and no zero order. On the right-hand side, note very sharp edges (for any type of pattern), appearing only in the far field, with some quantization noise (produced by the finite number of phase levels in a diffractive), and a more or less strong zero order (depending on the quality of the fabrication – especially the etch depth).

Figure 4.28 shows small arrays of such lenslet arrays produced by gray-scale lithography and etched in fused silica substrates (on the left, a hexagonal close-packed array; and on the right, a square array).

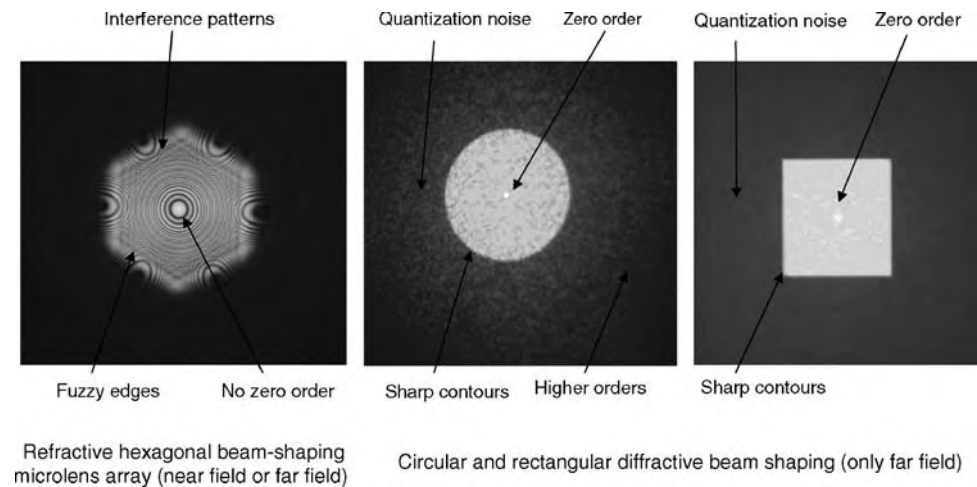


**Figure 4.27** Micro-refractive and -diffractive beam shapers (optical reconstructions)

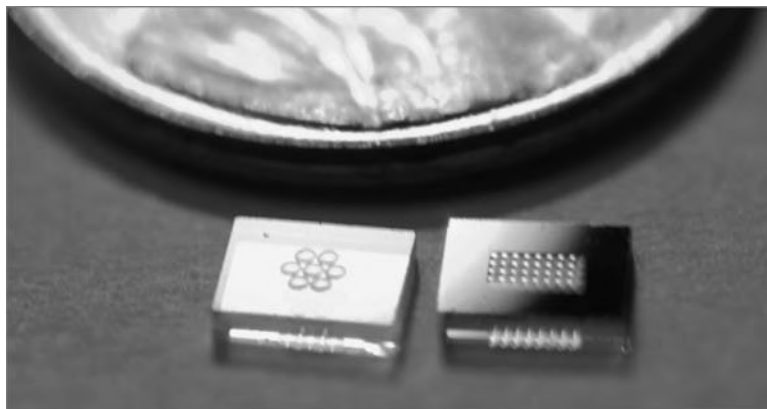


**Figure 4.28** Arrays of micro-refractive lenses in hexagonal and square array configurations

Both produce a fill factor of almost 100%, and work over the entire visible spectrum. The individual lenses in the hexagonal array are  $250\text{ }\mu\text{m}$  wide for a focal length of  $1.5\text{ mm}$ , and in the square array they are  $50\text{ }\mu\text{m}$  wide for a focal length of  $100\text{ }\mu\text{m}$ .

#### 4.4.4.5 Wavefront Sensor Lens Arrays

The field of adaptive optics was originally developed for astronomic observation through wavefront turbulences. This field is now becoming more democratized, and finds applications in everyday life, from ophthalmology to personal displays. The principle of a Shack–Hartmann lenslet array wavefront sensor is depicted in Chapter 16.

The more or less planar wavefront is focused on a focal plane array by the array of microlenses. Each shift in the position of the focal spot corresponds to a local phase gradient (a small prism effect), which relates to the local gradient in the wavefront curvature. One can therefore compute the entire wavefront map, and inject the inverse phase profile into an adaptive optical system to restore a perfect wavefront and therefore enable perfect imaging (perfect imaging means imaging without any phase perturbation in the optical path (such as atmospheric phase turbulences)).

Adaptive optics can be implemented by the use of pistons under a telescope main mirror, by the use of more complex DLP-based micromirror devices in personal display devices, or by the use of more complex arrays of LC devices.

## References

- [1] H.-P. Herzig, *‘Micro-optics: Elements, Systems and Application’*, Taylor and Francis, London, 1997.
- [2] S. Sinzinger and J. Jahns, *‘MicroOptics’*, Wiley-VCH, Weinheim, 1999.
- [3] K. Iga, Y. Kokubun and M. Oikawa, *‘Fundamentals of Micro-optics’*, Academic Press, Tokyo, 1984.
- [4] M.E. Motamedi, W.E. Tennant and R. Melendes, *‘FPAs and Thin Film Binary Optic Microlens Integration’*, SPIE 2687, 1996, 70–77.
- [5] D.A. Fletcher, K.B. Crozier, G.S. Kino, C.F. Quate and K.E. Goobon, *‘Micromachined scanning refractive lenses’*, Solid State Sensor and Actuator Workshop, June 4–8, 2000, 263–265.
- [6] R. Volker *et al.*, *‘Fabrication of non-conventional microlens arrays’*, EOS Topical Meeting on ‘Microlens Arrays’, at NPL Teddington, UK, May 11–12, 1995.

- [7] M. Eisner, K.J. Weible and R. Voelkel, 'Fabrication of Aspherical Microlenses in Fused Silica and Silicon', *Proceedings of SPIE*, **4440**, 2001, 40–43.
- [8] R. Göring, F. Wippermann and P. Dannberg, 'Replicated microlens arrays for optical fiber collimators', post-deadline paper at SPIE's International Symposium on Micromachining and Microfabrication, Technical Conference 'MOEMS and Miniaturized Systems', San Francisco, October 2001.
- [9] D. T. Moore, ed., '*Selected Papers on Gradient-index Optics*', SPIE Milestone Series, vol. MS 67, 1993.
- [10] Nippon Sheet Glass Ltd., product information.
- [11] Grintech GmbH (Germany), product information.
- [12] Y. Koike, 'Graded index materials and components', Chapter 3 in '*Polymers for Lightwave and Integrated Optics: Technology and Applications*', L.A. Hornak (ed.), Keio University, Yokohama, 1992.
- [13] R. Grunwald, S. Woggon, U. Griebner, R. Ehlert and W. Reinecke, 'Axial beam shaping with nonspherical microoptics', *Japanese Journal of Applied Physics*, **37**, 1998, 3701–3707.
- [14] R. Grunwald, S. Woggon, R. Ehlert and W. Reinecke, 'Thin-film microlens arrays with non-spherical elements', *Pure and Applied Optics*, **6**, 1997, 663–671.
- [15] D.H. Raguin, G. Grettton, D. Mauer *et al.*, 'Anamorphic and aspheric microlenses and microlens arrays for telecommunication applications', Technical Digest OFC 2001, paper MK 1.
- [16] W. Jark, 'Effective X-ray focusing with unconventionally shaped refractive/diffractive lenses', Workshop on Micro/Nano, Erice, Italy, June 15, 2007.
- [17] D. D'Amato and R. Centamore, '*Two Applications for Microlens Arrays: Detector Fill Factor Improvement and Laser Diode Collimation*', SPIE 1544, 1991, 166–171.
- [18] M.W. Farn, '*Micro-concentrators for Focal Plane Arrays*', Proc. SPIE 1751, 1992, 106–117.
- [19] M.F. Lewis and R.A. Wilson, 'The use of lenslet arrays in spatial light modulators', *Pure and Applied Optics*, **3**, 1994, 143–150.
- [20] N.F. Borelli, R.H. Bellman, J.A. Durbin and W. Lama, 'Imaging and radiometric properties of microlens arrays', *Applied Optics*, **30**, 1991, 3633–3642.
- [21] N. Davies, M. McCormick and M. Brewin, 'Design and analysis of an image transfer system using microlens arrays', *Optical Engineering*, **33**, 1994, 3624–3633.
- [22] W. Southwell, 'Focal-plane pixel energy redistribution and concentration by the use of microlens arrays', *Applied Optics*, **33**, 1994, 3460–3464.
- [23] M.E. Motamedi, 'Micro-optic integration with focal plane arrays', *Optical Engineering*, **36**, 1997, 1374–1381.
- [24] E.A. Wilson, D.T. Miller and K.J. Bernard, '*Analysis of Fill Factor Improvement Using Microlens Arrays*', Proc. SPIE 3276, 1998, 123–133.
- [25] B. Layet, I.G. Cormack and M. Taghizadeh, 'Stripe color separation with diffractive optics', *Applied Optics*, **38**, 1999, 7193–7202.





# 5

## Digital Diffractive Optics: Analytic Type

Chapter 2 proposed a classification of free-space digital optics into the following five types, which are differentiated by either their structural aspects or their optical functionalities:

- Type 1 – Holographic Optical Elements (HOEs);
- Type 2 – Diffractive Optical Elements (DOEs);
- Type 3 – Computer-Generated Holograms (CGHs);
- Type 4 – Sub-Wavelength digital Elements (SWEs); and
- Type 5 – Dynamic digital optics.

This classification is also chronological, with Type 1 (HOEs) being the first elements developed, and with Type 4 (especially photonic crystals and metamaterials) and Type 5 (dynamic diffractives) being the most recently developed (see Figure 5.1).

By the term ‘digital diffractive optics’, we actually mean diffractive optical elements calculated by computer and fabricated by lithographic techniques (using binary masks, as done for digital electronics) as planar elements.

Chapters 5 and 6 focus mainly on Types 2 and 3, which make up the vast majority of diffractive optics used in industry today. This chapter is longer than many of the others because of the importance of these elements today.

Holographic Optical Elements (HOEs, Type 1), dynamic digital optics (Type 5) and sub-wavelength digital elements (Type 4) will be covered in Chapters 8–10. Although HOEs and dynamic diffractives can be calculated by a computer and recorded from a digital CGH master, neither are directly fabricated by means of a binary lithographic. Photonic crystals and optical metamaterials, which are very similar to sub-wavelength digital elements in form (Type 4), will be analyzed in detail in Chapter 10, since both can actually be more complex than 2D planar structures and behave as real 3D structures.

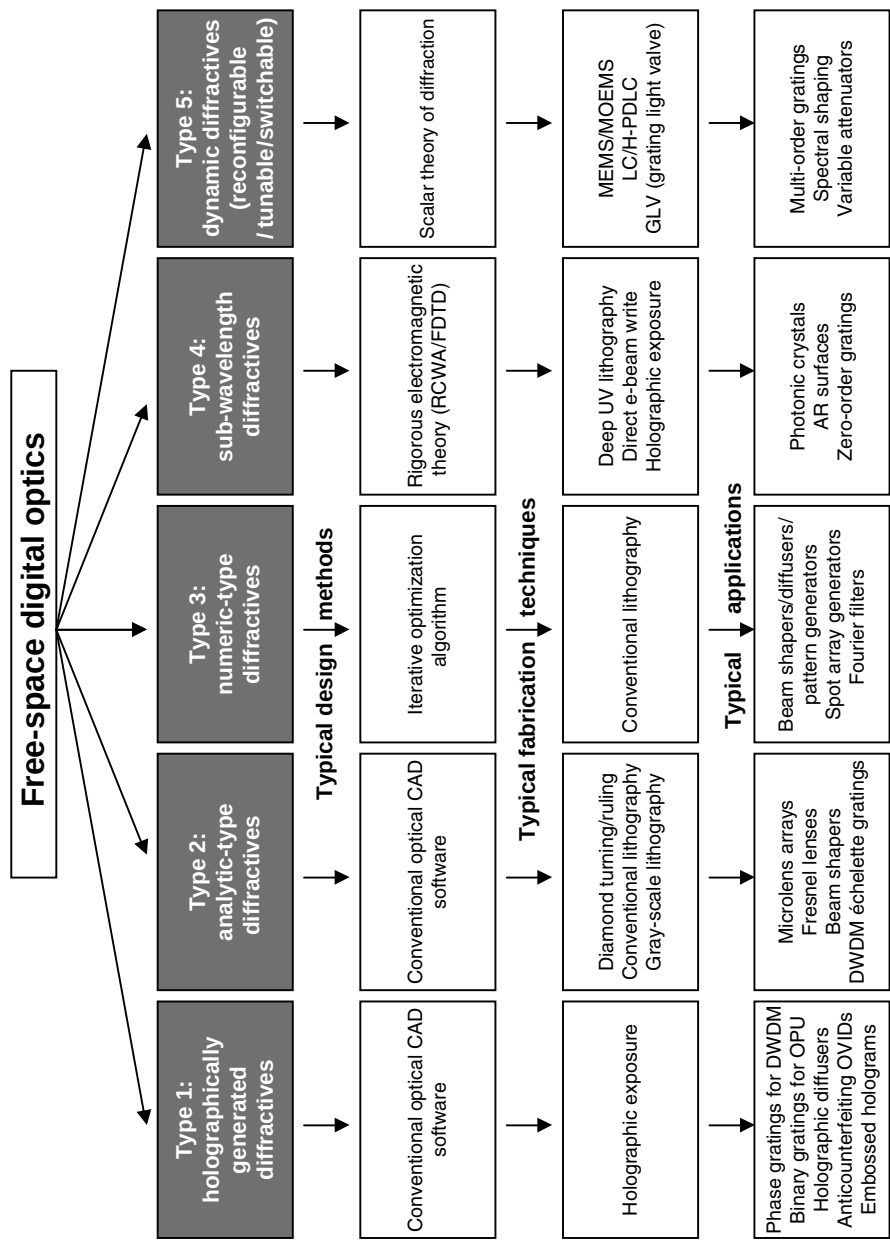


**Figure 5.1** The five main types of digital diffractive optics

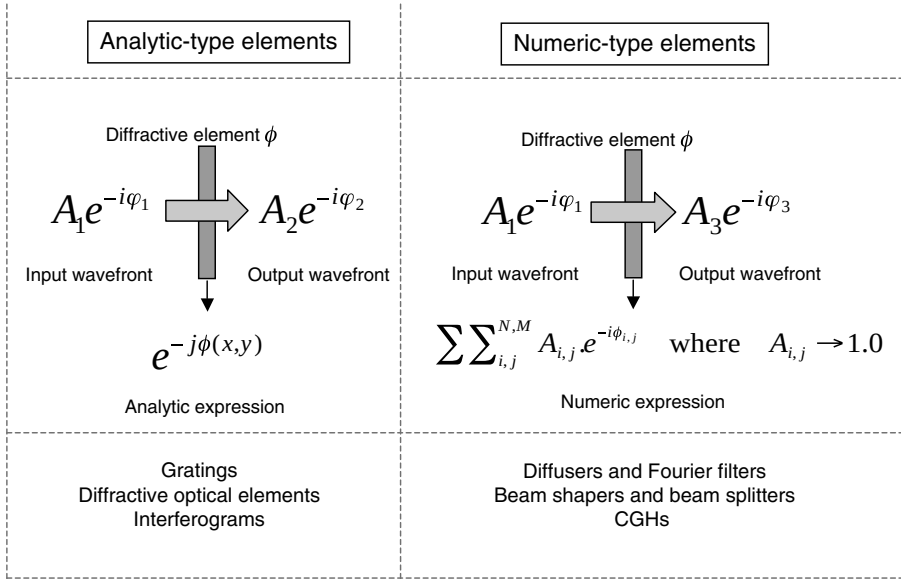


**Figure 5.2** Analytic- versus numeric-type digital diffractive elements

## 5.1 Analytic and Numeric Digital Diffractives

In order to further classify digital diffractive optics, two additional categories are used: analytic-type elements and numeric-type elements (see Figure 5.2):

- Analytic-type elements are elements that can be calculated in an analytic way, such as by solving an analytic equation (i.e. Type 2 elements – DOEs).
- Numeric-type elements are elements that cannot be calculated analytically, since the diffraction problem is usually much more complex than for analytic-type elements. Therefore, heavy numeric calculations are used to compute these elements on a computer, such as through the use of iterative optimization algorithms. Numeric elements include mostly Type 3 elements – CGHs.

This chapter will focus on analytic-type diffractives and Chapter 6 will focus on numeric-type diffractives.

## 5.2 The Notion of Diffraction Orders

As seen in Chapter 1, in order to produce a grating with an optimum diffraction efficiency, it is necessary to precisely etch down the grooves to a depth that would introduce a maximal local destructive interference, which in the blazed grating example yields a  $2\pi$  phase shift, and in the case of a binary phase grating would yield a  $\pi$  phase shift. So, when considering the fundamental positive diffraction order, where does the remaining light go? Well, most of the light that is not coupled into the positive fundamental order will leak into the conjugate order and the higher orders (direct or conjugates). Also, the light will leak into quantization noise and other optical noise artifacts linked to systematic fabrication errors (see also Chapter 13).

### 5.2.1 Fundamental and Higher Diffraction Orders

A diffraction order is triggered if the local constructive interference condition is satisfied (which is similar to a  $\phi = 2\pi$  phase shift for continuous structures or a  $\phi = \pi$  phase shift for binary structures – or, by

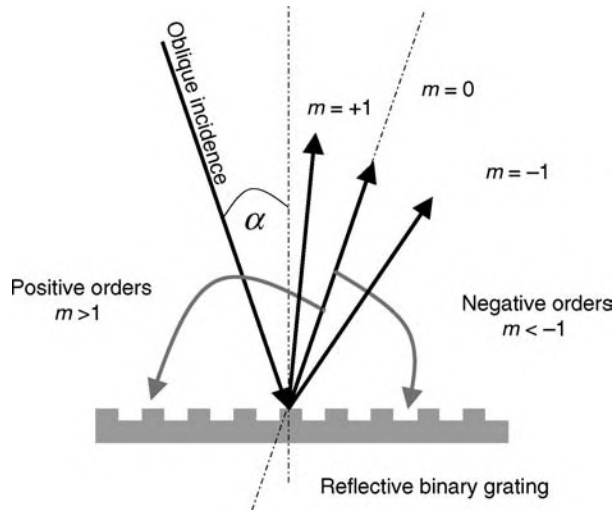


**Figure 5.3** Configuration of the various diffraction orders of a grating

extrapolation,  $\phi = 2\pi(N - 1)/N$  for a multilevel phase element with  $N$  surface steps). This condition is a function of the reconstruction wavelength  $\lambda$ . Therefore, if the wavelength is multiplied by an integer value  $m$ , this condition is again satisfied (because of the  $2\pi$  periodicity of the phase).

The physical significance of this is that the reinforcement of the wavelets diffracted by successive periods of the grating requires that each ‘ray’ be retarded (or advanced) in phase with every other; this phase difference must therefore correspond to a real distance (optical path difference) that equals a multiple of the wavelength. So, the diffraction orders are positive or negative depending on whether the integer is positive or negative (see Figure 5.3).

Therefore, it is important to note that a diffractive element (grating or lens) can be optimized for a specific wavelength  $\lambda_1$  to be diffracted in a specific direction (in a specific order  $m_1$ ). In addition, a diffractive element can also be optimized at the same time for another wavelength  $\lambda_2$  in another specific order  $m_2$ . This principle is used in optical DWDM channels multiplexing, through the use of reflection or transmission gratings (for more details, see Chapter 16).

Occasionally, the desired diffraction order is not the fundamental order, but a much higher order. For example, diffraction orders up to 15 are commonly used in DWDM gratings. Such gratings are easily fabricated by diamond tool ruling with a diamond tip that has the same geometry as the grooves.

In many cases, the application may not call for the highest diffraction efficiency, and diffractives might be tuned to a specific efficiency as seen in Section 16.8.2.2 (i.e. beam samplers and hybrid DVD/CD lenses). In other cases, it is best to have various diffraction orders rather than only one or two.

However, in most applications, a unique diffraction order with the highest efficiency is most desirable. Therefore, a quasi-continuous surface-relief profile is sought. However, as seen in Chapter 12, this is also the most difficult and most expensive diffractive profile to fabricate.

### 5.2.2 Diffraction Orders for Fresnel and Fourier Elements

Having introduced the concept of diffraction orders for generic diffractive elements, next we will discuss how these diffraction orders govern the behavior of diffractive optics, and how they differ between Fourier elements and Fresnel elements. Figure 5.4 shows both the Fourier and the Fresnel operating regimes (the far-field and near-field diffraction regimes) and associated diffraction order configurations.

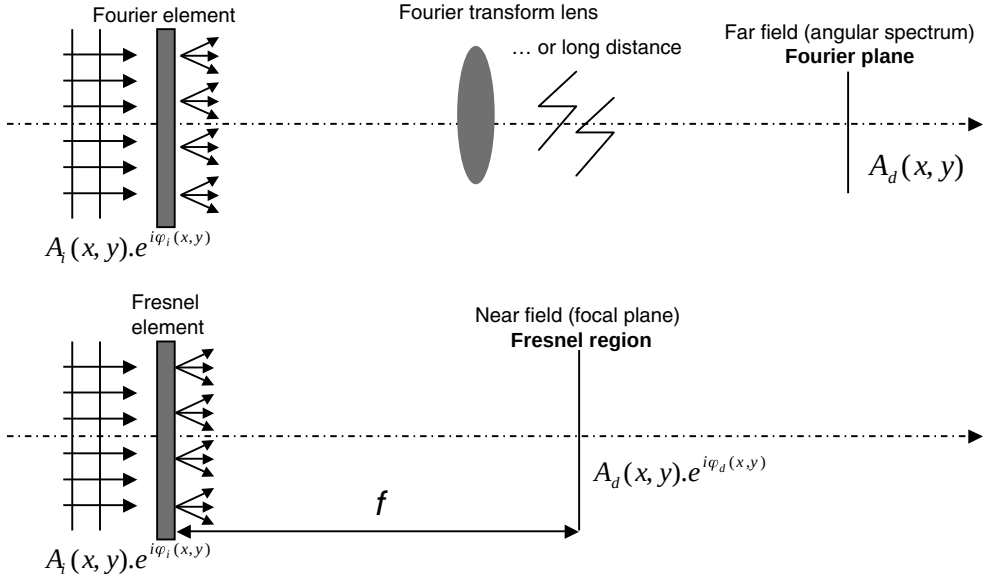


**Figure 5.4** The Fourier and Fresnel operating regimes

The Fourier and Fresnel diffractive elements are defined below:

- A Fourier diffractive element does not affect the convergence divergence or collimation degree of the incoming beam but, rather, affects the direction of propagation of this wavefront, in a single beam or multiple beams, all having the same characteristics as the incoming wavefront (beam). If the incoming beam is collimated, the diffracted beam(s) is (are) also collimated. Therefore, the reconstruction of a Fourier element (provided that the incoming beam is collimated) occurs in the far field (or angular spectrum, also called the Fourier plane). All the diffraction orders are ‘focused’ and conjugate orders are symmetric with respect to the optical axis. A simple example of a Fourier element is a grating (1D or 2D) or a 1 to  $N$  beam splitter. Usually, such elements are composed of periodic structures, which do not look like fringes in the 2D case but, rather, like complex shapes (see, e.g., Fourier-type CGHs).
- A Fresnel diffractive element, on the other hand, does affect the vergence of the incoming wavefront, in addition to introducing diffraction functionalities similar to those that a Fourier element could introduce. Such elements are considered to be lenses. These elements are composed of nonperiodic structures and usually yield fringe-type structures (as in a diffractive Fresnel lens). Note that a CGH can be either a Fourier or a Fresnel element.

The far-field angular spectrum of a Fourier element can also be projected into the near field by the use of a refractive (or diffractive) Fourier transform lens, as depicted in Figure 5.4. This is a more common set-up that is used in many applications (see Chapter 16).

Figure 5.5 shows the various diffracted wavefronts of a typical Fourier element (grating) in the far field, and Figure 5.6 shows the various diffracted wavefronts of a typical Fresnel element (lens) in the near field.

In the case of a Fresnel element, when illuminated by a collimated wavefront, the converging wavefronts constitute the positive diffraction orders and the diverging wavefronts constitute the negative diffraction orders.

Take the simple example of a binary-phase grating in normal incidence, which mainly produces two diffraction orders (the real and conjugate fundamental orders, yielding a theoretical 40.5% efficiency).

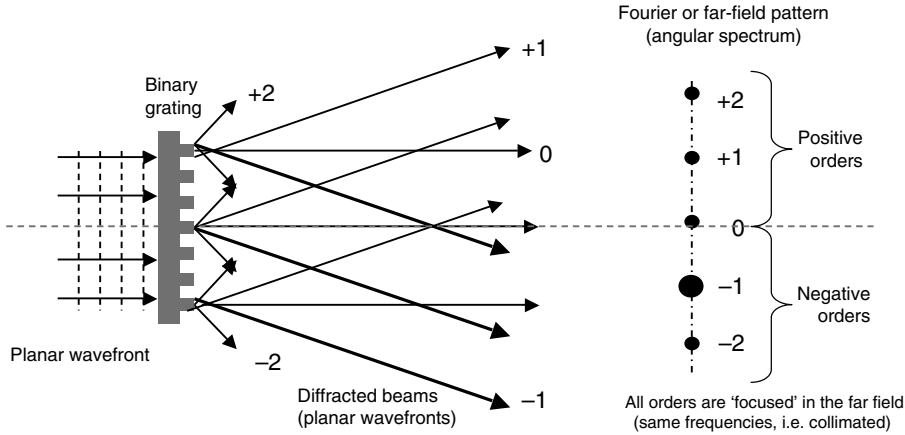


**Figure 5.5** The diffraction orders configuration in the far field for a Fourier element

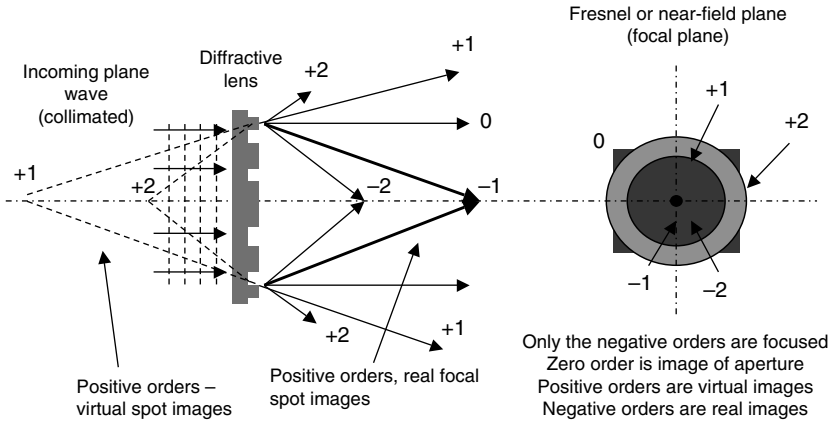


**Figure 5.6** The diffraction orders configuration in the near field for a Fresnel element

Such orders propagate with similar energies in symmetric spatial directions, thus producing in the far field (the angular spectrum of plane waves) two spots symmetrical to the zero-order spot (which sets the optical axis of the system, since a Fourier element does not have an optical axis).

In the case of a binary-phase diffractive lens, the exact same phenomenon is produced (since locally a Fresnel lens can be considered as a linear grating – see Section 11.1.3): mainly two diffracted wavefronts (the conjugate fundamental orders). However, here, the positive diffraction order produces a converging wave and the negative order produces a diverging wave, each with the same focal length. Therefore, a binary-phase Fresnel lens is at the same time a converging and a diverging lens. Such a binary lens has no counterpart in the refractive realm (see Chapter 1).

### 5.3 Diffraction Gratings

The simplest analytic elements are one-dimensional gratings [1]. The structure geometry can be easily calculated for a given diffraction angle by using the grating equation:

$$m\lambda = \Delta(n_1 \cdot \sin \alpha + n_2 \cdot \sin \beta) \quad (5.1)$$

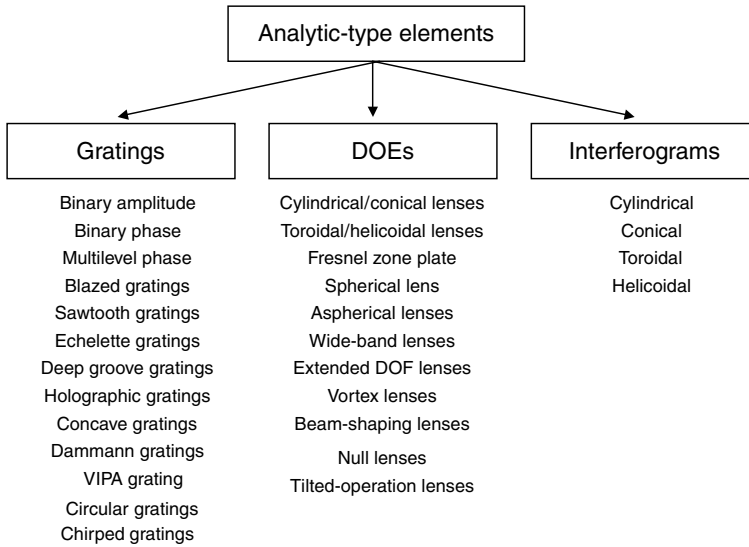


**Figure 5.7** The various analytic-type digital diffractive optics

where  $m$  is the diffraction order,  $\lambda$  is the reconstruction wavelength,  $\Lambda$  is the grating period,  $n_1$  and  $\alpha$ , respectively, are the index of refraction and the incident angle, and  $n_2$  and  $\beta$ , respectively, are the index of refraction and the diffracted angle in the grating material.

In the case of a reflection grating in air,  $n_1 = 1$  and  $n_2 = 1$ , and thus we obtain the simplified grating equation:

$$m\lambda = \Lambda(\sin \alpha + \sin \beta) \quad (5.2)$$

There are numerous ways to implement a grating (binary, multilevel, continuous profile, slanted, etc.) and in various materials (amplitude, phase or in combination). Two-dimensional gratings are mostly numeric elements (CGHs), with the exception of 2D Dammann gratings, which are treated as analytic elements.

After a description of the various gratings used in industry today, diffractive lenses (DOEs), which are an extrapolation of circular gratings (circular gratings with varying periods), will be considered next. Finally, interferogram-type elements, which can implement special effects (especially lenses, which cannot be implemented by DOEs) will be presented (see Figure 5.7).

Historically, linear diffraction gratings were the first type of diffractive element to be studied, fabricated and used successfully in industrial applications. Linear diffraction gratings still account for most of the diffractives used in industry today (spectroscopy, DWDM telecom applications, optical security devices, optical data storage, optical sensors etc.).

Most of the linear diffraction gratings used in industry today are designed with the grating equation in mind and thus are analytic-type elements (see Chapter 2). Due to their rather simple geometry, there is no need to use a special CAD tool to design such elements. Linear diffraction gratings can be fabricated by a wide variety of techniques and technologies, from diamond ruling to holographic exposure and microlithography (see Chapters 12–14).

### 5.3.1 Some Useful Grating Parameters

Historically, the primary purpose of a grating [2] was to disperse light spatially into its spectrum (Figure 5.8). The dispersion of a grating is the measure of the angular separation between the different wavelengths constituting the incoming beam.

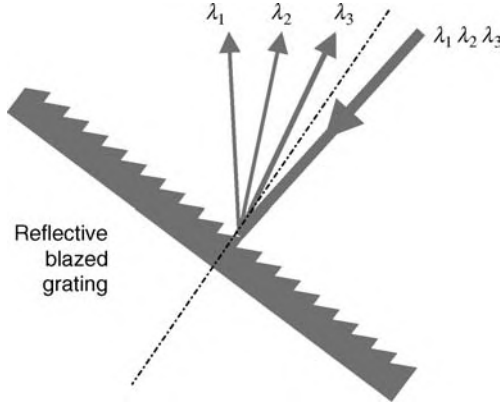


**Figure 5.8** The angular dispersion of a grating

### 5.3.1.1 Angular Grating Dispersion

Angular dispersion expresses the spectral range per unit angle. Linear resolution expresses the spectral range per unit length.

By differentiating the grating equation for a reflective grating (Equation (5.2)), the angular dispersion  $D$  can be obtained as follows:

$$D = \frac{\partial \beta}{\partial \lambda} = \frac{m}{\Lambda \cos \beta} = \frac{\sin \alpha + \sin \beta}{\lambda \cos \beta} \quad (5.3)$$

where  $\Lambda$  is the period of the grating,  $\lambda$  is the incident wavelength,  $\alpha$  is the incident angle,  $\beta$  is the diffracted angle, and  $m$  is the target diffraction order considered in the application.

### 5.3.1.2 Linear Grating Dispersion

The linear dispersion of a grating is the product of the angular dispersion  $D$  (see Equation (5.3)) and the effective focal length  $f$  of the system:

$$\begin{cases} f \cdot D = \frac{mf}{\Lambda \cos \beta} = \frac{Gmf}{\cos \beta} \\ P = \frac{1}{f \cdot D} = \frac{\Lambda \cos \beta}{mf} \end{cases} \quad (5.4)$$

where we have introduced the grating frequency  $G = 1/\Lambda$ .

### 5.3.1.3 Grating Resolving Power

Finally, the resolving power  $R$  of such a grating can be expressed as

$$R = \frac{\lambda}{\delta \lambda} = mN_0 = \frac{N_0 \Lambda (\sin \alpha + \sin \beta)}{\lambda} \quad (5.5)$$

where  $\delta \lambda$  is called the limit of resolution of the system and  $N_0$  is the number of grating grooves illuminated by the incoming beam of light.



If  $W$  is the width of the grating and if the maximum angles reachable are used ( $|\sin \alpha + \sin \beta| < 2$ ), the maximum resolving power  $R$  is derived as follows:

$$R_{\max} = \frac{2W}{\lambda} \quad (5.6)$$

### 5.3.1.4 Free Spectral Range

As the grating equation is satisfied for a different wavelength for each integral diffraction order  $m$ , light from different wavelengths can be diffracted along the same direction in space: light of wavelength  $\lambda$  in order  $m$  is diffracted along the same direction as light of wavelength  $\lambda/2$  in order  $2m$  and so on.

The range of wavelengths in a given spectral order for which superposition of light from adjacent orders does not occur is called the free spectral range  $F_\lambda$ , and is expressed as follows:

$$\begin{cases} F_\lambda = \Delta\lambda = \frac{\lambda_1}{m} \\ \lambda_1 + \Delta\lambda = \frac{m+1}{m} \lambda_1 \end{cases} \quad (5.7)$$

The concept of the free spectral range can be applied to any grating that can diffract in more than one order (holographic gratings, for example – see Chapter 8), but is best suited for *échelette* gratings (see the following sections). Such *échelette* gratings can diffract light from overlapping spectral orders and are very useful in DWDM telecom Mux/Demux applications.

### 5.3.2 Binary Gratings

The simplest example of a grating is the binary amplitude grating, composed of alternating opaque and transparent linear regions, which constitute the period of the grating (see Figure 5.4). Such gratings have very low diffraction efficiencies (see below). A binary-phase grating (see also Figure 5.9) can yield much

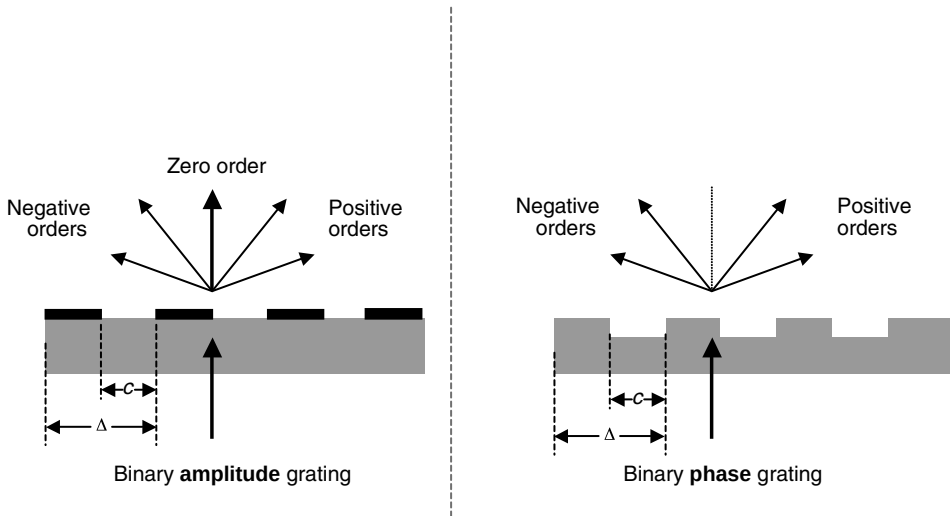


**Figure 5.9** Binary diffraction gratings

higher efficiency if the groove depths yield a phase shift of  $\pi$ . The maximum efficiency of such a grating in the fundamental order (negative or positive) is 40.5%, with the conjugate fundamental order having the same efficiency and the rest split over higher orders (the zero order has no intensity when the depth yields a phase shift of exactly  $\pi$ ).

The symmetrical geometry of binary gratings (amplitude or phase) cannot tell the incoming light in which direction to diffract, so the light diffracts in a symmetrical way for all the propagating diffraction orders. Therefore, a binary grating in normal incidence has similar energies in real and conjugate orders. By breaking the symmetry (going multilevel, for example, or having a tilted incidence), the energy balance in the real and conjugate orders is also broken.

Next, the diffraction efficiency for binary-amplitude gratings is derived in the various diffraction orders present.

### 5.3.2.1 Diffraction Efficiency Calculations for Binary Amplitude Gratings

The scalar theory of diffraction (see Appendix A) can be used effectively to predict the diffraction efficiency of binary amplitude or phase and multilevel phase gratings in the realm of validity of the theory. Based on Fraunhofer's formulation of the diffracted field far away from the diffractive element, the diffraction efficiency formulation can be derived for amplitude gratings (Figure 5.4).

The amplitude of the grating function can be expressed as a Fourier expansion:

$$a(x, y) = \begin{cases} 1 & \text{if } m\Lambda \leq x \leq c + m\Lambda \\ 0 & \text{elsewhere} \end{cases} \Rightarrow a(x, y) = \frac{c}{\Lambda} \sum_{m=-\infty}^{+\infty} e^{-i\pi(\frac{mc}{\Lambda})} \frac{\sin(m\pi\frac{c}{\Lambda})}{m\pi\frac{c}{\Lambda}} e^{2i\pi\frac{x}{\Lambda}} \quad (5.8, 5.9)$$

where each exponential term represents a plane wave, and  $m$  is the index of the different diffraction orders present.

The magnitude of this  $m$ th diffraction order can be expressed as follows:

$$\eta_m = \left| \frac{\sin(m\pi\frac{c}{\Lambda})}{m\pi} \right|^2 \quad (5.10)$$

For regularly spaced opaque/transparent lines (i.e. a grating with a duty cycle of 50%), the diffraction efficiency formulation in the  $m$ th order becomes

$$\eta_m = \left| \frac{\sin(\frac{m\pi}{2})}{m\pi} \right|^2 \quad (5.11)$$

### 5.3.3 Multilevel Gratings

If the final application uses both diffracted orders, the effective efficiency increases up to 81%. The efficiency of such binary-phase gratings depends greatly on the etch depth, as well as the quality of the side walls and the surface roughness.

Multilevel phase gratings (see Figure 5.10) attempt to reach the analog surface-relief gratings geometry (either blazed or échelette) and yield higher efficiency than binary-phase gratings.

#### 5.3.3.1 Multilevel Phase Gratings Efficiency Calculations

The diffraction efficiency for multilevel phase-relief gratings can be derived in a similar way to that of amplitude gratings. See, for example, the quadratic grating (four phase levels) depicted in Figure 5.10 (the center element).

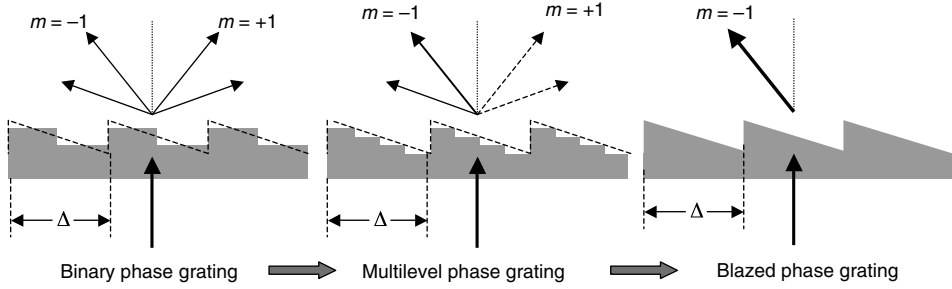


**Figure 5.10** From binary to multilevel to blazed surface-relief phase gratings

If the grating is etched so that the maximum optical path difference yields  $2\pi(N-1)/N$  in transmission, which maximizes the efficiency (see Section 5.1), the diffraction efficiency in the  $m$ th order for  $N$  phase levels is expressed as follows:

$$\eta_m^N = \left| \frac{\sin\left(\frac{m\pi}{N}\right)}{\frac{m\pi}{N}} \right|^2 \quad (5.12a)$$

When the depth of the structures is not optimal (due, for example, to etch depth errors or even to a design where a specific nonoptimal efficiency is required), the efficiency can be expressed as a function of the wavenumber  $\kappa$  (a wavenumber of  $\kappa = 1$  means perfect etch depth accuracy, e.g.  $2\pi(N-1)/N$  for  $N$  levels):

$$\eta_m^N = \left[ \frac{\sin(\pi(m-\kappa))}{\sin\left(\frac{\pi(m-\kappa)}{N}\right)} \cdot \frac{\sin\left(\frac{\pi m}{N}\right)}{\pi m} \right]^2 \quad (5.12b)$$

In fact, in many cases, it is more desirable to achieve a specific efficiency rather than the maximal reachable efficiency – for example, for the dual-focus hybrid lens of a CD/DVD read-out (see also Chapter 16).

Figure 5.11 shows the efficiency of a binary grating in various orders, where the efficiency is presented as a function of the wavenumber  $\kappa$ .

Table 5.1 summarizes the theoretical predicted values of diffraction efficiencies for both amplitude- and phase-relief diffractives at the optimum wavelength.

Note that the numbers given in Table 5.1 are theoretical values, derived from the scalar theory of diffraction through infinitely long and perfect gratings. In real life, these figures are never reached, since many effects will accumulate to reduce efficiency, for example:

- Fabrication-related effects:
  - fabrication errors (etch depth errors and lateral misregistrations);
  - the quality of the side walls;
  - surface roughness.
- Design-related effects:
  - the validity of scalar theory for a feature smaller than  $5\lambda$ .
- Operation-related effects:
  - the coherence of source (LED or laser?);
  - the collimation of source;
  - the incidence of the launch beam (shadowing effects);
  - the polarization state of the source.

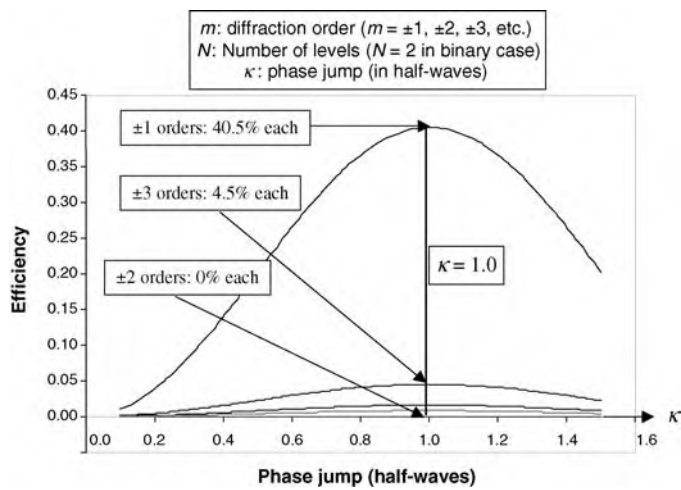


**Figure 5.11** The diffraction efficiency for binary grating in various orders, as a function of  $\kappa$

For details of how the fabrication errors affect the performance of the diffractive, see Chapter 12. Refer to Chapter 11 for details of how the design inaccuracies and source variations would affect the performance of the diffractive.

The results can be easily extrapolated in Table 5.1 from infinitely long linear gratings to more complex structures that are limited in space, such as Fresnel lenses and general CGHs (which can be approximated locally by such linear gratings).

When the number of phase levels grows to 16 or more, the diffractive element can easily be considered as a quasi-analog surface-relief element. In fact, it does not make sense to fabricate a diffractive element with more than 16 levels using conventional multilevel masking techniques, since the successive systematic lateral misalignment errors and cascaded etching depth errors would reduce the diffraction efficiency dramatically. The increase in the diffraction efficiency when moving from 16 levels to a higher number of levels (32, 64, 128 or 256), even for perfect fabrication (an impossible task), is infinitesimal (a fraction of a percent), while the fabrication effort is enormous and thus not practical.

**Table 5.1** The diffraction efficiency at optimum wavelength for various grating types

| Physical mode        | Zero order | Fundamental positive order + 1 | Fundamental negative order - 1 | + 2 order | - 2 order | + 3 order | - 3 order |
|----------------------|------------|--------------------------------|--------------------------------|-----------|-----------|-----------|-----------|
| Binary amplitude     | 25%        | 10.1%                          | 10.1%                          | 0%        | 0%        | 1.1%      | 1.1%      |
| Sinusoidal amplitude | 35%        | 6%                             | 6%                             | 0%        | 0%        | 5%        | 5%        |
| Two phase levels     | 0%         | 40.5%                          | 40.5%                          | 0%        | 0%        | 4.5%      | 4.5%      |
| Four phase levels    | 0%         | 81%                            | 0%                             | 0%        | 0%        | 0%        | 10%       |
| Eight phase levels   | 0%         | 94.9%                          | 0%                             | 0%        | 0%        | 0%        | 0%        |
| Sixteen phase levels | 0%         | 98.7%                          | 0%                             | 0%        | 0%        | 0%        | 0%        |
| Blazed phase         | 0%         | 100%                           | 0%                             | 0%        | 0%        | 0%        | 0%        |
| Sinusoidal phase     | 12%        | 34%                            | 34%                            | 0%        | 0%        | 0%        | 0%        |

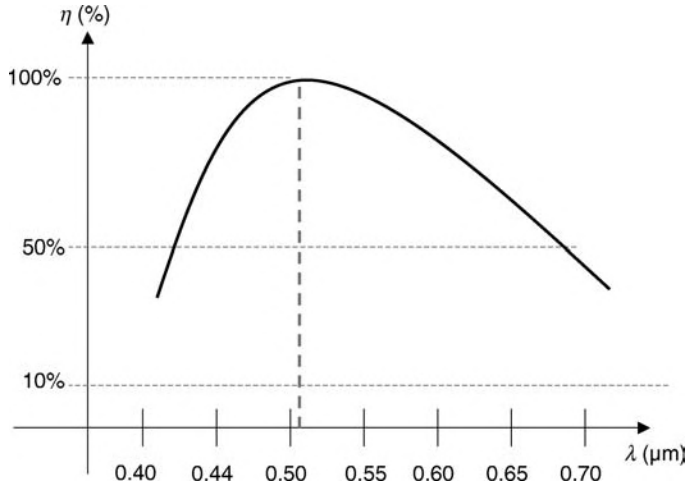


**Figure 5.12** The diffraction efficiency in the fundamental order for a quasi-analog surface-relief profile (lens or grating)

### 5.3.4 Blazed Gratings

Blazed gratings are, as depicted in Chapter 1, replicated linear micro-prism structures. Such blazed gratings have the highest efficiency in the scalar regime of diffraction when the maximum height of these structures reaches a phase shift of  $2\pi$  for the reconstruction wavelength.

A blazed grating can yield 100% efficiency theoretically (see Figure 5.12), but practically, due to side-wall inaccuracies and etch depth errors, as well as surface roughness, such gratings have real efficiencies of around 90–95%. The optimum technique to fabricate such gratings is by diamond ruling (see Chapter 12). The diffraction efficiency of a blazed grating element (for  $m = 1$  and  $N \rightarrow \infty$ ) is shown in Figure 5.12.

The design wavelength for the blazed grating in Figure 5.12 was 514 nm. Due to shadowing effects and the cosine effect when illuminating the grating at an angle, the best operating configuration for a blazed grating is actually normal incidence or incidence normal to the local slope.

Figure 5.13 shows the diffraction efficiency of a blazed grating (or, by direct extrapolation, a diffractive lens) for the diffraction orders  $+1$  (fundamental),  $+2$  and  $+3$ , as a function of the reconstruction wavelength.

The blazed element considered in Figure 5.13 is designed and fabricated to yield maximum efficiency in the fundamental order in the green region of the spectrum (550 nm). As the efficiency in the fundamental drops when going from green to lower wavelengths, the efficiency in the higher orders increases.

### 5.3.5 Grating Implementations

The various grating implementations that give rise today to numerous applications are reviewed in the sections that follow (for related applications, see Chapter 16).

#### 5.3.5.1 Sawtooth or Échelette Gratings

Sawtooth gratings are very similar to blazed gratings, the difference being that they do not yield straight side walls. This is actually a nice feature, since the shadowing effect discussed in the previous sections would be minimal here and is usually easier to fabricate [3].

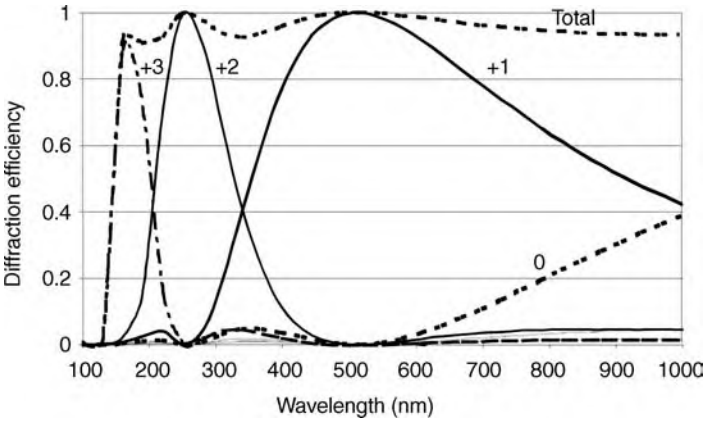


**Figure 5.13** The diffraction efficiency of a blazed grating for + 1, + 2 and + 3 orders

Figure 5.14 shows a reflective linear sawtooth grating (or *échelette* grating), which is used for its unique spectral dispersion characteristics (as in a DWDM Demux).

**5.3.5.2 Holographic Surface-relief Gratings**

Holographic surface-relief gratings are gratings recorded in photoresist as surface-relief elements. They are considered to be thin elements rather than thick elements, as are index-modulated holographic gratings (for definitions of thin and thick holograms, see also Chapter 8).

Usually, a holographic exposure would yield a sinusoidal profile, but it can also yield more complex profiles (see Chapter 8). In the case of a sinusoidal grating, the overall efficiency remains low and the zero order strong, even in the case of an optimum phase shift. The efficiency in the two conjugate first orders is 33%, whereas the rest of the light is located mainly in the zeroth order (see Figure 5.15).

The surface relief is usually produced in a photoresist layer, followed by a post-bake process. However, in some cases, where a high-power laser or stability over time is required, this profile can be etched down into the underlying substrate by proportional RIE etching (see Chapter 12).

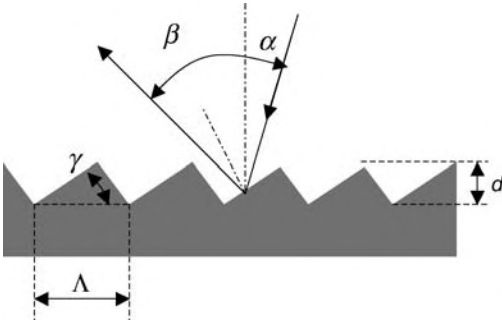


**Figure 5.14** The parameters of a reflective sawtooth linear grating

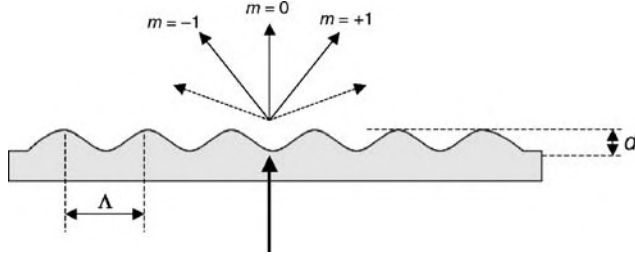


**Figure 5.15** The parameters of a holographic surface-relief grating

### 5.3.5.3 Concave Gratings

Concave gratings are used in reflection mode. The local blaze angle is constantly changing along the sag in order to accommodate the local curvature of the substrate (each facet should be perpendicular to the incoming light to produce maximum efficiency). Figure 5.16 shows a concave grating where monochromatic incident light diverges from a point source A and falls onto the grating. This beam is then diffracted to point B.

If  $r_A$  and  $r_B$  are the distances between the point sources and O is the center of the concave grating, the  $f$  numbers of the system can be written as follows:

$$\begin{cases} f_{input} = \frac{r_A}{W \cdot \cos \alpha} \\ f_{output} = \frac{r_B(\lambda)}{W \cdot \cos \beta} \end{cases} \quad (5.13)$$

The focal length of a concave grating is an important parameter when it comes to designing a grating spectrometer, since it governs the size of the optical system. The ratio of the input focal length to the output

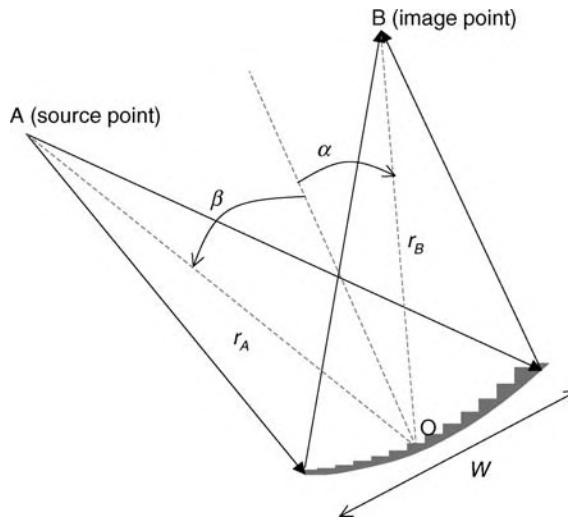


**Figure 5.16** The parameters of a concave grating configuration

focal lengths determines the projected width of the entrance slit that must be matched to the exit slit of the detector element size.

5.3.5.4 Dammann Gratings

Dammann gratings [4–6], are one- or two-dimensional separable phase gratings that may be used to split an incoming beam into a 1D or 2D array of beamlets. Being gratings, they have a periodicity. Within a single period, there are  $2N$  phase transition points, whose positions can be optimized in order to yield a set of  $2N-1$  diffracted orders with equal intensity, thus producing a set of 1D or 2D spots arrays in the far field (angular spectrum). Figure 5.17 shows several one-dimensional grating periods and their far-field reconstructions.

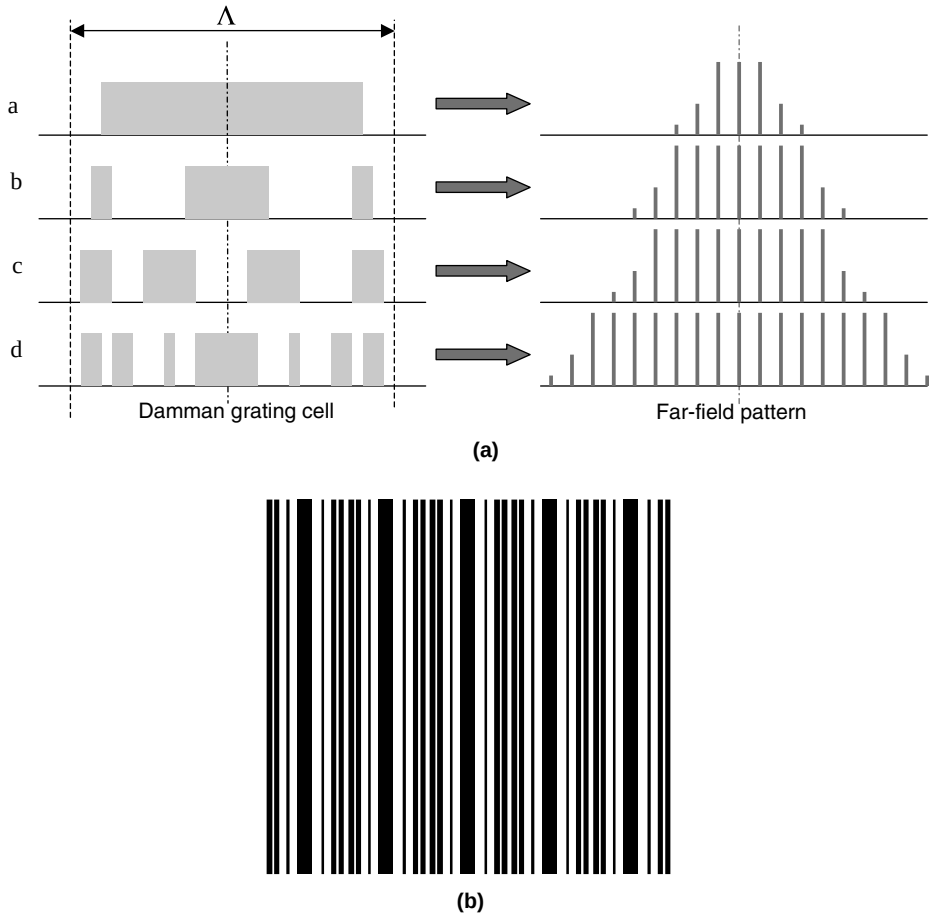


**Figure 5.17** (a) One-dimensional Dammann grating periods and far-field reconstructions in 1D (a, three beams; b, five beams; c, nine beams; d, 15 beams). (b) Replicated periods from Damman grating ‘d’ in (a)



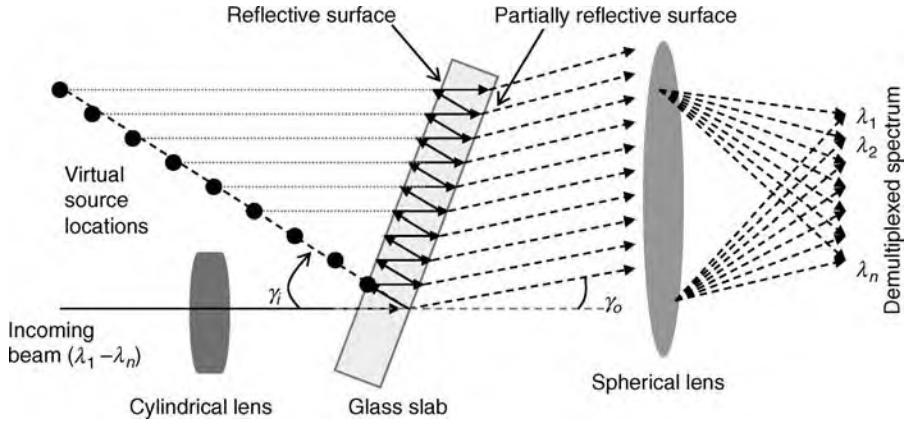


**Figure 5.18** The VIPA grating concept

Figure 5.17(b) shows a Damman grating formed by a linear succession of the basic periods. The grating pictured is 'c' in Figure 5.17(a), which produces a set of 15 beams in the far field. Applications using Damman gratings range from optical computing, image processing and fiber optic star coupling to coherent summation of incoherent laser beams and parallel laser processing

### 5.3.5.5 The Virtual Imaged Phase Array Grating

Virtual Imaged Phase Array (VIPA) gratings are very interesting in the sense that they are not real physical gratings like the ones described in the previous sections but, rather, they are virtual gratings.

Such a VIPA grating does not require any microstructuring of its surface and is a thin planar tilted glass slab, with one reflective surface and one partially reflective surface (see Figure 5.18). A VIPA grating is actually a virtual grating that can behave as a super-grating; namely, having very large angular dispersion and a low polarization sensitivity. This technology was developed especially for chromatic dispersion compensation applications in DWDM telecom networks.

The VIPA grating has about a 10–20 times larger angular dispersion than common gratings, which is proportional to  $\gamma_i$ . It also has a low polarization sensitivity, a simple structure, a low cost and a compact size (no microstructuring needed).

### 5.3.5.6 Arrayed Waveguide Gratings (AWG Gratings)

Arrayed waveguide gratings are similar to VIPA gratings in the sense that they do not need any grating structures. However, as opposed to VIPA gratings, the AWG gratings use real sources that come from waveguide structures. For this reason, AWG gratings are discussed in Chapter 3: for more insight on AWG gratings, see Chapter 4.

### 5.3.5.7 Arbitrarily Shaped Gratings

Circular gratings are simply linear gratings whose grooves have been curved in order to form circles or other shapes. Circular gratings can be implemented in many ways much like linear gratings (binary, multilevel, blazed, sawtooth etc.). Circular gratings have unique properties due to the fact that they diffract light into all directions of space (in the case of a  $360^\circ$  circular grating – see Figure 5.19).

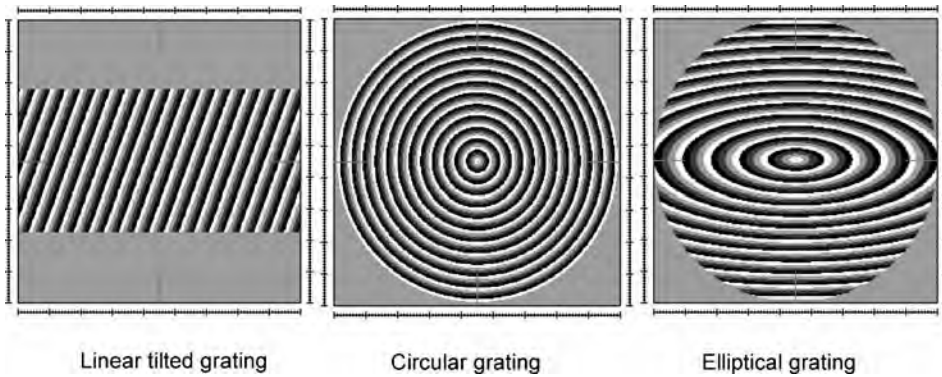


**Figure 5.19** Linear, circular and elliptical gratings

A circular grating would diffract an incoming beam into a cone of light, the angular value of the cone being a function of the wavelength.

A unique property of such circular gratings is their polarization insensitivity, which is the same for any incoming polarization. However, it is important to note that the losses of such gratings are high, and are actually due to the same polarization effects as for linear gratings, but these losses remain unchanged for any polarization, therefore introducing a virtual polarization insensitivity. Such gratings can also be implemented as axicons, which produce nondiffracting beams and are used in applications where extended depth of focus is needed.

### 5.3.5.8 Chirped Gratings

Chirped gratings are linear (or circular) gratings that have increasing periods along one or more directions. The period increases linearly. If the period increases in a nonlinear way, the grating becomes a lens (a highly aberrated cylindrical or spherical lens).

Chirped gratings are used in applications where the spectral dispersion or the deflection angle has to vary with the position of the beam along the grating. Examples are pulse compressors or chirped fiber Bragg gratings. When the periods are much smaller than the wavelength, chirped sub-wavelength gratings can be used as effective index media (see Effective Medium Theory – ELMT – in Chapter 11). There are many other grating types and grating concepts mentioned in the literature and used in industry, but the architectures described above are the ones that are most often used today.

By not introducing any power in the incoming beams, grating elements are considered Fourier elements, in opposition to diffractive lenses, which are considered Fresnel elements since they introduce power in the incoming beams (focusing, diverging etc.).

The sections that follow describe the various Fresnel digital diffractives, also known as diffractive lenses.

As a transition between gratings (Fourier elements) and diffractive lenses (Fresnel elements), it is interesting to note that when a nonlinear chirp is combined over a circular grating (either of the last two grating types described in the previous section), a Fresnel element is built (i.e. a spherical diffractive Fresnel lens). A nonlinear chirp in a grating, as discussed previously, actually destroys the Fourier nature of the element and introduces its Fresnel nature (lensing effect).

### 5.3.5.9 Optically Variable Imaging Devices (OVIDs)

Optically Variable Imaging Devices (OVIDs) are sets of spatially multiplexed gratings in any configuration and in any geometry. These elements are basically ‘eye candy’ and are used as anti-counterfeiting devices on consumer products such as garments, CDs or electronics, usually on the sticker or tags. Such elements can be originated holographically or lithographically. OVIDs have no other optical functionality

than to produce a nice spectral dispersion from sunlight or other sources of illumination. Curved gratings, chirped gratings and arbitrarily shaped grating geometries can produce very interesting pseudo-dynamic optical effects. When the observer moves his or her head, or if the OVID is moved around, the diffraction patterns hitting the observer's eye vary in intensity and color, thus providing a pseudo-dynamic effect.

### 5.3.5.10 The Talbot Self-imaging Effect

When illuminated with highly coherent light, two-dimensional gratings produce self-images and self-image-like field distributions in near-field planes behind the grating. The lens-like imaging property produced only by free-space propagation of diffracted light is known as the Talbot effect [7] (see Figure 5.20).

See Appendix B for the expression of the diffracted field at a distance, where  $z$  can be expressed by the diffractive element description  $H(x,y)$  multiplied by the incoming divergent wavefront (wavefront curvature  $R$ ) and convoluted with a quadratic function known as the Fresnel kernel:

$$U(x, y, z) = (U(x, y, 0) \cdot H(x, y) \cdot e^{-j\frac{\pi}{2R}(x^2 + y^2)}) \otimes e^{-j\frac{\pi}{2z}(x^2 + y^2)} \quad (5.14)$$

As the grating pattern is periodic, and is separable in the  $x$  and  $y$  directions into respective periods  $dx$  and  $dy$ ,  $H(x,y)$  can be expressed as a Fourier series, as follows:

$$H(u, v) = \frac{1}{d_x d_y} \sum_m \sum_n F_x\left(\frac{m}{d_x}\right) F_y\left(\frac{m}{d_y}\right) e^{-2j\pi\left(\frac{m}{d_x}u + \frac{m}{d_y}v\right)} \quad (5.15)$$

where  $F_x$  and  $F_y$  are the separable Fourier coefficients and related spatial frequencies  $u$  and  $v$  defined by

$$\begin{cases} F_x\left(\frac{m}{d_x}\right) = \int_{-\infty}^{+\infty} H_p(u, v) \cdot e^{-2j\pi\frac{m}{d_x}u} du \\ F_y\left(\frac{m}{d_y}\right) = \int_{-\infty}^{+\infty} H_p(u, v) \cdot e^{-2j\pi\frac{m}{d_y}v} dv \end{cases} \quad (5.16)$$

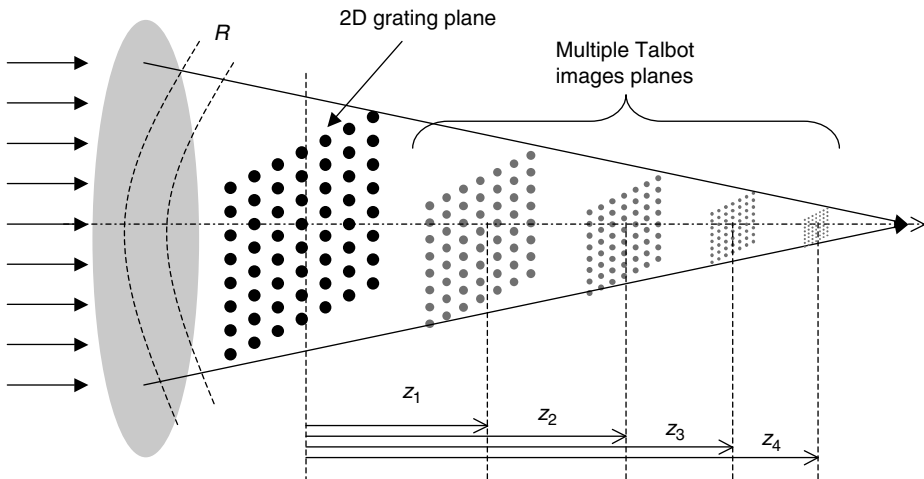


**Figure 5.20** The Talbot self-imaging effect

where  $H_p$  denotes a single two-dimensional period of the entire DOE. After development of the expression of Equation (5.14), the following convolution expression is obtained for the diffracted field at distance  $z$ :

$$U(x, y, z) = \frac{\alpha\gamma}{d_x d_y} \cdot \left( e^{-j\frac{\lambda}{R+z}x^2} \cdot H_p\left(\frac{x}{\gamma}\right) \right) \otimes \left( \sum_m \sum_n e^{-j\frac{\pi}{\lambda\gamma} \left( \frac{m^2}{d_x^2} + \frac{n^2}{d_y^2} \right)} e^{-j2\pi \left( \frac{m}{d_x\gamma}x + \frac{n}{d_y\gamma}y \right)} \right)$$

(5.17)

where  $\begin{cases} \gamma = \frac{(R+z)}{R} \\ \alpha = \frac{1}{\gamma} e^{-2j\frac{\pi z}{\lambda}} \end{cases}$

Now, it is interesting to note that the initial field distribution is reproduced integrally with a magnifying factor, if the quantity

$$e^{-j\frac{\pi}{\lambda\gamma} \left( \frac{m^2}{d_x^2} + \frac{n^2}{d_y^2} \right)} \quad (5.18)$$

is constant for any value of  $m$  or  $n$ . This condition is verified (in the one-dimensional case) if the following equation holds true:

$$z = \frac{2ld_x^2 R}{\lambda R - 2ld_x^2} = z_l \quad (5.19)$$

where  $l$  is an integer.

The reconstruction planes in the near field located at distances  $z_l$  (where  $l=1, 2, \dots$ ) are the fractional Talbot plane distance, and the corresponding effect is the fractional Talbot effect. The fractional Talbot effect has been applied for optical calculus and the implementation of logical operators (XOR, NAND etc.).

Multiple imaging is very useful in optical metrology applications and optical testing. A wafer-to-wafer alignment technique using this effect is described in Chapter 13.

## 5.4 Diffractive Optical Elements

Diffractive Optical Elements (DOEs) constitute the second group of analytic-type elements. The first type, gratings, have already been described in this chapter. Unlike gratings, DOEs have optical power and therefore can change the convergence of the incoming beam, and act as lenses.

The various elements defining the realm of DOEs, as seen in Figure 5.6, have been described earlier in this chapter. The simplest element is the well-known diffractive (spherical) Fresnel lens, or its basic amplitude implementation, the Fresnel Zone Plate (FZP).

However, the realm of DOEs is not limited to these two simple examples and includes numerous elements. The realm of DOEs can implement complex optical functionalities in order to replace refractive lens counterparts, to add functionality to refractives (see Chapter 7) or to implement new optical functionalities that cannot be implemented by refractive lenses (see Section 6.1.3).

### 5.4.1 The Fresnel Zone Plate

The diffractive Fresnel lens is the most straightforward example of a DOE. The diffractive Fresnel lens is calculated and optimized in an analytic way. In most cases, a classical optical CAD tool, based on ray tracing to optimize an aspheric phase profile within a plane or curve, can be used. It is interesting to note that unlike its refractive lens counterpart, fabricating a highly aspheric diffractive lens bears the same price tag as fabricating a simple spherical diffractive lens.

The phase profile used to describe a diffractive element can take on many forms, depending on the optical software used to design and optimize it. There are three main analytic descriptions used in today's optical design software tools to describe diffractive phase profiles:

1. The traditional sag equation:

$$\theta(r) = \frac{2\pi}{\lambda} \left( \frac{Cr^2}{1 + \sqrt{1 - (A+1)r^2C^2}} \right) + \sum_{i=1}^n C_i r^{2i} \quad (5.20)$$

2. Rotationally symmetric elements (simple polynomial in  $r$ ):

$$\theta(r) = \sum_{i=2}^n C_i r^i \quad (5.21)$$

3. Nonrotationally symmetric elements (general polynomial in  $x$  and  $y$ ):

$$\theta(x, y) = \sum_{j=1}^m \sum_{i=1}^n C_i x^i \cdot C_j y^j \quad (5.22)$$

Once the aspheric phase profile is defined in this infinitely thin surface (which can be planar or mapped on a refractive aspheric curvature surface), the phase profile is sliced into  $2\pi$  phase shift slices for the considered wavelength for maximum diffraction efficiency in the fundamental positive order, a process shown in Figure 5.21. This is analogous to the optimization of a blazed grating profile, which is described in Chapter 1 and earlier in this chapter (see Section 5.3.4).

A continuous-profile (blazed) Fresnel lens is actually an extension of a blazed grating where the blazing has been modulated, the groove geometries circumvolutated and the period axially modulated.

When no conventional optical CAD tool is available to design the lens phase profile, a simple alternative consists of computing the position and widths of the successive zones of a spherical Fresnel lens. The following equation and Figure 5.22 show how the fringe positions are computed – integer numbers of waves departing from the desired focal point are aimed at the DOE plane:

$$\phi(r, \lambda) = 2\pi \alpha \left( i - \frac{\sqrt{r^2 + f^2} - f}{\lambda_0} \right), \quad r_i \leq r \leq r_{j+1} \quad (5.23)$$

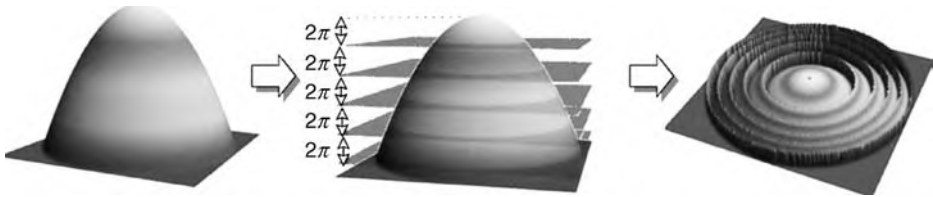


**Figure 5.21** From the initial general phase profile to the DOE fringes

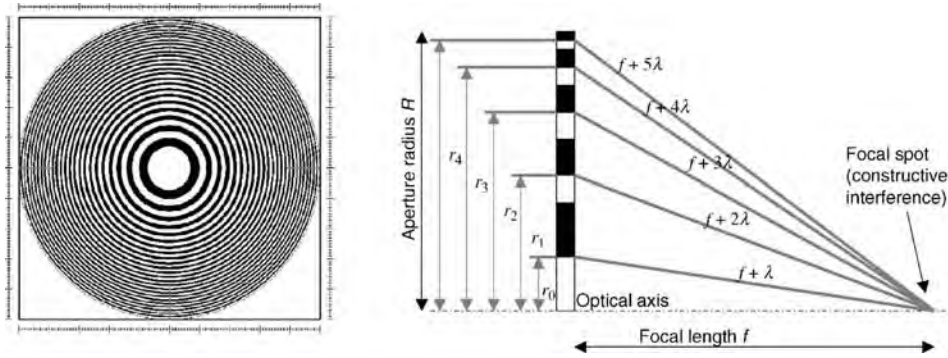


**Figure 5.22** The Fresnel lens (Fresnel Zone Plate)

where the coefficient  $\alpha$  (the detuning parameter for surface-relief elements) is defined as follows:

$$\alpha = \frac{\lambda_0}{\lambda} \left[ \frac{n(\lambda) - 1}{n(\lambda_0) - 1} \right] \tag{5.24}$$

Alternate black (opaque) and white (transparent) zones then form and the resulting binary fringe distribution (see the left-hand side of Figure 5.21) form the Fresnel Zone Plate. Such a lens (a binary amplitude lens) has a very low diffraction efficiency of around 10% (see also the diffraction efficiency figures in Figure 5.11).

5.4.2 The Diffractive Fresnel Lens

Similar to diffraction gratings, the maximum efficiency occurs when the grooves provide a continuous profile with a maximum  $2\pi$  phase shift (for total destructive local interference), both in reflection mode and transmission mode (see Figure 5.23). In the figure,  $n$  is the index of the material in the transmission operation case.

For a binary lens, the optimal etch depth (as for a grating) would be halved (yielding a phase shift of  $\pi$ ). For multilevel diffractive lenses, the optimal etch depth is given in Table 5.2, along with the efficiency of this lens in the fundamental negative order (first real focus or real image).

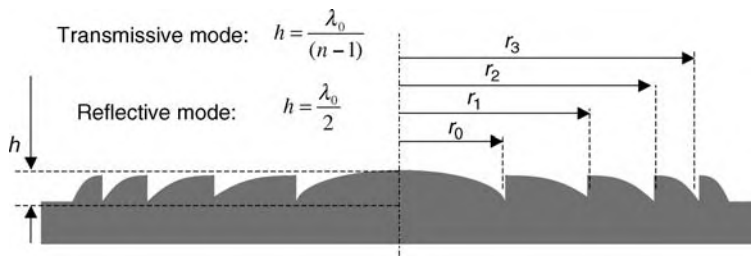


**Figure 5.23** Optimal etch depths for continuous surface-relief Fresnel lenses (transmission and reflective modes)

**Table 5.2** Optimal etch depths and diffraction efficiencies for a Fresnel lens

| Number of phase levels | Critical dimension         | Total groove depth (transmissive/reflective)            | Single step height (transmissive/reflective)        | Diffraction efficiency (%) |
|------------------------|----------------------------|---------------------------------------------------------|-----------------------------------------------------|----------------------------|
| 2 (amplitude)          | $\frac{\Delta_{\min}}{2}$  | NA                                                      | NA                                                  | 10%                        |
| 2                      | $\frac{\Delta_{\min}}{2}$  | $\frac{\lambda}{2(n-1)} \bigg/ \frac{\lambda}{4}$       | $\frac{\lambda}{2(n-1)} \bigg/ \frac{\lambda}{4}$   | 40.5%                      |
| 4                      | $\frac{\Delta_{\min}}{4}$  | $\frac{3\lambda}{2(n-1)} \bigg/ \frac{3\lambda}{8}$     | $\frac{\lambda}{4(n-1)} \bigg/ \frac{\lambda}{8}$   | 81%                        |
| 8                      | $\frac{\Delta_{\min}}{8}$  | $\frac{7\lambda}{8(n-1)} \bigg/ \frac{7\lambda}{16}$    | $\frac{\lambda}{8(n-1)} \bigg/ \frac{\lambda}{16}$  | 95%                        |
| 16                     | $\frac{\Delta_{\min}}{16}$ | $\frac{15\lambda}{16(n-1)} \bigg/ \frac{15\lambda}{32}$ | $\frac{\lambda}{16(n-1)} \bigg/ \frac{\lambda}{32}$ | 99%                        |
| Continuous surface     | $\Delta_{\min}$            | $\frac{\lambda}{(n-1)} \bigg/ \frac{\lambda}{2}$        | NA                                                  | 100%                       |

Figure 5.24 shows the same DOE lens encoded over increasing numbers of phase levels. The additional phase levels will lead to a continuous analog fringe pattern (100% efficient).

It is interesting to note that after eight phase levels, the efficiency no longer increases significantly. Ironically, fabrication errors add up as described in Chapter 12, and potentially decrease the efficiency

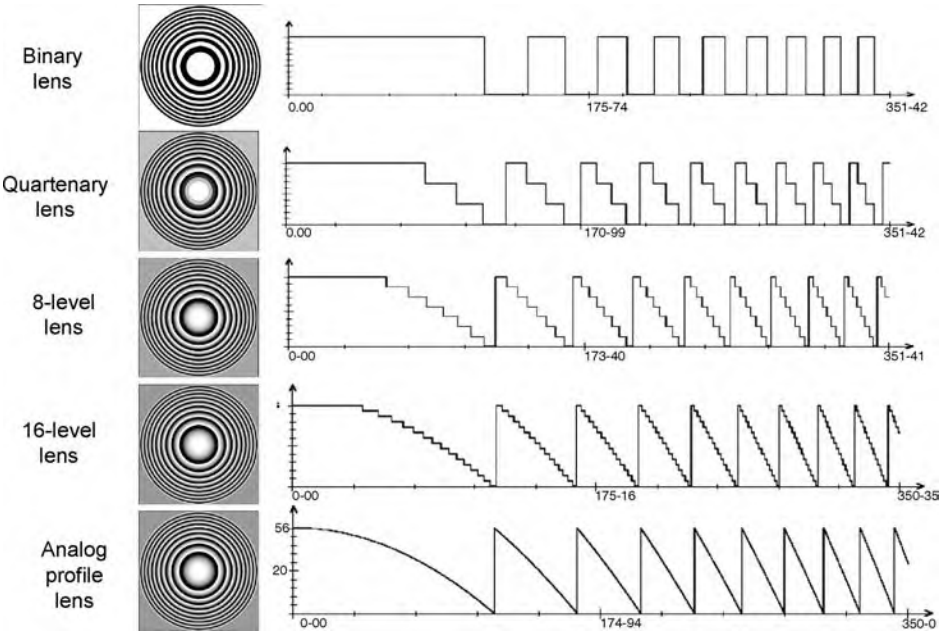


**Figure 5.24** The same DOE lens encoded over an increasing number of phase levels

after eight or 16 phase levels if traditional multilevel lithographic techniques are used. This is not the case with direct write techniques or gray-scale lithographic techniques.

#### 5.4.2.1 Efficiency Calculations for Multilevel Fresnel Lenses in the Real World

In order to calculate the diffraction efficiency of diffractive lenses, it is best to consider a diffractive lens as a linear grating (this is also done for simulating ray tracing through diffractives in conventional optical CAD tools). Therefore, the efficiency results versus the physical model (described in Figure 5.11 and Table 5.1) can be applied to diffractive lenses.

In many cases, the fabrication limitation (see Chapter 12) boils down to the lateral size of the smallest structure that can be etched into the final substrate (note that a similar limitation also applies to the validity of the scalar theory of diffraction – see Appendix B). As the fringe width (local period) of a Fresnel lens decreases radially, the maximum number of phase levels approximating the analog fringe relief surface also decreases. Therefore, a smart way to fabricate an efficient lens, considering the limitations of the fabrication tool, is to optimize the number of phase levels over each fringe. In such a lens, the efficiency is maximal in the center and decreases radially according to the number of phase levels present (see Figure 5.25).

For more details on optimal fabrication of Fresnel lenses, see Chapters 12–14.

#### 5.4.2.2 Remarks Concerning the Notion of Diffraction Efficiency for Real Applications

It is important to remember that diffraction efficiency values are theoretical values describing the efficiency in the  $m$ th diffraction order (e.g. the fundamental positive order, where  $m = 1$ ). However, for any given application, the diffraction efficiency is actually very often defined as a fraction of this value, or even as a combination of several diffraction orders. In order to illustrate this concept, two examples are given below where the efficiency linked to an application is, respectively, lower and higher than the theoretical efficiencies as calculated earlier in this chapter.

1. *A Fresnel lens for data storage:* In this case, while the imaging qualities of the lens are not important here, the amount of energy within the waist of the focal spot is (see Figure 5.26). The theoretical efficiency relates to the amount of energy at this spot (within the Airy disk) as well as the energy within the successive side lobes. Therefore, the efficiency linked to this application is lower than the predicted efficiency of the element. Figure 5.26 shows the focal spot and five different definitions of efficiency in this case.
2. *A binary grating:* In this case, the fact that in the Fourier plane (angular spectrum) of such an element, two main spots appear on-axis with the same efficiency (about 40%) is taken advantage of. These are the two conjugate fundamental orders (orders  $+1$  and  $-1$ ). If the desired reconstruction is symmetric in the Fourier plane (which it is with a linear grating,

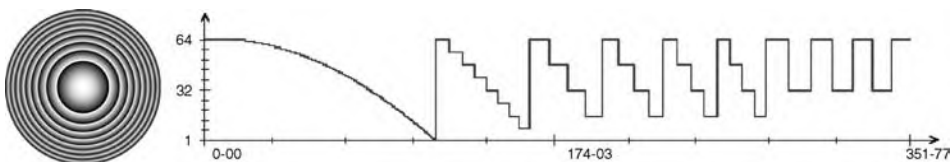


**Figure 5.25** The optimal number of phase levels considering a minimum feature size



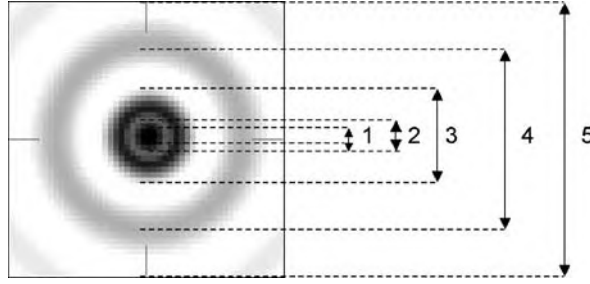


**Figure 5.26** The definition of application-oriented diffraction efficiency (1, FWHM; 2, waist; 3, Airy disk; 4, first side lobe; 5, second side lobe; etc.)

with two spots), these two reconstructions can overlap exactly and double the theoretical efficiency from 40.5% to 81%. This technique is often used for the binary (i.e. relatively cheap) fabrication of high-efficiency on-axis spot array generators, logo generators, beam shapers and so on (on Fourier-type CGHs, see also Chapter 6).

### 5.4.3 Some Useful Diffractive Lens Parameters

Summarized below are some useful lens parameters that are used as rules of thumb to assess whether a diffractive lens can actually be fabricated, without going through the whole design procedure.

#### 5.4.3.1 The $f$ Number ( $f\#$ )

The  $f$  number of a diffractive lens is defined in the same way as for its refractive counterpart, as

$$f\# = \frac{f}{D} \quad (5.25)$$

where  $f$  is its focal length and  $D$  is its full aperture (diameter).

#### 5.4.3.2 The Numeric Aperture (NA)

The numeric aperture is an important parameter of a diffractive lens and can be expressed as follows:

$$NA = \sin(\alpha_{\max}) = \frac{\lambda}{\Lambda_{\min}} = \sin\left(\arctan\left(\frac{D}{2f}\right)\right) \approx \frac{D}{2f} \quad (5.26)$$

where  $\alpha_{\max}$  is the maximum diffraction angle at the edges of the lens, where the minimum period ( $\Lambda_{\min}$ ) also occurs, and  $\lambda$  is the reconstruction wavelength.

#### 5.4.3.3 The Minimum Fringe Width (Critical Dimension – CD)

The minimum fringe width  $\delta$  of such a lens (or the minimal local grating period  $\Lambda_{\min}$ ) then becomes

$$\delta = \frac{\lambda}{p \cdot NA} = \frac{\Lambda_{\min}}{p} \quad (5.27)$$

where  $p$  is the number of phase levels used to fabricate that lens.

This smallest feature is also a very important parameter. This is usually the ‘go or no go’ flag when designing a lens to be fabricated by a specific technology, since ‘a lens is only useful if it can be actually fabricated’. For example, a 16 phase level Fresnel lens would require a minimum feature size,  $\delta$ , 16 times smaller than the smallest groove width  $\Delta_{\min}$ . So, if the minimum dimension printable on the wafer is  $1.0\ \mu\text{m}$ , the minimum printable period of the lens will be  $16\ \mu\text{m}$ . The efficiency of a lens is actually also a function of the local period. If the period gets too small (although printable), the efficiency can drop considerably and/or parasitic polarization effects can occur.

#### 5.4.3.4 The Depth of Focus of a Diffractive Lens (DOF)

The depth of focus is a function of the  $f$  number and the operating wavelength, as follows:

$$\text{DOF} = 2 \cdot \lambda \cdot f\#^2 \quad (5.28)$$

This expression is only valid for conventional Fresnel lenses with spherical or parabolic profiles. As seen in the following sections, the DOF can be increased by modulating the focal length of the lens radially or in a circular way. Although spherical and parabolic lenses can be easily implemented as refractives, more complex aspheric, anamorphic and especially focal length modulated phase profiles are costly to implement as refractive lenses.

#### 5.4.3.5 Change in Focal Length as a Function of Wavelength (Dispersion of a DOE Lens)

As diffractive gratings have a very strong spectral dispersion (see the previous subsections on grating dispersion), a diffractive lens would also yield a high dispersion and materialize the dispersion with a shift in focal length as a function of the operating wavelength, as follows (this describes the effect in one dimension):

$$\begin{cases} \phi(x) = \exp\left(-i\pi \frac{x^2}{\lambda_0 f_0}\right) \\ \lambda_0 f_0 = \text{constant} \end{cases} \Rightarrow f(\lambda) = \frac{\lambda_0 f_0}{\lambda} \quad (5.29)$$

This strong dispersion and focal shift is due to a small Abbe  $V$  number.

#### 5.4.3.6 The Abbe $V$ Number of a Diffractive Lens

Diffractives and refractives have opposite Abbe  $V$  numbers (spectral dispersion):

$$\begin{cases} \nu_{\text{refractive}} = \frac{n_{\lambda_2} - 1}{n_{\lambda_1} - n_{\lambda_3}} \\ \nu_{\text{diffractive}} = \frac{\lambda_2}{\lambda_1 - \lambda_3} = -3.452 \end{cases} \quad \begin{cases} \lambda_1 = 486.1\ \text{nm} \\ \lambda_2 = 587.6\ \text{nm} \\ \lambda_3 = 656.3\ \text{nm} \end{cases} \quad (5.30)$$

Refractive lenses have Abbe  $V$  numbers ranging from 35 (high positive dispersion) to 65 (low positive dispersion), whereas diffractives have a fixed Abbe  $V$  number of  $-3.452$  (very high negative dispersion). On top of this, diffractives have much stronger dispersion than refractives. These unique properties are used to compensate chromatic aberrations in hybrid singlets (refractive/diffractive) and athermalized lenses (similar but opposite effects), as seen in detail in Chapter 7.

### 5.4.3.7 The Strehl Ratio for Diffractive Lenses

When using diffractive lenses for imaging applications, the resolution is usually the more important feature. The Strehl ratio  $S$  defines the ratio between the normalized intensity at the point spread function of the lens and the normalized intensity of a perfect lens without any aberration ( $S = I_{\text{real}}/I_{\text{ideal}}$ ).

### 5.4.4 Super-zone Diffractive Lenses

Having introduced in the previous section the notion of smallest feature size (which can be a fraction of the smallest fringe width), there are techniques other than those described in Figure 5.25 to lower the burden associated with the minimum fabricable feature, such as introducing the concept of the super-zone Fresnel lens. As shown in Figure 5.18, the phase profile is sliced into  $2\pi$  phase shifts in order to produce the various fringes. If the profile is sliced into integer numbers of  $2\pi$  phase shifts, the same effects still exist; however, the grooves would need to be etched much more deeply. Figure 5.27 shows an example where a  $4\pi$  phase shift is used in order to increase the smallest lateral size of the last groove.

The super-zone diffractive lens uses this idea in order to increase the size of the smallest grooves in order to lower the burden on the fabrication. However, the added depth has to be accurately fabricated. Super-zone Fresnel lenses are best fabricated by diamond turning, rather than by lithography, where it is always costly to achieve additional (deeper) depths.

### 5.4.5 Broadband Diffractive Lenses

Diffractive Optical Elements, just like gratings, are very sensitive to wavelength changes (strong spectral dispersion), which produces high chromatic aberrations (lateral and longitudinal) and affects the diffraction efficiency (see Figures 5.12 and 5.13). Nevertheless, it is possible to optimize diffractives (e.g. Fresnel lenses) to function over a wider range of wavelengths or over a set of predetermined individual wavelengths. These lenses are commonly referred to as broadband, multi-order or harmonic diffractive lenses (see Figure 5.28).

The etch depths  $H_t$  or  $H_r$  (for transmission and reflective operation, respectively) of harmonic or multi-order DOE lenses are optimized so that the phase difference would yield an integer number of  $2\pi$  phase

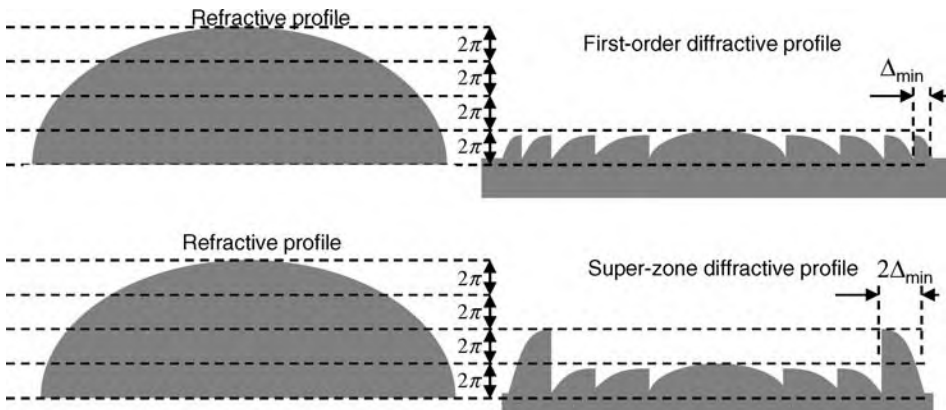


**Figure 5.27** The concept of the super-zone Fresnel lens

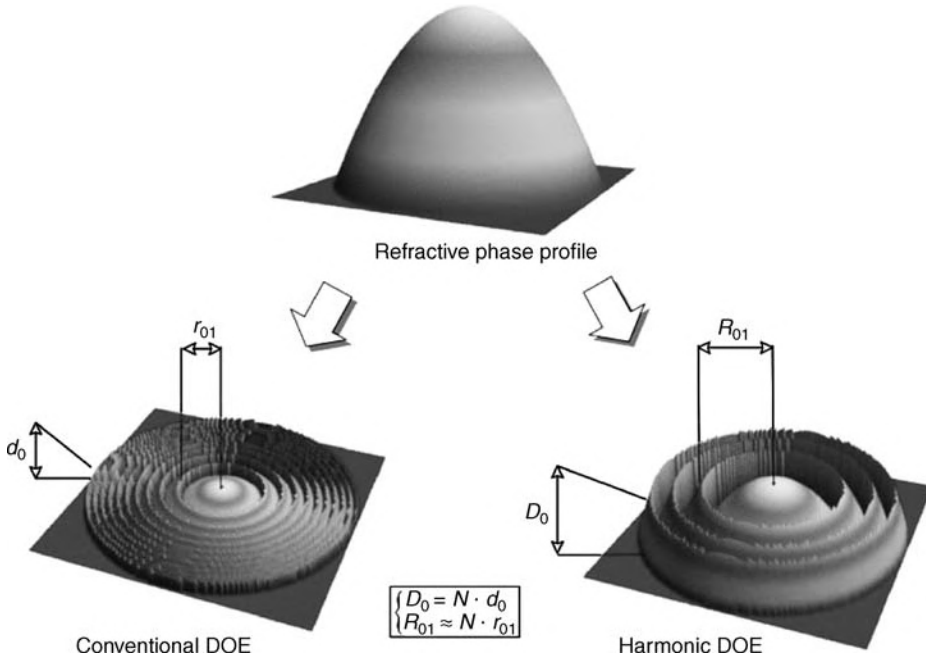


**Figure 5.28** The generation of narrowband and broadband diffractive lenses from the same phase profile

shifts for each wavelength (e.g. a modulo  $2\pi$  phase shift):

$$\begin{cases} H_T[\text{mod}_{(2\pi)}(\Delta\varphi = 2\pi)] = \frac{N_1\lambda_1}{(n-1)} = \frac{N_2\lambda_2}{(n-1)} = \frac{N_3\lambda_{31}}{(n-1)} \\ H_R[\text{mod}_{(2\pi)}(\Delta\varphi = 2\pi)] = \frac{N_1\lambda_1}{2} = \frac{N_2\lambda_2}{2} = \frac{N_3\lambda_{31}}{2} \end{cases} \quad (5.31)$$

where three different wavelengths are used and  $n$  is the index of refraction when using the transmission operation mode. Such a broadband diffractive lens is aimed at keeping the diffraction efficiency high over a broad range of wavelengths; however, it does not compensate for chromatic aberrations (for chromatic aberrations compensation, see Chapter 7).

Figure 5.29 shows how the diffraction efficiency of a harmonic DOE would behave when using two and then three harmonic conditions (see Equation (5.31)).

This technique, as with the previous super-zone Fresnel lens, has the advantage of widening the fringe width, but also the inconvenience of deepening the groove depth. So precautions have to be taken when using this technique. Also, this technique yields fabricable devices if the wavelengths for the diffractive elements are far apart.

First, consider a simple example: a binary DOE (or grating) etched for 850 nm (a VCSEL in the infrared region, used in optical telecom for 10 Gb/s Ethernet) would also yield decent efficiency results for a violet wavelength around 405 nm (the Nichia Violet laser used in Blu-ray applications). Since the wavelength is reduced by half from IR to violet, the phase shift is still close to  $2\pi$  (actually  $4\pi$  for violet and  $2\pi$  for IR).

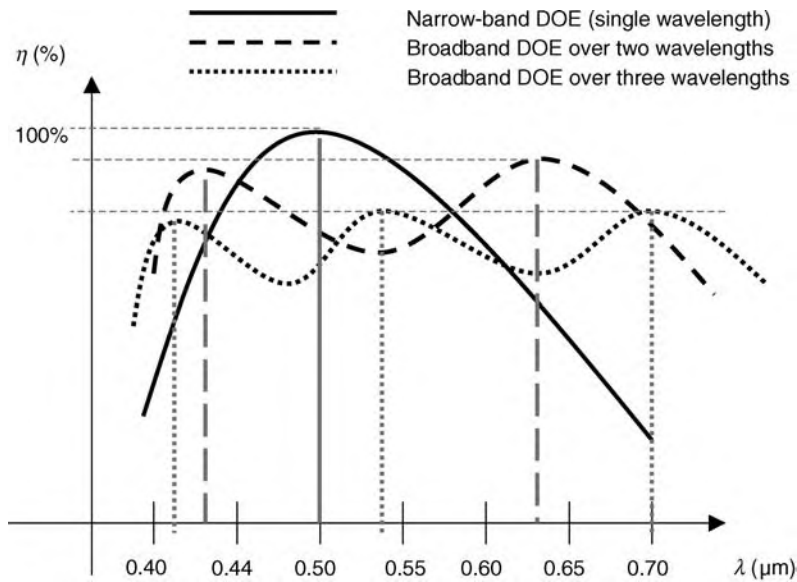


**Figure 5.29** The diffraction efficiency for narrowband DOEs and two- and three-mode harmonic DOEs

In another example, the grooves are etched to  $p \times 2\pi$ , instead of  $2\pi$ , and the diffraction order is  $k$ . The focal length expression in Equation (5.29) can be rewritten as follows:

$$f(\lambda) = \frac{p}{m} \frac{\lambda_0 f_0}{\lambda} \tag{5.32}$$

and the diffraction efficiency at order  $m$  for a  $p$ -order broadband diffractive becomes

$$\begin{cases} \eta = \text{sinc}(\pi(p\psi + k))^2 \\ \psi = \frac{\lambda_0}{n(\lambda_0) - 1} \cdot \frac{n(\lambda) - 1}{\lambda} \end{cases} \Rightarrow \eta \approx \text{sinc}\left(\pi\left(\frac{\lambda_0}{\lambda}p + k\right)\right)^2 \tag{5.33}$$

As an example, Table 5.3 lists the efficiency at various orders for a lens of broadband mode 10 over the visible spectrum (from 350 nm to 700 nm).

**Table 5.3** Broadband efficiency for order 10

| Broadband order, $p$ | Design, $\lambda_0$ (nm) | Diffraction order, $k$ | Reconstruction, $\lambda$ (nm) | Efficiency, $\eta$ (%) |
|----------------------|--------------------------|------------------------|--------------------------------|------------------------|
| 10                   | 550                      | −8                     | 700                            | 99%                    |
|                      |                          | −8                     | 650                            | 93%                    |
|                      |                          | −9                     | 600                            | 99%                    |
|                      |                          | −10                    | 550                            | 100%                   |
|                      |                          | −11                    | 500                            | 100%                   |
|                      |                          | −12                    | 450                            | 98%                    |
|                      |                          | −14                    | 400                            | 98%                    |
|                      |                          | −16                    | 350                            | 97%                    |

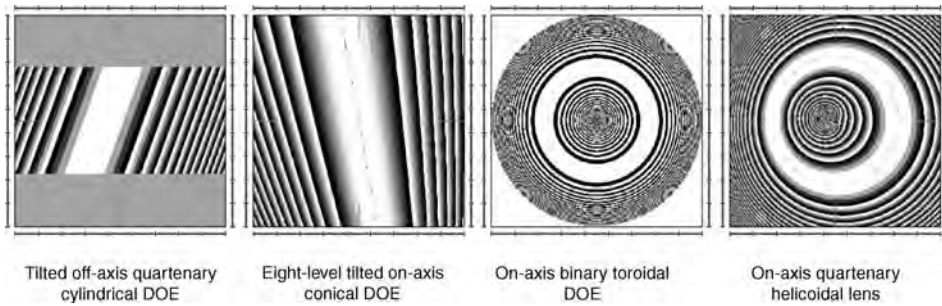


**Figure 5.30** Cylindrical, conical, toroidal and helicoidal diffractive lenses

### 5.4.6 *Diffractive Lens Implementations*

#### 5.4.6.1 Cylindrical, Conical, Toroidal and Helicoidal Diffractive Lenses

In the previous sections, DOEs were discussed that have rotationally symmetric phase profiles, such as spherical, parabolic or more complex aspheric profiles [8]. However, DOEs can take on much more complex phase functions, and thus implement cylindrical lenses, conical lenses, and even toroidal or helicoidal lenses (see Figure 5.30).

Just as a cylindrical lens would focus a collimated wavefront in a light segment parallel to the DOE plane, a conical lens would focus a segment tilted into that same plane. A toroidal lens would focus a ring of light parallel to the DOE plane and a helicoidal lens would focus a tilted ring of light.

Although cylindrical lenses can be easily implemented in a conventional refractive way, that is not the case for conical lenses, and especially not for toroidal and helicoidal lenses. While conical and toroidal lenses are commonly used in numerous applications, helicoidal lenses are not.

#### 5.4.6.2 OCDL Lenses

Orthogonal Cylindrical Diffractive Lens (OCDL) are produced when phase multiplexing two cylindrical lenses in order to implement a different focal length in the  $x$  and  $y$  directions. Such OCDL lenses have very unique characteristics and have no refractive counterparts. OCDLs have been used in applications where in one direction a Fourier transform function was required and in the orthogonal direction a 1D imaging task (a read-out head for a holographic optical data storage disk) was required.

Crossed cylindrical lenses are also often used to collimate strongly astigmatic laser diodes (rectangular exit waveguide section).

#### 5.4.6.3 Anamorphic Diffractive Lenses

Anamorphic DOE lenses are aspheric lenses that do not have circular symmetric fringes, but have circular fringes. The fringes of such anamorphic lenses can be elliptic, as depicted in Figure 5.31.

Such lenses are usually not used in imaging systems but, rather, in nonimaging systems (i.e. illumination etc.) where anamorphic scaling is required; for example, illuminating a specific aspect ratio rectangular screen or microdisplay from a different aspect ratio rectangular LED die or laser aperture. The fabrication of anamorphic refractive lenses by means of diamond turning requires a lathe with four axial degrees of freedom. An anamorphic DOE lens does not need any fancy fabrication technique.

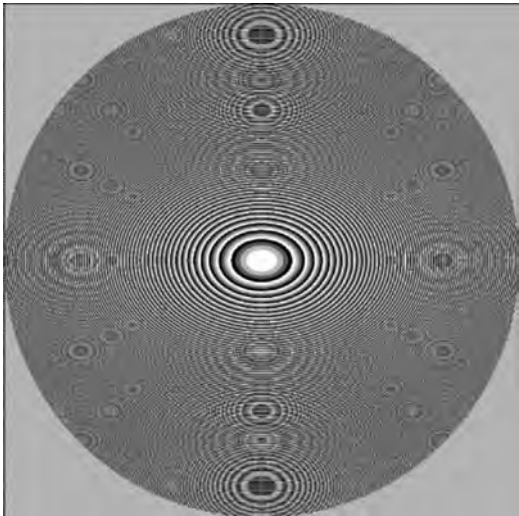


**Figure 5.31** An anamorphic DOE lens

**5.4.6.4 Extended Depth of Focus Diffractive Lenses**

One of the major advantages of using DOE lenses rather than refractive lenses in imaging and nonimaging systems is that DOE lenses can implement very complex functionalities, and are therefore no more complex to fabricate, unlike conventional refractive lenses. A very good example of the versatility of DOE lenses is the following depth of focus modulated lenses (see Figure 5.32).

The starting point for designing such modulated lenses is the spherical Fresnel lens. A radial modulation (on the left) with a sinusoidal modulation of the focal length (the modulation amplitude is  $\Delta f/f_0$ ) has

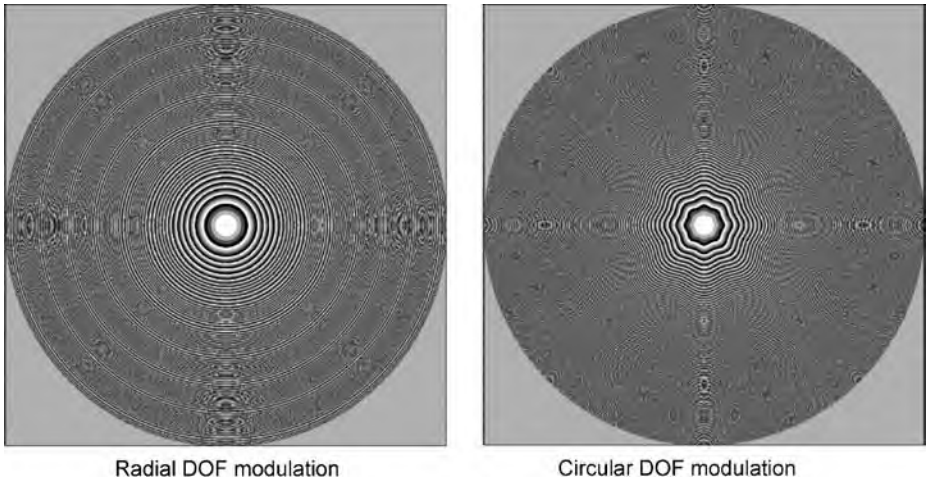


**Figure 5.32** Depth of focus modulated lenses (left: circular, right: radial) – eight modulations,  $\Delta f/f_0 = 1/50$

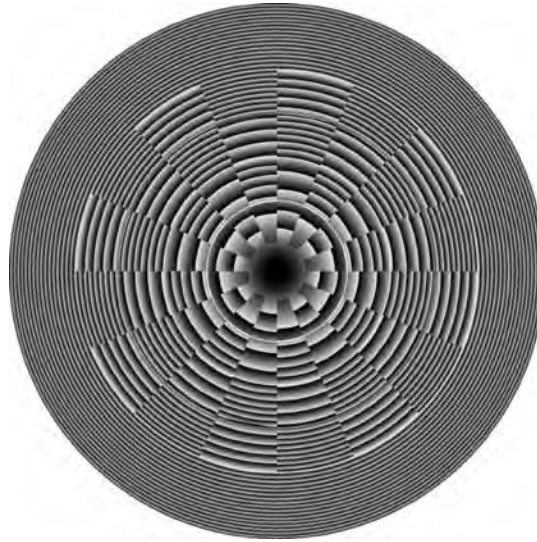


**Figure 5.33** A dartboard dual-focus lens

been imprinted on the circular fringes of the Fresnel lens. Radial modulation will not affect the circularity of the fringes, but the periods of the consecutive fringes, as they move along from the optical axis. One can note that the phase varies in a sinusoidal way, imprinted from the focal length variations  $\Delta f$ . Such a lens would have its focal length moving from  $f_0 - \Delta f$  to  $f_0 + \Delta f$  focal length  $N$  times from the optical axis to the edges of the lens. For an implementation example of such a lens, see Chapter 16.

Similar sinusoidal modulation can be applied circularly rather than radially and can produce a Fresnel lens that actually looks like the petals of a daisy (therefore, it is sometimes called the ‘Daisy lens’) – see the right-hand side of Figure 5.32. Such a lens would constantly move from  $f_0 - \Delta f$  to  $f_0 + \Delta f$  focal length over  $360/N$  degrees, where  $N$  is the number of modulations in that lens.

Although such a modulation will affect the Strehl ratio of the lens, and thus its imaging qualities, such lenses are best used for special applications such as laser cutting of 3D shapes (without constantly moving the workpiece or focusing lens), compensation for wobbling in CD/DVD drives (without having to constantly move the lens) and especially recent investigations in wavefront coding, for imaging applications in camera phones and IR missile trackers (see Chapter 16).

Spatial multiplexing of different lenses in a single circular aperture has been reported and used in various applications (see the multifocus lens used in intro-ocular ophthalmic lenses in Chapter 16). Spatially multiplexing can be simple (doughnut lens and center lens) or more complex. The lens shown in Figure 5.33 is a spatial combination of two different lenses, which are cut out in a dartboard format, so that each lens is encoded over a similar angular and radial value.

Such a spatially multiplexed lens has  $N$  foci,  $N$  being the number of different lenses encoded.

#### 5.4.6.5 Super-resolution Diffractive Lenses

Super-resolution lenses are simply lenses in which the central part is obstructed by an absorbing material (chrome on glass, for example). The central part of the lens (the phase profile of which is almost flat) does not contribute to high-frequency imaging. Therefore, when the lens needs to resolve higher frequencies, such an obscuration can be used. Figure 5.34 shows a dull Daisy lens and a super-resolution Daisy lens.



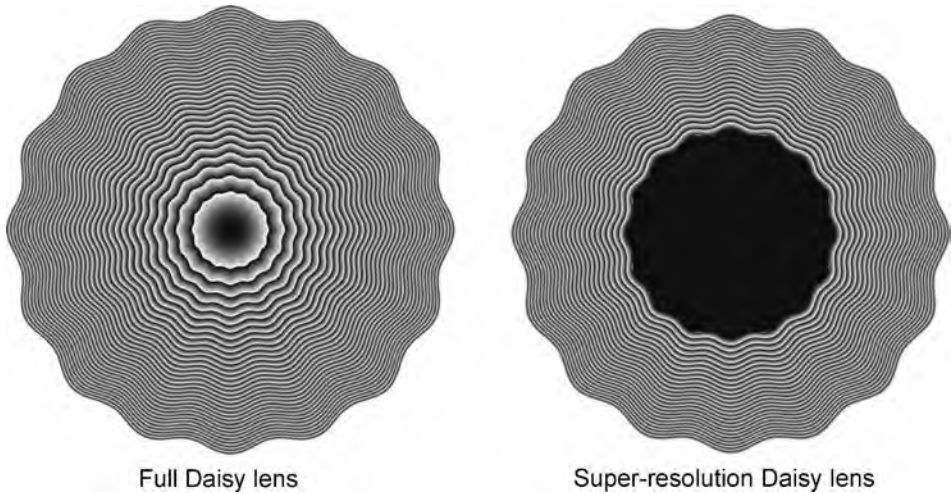


**Figure 5.34** Super-resolution diffractive lenses

There are two drawbacks to this technique:

- the light absorbed by the central shape is obviously lost; and
- the Strehl ratio is reduced, since a super-resolution lens creates larger side lobes (however reduced the Airy disk is).

Chapter 11 shows some modeling examples of such lenses used for DVD read-out.

**5.4.6.6 Vortex Diffractive Lenses**

Vortices are universal and robust states that exist in all realms of physics and, therefore, also in optics. Vortex diffractive lenses also have unique optical functionalities that make them suitable for very special applications. It is difficult to implement a vortex lens as a refractive lens, but quite easy to implement it as a diffractive lens.

Vortex lenses (in the diffractive implementation) create a vortex of  $k$  vectors propagating within a torus of light. The fringes composing such a lens are not circular but, rather, helicoidal (see Figure 5.35). Vortex

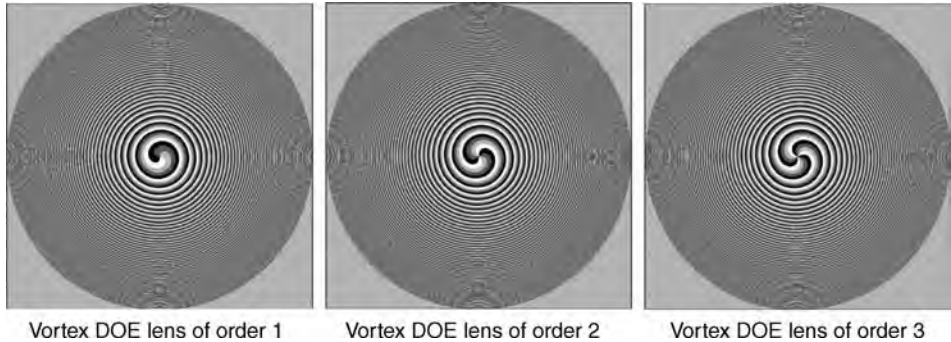


**Figure 5.35** Vortex lenses of different orders (orders 1, 2 and 3), encoded over four phase levels

lenses come in increasing modes, the central point of the lens producing a particularity in a diffractive (a nondiffracting high spatial frequency optical axis area).

Note that such a torus of light is very different from the torus of light produced by a toroidal lens (see previous sections). Such vortex lenses are, for example, used when coupling light into graded-index fibers and exciting a well-defined circular propagation mode in this type of fiber (see also Chapter 16). This because of its special central index variation, which prevents any other mode from propagating.

#### 5.4.6.7 Complex Aperture Lenses

Complex apertures can be implemented in diffractive lenses to spatially multiplex several lenses or to create a super-resolution lens (see previous sections). DOE lenses are very versatile elements and can be spatially multiplexed with other elements such as linear gratings.

In the example given in Figure 5.36, a series of Fresnel diffractive lenses are shown which are encoded over a thin rectangular shape (split shape) flanked by high spatial frequency gratings on each side.

Such a lens takes on the task of two bulk optical elements; namely, a spherical lens and a slit aperture stop. Such devices are used in optical encoders and other elements. As a slit is difficult to align to a lens, a lens with the slit functionality already incorporated in its structure and with the ability, therefore, to be replicated in volume as it is, would reduce footprint, volume and costs in applications requiring such a functionality. Note that here the light is not blocked as it would be with an amplitude slit, but diffracted into regions of no interest (regions far away from the detector).

#### 5.4.6.8 Beam-shaping Diffractive Lenses

Most of the beam shapers used in industry today are numeric-type diffractives (calculated by iterative algorithms, as detailed in Chapter 6). Beam shapers can be implemented as Fourier beam shapers (reconstruction in the far field) or near-field beam shapers (reconstruction in the near field).

However, analytic techniques can also be used to calculate simple beam shapers and lens beam shapers (Fresnel type). For example, in a typical Gaussian to rectangular top hat beam-shaping problem in the near field [9], although the DOE functionality can be quite complex, the pure phase profile to be encoded can be

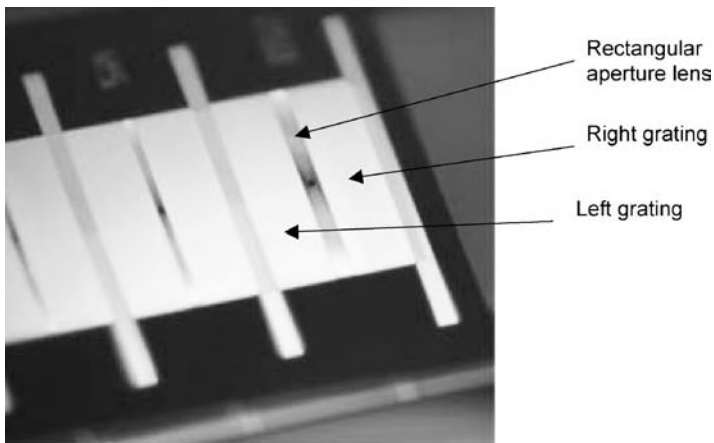


**Figure 5.36** Multiplexing of DOE lenses and gratings for diffractive slit operation

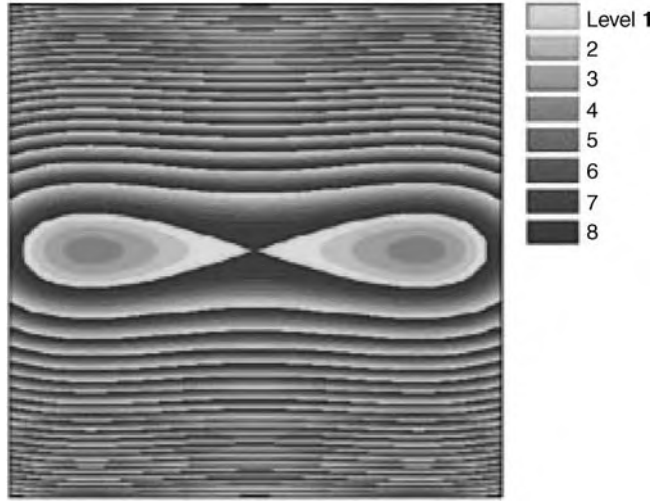


**Figure 5.37** An example of a diffractive beam-shaper lens (Gaussian to top hat)

obtained analytically by geometrical ray-tracing calculations:

$$\begin{cases} \varphi(x_1, y_1) = \varphi_x(x_1) + \varphi_y(y_1) \\ \varphi_x(x_1) = -\frac{\pi}{\lambda f} x_1^2 - \frac{2\pi a}{E_a f} \left[ x_1 \operatorname{erf} \left( \frac{x_1 \sqrt{2}}{\sigma} \right) - \frac{\sigma}{\sqrt{2\pi}} \left[ 1 - e^{-\frac{2x_1^2}{\sigma^2}} \right] \right] \end{cases} \quad (5.34)$$

where  $f$  is the focal length,  $a$  is the aperture size of the beam shaper in  $x$ ,  $\sigma$  is the laser beam waist and  $E_a$  is the spatial extent of the shaped beam in  $x$  at the focal plane.  $\operatorname{Erf}$  is the error function (incomplete Gamma function). In this case (rectangular Gaussian to top hat), the equation is derived from the analytic expression of the phase function in the plane of the beam shaper. Figure 5.37 shows an example of such a diffractive Fresnel analytic beam shaper encoded over eight phase levels.

Chapter 6 will discuss in more detail the vast variety of beam-shaper CGHs, calculated via numeric optimization algorithms.

#### 5.4.6.9 Null Lenses

Null lenses, or null CGHs, are diffractive phase plates that usually implement a complex aspherical phase function. They are used in interferometric systems in order to measure wavefront aberrations where the null lens is used to compensate for a theoretical phase profile. The remaining twiggles in the fringes would characterize the fabrication errors or inaccuracies of the mirror or lens under test (e.g. in astronomy, for testing large telescope mirror profiles).

The Zernike coefficients

$$\varphi(r, \phi) = \sum a_{m,n} Z_{m,n}(r, \phi) \quad (5.35)$$

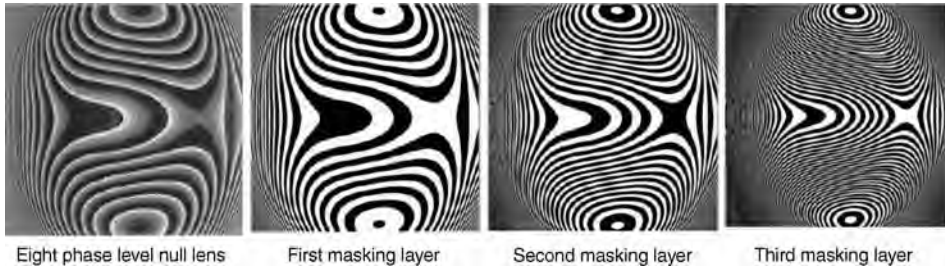


**Figure 5.38** A null lens for astigmatism correction, encoded over eight phase levels

are a convenient basis for expressing wavefront aberrations over a circular pupil. Zernike polynomials are orthogonal to each other:

$$\begin{cases} Z_{0,0} = 1 & \rightarrow \text{piston} \\ Z_{1,-1} = 2 \cdot r \cdot \sin(\phi) & \rightarrow \text{tip} \\ Z_{1,1} = 2 \cdot r \cdot \cos(\phi) & \rightarrow \text{tilt} \\ Z_{2,-2} = \sqrt{6} \cdot r^2 \cdot \sin(2\phi) & \rightarrow \text{astigmatism} \\ Z_{2,0} = \sqrt{3}(2r^2 - 1) & \rightarrow \text{focus} \\ Z_{2,2} = \sqrt{6} \cdot r^2 \cdot \cos(2\phi) & \rightarrow \text{astigmatism} \end{cases} \quad (5.36)$$

However, just as when using polynomial descriptions of aspheric phase profiles (see Equations (5.15), (5.16) or (5.17)), extreme care has to be taken when reading these coefficients from design software (CodeV, Zemax, ASAP, etc.). Each optical CAD design software package has its own way of normalizing these coefficients; the coefficients can also be expressed in many ways (radians, waves, differing dimensions etc.).

A typical null lens for astigmatic wavefront correction is depicted in Figure 5.40. The lens has eight phase levels and is thus fractured into three consecutive mask sets, as also depicted in Figure 5.38 (for fabrication details, see Chapter 12).

In the previous sections, mostly lenses with more or less circular symmetry were discussed. This one is very different and would be very difficult to implement as a refractive lens due to its complex profile.

#### 5.4.6.10 Tilted-operation Lenses

Tilted-operation lenses are very convenient when implementing an aspheric lens performing a specific imaging function in a tilted configuration. Tilting a lens can be very desirable, for example, in order to reduce the light back-reflected into the source (laser).

## 5.5 Diffractive Interferogram Lenses

The third group of diffractives constituting the realm of analytic-type diffractives are referred in the literature as interferogram diffractive lenses [8].

An interferogram phase function is directly calculated as the interference pattern of a given object wave and reference wave; therefore interferograms are very similar to holographic optical elements (HOEs – see also Chapter 8). As an example, consider a tilted planar reference wavefront that makes an angle  $\alpha$  with the DOE plane, and that interferes with a normal planar wavefront. The resulting interference pattern can

be expressed as follows:

$$t(x, y) = c + 2A_0(x, y) \cdot \cos \left[ \varphi_0(x, y) + 2\pi \frac{\sin(\alpha)}{\lambda} \cdot x \right] \quad (5.37)$$

Although the physical interference pattern is an intensity distribution, the resulting DOE interferogram can be recorded as a pure phase element (in phase or surface-relief modulation), or as an amplitude element. These elements have very remarkable properties when used in off-axis configurations as complex lenses (such as toroidal, conical or helicoidal lenses), but lack diffraction efficiency due to their typical sinusoidal phase profile (which limits their efficiency to 33% in sinusoidal mode or 40% in binary mode). In order to increase their efficiency as a surface-relief element, one can fabricate the calculated fringes as blazed structures instead of sinusoidal structures.

Interferograms may look like regular diffractive lenses, but have unique properties when operated in off-axis mode (see Figure 5.39).

For conventional DOE lenses, off-axis is defined as a shift of the DOE aperture along the infinite DOE fringe pattern. In an interferogram-type DOE, this offset is defined as the tilt in the reference wave that interferes with the on-axis lens wavefront. This does not shift the fringe pattern, as described in Figure 5.39, but literally shifts the reconstruction, as can be seen in Figure 5.40.

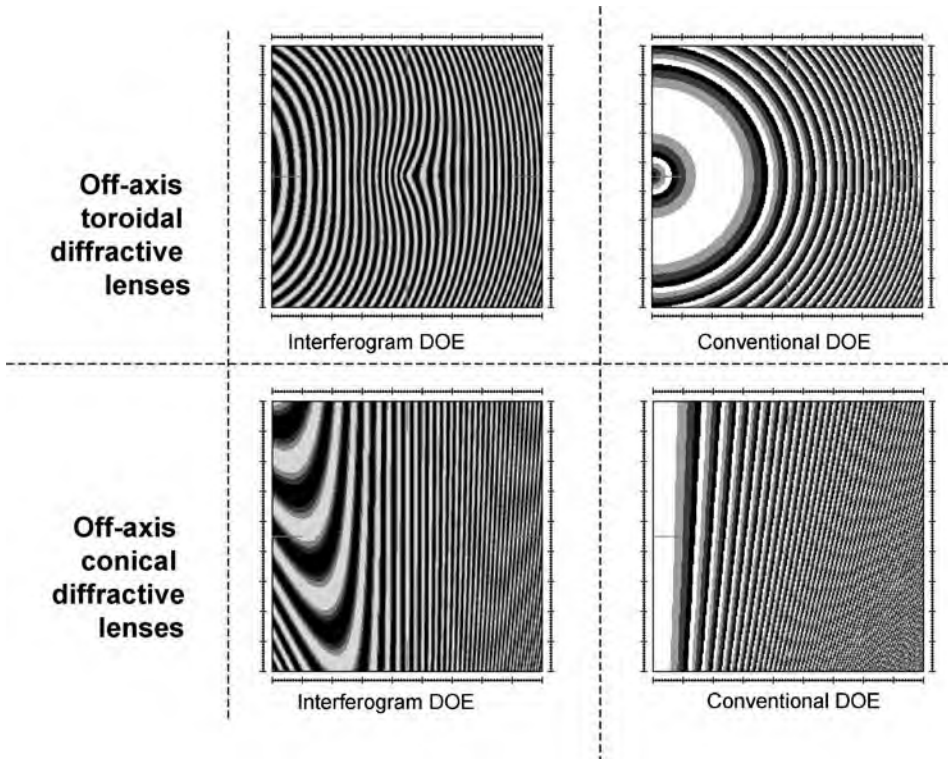


**Figure 5.39** Interferogram-type DOEs and conventional DOEs implementing similar lens functionalities

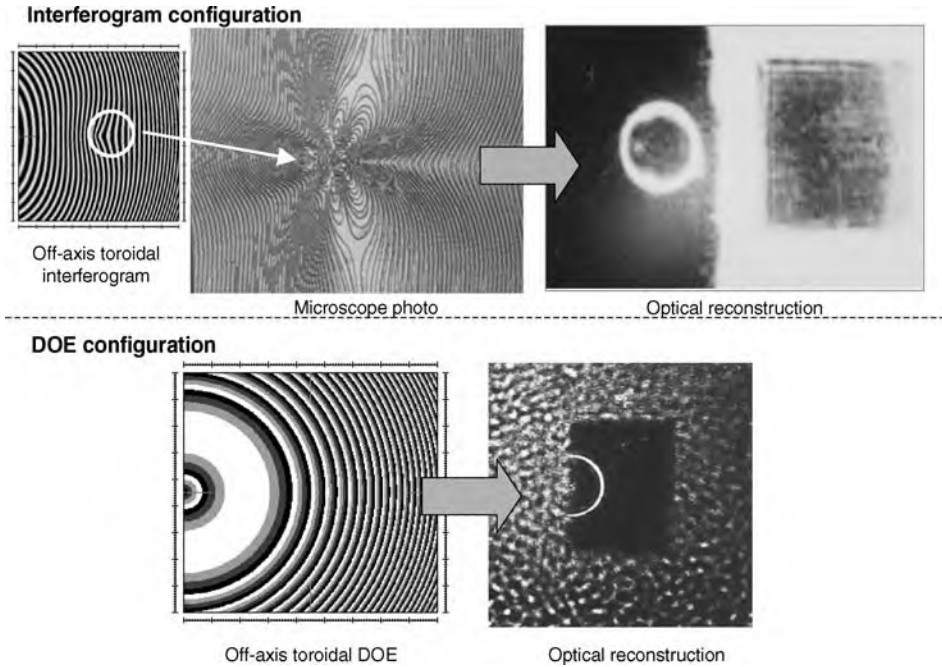


**Figure 5.40** Optical reconstructions from toroidal lenses, as conventional or interferogram DOEs

Figure 5.40 shows an implementation of a toroidal off-axis lens with conventional DOE coding and interferogram-type DOE coding. The optical reconstructions show that only part of the ring is focused in the focal plane of the conventional DOE lens, whereas the entire ring is focused for the interferogram-type DOE.

Such interferogram DOE lenses are used in very specific applications. For example, such diffractives can generate true 3D wire-framed structures in a true volume. Line segments and arc segments are focused from off-axis tilted conical interferogram lenses and partial toroidal or helicoidal off-axis interferogram DOEs. The position of the DOE in an interferogram is not directly linked to the position of the focal element (line, spot, circle), but linked to its tilt and encoded off-axis. Therefore, it is possible to generate a 2D matrix of square pupil interferogram elements and to associate a specific tilt/tip and focal position for each focal element generated by the regularly spaced matrix of interferogram DOEs. The most common focal elements are segments, generated by tilted off-axis conical interferograms.

As an example, a holographic animation has been produced composed of interferogram DOEs arranged in a matrix and displaying focal structures in a 2D plane (segments, curves and spots), as a pseudo-dynamic diffractive element compound. This example is shown in Chapter 9.

This chapter has reviewed the various analytic types of diffractive optical elements used in industry today. Such diffractives can be designed directly by analytic methods (e.g. solving an equation).

Analytic-type diffractives are composed of three main groups:

- grating-type elements (1D and 2D);
- diffractive lens elements (or DOEs); and
- interferogram-type diffractives.

In Chapter 6, diffractive optical elements will be discussed that do not have an analytic solution to the stated diffraction problem, and therefore require more complex design and optimization procedures, usually implemented numerically in a computer as an iterative optimization algorithm. The vast majority of these elements are referred to as Computer-Generated Holograms (CGHs).

## References

- [1] M.C. Hutley, '*Diffraction Gratings*', Academic Press, London, 1982.
- [2] G.N. Laurence, 'Using rules of thumb in the design of physical optics systems', *Optical Society of America Technical Digest*, **9**, 1993, 12–13.
- [3] R. Petit, '*Electromagnetic Theory of Gratings*', Springer-Verlag, Berlin, 1980.
- [4] H. Dammann, 'Blazed synthetic phase-only holograms', *Optik*, **31**, 1970, 95–104.
- [5] M. Goel and D.L. Naylor, '*Analysis of Design Strategies for Dammann Gratings*', Proc. SPIE Vol. 1689, 1996, 35–45.
- [6] H. Dammann and K. Görtler, 'High-efficiency in-line multiple imaging by means of multiple phase holograms', *Optics Communications*, **3**, 1971, 312–315.
- [7] X.-Y. Da, 'Talbot effect and the array illuminators that are based on it', *Applied Optics*, **31**(16), 1992, 2983–2986.
- [8] D. Leseberg, 'Computer generated holograms: cylindrical, conical and helical waves', *Applied Optics*, **26**(20), 1987, 4385–4390.
- [9] M. Duparré, M.A. Golub, B. Lüdge *et al.*, 'Investigations of computer-generated diffractive beam shapers for flattening of single-modal CO<sub>2</sub> laser beams', *Applied Optics*, **34**(14), 1995, 2489–2497.





# 6

## Digital Diffractive Optics: Numeric Type

Designing, calculating and optimizing numeric-type diffractive lenses essentially boils down to finding the best phase surface-relief profile (a black box) that will perform a specific wavefront transformation task [1–3], without having too many amplitude variations, as depicted in Figure 6.1.

### 6.1 Computer-generated Holograms

A Computer-Generated Hologram (CGH) is the typical numeric-type digital element. It is designed with a numeric optimization procedure and fabricated by digital microlithographic fabrication tools.

#### 6.1.1 *What is a CGH?*

A CGH is basically a complex wavefront processor. In the CGH design task, typically, an incoming wavefront description as well as the desired diffracted wavefront, or part of it, in the near or far field, are specified. The task in hand is then to find, by any means available, the best phase (or amplitude or even complex) mapping that will perform the diffraction job in an acceptable way considering the requirements of the final application.

Therefore, numeric-type elements can implement much more complicated optical functionalities than analytic-type elements are able to perform. Figure 6.2 shows some of the elements comprising the realm of numeric-type diffractives.

However, the fact that numeric-type elements can implement more complex optical functions than analytic types has to be weighed against the fact that analytic-type elements are designed in an absolute way (an analytic description of the position of fringes, which are not sampled) and numeric-type elements are already sampled in their existing form (see Figure 6.3). So, numeric-type elements lack the resolution that analytic-type elements have. Therefore, for imaging tasks or other highly demanding simple tasks, it is far better to use analytic-type elements, provided that the fabrication technology does not destroy this analytic behavior by digitizing or sampling the analytic element in a crude way. For example, a simple diffractive lens should always be designed as an analytic element and never sampled, especially in its design procedure, even when going to fabrication.

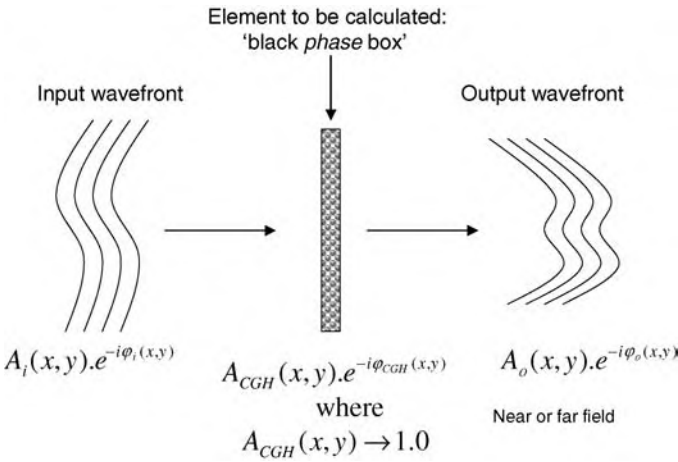


**Figure 6.1** The basic task of calculating a numeric-type diffractive

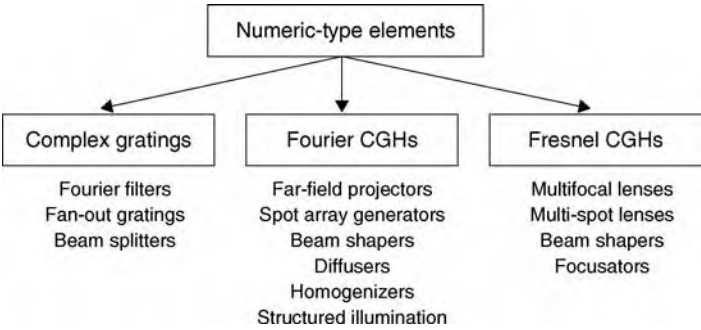


**Figure 6.2** The realm of numeric-type diffractives

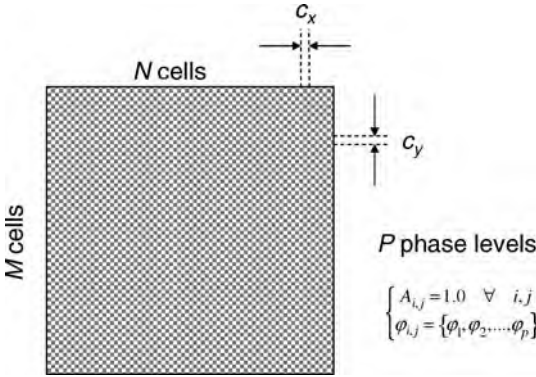


**Figure 6.3** The discrete numeric element phase matrix

### 6.1.2 The Space Bandwidth Product (SBWP)

In order to simplify the design task, the numeric-type element is usually considered as a pure phase element ( $A(x,y) = 1$ ), and is sampled in the  $x$  and  $y$  directions (see Figure 6.3) over discrete cells.

It is therefore interesting to derive a criterion – the Space Bandwidth Product, or SBWP – that can predict the ability to perform for a numeric-type element, even before this element is calculated and optimized. For an analytic-type element, this criterion would always be infinite, owing to the fact that analytic-type elements are not sampled in the design process [4]. The SBWP of numeric-type elements is given as follows:

$$\text{SBWP} = \frac{N \cdot M}{c_x \cdot c_y} \cdot P \quad (6.1)$$

The  $N \times M$  cells in Figure 6.2 can take on any phase value. However, in order for this element to be fabricable, the phase levels are constrained over the levels that can actually be fabricated with the target fabrication tool (e.g. 2, 4, 8 or 16) and computed by an adequate cost function that relates to aspects of the optical performance of the final system.

Calculating a numeric-type element is therefore a ‘black phase box’ optimization task, which optimizes a cost function given a set of constraints (the incoming and outgoing wavefront) and a set of CGH-related constraints (the number of cells, the size of the cells and the number of phase levels) [5]. Therefore, in the design step of numeric-type diffractives, ‘optimization’ is discussed rather than ‘calculation’.

Note that the cell sizes and the number of cells in each direction do not need to be equal (and can thus yield rectangular or even more complex shaped elements). In many cases, the cell size in one dimension can be orders of magnitude larger or smaller than the cell size in the orthogonal direction, depending on the application and the diffraction angles required in each direction.

### 6.1.3 Fourier and Fresnel CGHs

Computer-Generated Holograms (CGHs) are numeric-type elements, which are often used as free-standing optical elements – unlike analytic-type elements, which can be hybridized to refractives to increase their functionality.

The versatility of the numeric iterative optimization processes for CGHs allows the design of complex optical functionalities [6] that cannot be implemented by classical refractive or reflective optics [7]. CGHs are mainly used as complex optical wavefront processors and are the tools for the development of unconventional optical applications, especially in displays, as shown in the last section of this chapter.

There are two types of CGHs, the Fourier type and the Fresnel type. The issue of Fresnel and Fourier configurations, where applied to analytic-type elements (to gratings and lenses, respectively), has already been discussed in the previous chapter.

The Fourier type projects the desired pattern in the far field, whereas the Fresnel type reconstructs the desired pattern in the near field, and thus can be considered as a complex diffractive lens (bearing optical power as well as optical shaping/splitting).

Figure 6.4 depicts a Fourier and a Fresnel CGH, both of which implement almost the same optical functionality, namely splitting the incoming beam into an array of  $3 \times 3$  beams.

While the Fourier element would create nine collimated beams ( $3 \times 3$ ) from a single incoming collimated beam, the Fresnel element would create nine converging (and/or diverging) beams from the same collimated beam and, therefore, could be considered as a multifocus diffractive lens. Note the fringe-like pattern in the Fresnel-type element, which does not appear in typical Fourier-type elements.

The main difference between a standard array of  $3 \times 3$  diffractive microlenses and a  $3 \times 3$  multifocus diffractive lens is that the numeric aperture of the multifocus lens is much larger than for the standard microlens array (exactly nine times larger). Therefore, such a multifocus lens can be called a phase-multiplexed microlens array, as opposed to the conventional space-multiplexed microlens array.

Figure 6.5 shows the differences between a phase-multiplexed lens array (or a multifocus Fresnel CGH) and an array of Fresnel lenses (space-multiplexed).



In Chapters 4 and 16, imaging applications require mostly arrays of microlenses, whereas nonimaging applications require phase-multiplexed microlenses (e.g. clock broadcasting, multispot laser processing in the biomedical sector and industry etc.).

## 6.2 Designing CGHs

Before attempting to calculate or optimize a CGH, a number of parameters need to be set, such as the number of cells and cell sizes in both directions, the number of phase levels and so on, which are linked to the SBWP defined in Equation (6.1).

Once these parameters are defined, an appropriate cost function linked to the final application, an appropriate numeric propagator as well as an appropriate iterative algorithm are chosen. Then, the design task can be launched. In the sections that follow, these various steps involved in a CGH design task are reviewed.

### 6.2.1 Choosing the CGH Initial Parameters

#### 6.2.1.1 The Size of the CGH Aperture

When calculating a CGH, the first parameter to choose is the size of the CGH aperture itself. The effective exit pupil of the CGH is rarely the size of the CGH itself. The effective size of the CGH is the size of the incoming beam multiplied by the size of the CGH:

$$D_{eff}(x, y) = \text{rect}\left[x/M \cdot c_x, y/N \cdot c_y\right] \cdot e^{-\left(\frac{x^2}{\omega_x^2} + \frac{y^2}{\omega_y^2}\right)} \quad (6.2)$$

where  $\omega_x$  and  $\omega_y$  are the beam waists in both directions.

If the beam is larger than the CGH, an aperture stop should be inserted before or after the CGH, which has the size of the CGH or smaller, since any light falling onto a nonstructured part of the substrate (in either transmission or reflection mode) will contribute to the zero order.

Therefore, when using a Gaussian beam without an aperture stop and with a finite size CGH, even though the CGH has no zero order, a zero order would appear, which is just due to the amount of light from the Gaussian beam tail falling outside the CGH aperture (see Figure 6.2). This is a typical error that optical engineers make when using a CGH with a Gaussian beam.

Figure 6.6 shows three possible scenarios:

- The first uses a beam smaller than the CGH. However, a Gaussian beam has no physical lateral end, so the diffraction efficiency of the effective system '*beam + CGH*' is decent, with a small zero order coming from the small part of the beam tail hitting the edges of the CGH.
- The second uses a beam larger than the CGH. A large part of the incoming light falls outside the CGH and contributes to a large zero order, decreasing the effective efficiency although the CGH might have 100% efficiency.
- The third uses a beam larger than the CGH, but hard clips this beam to the exact size of the CGH through the use of an aperture stop. The effective system efficiency is thus maximal and patched to the efficiency of the CGH itself. However, due to the fact that the beam is Gaussian and not uniform, the uniformity of the reconstruction will also be affected by this beam. Later on in the chapter, a CGH optimized for a specific incoming beam (see Section 6.4.3) and thereby creating the exact intensity profile that is computed in the design process, can be seen.

The size of the beam thus has to be considered in choosing the size of the CGH (and its aperture stop if needed).

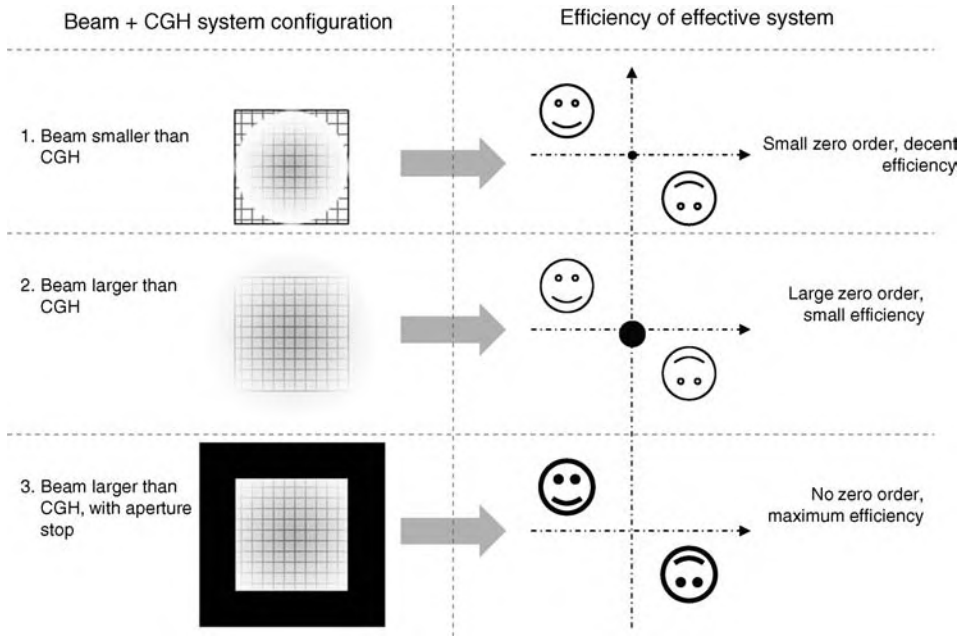


**Figure 6.6** The effective efficiency of a beam + CGH system: a Gaussian beam launched onto the same perfect binary CGH – with and without an aperture stop

Another criterion for choosing the size of the CGH is the pixel size (resolution) to be reached in the reconstruction plane (near field or far field). The reconstruction pixel size or resolution for circular beams is usually given by the diffraction-limited beam as

$$\begin{cases} R_x \approx 1.22 \frac{\lambda f}{D_{\text{eff}-x}} \\ R_y \approx 1.22 \frac{\lambda f}{D_{\text{eff}-y}} \end{cases} \quad (6.3)$$

When designing a Fresnel CGH, the focal length is the focal length of the Fresnel element. When designing a Fourier element, the focal length is the focal length of the Fourier transform lens used. In the case where no Fourier transform lens is used, Equation (6.3) cannot apply. Instead, we consider the natural divergence of the incoming beam and its size at the reconstruction distance to assess the size of the reconstruction pixels.

When the CGH size is chosen with reference to the size of the beam and the resolution in the reconstruction pixel, the basic cell sizes in both directions can be chosen.

### 6.2.1.2 Basic CGH Cell Sizes

As seen in Figure 6.3, the basic CGH cell size can be rectangular and the cell size multiplied by the number of cells gives the real pupil size of the CGH.

The basic cell size dictates the size of the principal reconstruction window – see Figure 6.5 (in which the fundamental diffraction orders will appear). In the case where an FFT-based optimization algorithm is

used, the number of sampled pixels in this window where energy can appear is exactly the number of cells in the CGH in the same direction. If a DFT-based algorithm is used, the number of pixels can be different.

In order to choose the right pixel size effectively, two main issues are considered:

1. The Shannon (or Nyquist) sampling criterion for the reconstructed pattern.
2. The realm of validity of scalar diffraction theory.

### ***The Shannon (or Nyquist) Sampling Criterion***

In order to fully reconstruct a pattern either in the near or the far field, the desired object (or pattern) should be sampled according to the Nyquist criterion, which states that the sampling frequency (which is twice the cell size in most binary cases) must be greater or equal to twice the highest frequency in the original signal (i.e. the pixel bearing a nonnull intensity located furthest away from the optical axis in the desired diffracted pattern):

$$\begin{cases} \frac{1}{P_{\min\_x}} = \frac{1}{2 \cdot c_x} \geq 2\nu_{\max\_x} \\ \frac{1}{P_{\min\_y}} = \frac{1}{2 \cdot c_y} \geq 2\nu_{\max\_y} \end{cases} \quad (6.4)$$

Now, this is a generic definition. Like any definition, it is not applicable to a specific problem without further analysis.

In order to fully understand the relation between the CGH pixel size (the smallest frequency in the CGH plane) and the extent of the reconstruction window (the largest possible angle in the reconstruction plane), the grating equation is applied and the smallest grating pitch should be able to diffract light at the edges of the reconstruction window (see Figure 6.7):

$$\begin{cases} \sin(\alpha_{\max\_x}) = \frac{\lambda}{P_{\min\_x}} = \frac{\lambda}{2c_x} \\ \sin(\alpha_{\max\_y}) = \frac{\lambda}{P_{\min\_y}} = \frac{\lambda}{2c_y} \end{cases} \quad (6.5)$$

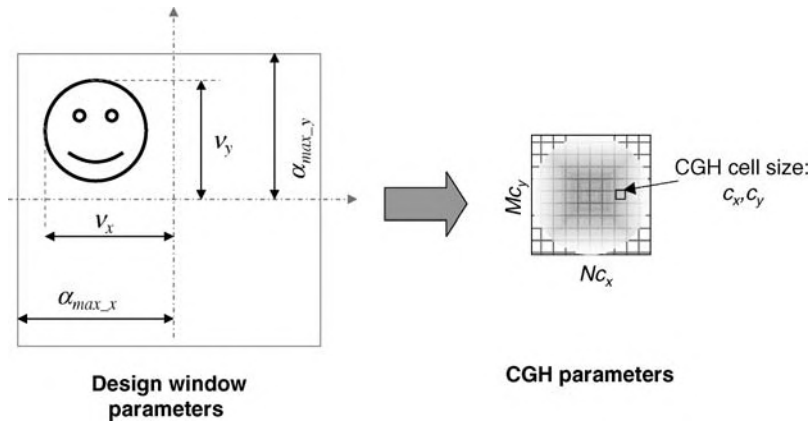


**Figure 6.7** The basic (fundamental) reconstruction window and the CGH plane

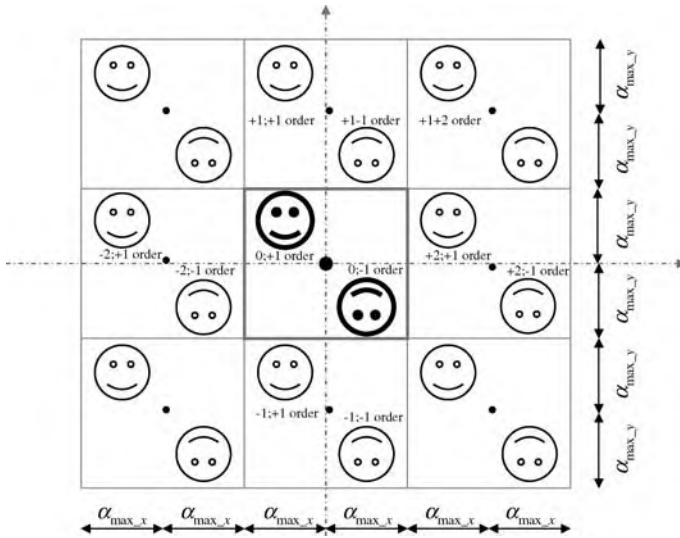


**Figure 6.8** The various reconstruction orders (in the Fourier plane) around the fundamental reconstruction window defined by the binary CGH cell size

Equation (6.5) actually shows the same thing as the Nyquist criterion, but from a grating equation point of view. The smallest grating period (twice the cell size) defines the reconstruction window, and therefore any point within the reconstruction itself is diffracted from gratings that have larger periods. This is true not only for binary elements but also for multilevel elements.

In Figure 6.7, note that the maximum frequency in the reconstructed signal (the smiling face) is smaller than the largest possible frequency owing to the smallest grating pitch (twice the basic CGH cell size). This gap is important for the quality of the reconstruction (and relates to the  $\geq$  sign in the Nyquist criterion), in order to increase the Signal-to-Noise Ratio (SNR) around the reconstruction. Many orders will appear in the reconstruction plane, as depicted in Figure 6.8 (especially for binary CGHs). The dark areas in the fundamental window help to reduce the interferences between higher and fundamental orders.

The position of the object within the basic reconstruction window is also important, and care should be taken about the position of the conjugate order within that same window (no higher orders will appear in a single basic window). Figure 6.9 shows a typical error when designing a CGH and then fabricating it as a binary element (binary phase elements yield the same efficiency in the fundamental negative and positive orders, of 40.5% – see Chapter 5).

Note that since the object pattern is symmetric and on-axis in the Fourier plane, the conjugate order is exactly the real image itself, so they overlap and produce 81% (effective) efficiency rather than 40.5% efficiency. This feature is used extensively for applications such as symmetric spot array generators, symmetric beam shaping and so on.

In Figure 6.10, note that in this case the reconstruction is not symmetric. A binary element is not sufficient to reproduce the pattern. Multilevel fabrication is necessary. Note that the zero order of some elements has almost vanished, owing to the quality and accuracy of the etching depth.

Figure 6.11 shows some symmetrical Fourier reconstructions. In this case, a binary or four phase level element generally produces the same efficiency: the efficiency increases when fabricating with more than four levels.

Actually, whether the element was a binary or multilevel element cannot be discerned using the optical reconstruction. The only difference is in the quality of the reconstruction (uniformity). However, due to



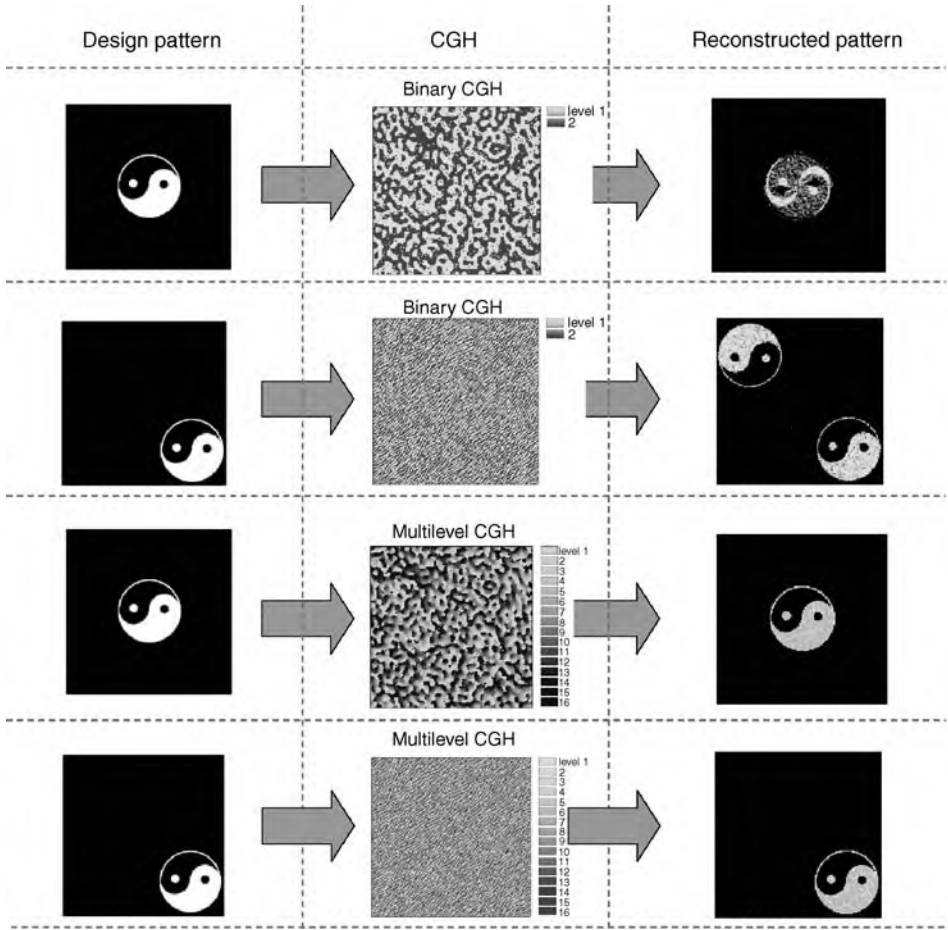


**Figure 6.9** The importance of positioning the object well within the reconstruction window

systematic fabrication errors (see Chapters 12 and 13), the multilevel optical reconstruction would, interestingly, be more noisier than the binary one. Note the higher orders appearing in higher-order windows in the central pattern (the grid).

Figure 6.12 is taken in the far field of a binary Fourier spot pattern generator. It represents two arrays of  $50 \times 50$  collimated beamlets.

If the desired reconstruction is only the array of spots on the right-hand side, this element is a binary element, splitting energy equally between both orders in the far field. If the desired pattern is two arrays of  $50 \times 50$  spots, it is uncertain whether this pattern is from a binary or a multilevel CGH, since both would yield the same type of reconstruction (namely, two arrays). In a binary element, there would be two diffraction orders reconstructing the two arrays (off-axis initial object) and in a multilevel element there would be a single diffraction order reconstructing the two arrays (on-axis initial object). It is most likely a binary element, since the higher orders on each side of the reconstruction are noted.

Now, Figure 6.13 shows a very interesting optical reconstruction of a Fresnel element in the near field. The element is binary and it is an off-axis  $50 \times 50$  multifocus lens (a phase-multiplexed lens – see also Figure 6.4).

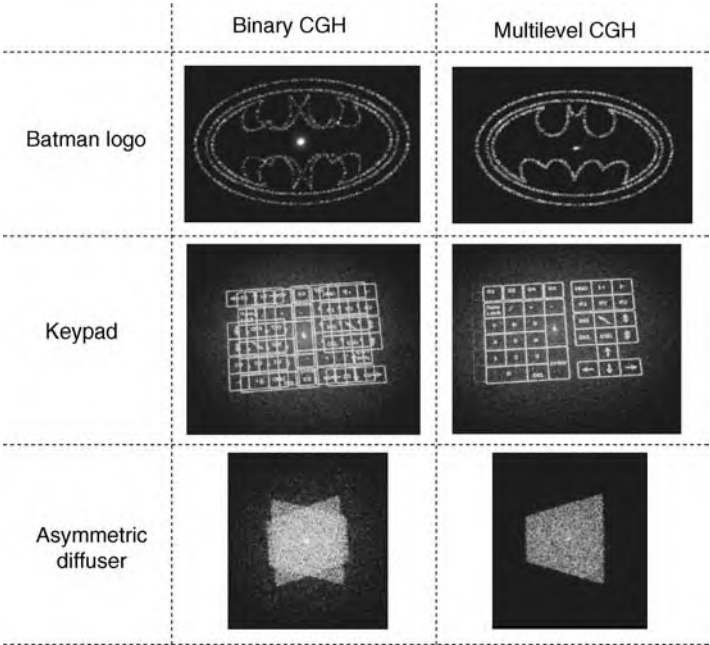


**Figure 6.10** An example of nonsymmetric Fourier patterns in binary and multilevel modes

In Figure 6.13, note the well-focused array of  $50 \times 50$  spots on the right-hand side and a less well-focused array of  $50 \times 50$  beams to the left of the central beam. The right-hand array is the converging array of beamlets arising from the fundamental positive diffraction order. The left-hand side is the conjugate order, namely the array of diverging beamlets arising from the fundamental negative order. The beam in the center is DC light (zero order). In this instance, the zero order is not focused (not in the far field).

There are  $2 \times 2500 = 5000$  beamlets in this photograph, but only two diffraction orders (the two fundamental orders and the zero order).

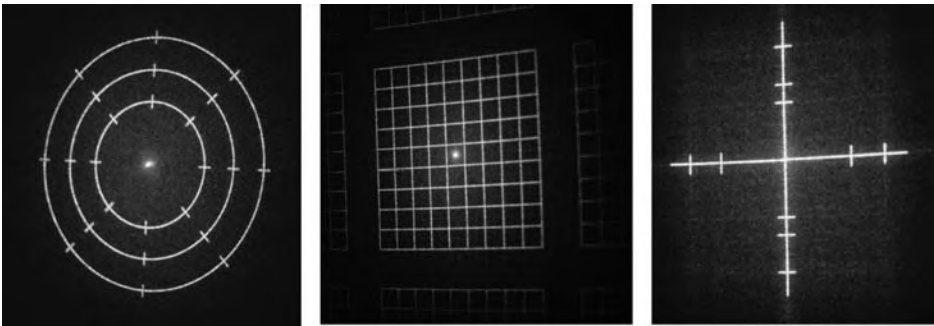


**Figure 6.11** On-axis symmetric Fourier patterns

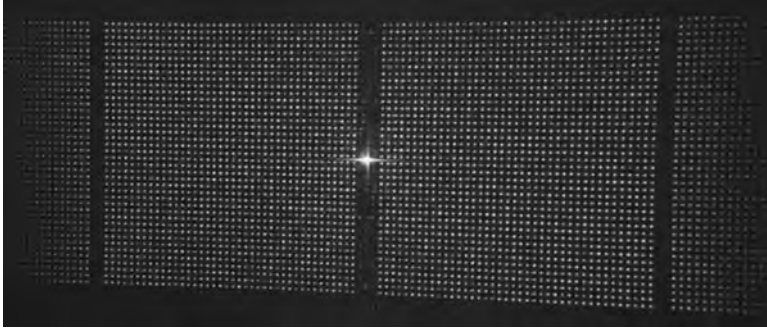


**Figure 6.12** The far-field pattern of a Fourier CGH (fan-out grating)

### *The Realm of Validity of Scalar Diffraction Theory*

Another important issue when choosing the basic CGH cell sizes is the validity of the scalar theory of diffraction. A numeric propagator based on scalar theory will work with any cell size, even if this cell size is much smaller than the wavelength (see Section 1.4). However, while scalar theory is only accurate for grating pitches that are larger than about five times larger than the wavelength, it is still more or less valid for grating pitches between twice and five times that wavelength:

$$\begin{cases} \text{Scalar theory accurate:} & P_{\min} > 5\lambda \\ \text{Scalar theory inaccurate:} & \lambda < P_{\min} < 5\lambda \\ \text{Scalar theory invalid:} & \lambda \leq P_{\min} \end{cases} \quad [\text{where } P_{\min} = 2c] \quad (6.6)$$

When choosing CGH cells close to the size of the reconstruction wavelength, scalar theory is no longer valid. Scalar theory predicts the amount of light in the various diffraction orders. It is important to understand that the diffraction angle calculations by means of the grating equation or the geometry of the reconstruction patterns calculated by the numeric propagators are not affected by reduction of the cell size. The amount of light diffracted in these orders will change, and there might no longer be any light in

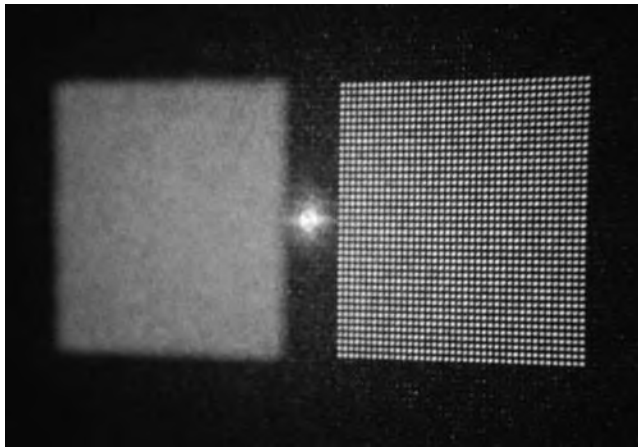


**Figure 6.13** A binary Fresnel  $50 \times 50$  off-axis multipot lens – (near-field location)

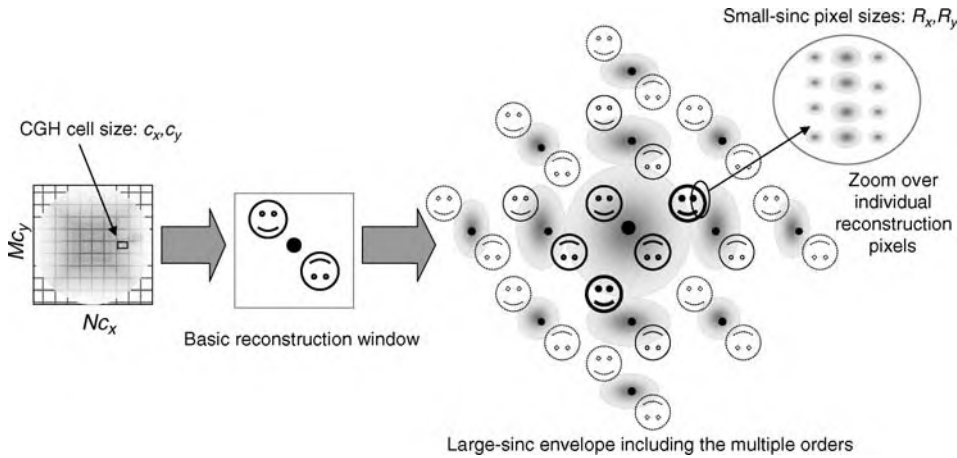


**Figure 6.14** The global reconstruction plane, which is infinite in both directions, comprising all diffraction orders

the fundamental order, even though the depth of the structures might be still optimal according to scalar theory. In order to push more light into these orders when reducing the cell size, rigorous electromagnetic and vector diffraction theories can be used to re-optimize the depth of these smaller structures, but this depth might actually become a function of the local grating pitch in the CGH, and therefore increase the complexity of the fabrication dramatically.

### 6.2.1.3 The Diffraction Orders Envelope in the CGH Reconstruction Plane

Once the pixel size ( $c_x, c_y$ ) and the number of cells ( $N, M$ ) in the CGH are set, along with the reconstruction wavelength  $\lambda$  (and, in the case of a Fresnel element, the focal length  $f$ ), the reconstruction plane is then defined – see Figure 6.14. This reconstruction plane comprises two main features:

- a large-sinc envelope, under which the various reconstructed diffraction orders appear; and
- a small-sinc envelope, which defines the size of the individual reconstruction pixels.

Figure 6.8, in the previous section, shows the location of the various diffraction orders around the main reconstruction window. Figure 6.14 now shows the intensity modulation over these various orders. In the case of a rectangular CGH cell, this envelope is a sinc function. This sinc function is actually the Fourier transform of the basic cell size, and the reconstruction pixel is the Fourier transform of the CGH aperture. So, depending on the cell geometry, the sinc function can be symmetric or asymmetric.

Figure 6.15 shows a numeric reconstruction of the CGH designed in Figure 6.9. This numeric reconstruction uses both the oversampling and embedding processes defined in Chapter 11 in order to show clearly the various reconstructed orders and how they fit into the 2D sinc envelope.

#### The ‘Sinc’ Compensation Method

Although the fundamental orders are placed within the main lobe of the sinc function (Figure 6.15) before the main lobe hits a null value, their intensity is still modulated by that envelope, depending on their position (diffraction angles) within the main reconstruction window (which is also linked to the main sinc lobe, as will be seen in Chapter 11).

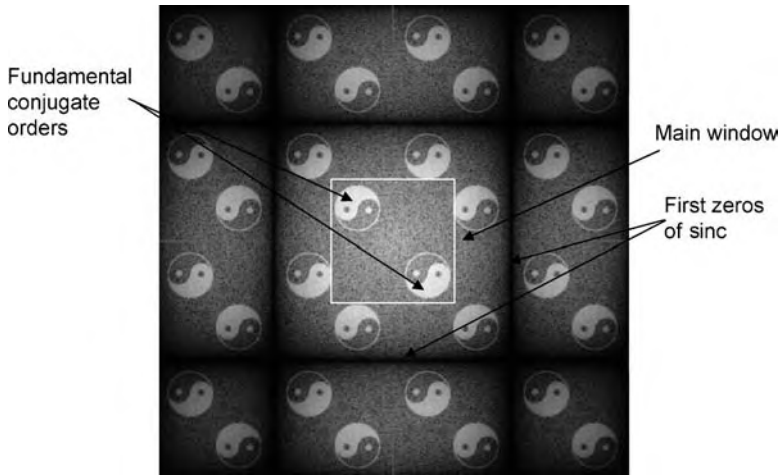


**Figure 6.15** A numeric reconstruction of a binary Fourier CGH (log intensity scale)

If the uniformity of the reconstruction is a critical feature for the application (see the cost function definitions in Section 6.2.3), it is therefore very desirable to compensate for this sinc modulation effect by a pre-emphasis on the pattern, and modulate that desired pattern by the inverse sinc envelope. In the design of diffractive optics, this pre-emphasis is called the sinc compensation process.

## 6.2.2 Constraints and Degrees of Freedom in CGH Optimization

### 6.2.2.1 CGH-related Degrees of Freedom

In order to optimize a CGH element for a specific task, we consider what can actually change in the design process; namely the degrees of freedom of the CGH. The degrees of freedom are directly related to the space bandwidth product (Equation (6.1)) of that CGH. The SBWP is defined prior to the optimization process and describes a set of constraints related to CPU time consumption, the geometry of the CGH reconstruction type and fabrication constraints. Usually, the degrees of freedom are the number of CGH cells, the size of these cells (and thus the CGH's aperture) and the number of phase levels under which to fabricate the CGH.

An algorithm that tries to optimize a binary CGH over  $32 \times 32$  cells may eventually not converge, since the degrees of freedom are very few, whereas the same optimization process may be able to lead to an adequate solution to the problem by increasing the number of cells and/or the number of phase levels.

### 6.2.2.2 Reconstruction Plane Degrees of Freedom

The reconstruction plane (or volume, for this can be a 3D reconstruction in the near field) represents the desired object (phase, amplitude or a combination thereof), which is a constraint on the design procedure. The optimization algorithm will thus try to find a CGH that would reconstruct that object as accurately as possible while reducing the amplitude variations in the CGH plane to unity.

To relax the degrees of freedom and allow the algorithm to converge smoothly, use is made of the fact that the surrounding area of that object may or may not be of importance to the final application. In some cases, a degree of freedom can be relaxed in the reconstruction plane by allowing noise to evolve and concentrate in regions of no interest to the application.

Another degree of freedom can be relaxed over the object itself, when the application does not require high uniformity. Note that the decrease in uniformity does not imply a decrease in diffraction efficiency, but often has an inverse effect: it increases the overall efficiency of the CGH. Reconstruction quality considerations can be expressed accurately by the diffraction efficiency, the Root Mean Square (RMS) error in reconstruction, the Signal-to-Noise Ratio (SNR), the uniformity of the reconstruction and several other parameters. These criteria give an idea of the quality of the reconstruction, and thus can drive the optimization algorithm. These criteria are the basis for defining the appropriate cost function.

### 6.2.3 Cost Function Definitions

The cost function is the global function that has to be optimized by the iterative optimization algorithm [8]. Such a function could be simply the diffraction efficiency, or the uniformity, or it could be a linear combination of several criteria [9].

An efficient cost function definition is a linear combination of several of the criteria described below, each weighted by a coefficient related to the importance of that particular criterion:

$$\Psi = \sum_{i=1}^N (1 - c_i) \cdot (1 - \wp_i) \quad (6.7)$$

The cost function  $\Psi$  is defined in Equation (6.7) with normalized coefficients  $c_i$  and normalized criterion  $\wp_i$ . The task is to minimize  $\Psi$ .

Defined below are some of the criteria that can be used in the cost function definition. In order to define the following quality criteria or parameters, the following functions and parameters need to be defined first (see Figure 6.16):

Here, we assume perfect fabrication of the DOE and perfect optical materials. The diffraction efficiency is usually a major ingredient for most of the cost functions, whatever the intended application.

#### 6.2.3.1 Diffraction Efficiency

The theoretical diffraction efficiency using scalar diffraction theory is described in Chapter 5 and is developed for simple linear gratings. As the diffraction pattern for CGHs is often no longer a couple of spots in the far-field (binary grating) or a plane uniform wave in the near field, calculation of the exact theoretical diffraction efficiency can be very complex.

In a straightforward way, the diffraction efficiency can be described as the ratio of the amount of light in the first positive fundamental negative order (within the region of interest A in Figure 6.16) divided by the

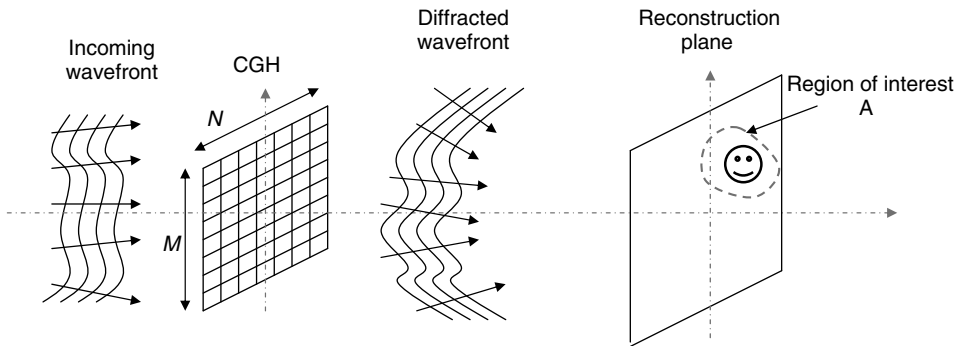


**Figure 6.16** The CGH optimization task

amount of light launched onto the DOE (any other energy losses are not considered):

$$\eta = \frac{\sum_{(n,m) \in A} |h_{m,n}|^2}{\sum_{n=1}^N \sum_{m=1}^M |h_{m,n}|^2} \quad (6.8)$$

The above expression is valid when the fundamental negative order is spatially isolated from all the other orders in the desired reconstruction plane; this is actually seldom the case in the near field, especially for binary Fresnel CGHs, where a large amount of light is diffracted in diverging negative orders, and for realistic elements with etch depth errors that produce zero-order waves that submerge the whole near field.

Finally, diffraction efficiency is a notion that is very tightly related to the type of application, as seen previously.

Later on, in practical examples, if optimization is attempted, a CGH with only diffraction efficiency in mind, other criteria might go downhill, such as uniformity in the reconstruction, the SNR or the RMS error in the reconstruction plane.

### 6.2.3.2 Root Mean Square (RMS) Error

The RMS error in the reconstruction plane is a good parameter to evaluate the quality of the numeric reconstruction and relates to the convergence rate of the optimization algorithm. This convergence rate is very closely linked to the number of degrees of freedom allowed by the design (SBWP and others). The complex error  $\text{RMS}_c$  in the reconstruction plane is defined as follows:

$$\left\{ \begin{array}{l} \text{RMS}_c = \frac{1}{NM} \sum_{n \in A} \sum_{m \in A}^{M_A} (f_{n,m} - a_c \cdot h_{m,n})^2 \\ \text{where } a_c = \frac{\sum_{n \in A} \sum_{m \in A}^{M_A} (f_{n,m} - h_{m,n}^*)^2}{\sum_{n \in A} \sum_{m \in A}^{M_A} |h_{m,n}|^2} \end{array} \right. \quad (6.9)$$

where  $a_c$  is a constant that normalizes the numerical reconstruction in order to obtain coherent error values;  $N$  and  $M$  are the numbers of square cells in the DOE;  $N_A$  and  $M_A$  are the numbers of pixels in the considered area  $A$  within the reconstruction plane, an area that is important to the application ( $N_A \cdot M_A < N \cdot M$ );  $h_{n,m}$  is the numerically reconstructed wavefront's complex amplitude, sampled at a location indexed by  $n$  and  $m$  in the reconstruction plane (either near-field or far-field); and  $f_{n,m}$  is the desired reconstruction in the reconstruction plane.

The normalized complex  $\text{RMS}_{CN}$  considers the differences in both phase and amplitude between the desired and actual reconstructions, and is defined as follows:

$$\text{RMS}_{CN} = \frac{\sqrt{\text{RMS}_c}}{\max_{n,m} (|f_{n,m}|)} \quad (6.10)$$

If it is desirable that no energy at all should be present in the reconstruction plane outside the area of interest  $A$  (a high constraint on the optimization algorithm), the optimization criterion becomes:

$$\sum_{n \in A} \sum_{m \in A}^{M_A} (h_{m,n} f_{n,m})^2 = \sum_{n=1}^N \sum_{m=1}^M (h_{m,n} f_{n,m})^2 \quad (6.11)$$

The RMS error in amplitude (or even in intensity) can also be defined, and is actually more often used than the complex error, since the numeric-type CGH-related applications are often only concerned about intensity distributions, and much less about phase distributions in the reconstruction plane.

Thus, the errors in amplitude  $\text{RMS}_A$  and intensity  $\text{RMS}_I$  are defined as follows:

$$\left\{ \begin{array}{l} \text{RMS}_A = \frac{1}{NM} \sum_{n \in A} \sum_{m \in A}^{M_A} (f_{n,m} - a_c \cdot h_{m,n})^2 \\ \text{where } a_A = \frac{\sum_{n \in A} \sum_{m \in A}^{M_A} (f_{n,m} - h_{m,n}^*)^2}{\sum_{n \in A} \sum_{m \in A}^{M_A} |h_{m,n}|^2} \end{array} \right. \quad \left\{ \begin{array}{l} \text{RMS}_I = \frac{1}{NM} \sum_{n \in A} \sum_{m \in A}^{M_A} (f_{n,m} \cdot f_{n,m}^* - a_c \cdot h_{m,n} \cdot h_{m,n}^*)^2 \\ \text{where } a_I = \frac{\sum_{n \in A} \sum_{m \in A}^{M_A} (f_{n,m} \cdot f_{n,m}^* - h_{m,n} \cdot h_{m,n}^*)^2}{\sum_{n \in A} \sum_{m \in A}^{M_A} |h_{m,n} \cdot h_{m,n}^*|^2} \end{array} \right. \quad (6.12)$$

For most of the applications described in this chapter and in the following chapters (including Chapter 16), the amplitude and especially the intensity-based RMS definition are used.

### 6.2.3.3 The Signal-to-Noise Ratio (SNR)

If it is assumed, in this section, that the fabrication of the CGH is perfect, and if the positive fundamental diffraction order is considered to be the desired signal, the only remaining noise in the reconstruction plane is the multiple diffraction orders that are present. If the signal is only a part of the actual first-order diffraction (i.e. 'noise' introduced in the first order by encoding and quantization errors), this noise cloud adds to the multiple diffraction orders. Note that although the different orders may be well separated spatially in the far-field pattern, overlapping may be observed in the near field between different orders (especially diverging negative diffraction orders overlapping converging positive diffraction orders, and thus the first positive fundamental order).

The SNR is defined as follows:

$$\text{SNR} = -20 \log_{10} \left( \frac{S}{N} \right) \quad (6.13)$$

If a perfect surface-relief phase grating is considered, the signal is the diffraction efficiency in the first order and the noise is the efficiency in the remaining higher positive and negative orders (indexed by  $p$ ); thus

$$\text{SNR} = -20 \log_{10} \left( \frac{\eta_{+1}}{\eta_{-1} + \sum_{p=2}^{+\infty} \eta_p + \sum_{p=-2}^{-\infty} \eta_p} \right) \quad (6.14)$$

Equation (6.13) expresses the noise as the summation of all other diffraction orders. This is a theoretical definition that cannot easily be applied on a case-by-case basis with numeric reconstructions. Therefore, in most practical cases, the following definition of the SNR is used:

$$\text{SNR} = -20 \log_{10} \left( \frac{\sum_{n \in A} \sum_{m \in A}^{M_A} f_{n,m} \cdot f_{n,m}^*}{\sum_{n=1}^N \sum_{m=1}^M f_{n,m} \cdot f_{n,m}^*} \right) \quad (6.15)$$

This is the definition of the SNR that that will be used for most of the designs shown throughout this book.



For applications in optical interconnections, a very desirable characteristic is the reliability of the connections. This reliability of transmission is governed by the Bit Error Rate (BER) criterion, which also depends on the value of the SNR.

### 6.2.3.4 The Bit Error Rate (BER)

The Bit Error Rate (BER) is an important characteristic for analyzing and evaluating image planes from arrays of multiple fan-out Fresnel or Fourier DOEs or Fresnel lenses, especially in optical computing applications such as optical interconnects or optical signal broadcasting in a fiber. For a symmetric system with Gaussian noise, the BER can be related to the SNR by

$$\text{BER} \cong \frac{1}{2} \left( 1 + \frac{\text{erf}(\text{SNR})}{2\sqrt{2}} \right) \quad (6.16)$$

where  $\text{erf}$  is the standard error function.

In order to achieve  $\text{BER} < 10^{-12}$  (which is a typical value for optical interconnects within optoelectronic devices), Equation (6.16) shows that the SNR value has to be at least 14.

In optical data storage, the BER value has to be larger. However, this can also be controlled by software (error correction algorithms). For example, in holographic page data storage, the image crosstalk due to the limited angular bandwidth of angularly multiplexed holograms (see Chapters 8 and 16) is the principal cause of deterioration in the BER.

### 6.2.3.5 The Strehl Ratio

The Strehl ratio is a criterion linked to the ability of lenses (refractive, reflective or diffractive) to focus energy in the Airy disk. The larger the side lobes, the smaller is the Strehl ratio. The Strehl ratio for a diffractive can be defined as follows:

$$\gamma = \frac{\eta_{+1}^{+\infty}(\phi_0 = 1)}{\eta_{+1}^N(\phi_0 \neq 1)} \quad (6.17)$$

where  $\eta_{+1}^{+\infty}(\phi_0 = 1)$  is the efficiency of the perfect lens (detuning factor  $\phi_0 = \text{unity}$ ) in the fundamental order as an analog surface-relief element ( $N \rightarrow \infty$ ) and  $\eta_{+1}^N(\phi_0 \neq 1)$  is the efficiency in the fundamental order of the actual imperfect fabricated lens with a limited number  $N$  of phase-relief levels ( $N = 2, 4, 8, \dots$ ) and the actual detuning factor ( $\phi_0 \neq 1$ ) linked to systematic fabrication errors (such as etch depth errors).

## 6.2.4 The Optimization Task in the Complex Plane

Before choosing an adequate numeric propagator associated with an adequate iterative optimization algorithm (see the following sections), the optimization tasks at hand need to be defined. Previously, it was seen that an amplitude diffractive element yields very low efficiency. Therefore, the optimization procedure will be asked to constrain the amplitude function to unity over the entire diffractive element. Second, the fabrication tools on hand, which can only fabricate phase levels over a limited number of steps (2, 4, 8, 16), need to be considered.

Therefore, turning to the complex plane, the CGH complex information in the complex unity circle (phase of the pixel being the angle of the vector, and the amplitude of the pixel being the length of the vector) is described. The optimization task in that complex plane can thus take the following form:

*Find an optimal sampled complex matrix that will produce the desired pattern in the near or far field, with an amplitude as close as possible to the unit circle and such that the phases of the various pixels are as close as possible to the  $N$  directions of space linked to the  $N$  allowed fabrication steps.*

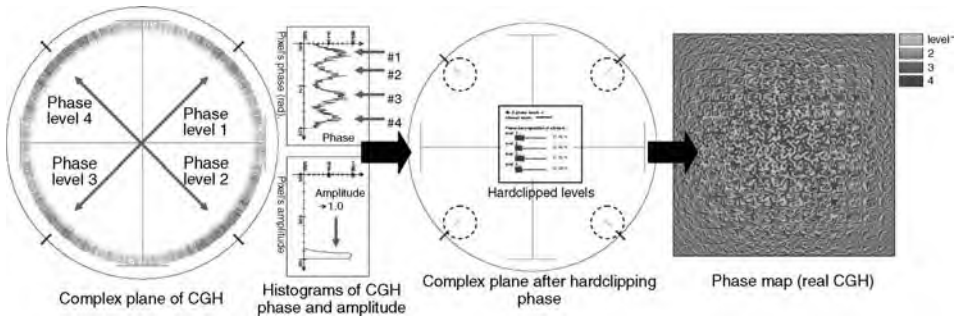


**Figure 6.17** An example of complex plane CGH optimizations

Figure 6.17 shows examples of a Fresnel CGH optimized in the complex plane as four phase level elements.

Figure 6.17 shows, on the left-hand side, the phase map of the CGHs (without amplitude information) and, on the right-hand side, the associated complex CGH plane. The complex plane shows the resulting function after the algorithm has stopped, before hard clipping the pixels to the four allowed phase levels. There is a relatively large amount of information that has been lost when hard clipping this function of a four phase level element: loss of amplitude information (which has been pushed to the unit circle but has still some remaining modulations) and of the phase, which is grouped over the eight phase poles but not perfectly set. Nevertheless, such an element can yield an efficiency of almost 80%. This is why the optimization of a CGH is being discussed, rather than calculation. The perfect point cannot be reached at which the optical functionality matches exactly what the fabrication technology can do (analog phase and analog amplitude).

### 6.2.5 Choosing an Adequate Numeric Propagator

The various numeric propagators that can be used to implement either a near-field propagation or a far-field propagation are listed in Chapter 11. There are mainly two groups:

- FFT-based propagators; and
- DFT-based propagators.

The first set of propagators are based on the Fast Fourier Transform algorithm (see Appendix C), which can quite easily implement Fraunhofer propagation (basically a complex 2D FFT) and two sets of near-field propagations (Fresnel propagation): the direct one and the convolution-based one.

The three popular FFT-based scalar propagators – (A) Fraunhofer, (B) Fresnel direct and (C) Fresnel convolution – can be summarized as follows:

$$\left\{ \begin{array}{l} \text{(A)} \quad U'(u_n, v_m) = \text{FFT} [U(x_n, y_m)] \\ \text{(B)} \quad U'(x'_{1n}, y'_{1m}) = e^{\frac{j\pi}{\lambda d}(x_{1n}^2 + y_{1n}^2)} \cdot \text{FFT} \left[ U(x_n, y_m) \cdot e^{\frac{j\pi}{\lambda d}(x_n^2 + y_n^2)} \right] \\ \text{(C)} \quad U'(x'_{2n}, y'_{2m}) = \text{FFT}^{-1} \left[ \text{FFT} [U(x_n, y_m)] \cdot \text{FFT} \left[ e^{\frac{j\pi}{\lambda d}(x_n^2 + y_n^2)} \right] \right] \end{array} \right. \quad (6.18)$$

The sizes of the reconstruction windows, set to the fundamental order window, are very rigid; however, the sizes of the reconstruction windows give fast and reliable results.

The second set of propagators are based on the Discrete Fourier Transform, and are more versatile in implementations (especially in reconstruction window resolution and size), but are very slow, and therefore are not often used. They also need more design rule checks than the FFT-based propagators.

For more insight on these various numeric propagators, see Chapter 11.

### 6.2.6 Choosing an Adequate Optimization Algorithm

Iterative optimization algorithms have been developed by engineers in various fields since the day computers became available to solve problems that are too complicated to be solved using a pen, a piece of paper and even the mind of a genius.

According to Darwin's theory of evolution, for a long time now nature has been using its own iterative optimization algorithm in order to develop complex solutions to a given problem: how to create and sustain life on Earth and allow our human nature to evolve, by using one of the groups of algorithms presented below (genetic algorithms).

Three different groups of algorithms are presented:

- IFTA-type algorithms;
- steepest descent algorithms; and
- genetic algorithms.

For each of these groups of algorithms, the task at hand is to minimize the cost function defined earlier. In Figure 6.18, a one-dimensional problem is depicted. The real task is, however, multi-dimensional.

If the cost function is a monotonous function, finding the global minimum is an easy task. If the cost function has several local minima (as depicted on the left-hand side of Figure 6.18), the deterministic iterative algorithm could get stuck in one of these local minima and then it would be impossible to find an acceptable solution. Therefore it is important to allow the algorithm to get out of these traps in order to converge to an acceptable solution, even though it might not be the global minimum (the right-hand side of Figure 6.18). Such a process is called a stochastic process, and simulated annealing is one of the potential algorithms that can produce such an effect.

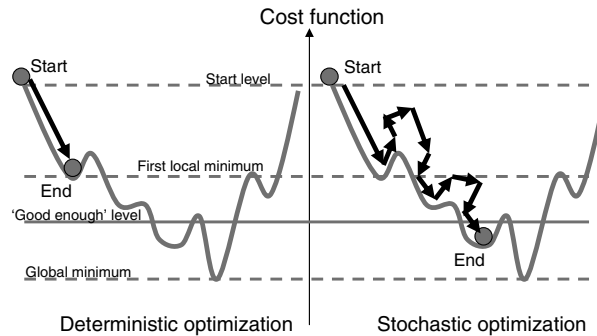


**Figure 6.18** Minimizing the cost function

### ***When is the Cost Function Low Enough to Stop the Algorithm?***

This is a very pertinent question, since an iterative algorithm can run for an infinite time. There are four ways to stop an iterative algorithm:

1. The most straightforward way is to stop the algorithm when the cost function no longer changes (trapped in a local or global minimum).
2. The most common way is to define an acceptable cost function (see Figure 6.18). When the algorithm reaches this cost function, it stops. It is tempting to set the cost function to zero, which would correspond to a perfect element. However, such a cost function reduction would never be reached. So, care has to be taken when defining the end point.
3. The other popular way to end an iterative algorithm is to measure the convergence rate of the cost function. If the cost function does not improve by more than a certain percentage per iteration, one can decide to stop the algorithm.
4. The last method is to stop the cost function after a fixed amount of time or after a fixed set of iterations. This method is frequently used for IFTA optimization algorithms.

#### **6.2.6.1 How to Start an Iterative Algorithm**

In the previous section, it has been shown how to stop an algorithm. Now it is important to know how to jump-start such an algorithm.

There are many ways to start the iterative process, one of which is to have a uniform array of amplitude and phase values, while another way is to have a complete set of random complex pixels filled out in the original CGH plane. Yet another approach, which is closer to the final solution, is to produce a set of complex clouds around the various allowed phase values, and thus also set the intensity constraints to the unit circle (see Figure 6.19).

In a general way, it is always interesting to get closer to the final solution (a set of phase levels in a uniform amplitude bath) and, at the same time, allow some dynamic range in the CGH (by randomizing the phase around these poles). Such a randomization has been done in the clouds of Figure 6.19.

#### **6.2.6.2 IFTA Optimization Algorithms**

An Iterative Fourier Transform Algorithm (IFTA) is an algorithm that constantly bounces back and forth between the CGH plane and the reconstruction plane by constraining the amplitude to a set of constraints and by letting the phase map converge to a solution.

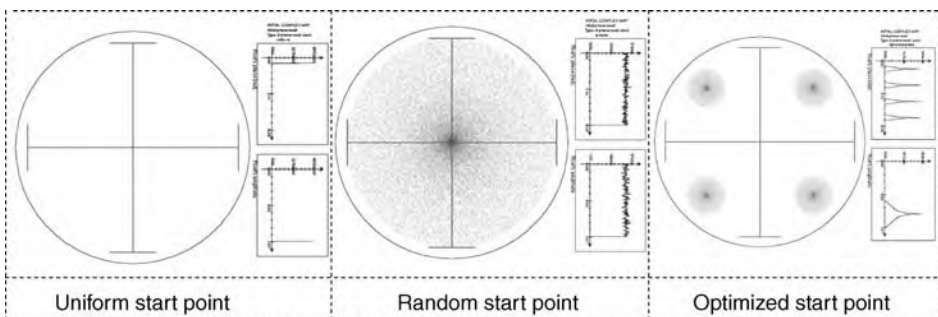


**Figure 6.19** The various starting positions in the complex plane for an iterative optimization algorithm

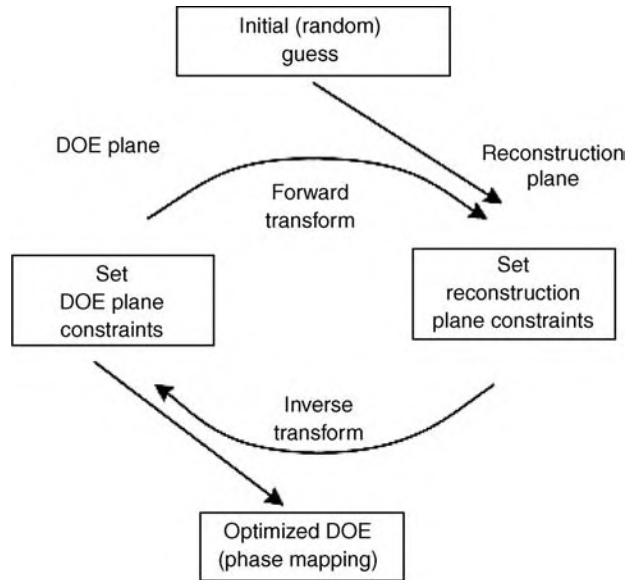


**Figure 6.20** The Gerchberg–Saxton iterative CGH optimization algorithm (IFTA type)

The reconstruction plane can be located either in the near or the far field to produce Fourier or Fresnel CGHs, respectively. There are many IFTA algorithms in the literature, all based on the standard Gerchberg–Saxton phase retrieval algorithm.

### ***The Gerchberg–Saxton Algorithm***

The most commonly used algorithm in industry to design CGHs is the Gerchberg–Saxton (G–S) algorithm [10–12]: an IFTA-type algorithm. The G–S algorithm was originally developed for e-beam microscopy phase retrieval in the far field. It is fast, very reliable (i.e. it converges in most cases), and it is very easy to implement and to use [13]. However, although the diffraction efficiency criterion is optimized quite well using the G–S algorithm, this is not the case for the uniformity criterion. Figure 6.20 shows the flow chart of the G–S algorithm.

Once the complex data of the CGH have been calculated (optimized), the optical design engineer has to choose a specific encoding technique in order to encode the phase, amplitude or complex information into a substrate.

The standard G–S algorithm is often used with an FFT-based propagator. When using a Fresnel FFT-based propagator rather than a far-field Fraunhofer propagator, the ‘Ferwerda’ algorithm is often mentioned, which basically just means a G–S algorithm in the near field.

### ***The Ping-Pong Algorithm***

There are many other IFTA-type algorithms that have been developed in the literature [14, 15]. One of these is the Ping-Pong algorithm, which bounces back and forth between  $N$  planes, rather than two planes, as in the case of the simple G–S algorithm.

As with the Ferwerda algorithm, the Ping-Pong algorithm is based on the Gerchberg–Saxton phase retrieval algorithm. It can optimize a CGH that performs a specific optical function within a first near-field plane and a different optical function within a second near-field plane. This seems to be a very attractive way to generate powerful elements that would integrate several different functions, but the actual

convergence is very strongly related to the complexity of the reconstructions and the differences between the various optical functions that need to be performed. Actually, the Ping-Pong algorithm only converges for a very small number of applications.

This algorithm is especially suited for 3D display Fresnel DOEs where the intensity is smoothly distributed over the entire near field. It is very difficult to make such an algorithm converge for off-axis isolated sharp intensity irradiances, since it is quite complicated to ask light to disappear and reappear somewhere else in space (at least, not with structures larger than the wavelength – see, for example, Section 10.9).

### The Yang–Gu Algorithm

The Yang–Gu algorithm also optimizes multifunctional CGHs, but here the DOE has a specific and single optical function for a specific illumination scheme [16, 17]. The different optical functions here are not triggered at the same time (as in the previous Ping-Pong algorithm), but one at a time, depending on how the CGH's aperture is illuminated and, in particular, on the wavelength with which it is illuminated. Typical applications for such an algorithm are combined de-multiplexing and focusing properties. The CGH works as a range of different off-axis Fresnel lenses that can be triggered one at a time, by tuning the incoming laser beam wavelength.

The Yang–Gu algorithm can deal with nonunitary linear transforms, whereas the original G–S requires unitary linear transformations. A nonunitary linear numeric transformation or propagator  $\hat{G}$  implies the following:

$$\hat{G}^+ \hat{G} = \hat{W} \neq \hat{I} \quad (6.19)$$

where the superscript  $+$  stands for Hermitian conjugation (i.e. it reverses the propagation direction for Fresnel propagators, and it can also be thought as the complex conjugate of the associated numeric propagator),  $\hat{G}$  is the Hermitian operator and  $\hat{I}$  is the identity transform.

A simple example of a nonunitary transform is any numeric propagator propagating a wavefront through an aperture of finite size, thus introducing diffraction losses (which is generally the case). Therefore, the nonunitary behavior increases as the aperture size decreases. If  $N$  different wavelengths are considered in the design procedure in order to implement  $N$  different reconstruction figures  $U_{0,\lambda\alpha}$  ( $1 \leq \alpha \leq N$ ), one obtains

$$\begin{cases} U_{1,\lambda\alpha}(x_1, y_1) = A_{1,\lambda\alpha}(x_1, y_1) \cdot e^{j\varphi_{1,\lambda\alpha}(x_1, y_1)} \\ U_{0,\lambda\alpha}(x_1, y_1) = A_{0,\lambda\alpha}(x_1, y_1) \cdot e^{j\varphi_{0,\lambda\alpha}(x_1, y_1)} \\ U_{0,\lambda\alpha}(x_0, y_0) = \hat{G}_{\lambda\alpha} \cdot U_{1,\lambda\alpha}(x_1, y_1) \end{cases} \quad (6.20)$$

where  $U_{0,\lambda\alpha}$  describes the complex wavefronts launched onto the multifunctional CGH. Since the operator considered here is nonunitary, the phase distributions in each plane can be expressed as follows:

$$\begin{cases} \varphi_1(x_1, y_1) = \arg \left[ \sum_{\alpha=1}^N \hat{G}_{\lambda\alpha}^+ \cdot U_{0,\lambda\alpha}(x_0, y_0) - \hat{W}_{\lambda\alpha} \cdot U_{1,\lambda\alpha}(x_1, y_1) \right] \\ \varphi_{0,\lambda\alpha}(x_0, y_0) = \arg[\hat{G}_{\lambda\alpha}^+ \cdot U_{1,\lambda\alpha}(x_1, y_1)] \end{cases} \quad (6.21)$$

The particularity of the Yang–Gu algorithm lies in a particular iteration loop that is inserted within the general Gerchberg–Saxton iteration loop (see Figure 6.22).

This particular loop tends to accommodate the lack of unitary propagators. During the loop referred to by the superscript  $n$  of the general G–S algorithm, the daughter loop iterates with superscript  $m$  and starts by generating a random phase distribution  $\varphi_1^{n,m=0}(x_1, y_1)$ . If it is injected together with the known amplitude distribution  $A_1(x_1, y_1)$  in Equation (6.20), one can get a feeling of the phase distribution in the near field for a particular wavelength.

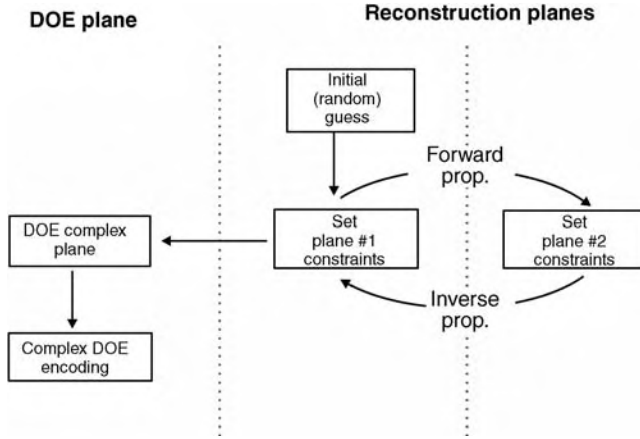


**Figure 6.21** The Ping-Pong iterative CGH optimization algorithm (IFTA type)

Hence, one can propagate backwards to the CGH plane by re-injecting the various phase and amplitude information  $A_{0,\lambda_a}(x_0, y_0)$ , in Equation (6.20), and thus retrieving the amplitude  $A_1(x_1, y_1)$  and phase  $\varphi_1^{n,m=0}(x_1, y_1)$  distributions. Another loop then begins, and so on until the following condition is satisfied:

$$\sum_{x_1, y_1} \frac{|\varphi_1^{n,m}(x_1, y_1) - \varphi_1^{n,m+1}(x_1, y_1)|}{|\varphi_1^{n,m}(x_1, y_1)|} \leq \varphi_1 \quad (6.22)$$

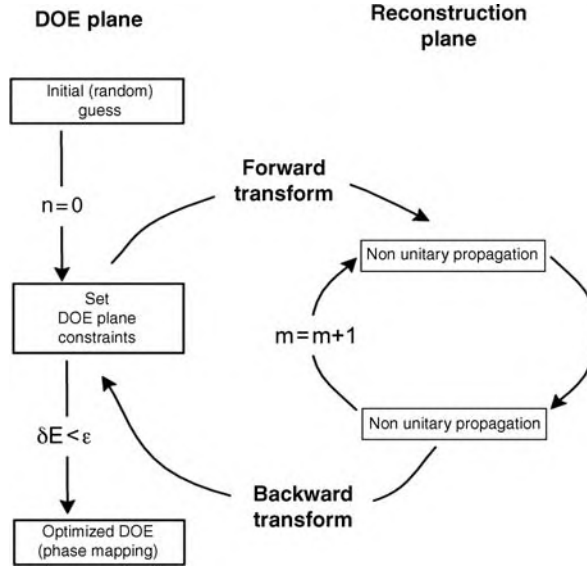


**Figure 6.22** The Yang-Gu optimization algorithm

where  $\varphi_1$  is a given small value that indicates a satisfactory convergence rate of the phase distribution. Eventually, the general G–S algorithm iterates again, and is ruled by the desired intensity cost function over the reconstruction plane, as discussed in the previous paragraphs.

As the resulting phase mapping will be implemented as a multilevel surface-relief phase DOE, a wavelength has to be chosen to implement the thickness of the relief. If  $N$  wavelengths have been used to design the DOE, a solution is to choose the mean wavelength to encode the surface relief.

Hence, the corresponding surface-relief profile  $\delta d$  is obtained as follows:

$$\delta d(x_1, y_1) = \lambda_0 \cdot \frac{\varphi_1(x_1, y_1)}{2\pi} \quad (6.23)$$

where  $\varphi_1$  is related to  $\varphi_{1,\lambda_\alpha}$  by

$$\varphi_{1,\lambda_\alpha}(x_1, y_1) = \left( \frac{\lambda_0}{\lambda_\alpha} \right) \varphi_1(x_1, y_1) \quad (6.24)$$

In Figure 6.23 some examples of complex plane optimization using IFTA algorithms are presented. The left-hand side shows the optimization procedure of a Fourier CGH and the right-hand side a Fresnel CGH optimization. The complex planes are plotted before optimization, after two, five and 20 iterations, and after the final hard clipping to four fabricable phase levels.

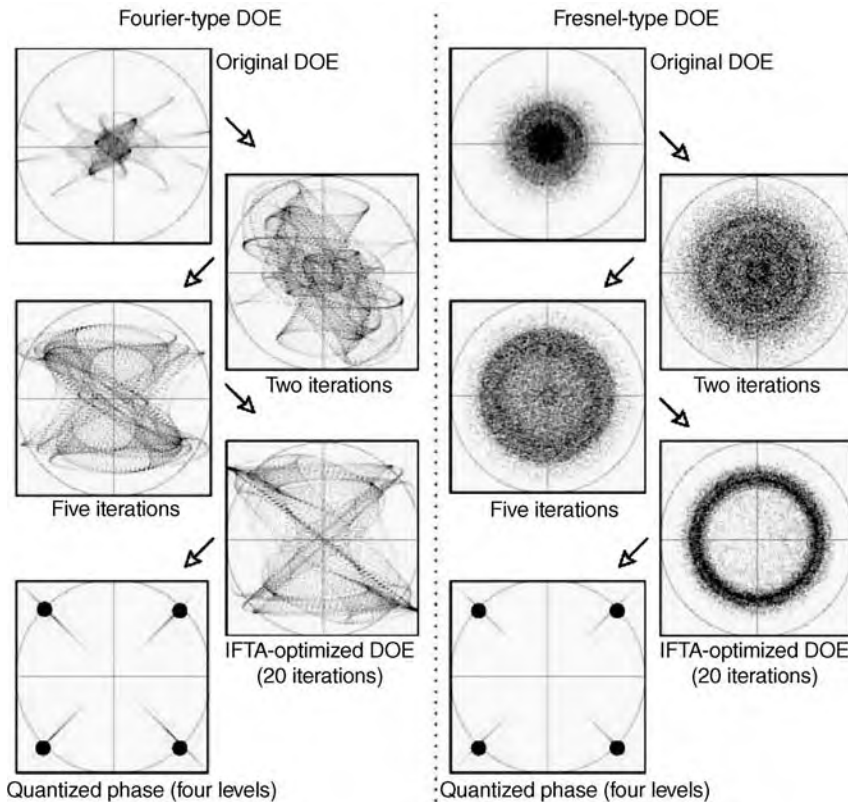


**Figure 6.23** Examples of IFTA optimization procedures in the complex plane



Note how the various complex pixels tend to regroup around the four allowed fabrication levels, and how the amplitude tends to diminish its variation, and to converge to a ring (constant amplitude in the complex plane). In the case of a Fresnel element, the remnant amplitude variations are more rare than in the case of a Fourier element. However, this also depends on the complexity of the optical functionality (e.g. a set of beam splitters – fan-out gratings and multifocus lenses).

As a practical example, in Section 6.3 some application-oriented CGHs that have been optimized with various IFTA algorithms are shown.

### 6.2.6.3 Steepest Descent Algorithms

Steepest descent algorithms (or input–output algorithms) are traditional optimization algorithms as found in numerous scientific fields (mathematics, economics, biotechnology, medicine, electronics etc.).

A steepest descent algorithm, as opposed to an IFTA algorithm, changes a single pixel at each iteration, while the IFTA can change the whole pixel map of the CGH. However, in both iterations, an entire field propagation process is needed. Therefore, steepest descent algorithms are usually much slower than IFTA algorithms, but can yield better results – particularly in terms of uniformity.

#### *Direct Binary Search (DBS)*

Direct Binary Search (DBS) was one of the first iterative algorithms to be successfully implemented for the optimization process of CGHs [18–20]. This algorithm minimizes a cost function (also often called an error function) by directly manipulating the CGH's complex data one by one and observing the effects on the numeric reconstruction.

DBS is a simple unidirectional optimization algorithm that can be applied to CGH synthesis and quantization [21], either as phase or amplitude modulation. If the change made to one of the CGH's pixel has positive effects (i.e. it decreases the cost function), the change is kept; on the other hand, if the change induces negative effects (i.e. it increases the cost function), the previous CGH configuration is restored and other changes are made. Figure 6.24 shows the main DBS flow chart.

In this sense, DBS can be considered as a straightforward steepest descent algorithm [22]. When all the pixels have been changed within the DOE matrix, another loop is engaged, and this process is repeated until the cost function reaches one of the conditions described in the previous section, or until there are no more changes to be accepted within an entire loop: the algorithm has converged to a local minimum. With DBS, this minimum has an infinitesimal chance of being a global minimum (local minimum). DBS would therefore produce the optimization shown on the left-hand side of Figure 6.18.

The CPU time required to optimize a CGH grows linearly with the CGH SBWP. If an FFT propagator (or worse, a DFT-based propagator) is used to evaluate the reconstruction for each pixel change, the algorithm becomes obsolete, as the CPU time would be way beyond the designer's patience. To overcome this difficulty, a novel approach of computing the updated error function without directly calculating the entire CGH's reconstruction has been proposed. This method, termed the Recursive Mean-Squared Error (RMSE) method, decreases the computation time significantly for off-axis images produced from binary amplitude CGHs. It has also been applied to multilevel phase CGHs. This method updates the reconstruction rather than recalculating it entirely at each iteration.

#### *The Simulated Annealing Algorithm (SA)*

The Simulated Annealing (SA) algorithm has been derived from the DBS algorithm, with a slight but yet very noticeable change, in order to be able to get away from the local minima of the global cost function [23–25]. This change lies in the way the pixel change is considered (see the flow chart in Figure 6.25).

A probability function  $P(\Delta E)$  described by Boltzman's distribution is used when the pixel change introduces a negative energy variation (i.e. increases the cost function), to decide whether the

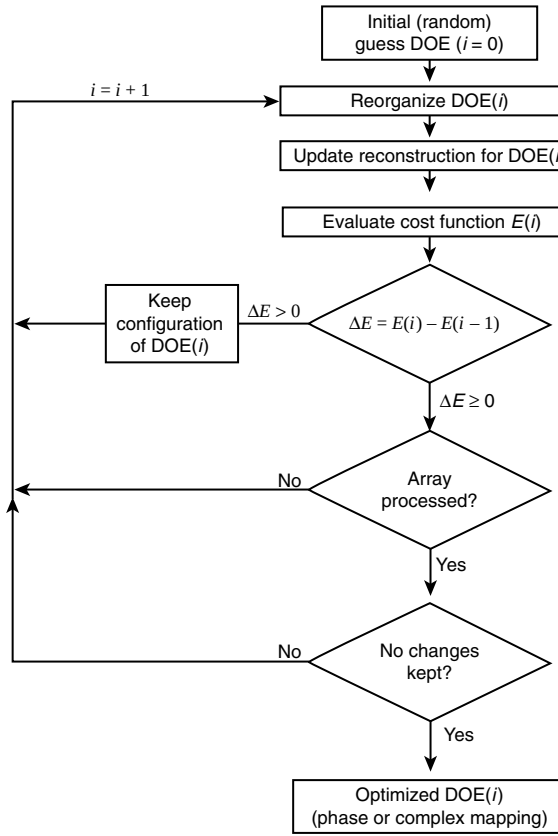


**Figure 6.24** The DBS algorithm flow chart

configuration is to be kept or to be reorganized:

$$P(\Delta E) = e^{-\frac{\Delta E}{T}} \quad (6.25)$$

where  $T$  is the temperature parameter of the SA algorithm and  $\Delta E$  is the variation of the energy (or the cost function) from the previous state to the current one.

In this way, ‘noise’ is introduced in the algorithm, and the convergence may not be ‘trapped’ within a local minimum, as would happen with DBS (thus producing the stochastic optimization presented on the right-hand side of Figure 6.18). The amount of ‘noise’ or ‘shaking’ depends on the initial temperature used. Besides, at each iteration, the temperature is decreased in order to tune the algorithm more finely as it begins to converge to the global minimum. A classical way to decrease the temperature at each iteration is as follows:

$$T_i = \frac{T_0}{1 + i} \quad (6.26)$$

where  $i$  is the iteration index and  $T_0$  is the initial temperature.

Again, the final application’s specifications are used for fine-tuning of the initial temperature and the cooling rate. Eventually, the algorithm has converged when no pixel change is accepted during a whole iteration, or when using any of the criteria listed in the previous section.

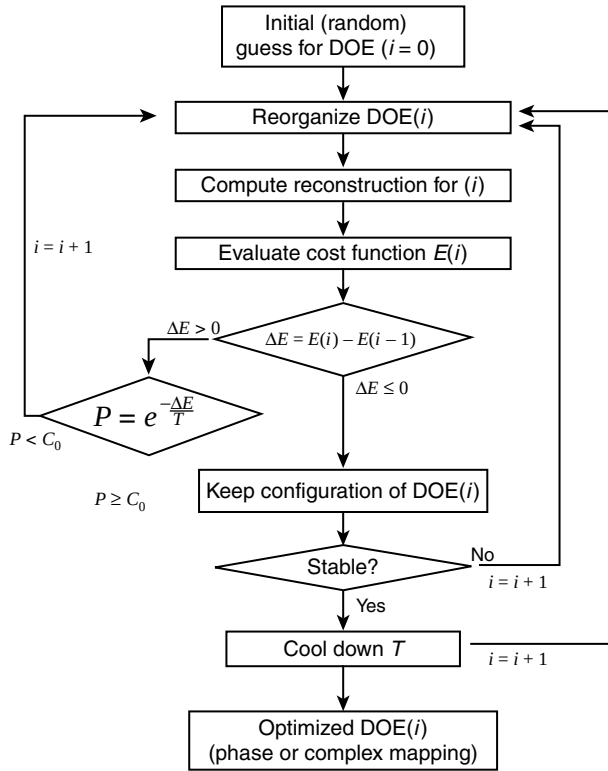


**Figure 6.25** The Simulated Annealing (SA) flow chart

As the SA and DBS algorithms require a large amount of computing power (when applying RMSE update techniques), they are best suited for the fine optimization of a small number of pixels. Small Fourier elements incorporating a small number of pixels (e.g.  $32 \times 32$  cells) can act like larger complex cell that can be replicated in the  $x$  and  $y$  directions to form a decent CGH aperture. The SA or DBS optimization procedure is then performed only over one two-dimensional period of the final Fourier CGH aperture. Thus, the SA and DBS algorithms are mostly used for the optimization of spot-array generators, 2D Fourier display elements [26] or Fourier filters [27]. Optimization for a Fresnel CGH is of course entirely possible, though replication is not possible and the convergence might take a long time.

#### **The Iterative Discrete On-axis (IDO) Algorithm**

The IDO encoding algorithm was developed for on-axis CGH configurations that are designed to work either as binary amplitude or multilevel phase-relief elements [28]. IDO is especially well suited for Fresnel CGH (e.g. on-axis Fresnel lenses etc.), where the quantization levels have to be chosen carefully in order to be able to fabricate the lens (see also optimal fracture of diffractive lenses in Chapter 12).

The IDO and SA algorithms are very similar. Both are ruled by the flow chart described in Figure 6.25. The IDO encoding algorithm deals especially with on-axis configurations for CGHs, which are designed to work either as binary amplitude or multilevel phase-relief elements. A RMSE method has also been proposed that is very similar to the DBS one, and that reduces the computation time by only updating the error function, rather than recalculating it at every iteration.

When optimizing a Fresnel diffractive lens, for example, with IDO [29], as the pixel size is fixed to a certain value prior to the optimization process, IDO can find a good compromise between SBWP and the minimum feature size (the pixel size), even if the straightforward design for Fresnel lenses asserts that the minimum feature size of that lens is too small (i.e. it is simply the minimum period divided by the number of phase levels – see Chapter 5). Hence, IDO will choose which phase level(s) are to be assigned to the outer fringes (the smallest periods to encode).

### The Blind Algorithm

The ‘blind’ algorithm can be applied on either the DBS or the SA algorithm. The blind algorithm, as its name suggests, does not see the entire reconstruction field, and reconstructs only an area of interest. As the reconstruction does not involve the entire field, the constraint of the intensity can only be applied to locations of the space where the numeric reconstruction is performed (i.e. in the area of interest). And you need to keep your fingers crossed in the hope that no other intensity hot spot is produced by this algorithm. Contrary to expectations (a pessimist’s), the blind algorithm actually does not produce any unwanted hot spots other than the ones that are constrained to the desired values with the DFT-based propagators, and therefore serves as a facilitator of the DBS or SA and speeds up their convergence. It is a still a slow algorithm when compared to an IFTA algorithm, but it produces much better results in terms of uniformity. As this algorithm puts an emphasis on uniformity and/or the SNR rather than brute efficiency, as IFTAs would do, it is best suited for use in designing spot array generators, or multifocus lenses, since in these cases the pixels bearing light and the pixels bearing an absence of light are minimal (and thus the reconstruction over these limited points in space can be fast).

The convergence rate for various criteria, such as the diffraction efficiency, the uniformity and the SNR, are shown in Figure 6.26, for a typical IFTA algorithm (GS) and a typical steepest descent algorithm (SA).

It is interesting to note that the IFTA algorithm provides optimum efficiency very quickly (after three iterations), but the uniformity and RMS criteria tend to be trapped in an early local minimum (after about five iterations). In the SA algorithm, efficiency as well as RMS increase monotonously, and try to get away from the local minima. The convergence rate is slow, but steady. Note that uniformity is very fast to converge, since a spot array generator is present and only a few spots are ON. For a nonuniform pattern

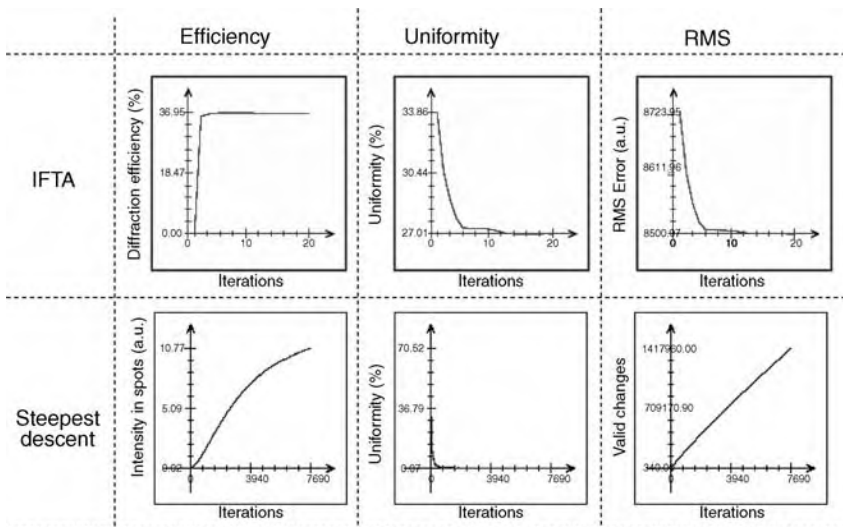


**Figure 6.26** Example of the convergence of an IFTA algorithm

(such as a gray-scale pattern), the uniformity is not of interest. The RMS criterion is more interesting. In Figure 6.26, the SA shows very uniform RMS convergence, which is a desirable behaviour.

#### 6.2.6.4 Genetic Optimization Algorithms

Genetic algorithms (GA) are categorized as global search heuristics [30–32]. Genetic algorithms are a particular class of evolutionary algorithms (also known as evolutionary computation) that use techniques inspired by evolutionary biology, such as inheritance, mutation, selection and crossover (also called recombination).

Unlike the other previously reported stochastic optimization algorithms, GAs operate on a large set from a population of  $N$  possible solutions, and are therefore best suited for relatively small SBWP Fourier-type CGHs. A typical GA uses a large number of CGH populations, performs breeding, calculates fitness functions for each CGH and keeps the fittest. Random mutations are inserted to maintain diversity within the successive generations. The mutation rate, the population size, the fitness function and the crossover parameters control the convergence of the GA towards its eventual solution. Note that the amount of number crunching, in this instance, can quickly become highly prohibitive when the pixel array increases (i.e. when the SBWP increases).

The variables are encoded as binary strings (a 1D version of the 2D CGH matrix) and these strings are treated as chromosomes to define the population of possible solutions. The chromosomes are then evaluated by the previously defined cost functions and a fitness criterion is assigned to each of them. The fittest members of the population are favored for recombination to produce a new population. Figure 6.27 depicts the GA flow chart for one generation: encoding, evaluation, recombination and mutation.

In GAs, recombination (breeding) is achieved by crossover – swapping of portions of the chromosome chains of the two parent solutions. As successive generations are formed, by survival of the fittest, the algorithm converges towards an optimum. To maintain diversity within the different generations, random mutations perturb the values in the chromosome strings.

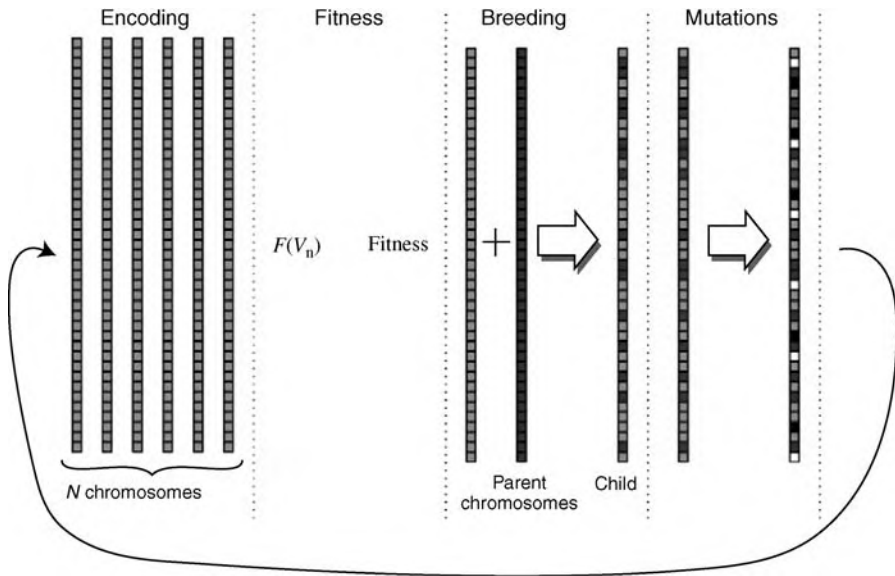


**Figure 6.27** The Genetic Algorithm used for CGH optimization

### 6.2.6.5 Other Optimization Algorithms

Several other iterative methods have been developed for specific applications. For example, the *Global Iterative Encoding Algorithm* [33] optimizes an already encoded CGH over a dynamical device such as an SLM. The three design processes, namely optimization, quantization and encoding, are performed at the same time within the main iterative loop.

*Generalized Error Diffusion* is a global optimization algorithm very similar to DBS that optimizes a CGH by calculating the error in the CGH plane rather than in the object plane, by means of additional filters. This algorithm can be used for the optimization of either Fourier- or Fresnel-type CGHs, and for either binary amplitude or multilevel phase fabrication.

Table 6.1 summarizes and compares the various algorithms presented here, and their respective characteristics.

### 6.2.6.6 So, Which Algorithm Should You Use?

As reading through the various iterative optimization algorithms used in the literature and presented above might result in a headache, the eager potential CGH designer might consider a fit-all algorithm, which takes the best from the previous various algorithms.

The proposed fit-all CGH task optimization algorithm uses four basic building blocks:

- it is based on a IFTA loop;
- it includes a simulated annealing procedure;
- it uses an optimized first guess; and
- (optional) it fine-tunes the result with a DBS procedure.

The standard IFTA loop is designed so that the error is modulated prior to being re-injected into the reconstruction plane. This modulation is a simulated annealing process, which can work on several iterations.

Such a fit-all-tasks algorithm has been applied to the design of a  $16 \times 16$  regularly spaced spots fan-out grating, as shown in Figure 6.28.

The cost function used in this example was a linear combination of the efficiency, the uniformity and the SNR (see the previous section defining these parameters). If the algorithm had not been annealed by the error re-injection modulation, the efficiency would have been acceptable but the uniformity and the SNR would not have been as good. The error re-injection happened after 10 iterations, and the error re-injection annealing lasted for 40 more iterations before stopping the algorithm.

The dotted lines in Figure 6.28 show the performance values if simulated annealing is not used in the IFTA algorithm (trapped in local minima); they are especially bad for uniformity, which is a very important parameter in a fan-out grating (see, e.g., Chapter 16).

## 6.2.7 The Physical Encoding Scheme

Once the complex data have been optimized by the iterative algorithm, a physical encoding scheme has to be chosen in order to encode the resulting data [34], which can be pure phase or complex (amplitude and phase).

If the resulting CGH function is a pure phase element, such as with the steepest descent or GA algorithms, the encoding into the surface-relief phase is straightforward.

In many cases, especially with IFTA algorithms [35, 36], if the resulting complex CGH data are not pure phase and some amplitude information has to be encoded, two choices are left:

- either encode the phase and leave the amplitude out of the picture; or
- encode both the phase and the amplitude on each pixel – this, however, will decrease the efficiency of the CGH, but produce a very accurate reconstruction.

**Table 6.1** A summary of the various CGH optimization algorithms

| Type                        | Algorithm | Propagator | Fourier<br>CGH | Fresnel<br>CGH | Pure phase | CGH size | Efficiency | Uniformity | CPU time  | Convergence |
|-----------------------------|-----------|------------|----------------|----------------|------------|----------|------------|------------|-----------|-------------|
| IFTA                        | GS        | FFT-based  | Yes            | Yes            | No         | Large    | Very good  | Low        | Fast      | Easy        |
|                             | Ferwerda  | FFT-based  | No             | Yes            | No         | Large    | Very good  | Low        | Fast      | Easy        |
|                             | Ping-pong | FFT-based  | No             | Yes            | No         | Large    | Good       | Low        | Slow      | Difficult   |
|                             | Yang–Gu   | FFT-based  | Yes            | Yes            | No         | Medium   | Good       | Low        | Slow      | Difficult   |
| Steepest                    | DBS       | FFT/DFT    | Yes            | Yes            | Yes        | Medium   | Medium     | Good       | Medium    | Medium      |
| Descent                     | SA        | FFT/DFT    | Yes            | Yes            | Yes        | Small    | High       | Very good  | Slow      | Easy        |
|                             | IDO       | FFT/DFT    | Yes            | Yes            | Yes        | Small    | Medium     | Good       | Slow      | Medium      |
|                             | Blind     | DFT        | Yes            | Yes            | Yes        | Large    | Medium     | Very good  | Medium    | Medium      |
| Evolutionary<br>programming | GSA       | FFT        | Yes            | Yes            | Yes        | Small    | Good       | Good       | Very long | Easy        |

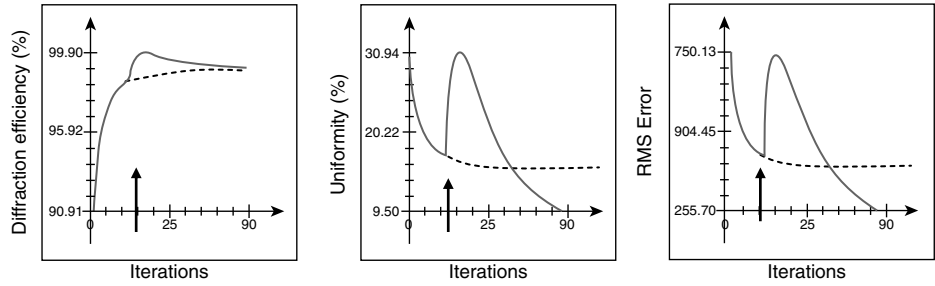


**Figure 6.28** CGH optimization of a  $16 \times 16$  fan-out grating

Therefore, it all comes down to the final application: Is it better to trade efficiency for functionality, or is the efficiency the core criterion?

As an example, Figure 6.29 shows the complex planes of two CGHs optimized by the GS algorithm, in Fourier and Fresnel form, over 16 phase levels. Although the Fourier CGH regroups the phase values around the 16 allowed phases to be fabricated much better than the Fresnel CGH, its amplitude variations are greater. This is typical of a Fourier CGH.

The resulting amplitude variations can be encoded into the final element by either a phase detour technique or an error diffusion technique [37, 38].

This time, Figure 6.30 shows both the phase and amplitude maps after a GS optimization process, for Fourier and Fresnel elements [39].

If the CGH data have to be implemented (encoded) as complex data, a complex encoding technique has to be used. Several of these techniques have been developed in industry.

Figure 6.31 shows five different data encoding methods that have been used in the literature:

- Lohmann encoding;
- Burch encoding;
- kinoform encoding;
- complex kinoform encoding; and
- error diffusion encoding (real or complex error diffusion).

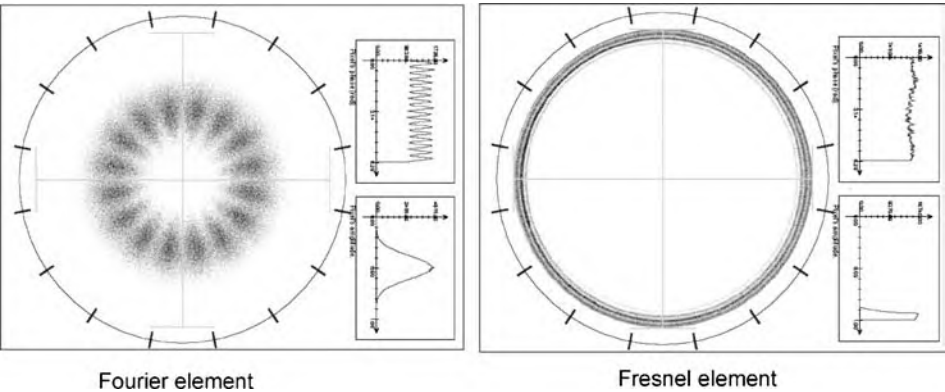


**Figure 6.29** Complex maps of a Fourier and Fresnel CGH optimized by GS over 16 phase levels



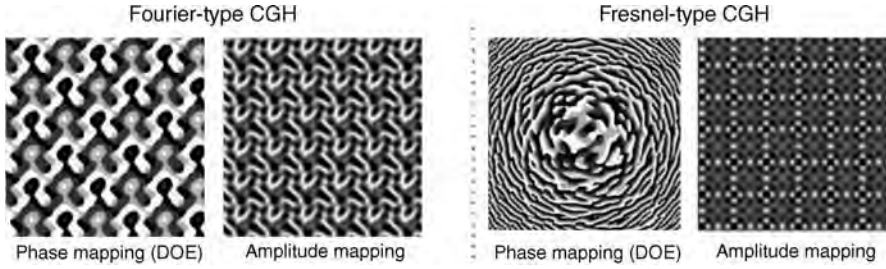


**Figure 6.30** The resulting phase and amplitude maps for Fourier and Fresnel CGHs

It is interesting to note that although the five CGHs described in Figure 6.31 look very different, each actually encodes the same optical functionality, which is a Fourier pattern projector CGH.

As seen numerous times in the previous chapters, it is never a good idea to encode amplitude in a diffractive element, or to use a physical amplitude encoding method, since the diffraction efficiency drops dramatically. There have been extensive publications on Lohmann and Burch encodings, which were the first techniques to encode CGH data in the late 1960s. However, fabrication technologies have developed considerably, and these amplitude-encoding schemes are no longer used today. Besides, the minimum feature size in a cell-oriented method is much smaller than the actual cell itself. So this is a sort of a waste of the SBWP of the CGH.

Today, basic kinoform encoding is the *de facto* encoding method and it is explained in detail in Chapter 12. The complex encoding methods are included for their academic interest and to pay respect to the various pioneers in the field of digital optics. However, these complex encoding methods do not generate much interest in industrial applications, for the reasons mentioned.

### 6.2.7.1 Detour Phase Encoding Methods

#### *Lohmann Encoding*

The Lohmann [3, 6] and Burch encoding methods are also called cell-oriented or detour phase encoding methods. These methods are encoded over an amplitude substrate (chrome on glass, for example).

The basic approach to a detour phase encoding method consists of modulating each period of a binary grating with local shifts in the grating line and in the grating line width, to match the local complex value.

The local shift of the grating line  $P_{n,m}$  within each period is proportional to the phase value  $\phi_{n,m}$  and the width of the line  $W_{n,m}$  is proportional to the amplitude  $A_{n,m}$ :

$$\begin{cases} W_{n,m} = \arcsin\left(\frac{A_{n,m}}{\pi N}\right) \\ P_{n,m} = \frac{\phi_{n,m}}{2\pi N} \end{cases} \quad \text{where } N = x_0 \cdot \delta u \quad (6.27)$$

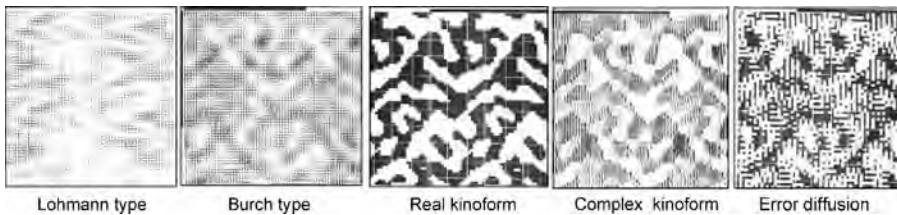


**Figure 6.31** Various CGH encoding methods used in the literature

where  $x_0$  is the spatial distance of the first diffraction order (where the reconstruction is obtained) and  $\delta u$  is the resolution of the reconstruction.

As noted by Lohmann and Brown, since the apertures are not generally placed at the center of the cells, any errors will be errors in the optical reconstruction of the CGH. Therefore, a revised version of the encoding method has been proposed, so that the complex data are cubically interpolated to represent the complex wavefront as it is sampled at the center of the aperture, as opposed to the center of the cell.

### Lee Encoding

In the Lee encoding method [40], the projections of the complex wavefront onto the positive real, negative real, positive imaginary and negative imaginary axes of the complex plane are calculated. Obviously, at the most, only two of the projected values are nonzero. The cell is then divided into four sub-cells corresponding to the four projections, and an aperture is placed in the sub-cells that correspond to the nonzero projections. The height of the sub-cell is proportional to the value of the projection. Here also, the wavefront is sampled at the center of each cell and a revised method has been developed for data to be interpolated at the center of each sub-cell.

### Burch Encoding

In the Burch encoding method [41], the transmittance function is calculated in the same way as for the interferogram design method described previously. The transmittance function is then sampled at regularly spaced intervals in  $x$  and  $y$ , paces that are the lateral dimensions of the equivalent cells. Transparent square apertures centered onto these square cells are then generated, whose area is proportional to the transmittance function sampled at the center of the cell.

Figure 6.32 depicts the Lohmann, Lee and Burch cell-oriented binary amplitude detour phase encoding methods.

### Complex and Real Kinoform Encoding Methods

The complex kinoform encoding method is described in Figure 6.33, along with the real kinoform encoding method. The wavefront is sampled at the center of the cells. The etch depth of the basic cell encodes the phase information, whereas the amplitude information is proportional to the width of a square or rectangular sub-cell, centered to the basic cell, and whose etch depth introduces a phase shift of  $\pi$ . Local destructive interference produces local amplitude variations.

Figure 6.34 shows such a complex kinoform fabricated in a photoresist layer.

Although complex kinoform encoding is a very elegant method, as soon as the amplitude information is encoded, the diffraction efficiency drops dramatically. Besides, the complex kinoform encoding requires further fracture of the basic cell into sub-cells (windows); this leads to an increase in the final fabrication

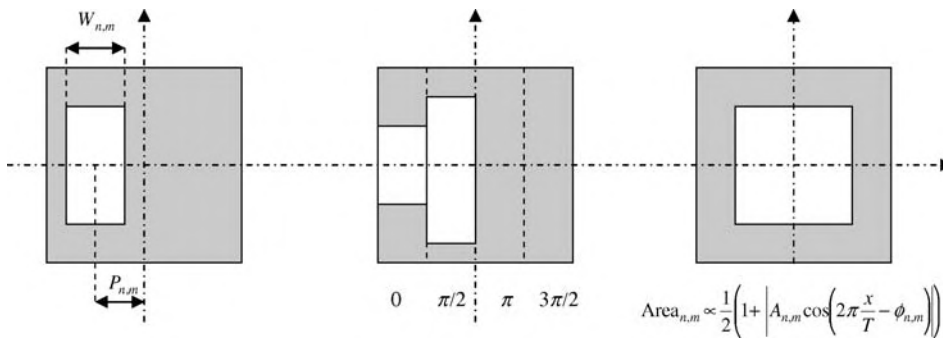


**Figure 6.32** Binary amplitude detour phase encoding methods

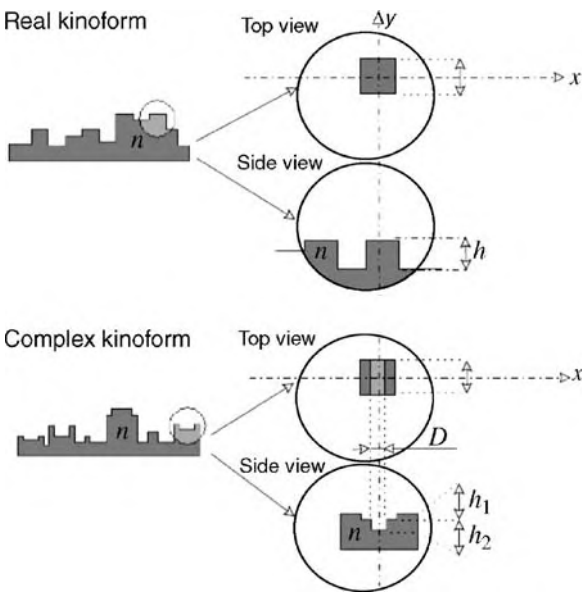


**Figure 6.33** Complex and real kinoform encoding methods

file and, most of all, it decreases the minimum feature size required to fabricate the CGH. This is why this latter method is used less than the real kinoform method.

**6.2.7.2 Error Diffusion Encoding Methods**

Error diffusion encoding methods [42, 43] are not detour phase methods. The smallest resolution required is the pixel itself (cell), whereas in the detour phase methods this resolution needs to be much smaller.

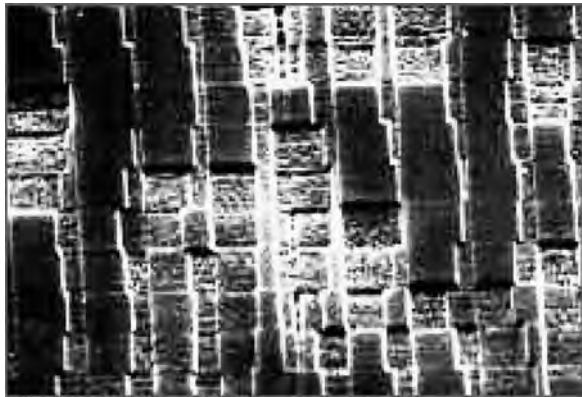


**Figure 6.34** A complex kinoform encoding in photoresist

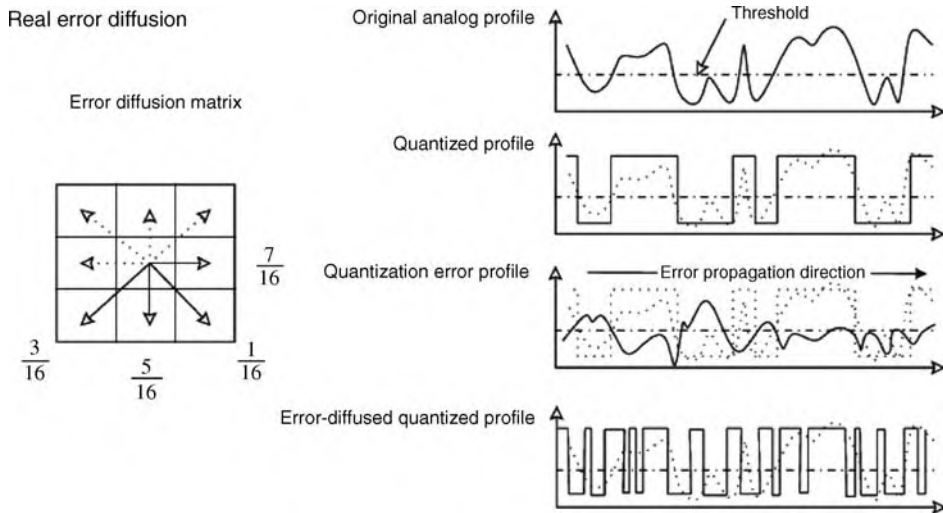


**Figure 6.35** The real error diffusion encoding method (SEM and optical reconstruction)

Error diffusion encoding methods attempt to encode a gray-scale phase or amplitude as a binary phase or amplitude. However, this method produces lots of scattering and unwanted noise, which reduces the reconstruction SNR.

#### **Real Error Diffusion Method**

During the quantization process, the error due to the phase or amplitude quantization is propagated to the neighboring pixels (in real or complex planes) [44, 45]. The extension to real or complex multilevel DOE encoding is straightforward. The real error diffusion technique is described in Figure 6.35.

A diffusion matrix is used to propagate the error throughout the whole CGH aperture, and the different weights help to define the main propagation directions. Note that the quantization threshold is not necessarily a constant function. It can be a slowly varying function for additional modulation of the already optimized CGH data.

When applied to CGH encoding, this technique not only has the capability of gray-tone encoding, but is also able to shift quantization irradiance clouds in the reconstruction planes to areas where the SNR is not of much interest. Figure 6.36 shows an example of real error diffusion over a toroidal binary phase diffractive lens that was fabricated by direct LBW write.

The DOE has been fabricated by binary LBW with a  $3\text{ }\mu\text{m}$  spot size. The SEM photograph shows part of the central fringe. Figure 6.36 shows an optical reconstruction from a binary Fourier CGH that was encoded with real error diffusion, where the quantization and noise clouds are pushed outside the region of interest.

#### **Complex Error Diffusion Encoding Method**

*Complex error diffusion* has similar capabilities [46], but makes use of the fact that the quantization error for both amplitude and phase can be diffused over the complex plane. Although complex error diffusion diffuses the complex quantization error from and to complex data, the resulting quantized pixels are phase only (situated on the unitary circle in the complex plane). This method is therefore very attractive, since it is possible to encode amplitude information (via complex error diffusion) into a phase-only CGH [47]. Although complex information is kept, the diffraction efficiency does not drop because the final element is

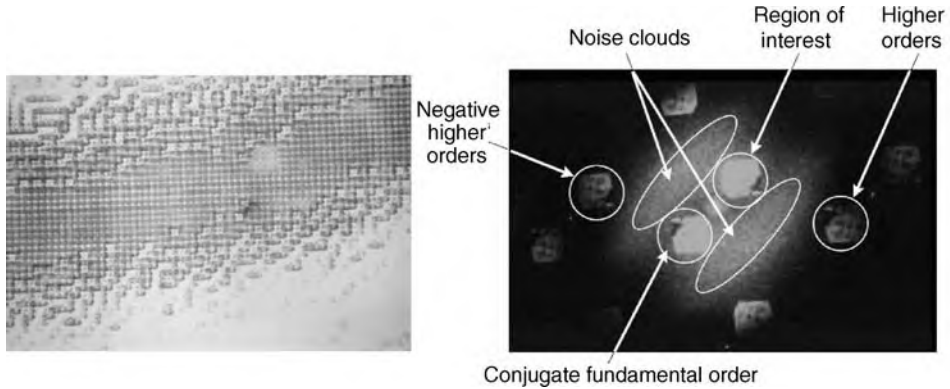


**Figure 6.36** Real error diffusion over a toroidal interferogram lens (detail of a fringe)

pure phase and, furthermore, without local destructive interference effects (unlike the case of amplitude encoding methods) [48].

#### **Other Encoding Methods**

There are several other cell-oriented and pixel-oriented encoding methods described in the literature, such as the *Burckardt encoding method*, *ROACH encoding*, the Huang and Prasada method and the Hsueh and Sawchuck method. These methods are more or less variations of the ones discussed above, and are not widely used by researchers in the field. In fact, none are used in realistic applications, where SBWP and diffraction efficiency are key features.

### **6.2.8 Going the Extra Mile**

There are ways to increase the efficiency of a CGH by twisting the theoretical principles depicted in this chapter [39, 49]. For example, it has been said that a fundamental order cannot diffract to angles larger than what the grating equation dictates.

Having said that, make sure that there is no light in the fundamental order and push a little light into the second diffracted orders, therefore multiplying the diffraction angles by two without reducing the CGH cell sizes. This can be done by modulating the etch depth and the lateral features (by actually etching the GCH in a shallower way), as seen in the optical reconstructions in Figure 6.37.

Thus, by modulating the etch depth and the lateral features, the light can be tricked into diffracting into higher orders, much like when designing a broadband diffractive lens (see Chapter 5). However, in the case of a broadband or multi-order) diffractive lens, the etch depth has to be increased to integer values of  $2\pi$ .

### **6.2.9 Design Rule Checks (DRCs)**

Design rule checks (DRCs) are an important part of the CGH design procedure. DRCs are effective tools that are implemented in much industrial design software, especially in IC design software and mask layout design software (see the mask-related DRCs presented in Chapter 12). Listed below are some of the DRCs that are implemented in CGH design procedures today.

#### **6.2.9.1 The Validity of Scalar Theory**

Such a DRC should check if the designer is still in the realm of validity of scalar theory, by computing the smallest  $\Lambda/\lambda$  ratio (see also Appendices A and B, and Chapter 11). This seems to be obvious for a CGH

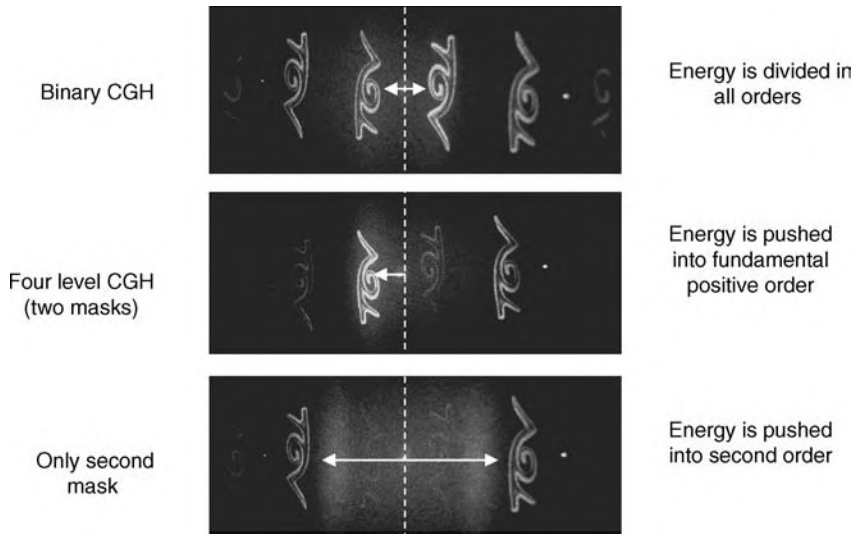


**Figure 6.37** Pushing light into the second diffraction orders

where the pixel size is more or less the same everywhere, or for a spherical lens where the smallest feature size is well defined (see Chapter 5).

However, this is a less evident task when considering an aspheric lens (astigmatic) or complex grating, where the constituting fringes get larger and smaller in a nonmonotonous way. Consider, for example, a helicoidal lens. The lens could work properly over a certain area and then less well where the fringes get closer to the size of the wavelength. Also, an adverse polarization effect can kick in, where a spherical lens can produce an elliptical diffracted beam with a linear polarized incoming beam, since the efficiency is a function of the local orientation of the grating in regard to that polarization direction (the analyzer effect).

### 6.2.9.2 Reconstruction Windows

First, when choosing a CGH with a specific cell size, it is important to make sure that the desired object fits into this window properly. For example, a Gaussian function has infinite extent, and so will not fit into such a window.

Second, when setting a desired reconstruction in the reconstruction window, make sure there is plenty of free space around the object, since the higher orders will be stitched to this fundamental window. A good example is described earlier in this chapter.

### 6.2.9.3 The Absolute Position of Spots

As the reconstruction space is sampled to a fixed grid, the diffracted spots (or beams) can only be carried in very specific directions in space (quantized directions).

When designing a Fourier fan-out grating, and having made sure that the previous condition (Section 6.2.8.2) is acceptable, this does not mean that it is possible to diffract the beamlets in any direction below the maximum allowed diffraction angle.

For example, if a CGH with 256 cells in each direction, with a cell size of  $2.0\text{ }\mu\text{m}$  (thus creating a  $0.5 \times 0.5\text{ mm}$  CGH), to be used with 632 nm laser light is defined, then the fundamental reconstruction

window thus has an angular extent of  $18^\circ$  full angle (actually,  $18.1817^\circ$  exactly). Such a CGH can theoretically diffract spots in any angle within this cone. For example, for an optical clock broadcasting application, the CGH has to diffract an array of  $16 \times 16$  beamlets each spaced by  $0.52^\circ$  precisely, which makes up a cone of  $7.50^\circ$ , well within the  $18^\circ$  cone.

However, in this configuration, the smallest angular step would be  $0.0713^\circ$ ; thus the closest appropriate angular beamlet spacing would be either  $4.991^\circ$  (seven spacings) or  $0.5704^\circ$  (eight spacings). Obviously,  $4.991^\circ$  is the closest to  $0.52^\circ$ . The cell size can thus be changed in order to produce the desired  $0.52^\circ$  beamlet spacing. The optimum and closest cell size candidate to be able to produce the  $0.52^\circ$  beamlet spacing would thus be  $1.92 \mu\text{m}$  instead of the original  $2.0 \mu\text{m}$ . As the cell size can be rectangular, if the angular spacings are different in the  $x$  and  $y$  directions, such a DRC can be applied to both dimensions.

### 6.3 Multiplexing CGHs

The Ping-Pong and Yang–Gu algorithms reviewed in the previous sections lead to the definition and classification of the different types of multifunctional CGHs. There are three main groups of multifunctional CGHs, each of which uses a different design method and a different operation mode to trigger the different optical functions.

#### 6.3.1 Spatial Multiplexing

The first group is the most straightforward one in the sense that it uses simple spatial multiplexing techniques (i.e. several different CGHs are generated independently and are recombined side by side on a same substrate to build up a global CGH aperture). The different sub-apertures may not be of the same geometry or size, and the CGHs may or may not be of same type, but they are certainly of the same physical aspect, since the entire mask is processed in one step.

Hence, the different CGHs (and thus the different optical functions) can be triggered by launching light only over the desired sub-apertures. Chapter 9 shows such a multifunctional element composed of 27 different diffractive lenses, which is used in combination with a ferro-electric amplitude SLM to perform a dynamical 3D reconstruction (i.e. a very rough version of a ‘three-dimensional video’).

#### 6.3.2 Phase-multiplexing

The second group deals with multifunctional CGHs where all the optical functions are triggered at the same time for a specific and unique illumination scheme. It includes CGHs generated by phase-multiplexing of different diffractive lenses or Fourier elements. The resulting DOE then incorporates amplitude and phase information, even when the constituting CGHs are pure phase elements.

Another example consists of CGHs generated by the Ping-Pong algorithm. Methods that use nonlinear quantization processes have been also been used to design multifunctional DOEs.

When multiple-focusing CGHs are required for an application, a design technique that uses nonlinear quantization of the continuous optimized phase profile can be used. The resulting CGH acts as a multi-order CGH (see Chapter 5) in the way in which it triggers different higher (negative and positive) diffraction orders for the same illumination scheme. The nonlinear quantization process [50, 51] is depicted in Figure 6.38.

Note that although this process can synthesize an off-plane multi-focus CGH, there are strong restrictions on the locations of these focal spots in space as they are higher orders of the same fundamental, rather than several different orders or just one fundamental.

Finally, there are also some complex design techniques and algorithms that have been developed to design and optimize elements that need to incorporate several optical functions. These functions can be triggered one by one depending on how the illumination scheme is defined. Typical design techniques include Yang–Gu algorithms and CGHs designed for broadband illumination.

The different configurations of multifunctional CGHs are summarized in Figure 6.39.

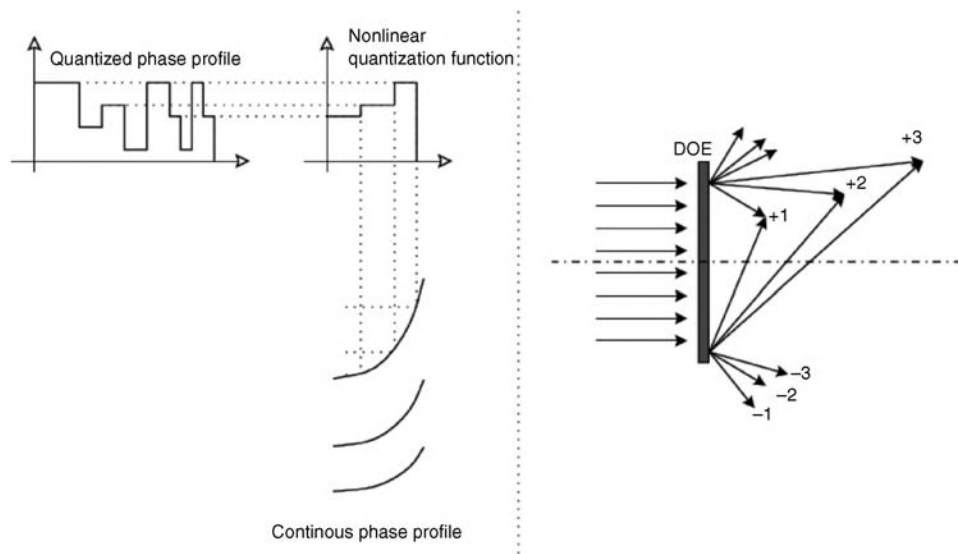


Figure 6.38 The nonlinear quantization process

| Operational specifications                                                                                     | Physical type (aperture) | Physical type | Crosstalk                                            |
|----------------------------------------------------------------------------------------------------------------|--------------------------|---------------|------------------------------------------------------|
| Spatial multiplexing:<br>requires spatial combinations and discriminations of incoming laser beams             |                          |               | Depends on laser beam alignment and size             |
| Phase multiplexing:<br>all the information is triggered at each time<br>Only one type of illumination required |                          |               | Impossible to use the optical functions individually |
| Wavelength multiplexing<br>Broadband DOE<br>Requires wavelength tuning of laser source                         |                          |               | Depends on wavelength selectivity of DOE             |

Figure 6.39 Multifunctional CGH implementation techniques



### 6.3.3 Combining Numeric and Analytic Elements

The numeric and analytic (or Fourier and Fresnel) diffraction regimes can be easily combined in order to produce elements with more complex optical functionalities, or multiplexed optical functionalities. The combination can either be performed in the complex plane (complex amplitude and phase-multiplexing) or simply by adding their phases.

A typical example is a Fourier CGH that has been optimized by a G–S algorithm to produce a reconstruction in the far field. This far field should be brought back into the near field by carefully controlling the aberrations. When designing a Fresnel CGH, the phase function introduced in the Fresnel propagation process is spherical, and thus does not control any aberrations.

A Fourier CGH can be combined with an analytic Fresnel diffractive lens with a prescribed set of aberrations to produce the desired reconstruction in the near field, with extended depth of focus, or in a volume in the near field, and so on (depending on the complexity of the Fresnel lens – see the various diffractive Fresnel lens implementations in Chapter 5).

## 6.4 Various CGH Functionality Implementations

Some practical examples of CGHs designed using the techniques discussed in the previous sections are described below.

### 6.4.1 Beam Splitters and Multifocus Lenses

Beam splitters are very popular CGH. Such beam splitters or fan-out gratings can be applied to numerous applications, including:

- metrology;
- optical interconnections;
- signal broadcasting (in free space and fibers);
- illumination (DNA assays etc.);
- multispot laser drilling and welding; and
- medical treatment (skin etc.).

Figure 6.40 shows some examples of fan-out gratings and multispot lenses calculated by DBS, G–S and the Ferwerda algorithm.

For a multispot binary off-axis Fresnel CGH example, see also Figure 6.13.

#### 6.4.1.1 Entropy and the Myth of Perfect Beam Combination

When discussing beam splitting, think about the reverse functionality, namely beam combination. Beam combination can be performed when the beams to be combined are different; for example, through the use of polarization combiners (see Chapter 9) or wavelength division combination (see Chapter 5). However, a fan-out grating cannot work in reverse mode and attempt to combine several beams with same wavelength and same polarization. This would actually contradict the third principle of entropy.

### 6.4.2 Far-field Pattern Projectors

To the general public, far-field pattern projectors (Fourier CGHs) are perhaps the best known diffractives, through the use of laser point-pattern projectors. Such CGHs are actually fan-out gratings, where the spots form an image rather than a regular matrix as presented previously. Several pattern generators have been presented in this chapter. Figure 6.41 presents a quite large one, which is formed by 16 million spots (this is a  $4096 \times 4096$  pixels 16 phase level Fourier CGH, and it represents the Rosace of the gothic cathedral of Strasbourg).

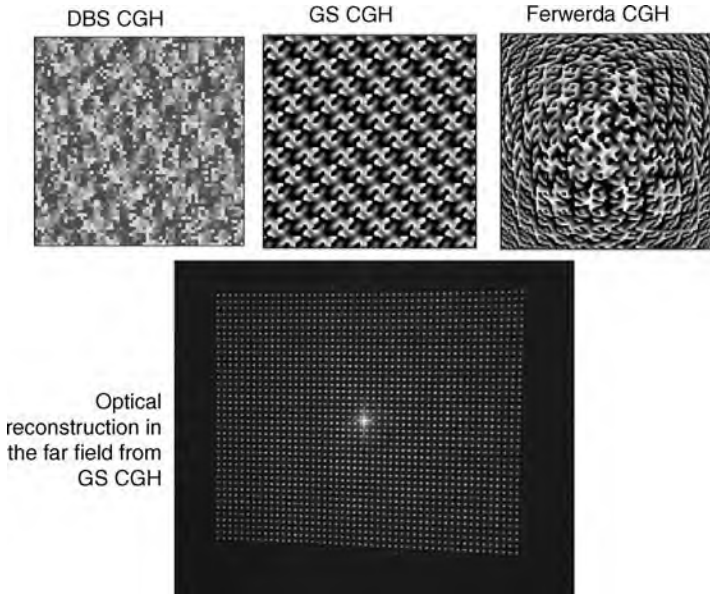


**Figure 6.40** Examples of fan-out gratings and multifocus lenses

A pattern projector can also be designed as a Fresnel element; that is, a diffractive lens that does not focus into a single spot but, rather, into a set of four million spots.

Figure 6.41 also shows the corresponding Fourier and Fresnel CGHs producing the central diffraction pattern. The Fresnel element on the right shows fringe-like structures in its lower right corner, informing us that this lens might be an off-axis lens. The Fourier CGH does not yield any fringe-like structures, but is formed by two-dimensional repetitive structures.

### 6.4.3 Beam Shaping and Focusators

Beam shaping is also one of the major successes for CGHs today. Such beam shapers can shape the beam and homogenize the beam at the same time, either in the far field (Fourier beam shapers) or in the near field

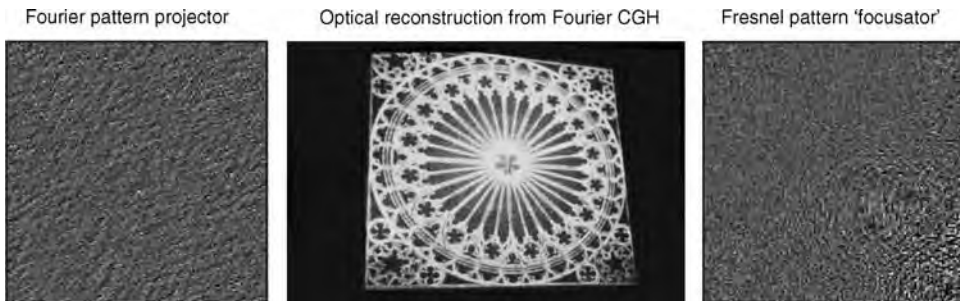


**Figure 6.41** Fourier and Fresnel pattern projectors

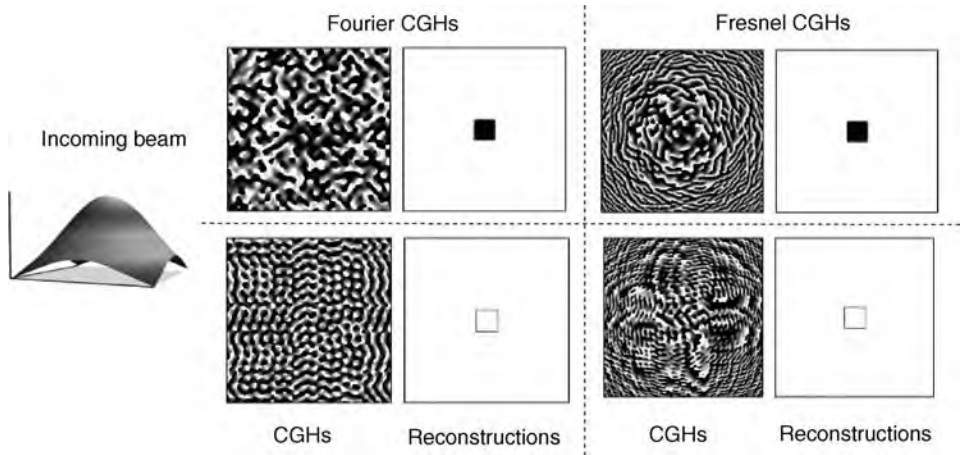


**Figure 6.42** An example of beam-shaping CGHs

(beam-shaping lenses). Figure 6.42 shows such beam shapers, which are optimized for an incoming Gaussian  $TEM_{00}$  beam, and that shape it into a variety of different shapes (full or empty) in the near and far fields.

The well-known ‘top hat’ beam shaper is depicted in the upper row, both in the traditional far field, and also in the near field as a ‘top hat’ beam-shaping lens. Fresnel beam-shaping CGHs are often also called focusators, and are used in laser material processing applications. This is a very desirable feature for engraving a logo, drilling a set of holes, welding complex shapes and so on.

**6.4.4** *Diffractive and Holographic Diffusers*

Many examples of diffusers have been described and shown in this chapter (see, e.g., Figures 6.10 and 6.15). In Figure 6.43, it is shown that it is possible to use the previous algorithms in order to design and

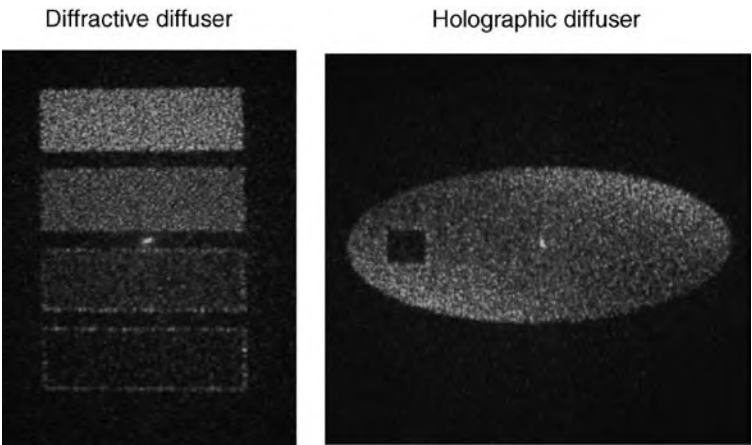


**Figure 6.43** Examples of complex diffractive diffusers

fabricate more complex diffusers that can have a very specific and controlled intensity level over a wide range of angles and that are also asymmetric. Such diffusers might have interesting applications in LED and laser lighting.

The differences between a diffractive and a holographic diffuser are very subtle. A diffractive diffuser reconstructs a beam in the near or far field, having a finite spatial extent, thus yielding relatively large structures (cell sizes).

A holographic diffuser is a far-field diffuser that does not have a finite extent of its beam (since, in theory, it generates a Gaussian diffusing beam profile that extends to infinity). Therefore, although most of the energy is diffused in a given cone, it still yields very fine structures that are used to diffract some of the incident beam energy into large angles.

Although the optical reconstruction seems to be identical for a Fourier diffractive diffuser and a holographic diffuser, the microstructures can be quite different, and thus can yield problems when it comes to replicating these structures in plastic or even in other materials (holographic by CGH exposure).

As an example, the diffuser on the left-hand side in Figure 6.43 is a diffractive Fourier diffuser, whereas the diffuser on the right-hand side resembles more a Gaussian holographic diffuser, although it shows an asymmetry on the left-hand side (the empty square-section cone of light within the elliptical Gaussian diffusion cone). These elements have been fabricated over four phase-relief levels.

This chapter has addressed the various issues related to numeric-type diffractive elements and especially to Computer-Generated Holograms (CGHs). The various design techniques and data encoding techniques used in industry have been reviewed, as well as the various high-level operations that one can perform on such CGHs (multiplexing, for example). Finally, some typical optical functionalities that are usually implemented in CGHs have been presented.

## References

- [1] B. Lesem and P.M. Hirsch, 'The kinoform: a new wavefront reconstruction device', *IBM Journal of Research and Development*, **13**, 1969, 150–155.
- [2] W.-H. Lee, 'Binary computer-generated holograms', *Applied Optics*, **18**(21), 1979, 3661–3669.
- [3] B.R. Brown and A.W. Lohmann, 'Computer-generated binary holograms', *IBM Journal of Research and Development*, **13**, 1969, 160–168.
- [4] A. Vander Lugt and A. Cozma, 'A computer algorithm for the synthesis of spatial frequency filters', *Proceedings of the IEEE*, **55**, 1967, 599–601.
- [5] D.C. Chu, J.R. Fienup and J.W. Goodman, 'Multi-emulsion on-axis computer generated hologram', *Applied Optics*, **12**(7), 1973, 1386–1388.
- [6] A.W. Lohmann and D. Paris, 'Binary Fraunhofer holograms, generated by computer', *Applied Optics*, **6**(10), 1967, 1739–1748.
- [7] P.F. MacKee, J.R. Towers, M.R. Wilkinson and D. Wood, 'New applications of optics from modern computer design methods', *British Telecom Technology Journal*, **11**(2), 159–169.
- [8] J.R. Fienup, 'Iterative method applied to image reconstruction and to computer generated holograms', *Optical Engineering*, **19**(3), 1980, 297–305.
- [9] B.K. Jennison, 'Iterative approaches to computer-generated holography', *Optical Engineering*, **28**(6), 1989, 629–637.
- [10] R.W. Gerchberg and W.O. Saxton, 'A practical algorithm for the determination of phase from image and diffraction plane pictures', *Optik*, **35**(2), 1972, 237–246.
- [11] X. Tan, B.-Y. Gu, G.-Z. Yang and B.-Z. Dong, 'Diffractive phase elements for beam shaping: a new design method', *Applied Optics*, **34**(8), 1995, 1314–1320.
- [12] G.-Z. Yang, B.-Z. Dong and B.-Y. Gu, 'Gerchberg–Saxton and Yang–Gu algorithms for phase retrieval in a nonunitary transform system: a comparison', *Applied Optics*, **33**(2), 1994, 209–218.
- [13] W.-H. Lee, 'Method for converting a Gaussian laser beam into a uniform beam', *Optics Communications*, **36**(6), 1981, 469–471.

- [14] R.G. Dorsch, A.W. Lohmann and S. Sinzinger, 'Fresnel Ping-Pong algorithm for two-plane computer generated hologram display', *Applied Optics*, **33**(5), 1994, 869–875.
- [15] U. Mahlab and J. Shamir, 'Iterative optimization algorithms for filter generation in optical correlators: a comparison', *Applied Optics*, **31**(8), 1992, 1117–1125.
- [16] G.-Z. Yang, B.-Y. Gu, X. Tan *et al.* 'Iterative optimization approach for the design of diffractive phase elements simultaneously implementing several optical functions', *Journal of the Optical Society of America A*, **11**(5), 1994, 1632–1640.
- [17] B.-Y. Gu, G.-Z. Yang, B.-Z. Dong, M.-P. Chang and O.K. Ersoy, 'Diffractive-phase-element design that implements several optical functions', *Applied Optics*, **34**(14), 1995, 2564–2570.
- [18] M.A. Seldowitz, J.P. Allebach and D.W. Sweeney, 'Synthesis of digital holograms by direct binary search', *Applied Optics*, **26**(14), 2788–2798.
- [19] M. Clark, 'Enhanced Direct-search Method for the Computer Design of Holograms Using State Variables', *'Diffractive and Holographic Optics Technology III'*, Proc. SPIE Vol. 2689, 1996, 24–34.
- [20] V.V. Wong and G.J. Swanson, 'Design and fabrication of a Gaussian fanout optical interconnect', *Applied Optics*, **32**(14), 1993, 2502–2511.
- [21] M. Clark, 'A direct search method for the computer design of holograms for the production of arbitrary intensity distributions', in *'Diffractive Optics: Design, Fabrication and Applications'*, *Optical Society of America Technical Digest* special issue, 1994.
- [22] S. Teiwe, B. Schillinger, T. Beth and F. Wyrowski, 'Efficient design of paraxial diffractive phase elements with descent search methods' in *'Diffractive and Holographic Optics Technology II'*, I. Cindrich and S.H. Lee (eds), SPIE Vol. 2404, SPIE Press, Bellingham, WA, 1995, 40–49.
- [23] N. Yoshikawa and T. Yatagai, 'Phase optimization of a kinoform by simulated annealing', *Applied Optics*, **33**(5), 1994, 863–868.
- [24] M.S. Kim and C.C. Guest, 'Simulated annealing algorithm for binary phase only filters in pattern classification', *Applied Optics*, **29**(8), 1990, 1203–1208.
- [25] M.T. Eismann, A.M. Tai and J.N. Cederquist, 'Holographic beam former designed by an iterative technique', in *'Holographic Optics: Optically and Computer Generated'*, SPIE Vol. 1052, 1989, 10–17.
- [26] K.S. Urquhart, R. Stein and S.H. Lee, 'Computer generated holograms fabricated by direct write of positive electron-beam resist', *Optics Letters*, **18**, 1993, 308–310.
- [27] A. Larsson, 'Holograms can transform guided beams', *Optics and Laser Europe*, OLE magazine issue, September 1996, 67–69.
- [28] J.D. Stack and M.R. Feldman, 'Recursive mean-squared-error algorithm for iterative discrete on-axis encoded holograms', *Applied Optics*, **31**(23), 1992, 4839–4846.
- [29] T. Dresel, M. Beyerlein and J. Schwider, 'Design and fabrication of computer generated beam shaping holograms', *Applied Optics*, **35**(23), 1996, 4615–4621.
- [30] U. Mahlab, J. Shamir and H.J. Caulfield, 'Genetic algorithm for optical pattern recognition', *Optics Letters*, **16**(9), 1991, 648–650.
- [31] E. Johnson, M.A.G. Abushagar and A. Kathman, 'Phase grating optimization using genetic algorithms', *Optical Society of America Technical Digest*, **9**, 1993, 71–73.
- [32] R. Brown and A.D. Kathman, 'Multielement diffractive optical designs using evolutionary programming', in *'Diffractive and Holographic Optics Technology II'*, I. Cindrich and S.H. Lee (eds), SPIE Vol. 2404, SPIE Press, Bellingham, WA, 1995, 17–27.
- [33] A. Bergeron and H.H. Arsenault, 'Computer generated holograms optimized by a global iterative coding', Prepublication communications, 1993.
- [34] U. Krachardt, 'Optimum quantization rules for computer generated holograms', *Optical Society of America Technical Digest*, **11**, 1994, 139–142.
- [35] F. Wyrowski, 'Iterative quantization of digital amplitude holograms', *Applied Optics*, **28**(18), 1989, 3864–3870.
- [36] R. Bräuer, F. Wyrowski and O. Bryngdahl, 'Binarization of diffractive elements with non-periodic structures', *Applied Optics*, **31**(14), 1992, 2535–2540.
- [37] J.W. Goodman and A.M. Silvestri, 'Some effects of Fourier-domain phase quantization', *IBM Journal of Research and Development*, **14**, 1970, 478–484.
- [38] U. Kristopher and S.H. Lee, 'Phase only encoding method for complex wavefronts', in *'Computer and Optically Generated Holographic Optics'*, SPIE Vol. 1555, 1991, 13–22.

- [39] M. Bernhardt, F. Wyrowski and O. Bryngdahl, 'Coding and binarization in digital Fresnel holography', *Optics Communications*, **77**(1), 1990, 4–8.
- [40] C.B. Burckhart, 'A simplification of Lee's method for generating holograms by computer', *Applied Optics*, **9**, 1970, 1949–1965.
- [41] C.K. Hsueh and A.A. Sawchuk, 'Computer generated double phase holograms', *Applied Optics*, **17**, 1978, 3874.
- [42] A. Kirk, K. Powell and T. Hall, 'Error diffusion and the representation problem in computer generated hologram design', *Optical Computing and Processing*, **12**(3), 1992, 199–212.
- [43] S. Weissbach, F. Wyrowski and O. Bryngdahl, 'Digital phase holograms: coding and quantization with an error diffusion concept', *Optics Communications*, **72**(2), 1989, 37–41.
- [44] M.P. Chang and O.K. Ersoy, 'Iterative interlacing error diffusion for synthesis of computer generated holograms', *Applied Optics*, **32**(17), 1993, 3122–3129.
- [45] S. Weissbach, 'Error diffusion procedure: theory and applications in optical signal processing', *Applied Optics*, **31**(14), 1992, 2518–2534.
- [46] H. Farhoosh, M.R. Feldman and S.H. Lee, 'Comparison of binary encoding schemes for electron-beam fabrication of computer generated holograms', *Applied Optics*, **26**(20), 1987, 4361–4372.
- [47] R. Eschbach, 'Comparison of error diffusion methods for computer generated holograms', *Applied Optics*, **30**(26), 1991, 3702–3710.
- [48] R. Eschbach and Z. Fan, 'Complex valued error diffusion for off-axis computer generated holograms', *Applied Optics*, **32**(17), 1993, 3130–3136.
- [49] D. Just, R. Hauck and O. Bryngdahl, 'Computer-generated holograms: structure manipulations', *Journal of the Optical Society of America A*, **2**(5), 1985, 644–648.
- [50] M.A. Golub, L.L. Doskolovich, N.L. Kazanskiy, S.I. Kharitonov and V.A. Soifer, 'Computer generated multi-focal lens', *Journal of Modern Optics*, **34**(6), 1992, 1245–1251.
- [51] O. Manela and M. Segev, 'Nonlinear diffractive optical elements', *Optics Express*, **15**(17), 10 863–10 868.

# 7

## Hybrid Digital Optics

### 7.1 Why Combine Different Optical Elements?

Combining different optical elements in different implementations might at first sight be a ubiquitous task, since optics with similar physical implementation might work better together than hybrid optics. There are three main reasons to do so:

- First, improvement of existing functionalities of standard optics when introducing other optical elements and mixing them in a single system. The first section of this chapter will review such elements, which are mainly hybrid refractive/diffractive elements. These will be labeled ‘elements’.
- Second, the generation of new optical functionalities that cannot be implemented with a single optical element, or multiplexed optical elements. Such elements are not simply spatially multiplexed but, rather, integrated together to produce the new functionality. Such typical elements are integrated waveguide gratings.
- Third and last, but not least, if no optical functionality is to be optimized or even generated, the reduction of the size, weight or cost of a given optical system by hybridizing various optical elements. The planar optical bench is one such example. Chapter 16 provides many more examples.

As most of these hybrid examples are linked to imaging systems [1–3], the various aberrations, expressed for thin lenses, diffractive surface-relief lenses and holographic lenses, will be reviewed. The various functionality enhancement applications using hybrid refractive/diffractive lens configurations will also be reviewed [4, 5].

The hybrid waveguide/free-space optics are a field apart: therefore, an entire section will be allocated to such hybrid optical elements. These have numerous applications in signal processing and telecoms. Finally, a practical method for optimizing a parametric hybrid lens system by using only ray-tracing models will be described.

### 7.2 Analysis of Lens Aberrations

This analysis of lens aberrations is based on three different optical elements:

- the thin refractive lens mode;
- the surface-relief diffractive element; and
- the holographic lens.

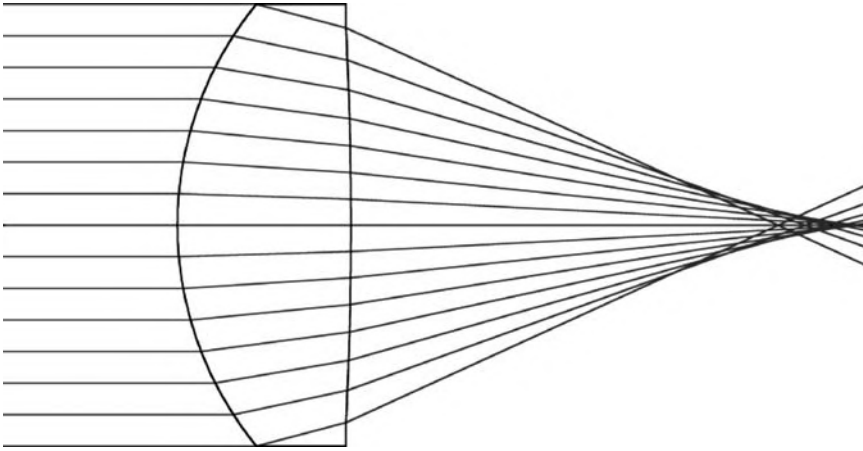


**Figure 7.1** The operation of a plano-convex spherical lens

### 7.2.1 *Refractive Lens Aberrations*

Geometrical aberrations are due entirely to the failure of the optical system to produce a perfect or point image. The geometry of focusing light with spherical surfaces is mathematically imperfect (see Figure 7.1).

Such spherical refractive surfaces are used almost exclusively, due to their inherent ease of fabrication, as shown in Figure 7.2.

Further, the refractive index (or bending power) of glass and other transmitting materials changes as a function of wavelength. This produces changes in the aberrations at each wavelength. The rate of change of slope is constant everywhere on a spherical surface. For two different locations on the surface, a rotation of  $\Theta$  yields surface tangents that are  $\Theta$  apart, regardless of the position on the sphere (see Figure 7.3).

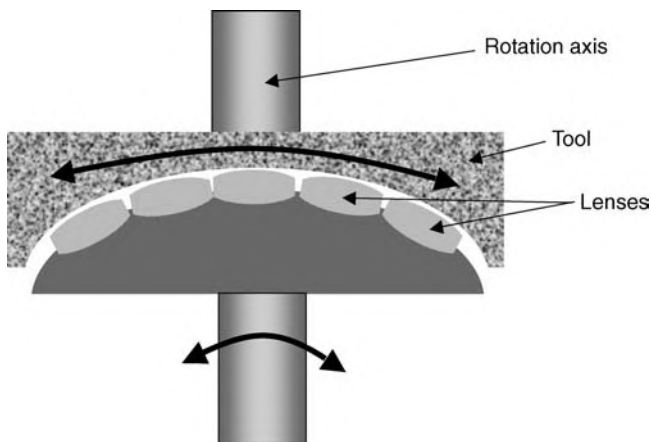


**Figure 7.2** The conventional fabrication of spherical refractive optics



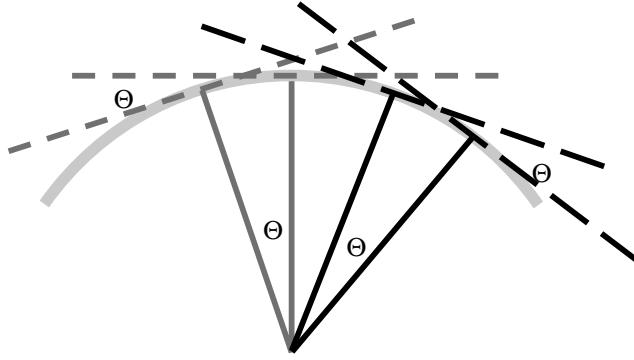


**Figure 7.3** A spherical wavefront

In order to correct aberrations in a single lens, the spherical surface has to be tweaked and rendered aspheric. The standard equation describing rotationally aspheric refractive surfaces is as follows:

$$\text{Sag} = \frac{cy^2}{1 + \sqrt{1 - (k+1)c^2y^2}} + Ay^4 + By^6 + Cy^8 + Dy^{10} + \dots \quad (7.1)$$

where  $c$  is  $(1/\text{radius})$  at the surface vertex,  $k$  is the conic constant:

- if  $k < -1$ , the surface is a hyperboloid,
- if  $k = -1$ , the surface is a paraboloid,
- if  $-1 < k < 0$ , the surface is an ellipsoid,
- if  $k = 0$ , the surface is spherical,
- if  $k > 0$ , the surface is an oblate spheroid, and

$A, B, C, D, \dots$  are the fourth-, sixth-, eighth-, tenth-, ... order aspheric terms.

In some cases, the constraint on the aspheric surface is such that the lens no longer keeps its rotational symmetric profile and becomes an anamorphic lens. Anamorphic refractive lenses are more expensive to produce than standard lenses. For example, a lathe with four (or five) degrees of freedom is needed to turn an anamorphic refractive lens. Anamorphic lenses are mostly used in nonimaging tasks; for example, illumination, beam shaping and so on.

Aberrations are basically the failure of the rays from a given object point to come to a single or common image point.

Real rays obey Snell's law ( $n \sin \theta = n' \sin \theta'$ ), and paraxial rays obey  $n\theta = n'\theta'$ . Paraxial optics have zero aberrations (see Figure 7.4).

The Optical Path Difference (OPD), or the wave aberration function, can be mathematically expressed in the form of a polynomial for rotationally symmetric optical systems. In Seidel aberrations, each term of the polynomial is identified with a particular type of aberrations (third order). The Nijboer criterion expresses the wavefront aberration as a function of  $h$ ,  $\rho$  and  $\cos \theta$ :

$$\begin{aligned} W(h, \rho, \cos \theta) = & \frac{1}{8}S_I\rho^4 + \frac{1}{2}S_{II}h\rho^3\cos \theta + \frac{1}{2}S_{III}h^2\rho^2(\cos \theta)^2 \\ & + \frac{1}{4}(S_{III} + S_{IV})h^2\rho^2 + \frac{1}{2}S_Vh^3\rho(\cos \theta) + \dots \text{ (higher order terms)} \end{aligned} \quad (7.2)$$

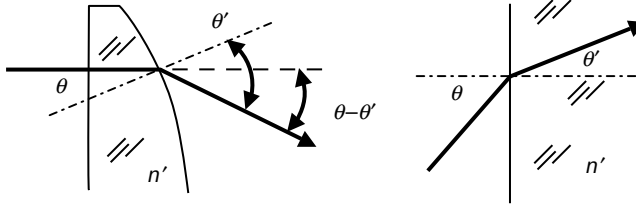


**Figure 7.4** Snell's law and paraxial rays

where  $h$  is the height of the image,  $\rho$  is the aperture (the radial position in the lens plane) and  $\theta$  is the angular position of that point in the lens.

As an example, on the left-hand side of Figure 7.5, a plot of the OPD for a lens with defocus and third-order spherical aberration is shown. On the right-hand side, the data for a lens with third-, fifth- and seventh-order aberrations is shown:

- Spherical aberration is an axial aberration, and is generally cubic with aperture ( $\propto \rho^4$ ). Therefore, a given lens with an image blur of 0.01 inches would have a 0.00125 inch blur at half of its aperture ( $0.5^3 = 0.125$ ). Spherical aberrations can be controlled by varying lens bendings (see Figure 7.6(a)) or by adding lenses (or splitting the optical power) (see Figure 7.6(b)).
- Coma depends on  $h$  and changes sign with  $\theta$ . It is proportional to  $\propto h\rho^3 \cos \theta$ .
- Astigmatism is proportional to  $\propto h^2 \rho^2 \cos^2 \theta$ .
- Petzval curvature is proportional to  $\propto h^2 \rho^2$ .
- Distortion is proportional to  $\propto h^3 \rho \cos \theta$ .
- Chromatic aberrations are basically variations of the position of the focal spot as a function of wavelength, which is mostly linked to the index of refraction changing with the wavelength. Note that in the case of a refractive lens, only the position of the spot varies, not its efficiency (as opposed to diffractive lenses).

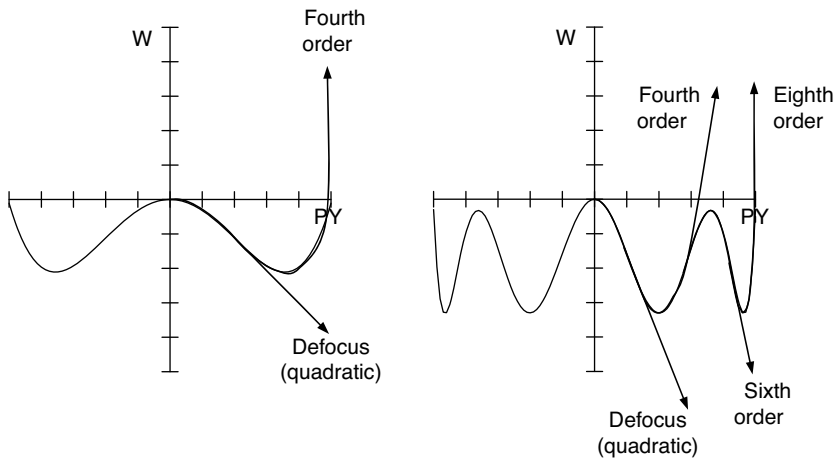


**Figure 7.5** Lenses with third-order and seventh-order aberrations

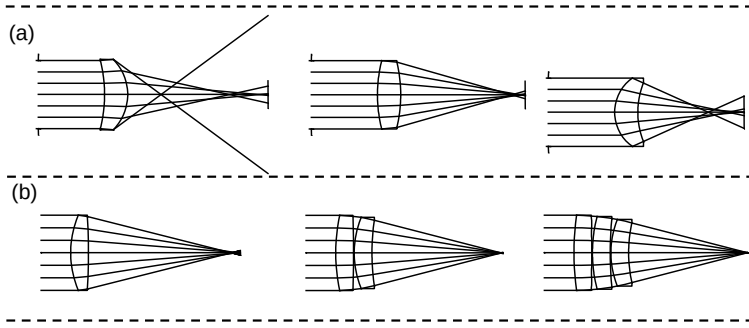


**Figure 7.6** Controlling spherical aberration in refractive lenses

The Abbe  $V$  number expresses the spectral dispersion of lenses, and in the case of refractive lenses is defined as follows:

$$V_{ref} = \frac{n_{med} - 1}{n_{short} - n_{long}} \approx 50 \quad (7.3)$$

The higher this number is, the lower the spectral dispersion and, therefore, also the chromatic aberrations of such a lens. For example, Abbe  $V$  numbers of diffractives are much smaller, translating into a strong spectral dispersion (and chromatic aberrations). Refractives and diffractives have opposite Abbe  $V$  numbers, and thus opposite spectral dispersions, which is the key to most of the hybrid solutions presented in this chapter.

The shift introduced in the focal length for a refractive lens, as a function of the wavelength (only related to the change in the index of refraction) is as follows:

$$\frac{1}{f(\lambda)} = (n(\lambda) - 1) \cdot \left( \frac{1}{R_1} - \frac{1}{R_2} \right) \quad (7.4)$$

### 7.2.2 Aberrations of Diffractive Lenses

In order to be able to model a diffractive lens and derive a list of aberration coefficients, the Sweatt model (see also Chapter 11) must be introduced, which involves the modeling of such a diffractive lens as a very thin refractive lens with an extremely high refractive index (e.g. 10 000). Some of the aberrations linked to diffractives, under the Sweatt model assumptions, are derived below.

Assuming that a diffractive lens can be modeled as a refractive lens with two curvatures,  $c_1$  and  $c_2$ , as the index of the material composing this lens increases, the curvature decreases, and converges towards the diffractive lens curvature (the underlying surface, which can be planar or curved in the case of a hybrid element): see Figure 7.7.

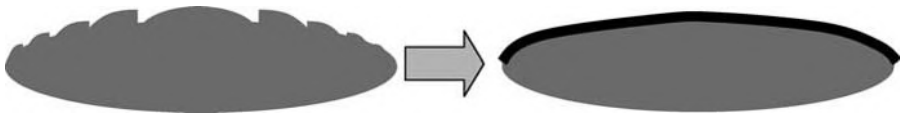


**Figure 7.7** The Sweatt model for a diffractive lens on top of a refractive lens

The bending parameters of a diffractive lens can be written as

$$B = \frac{c_1 + c_2}{(n-1)(c_1 - c_2)} = \frac{c_2 + c_2}{\theta} \rightarrow \frac{c_s}{\theta} \quad (7.5)$$

Similarly to refractive lenses, the wavefront aberration polynomial can be written as follows:

$$W_D(h, \rho, \cos \theta) = \frac{1}{8} S_{D-I} \rho^4 + \frac{1}{2} S_{D-II} h \rho^3 \cos \theta + \frac{1}{2} S_{D-III} h^2 \rho^2 (\cos \theta)^2 \\ + \frac{1}{4} (S_{D-III} + S_{D-IV}) h^2 \rho^2 + \frac{1}{2} S_{D-V} h^3 \rho (\cos \theta) + \dots \text{ (higher order terms)} \quad (7.6)$$

Based on the wavefront aberration polynomials (in Equations (7.2) and (7.6)), Table 7.1 summarizes the differences between aberrations in refractive and diffractive lenses. In the table,  $y$  denotes the position of the marginal ray and  $H$  is the Lagrange invariant.

In Chapters 5 and 6, a diffractive lens that has strong chromatic aberrations (i.e. it has a strong spectral dispersion) is shown.

The equivalent Abbe  $V$  number of a diffractive lens is as follows:

$$V_{dif} = \frac{\lambda_{med}}{\lambda_{short} - \lambda_{long}} \approx -3.5 \quad (7.7)$$

Similar to refractive lenses, the shift in focal length can be expressed as a function of wavelength (Equation (7.8)).

$$f(\lambda) \approx f(\lambda_0) \cdot \frac{\lambda_0}{\lambda} \quad (7.8)$$

Typical values for diffractive lens Abbe  $V$  numbers are  $-3.45$  for visible wavelengths,  $-1.9$  for near-infrared wavelengths (around  $4 \mu\text{m}$ ) and  $-2.4$  for far-IR wavelengths (around  $10 \mu\text{m}$ ).

**Table 7.1** Lists of the aberrations produced by refractive and diffractive lenses

|                     | Refractive lens                                                                                                                                                              | Diffractive lens                                                       |
|---------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------|
| Spherical           | $S_I = \frac{y^4 \theta^4}{4} \left( \left( \frac{n}{(n-1)} \right)^2 + \frac{n+2}{n(n-1)^2} E^2 \right. \\ \left. + 4 \frac{(n+1)}{n(n-1)} ET + \frac{3n+2}{n} T^2 \right)$ | $S_I = \frac{y^4 \theta^4}{4} (1 + B^2 + 4BT + 3T^2) - 8\lambda B y^4$ |
| Coma                | $S_{II} = \frac{-y^2 \theta^2 H}{2} \left( \frac{n+1}{n(n-1)} E + \frac{2n+1}{n} T \right)$                                                                                  | $S_{II} = \frac{-y^2 \theta^2 H}{2} (B + 2T)$                          |
| Astigmatism         | $S_{III} = H^2 \theta$                                                                                                                                                       | $S_{III} = H^2 \theta$                                                 |
| Petzval curvature   | $S_{IV} = \frac{H^2 \theta}{n}$                                                                                                                                              | $S_{IV} = 0$                                                           |
| Distortion          | $S_V = 0$                                                                                                                                                                    | $S_V = 0$                                                              |
| Bending parameter   | $E = \frac{c_1 + c_2}{c_1 - c_2}$                                                                                                                                            | $B = \frac{2c_{sub}}{\theta}$                                          |
| Conjugate parameter | $T = \frac{u + u'}{u - u'} = \frac{m + 1}{m - 1}$                                                                                                                            | $T = \frac{u + u'}{u - u'} = \frac{m + 1}{m - 1}$                      |

### 7.2.3 Aberrations of Holographic Lenses

Since holograms can take on the functionality of lenses (as sets of circular gratings with variable periods), sets of optical aberrations can therefore be derived in a similar way to how they are derived for conventional refractive lenses.

If the hologram is recorded in the configuration shown in Figure 8.21 (see Chapter 8), and if a third wavefront is also incident on the hologram (e.g. a wavefront of radius of curvature  $R_i$  and wavelength  $\lambda_i$ ), the radius of curvature of the resulting diffracted wavefront in the  $m$ th order can be expressed as follows:

$$\frac{1}{R_m} = m \frac{\lambda}{\lambda_i} \cdot \left( \frac{1}{R_O} - \frac{1}{R_R} \right) + \frac{1}{R_i} \quad (7.9)$$

From Equation (7.9), the total aberrations  $\Delta r$  produced by an HOE lens can be derived in the fundamental negative order ( $m = -1$ ) as a summation of the  $\Delta_S$  spherical (S),  $\Delta_A$  astigmatic (A) and  $\Delta_C$  comatic (C) aberrations:

$$\Delta_R = \Delta_S + \Delta_C + \Delta_A$$

$$\text{where } \begin{cases} \Delta_S = -\frac{1}{8\lambda} x^4 \left( \frac{1}{R_i^3} - \frac{1}{R_m^3} + m \frac{\lambda}{\lambda_i} \left( \frac{1}{R_O^3} - \frac{1}{R_R^3} \right) \right) \\ \Delta_C = \frac{1}{2\lambda} x^3 \left( \frac{\sin(\alpha)}{R_C^2} - \frac{\sin(\theta_m)}{R_m^2} + m \frac{\lambda}{\lambda_i} \left( \frac{\sin(\theta_O)}{R_O^2} - \frac{\sin(\theta_R)}{R_R^2} \right) \right) \\ \Delta_A = -\frac{1}{2\lambda} x^2 \left( \frac{\sin^2(\alpha)}{R_C} - \frac{\sin^2(\theta_m)}{R_m} + m \frac{\lambda}{\lambda_i} \left( \frac{\sin^2(\theta_O)}{R_O} - \frac{\sin^2(\theta_R)}{R_R} \right) \right) \end{cases} \quad (7.10)$$

The analysis of these aberrations is used to fine-tune the position of the various point sources in order to minimize a particular aberration. Usually, the first step is to reduce the comatic aberrations for all values of  $x$  ( $\Delta_C = 0$ ), and then the spherical and astigmatic aberrations are set to cancel each other at the edges of the hologram ( $\Delta_A = -\Delta_S$ ).

Other parameters also enter into the optimization of the exposure set-up, such as the optimal grating period for the holographic material used, the optimal wavelength of this material, the lateral extent of this material and so on.

## 7.3 Improvement of Optical Functionality

In this section, two examples of an improvement in optical functionality by the use of hybrid refractive/diffractive lenses are presented:

- lens achromatization; and
- lens athermalization.

### 7.3.1 Achromatization of Hybrid Refractive/Diffractive Lenses

Using hybrid optics to achromatize a lens is now becoming a conventional technique [6–14]. This technique consists of inserting a diffractive structure that is carefully designed so to balance the spectral dispersion, as the Abbe  $V$  numbers of refractives and diffractives have opposite signs. As the spectral dispersion of diffractives is much stronger than for refractives (see previous sections), an achromat singlet typically has a much stronger refractive power than its diffractive power – about 10 times or more (see Figure 7.8).

In the previous sections, the change in focal length as a function of the wavelength for refractive and diffractive lenses was derived. When achromatizing a refractive doublet, or here a hybrid singlet for two

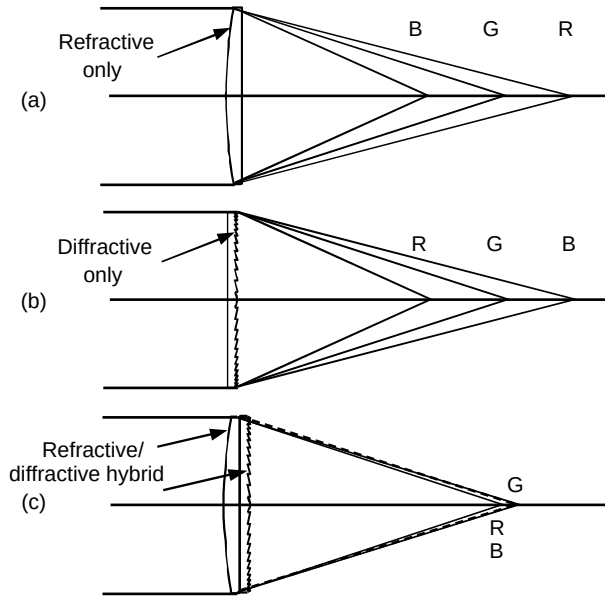


**Figure 7.8** Achromatization with a hybrid diffractive/refractive singlet

wavelengths  $\lambda_1$  and  $\lambda_2$ , one of the following conditions must be satisfied:

$$\frac{C_1}{V_1} + \frac{C_2}{V_2} = 0 \quad \text{where } C_{total} = C_1 + C_2 \quad (7.11)$$

where  $V_1$  and  $V_2$  are, respectively, the Abbe  $V$  numbers of the first and second lens, and  $C_1$  and  $C_2$  are, respectively, the lens powers of the two separate refractive lenses (doublet) or of the refractive and diffractive profiles of a single lens (singlet).

The solution is either to use opposite signs of lens powers when using refractives (an achromatic doublet) or to have the same lens power signs and use a hybrid lens (the Abbe  $V_1$  and  $V_2$  numbers will thus have opposite signs and satisfy the condition). Note that the drop in diffraction efficiency that occurs when another wavelength, rather than the wavelength used for the design and fabrication of the diffractive lens (the depth of the grooves), is used to illuminate the hybrid lens must be taken into consideration. Figure 7.9 shows such a hybrid achromatic singlet. Such lenses are typically fabricated by diamond turning.

When using diamond-turning fabrication, the underlying aspheric preform has to be turned prior to the actual diffractive fringes, since the diamond tool geometry and sizes used to carve out the aspheric



**Figure 7.9** A hybrid achromatic singlet

preform and then the very thin and narrow blazed fringes are not the same. A lot of material is machined out of the substrate when turning the preform and a very low amount of material is machined out to form the fringes on top of it. Typical diamond-turning materials are acrylic and PCB, with refractive indices of about 1.49.

### 7.3.2 Athermalization of Hybrid Refractive/Diffractive Lenses

Thermal effects on a lens decrease or increase its focal length, therefore imposing a burden on the use of refractive lenses in systems that have to undergo large temperature swings. (For example,  $-30^{\circ}\text{C}$  to  $+50^{\circ}\text{C}$  is a normal temperature swing in the Northern Hemisphere, especially if the optical system is located in machinery that has to rely heavily on the optics. Such typical applications are military – IR lenses for missile laser tracking, IR lenses for IR imaging and so on.) Such IR lenses are usually fabricated using exotic materials such as Ge, ZnS or ZnSe, since glass or fused silica is no longer transparent to these wavelengths.

Similar to the design of a hybrid singlet achromat lens, the design procedure for a singlet athermal lens is derived here [15–17]. First, the thermal effects on both refractive and diffractive lenses are derived.

The thermal effects ( $T$ ) on a refractive lens can be expressed by the opto-thermal expansion coefficient, as described in Equation (7.12):

$$f = \frac{1}{(n-1)c} \Rightarrow x_{f,ref} = \frac{1}{f} \frac{\partial f}{\partial T} \approx \frac{1}{n-1} \frac{\partial n}{\partial T} \quad (7.12)$$

For example, an IR lens built in Ge would have  $\partial n / \partial T = 390 \times 10^{-6}$ .

When  $T$  increases, the index also increases, the focal length decreases and  $\partial f / \partial T$  is negative. Therefore an IR lens (KRS-5) has an opto-thermal expansion coefficient of  $-234 \times 10^6$ .

The thermal effects on a diffractive lens can be expressed based on the thermal expansion coefficient of the material in which the diffracted is fabricated. When the temperature increases, the lens is thermally expanded.

A diffractive lens can be expressed as a chirped grating in which the transition points are defined at  $\Phi(r) = 2m\pi$ , where  $m$  is the order of the lens (see also Chapter 5):

$$r_m = \sqrt{\frac{2m\lambda_0 f}{n_0}} \quad (7.13)$$

Thus, expressed as

$$r_m(T) = r_m(1 + \alpha_g \cdot \Delta T) \quad (7.14)$$

the focal length expression for the diffractive lens then becomes

$$f(T) = \frac{n_0 r_m^2}{2m\lambda_0} = \frac{1}{2m\lambda_0} r_m^2 (1 + \alpha_g \cdot \Delta T)^2 \cdot n_0 \quad (7.15)$$

The opto-thermal expansion coefficient for diffractive lenses is extracted as follows:

$$x_{f,dif} = \frac{1}{f} \frac{\partial f}{\partial T} \approx 2\alpha_g + \frac{1}{n_0} \cdot \frac{\partial n_0}{\partial T} \quad (7.16)$$

Therefore, the opto-thermal expansion coefficient is a positive value, with an opposite sign from that of the refractive opto-thermal expansion coefficient.

For IR material, this coefficient is much smaller in amplitude than for refractive lenses:

$$|x_{f,dif}| \ll |x_{f,ref}| \quad (7.17)$$

For a hybrid athermal lens, the expansion coefficient of the doublet should equal that of the lens mount, and is

$$x_{f,doublet} = x_{f,mount} \quad (7.18)$$

The derivative of the effective focal length  $f_{doublet}$

$$\frac{1}{f_{doublet}} = \frac{1}{f_{ref}} + \frac{1}{f_{dif}} \quad (7.19)$$

becomes

$$\frac{1}{f_{doublet}^2} \frac{\partial f_{doublet}}{\partial T} = \frac{1}{f_{ref}^2} \frac{\partial f_{ref}}{\partial T} + \frac{1}{f_{dif}^2} \frac{\partial f_{dif}}{\partial T} \quad (7.20)$$

which becomes

$$\frac{x_{f,doublet}}{f_{doublet}} = \frac{x_{f,dif}}{f_{dif}} + \frac{x_{f,ref}}{f_{ref}} \quad (7.21)$$

and when Equation (7.18) is considered again, we can write

$$f_{dif} = \frac{x_{dif}}{\frac{x_{mount}}{f_{doublet}} - \left(\frac{x_{ref}}{f_{ref}}\right)} \quad (7.22)$$

which becomes the condition for an athermalized hybrid refractive/diffractive achromat.

Note that holograms are rarely used along with refractives, since holograms cannot be fabricated in the same material as the lens, although theoretically it would be possible to achromatize or athermalize hybrid holographic refractives.

Diffractives can be directly etched into refractive lenses by diamond turning or curved lithography plus RIE.

## 7.4 The Generation of Novel Optical Functionality

### 7.4.1 Hybrid Holographic/Reflective/Refractive Elements

Previously, it was shown that holograms are rarely used in conjunction with refractives, due to material issues. However, holographic replicated gratings are very often used with refractives, for their spectral dispersion characteristics or for their beam-splitting effects.

#### 7.4.1.1 Concave Spectroscopic Gratings

A holographic spectroscopic grating disperses the spectrum, and a curved mirror focuses light. These two optical functionalities can be integrated in one single functionality when fabricating a concave grating. Such a grating is depicted in Figure 5.16 (see Chapter 5).

#### 7.4.1.2 Holographic OPU Lens

Replicated holograms (etched in plastic or quartz) are also often used as beam splitters in conjunction with objective lenses in Optical Pick-up Units (OPUs) for focus/track control.

Today, most of the 780 nm lasers used in CD/DVD OPUs have a hologram etched into the laser window for direct focus/track control (see Chapter 16).



## 7.4.2 Hybrid Diffractive/Refractive Elements

In the previous section, it was noted that hybrid refractive/diffractive lenses can compensate chromatic and thermal aberrations. Now, how they can add functionality to existing refractives – in a way that refractives alone could not do, even with multiple refractives – will be described. The hybrid best seller for many years now is Matsushita's hybrid multifocus lens.

### 7.4.2.1 The Hybrid Multifocus Lens

Another very promising application of hybrid refractive/diffractive optics is the dual-focus Optical Pick-up Unit (OPU) lens, which is used in most of the CD/DVD drives available on the market today. Such a lens has a convex/convex refractive, with one bearing a diffractive profile. The diffractive profile is intentionally detuned so that it produces only 50% efficiency in the fundamental positive order, leaving the rest of the light in the zero order. Such a lens therefore creates two wavefronts to be processed by the refractive profiles (which are always 100% efficient). When the various profiles are carefully optimized so that the zero order combined with the refractive profiles would compensate for spherical aberrations at 780 nm wavelength through a CD disk overcoat, and the diffracted fundamental order combined with the same refractive profiles would compensate for spherical aberrations at 650 nm wavelength through a DVD overcoat, this lens is ready to pick up either CD or DVD tracks.

Figure 7.10 describes how such lenses have been developed, from a spatially multiplexed lens couple to a compound bifocal multiplexed refractive lens to a hybrid dual focus lens.

The compound bifocal microlens shown in the center of Figure 7.10 is a dual-focus lens, but the dual NAs of such a lens cannot be optimized correctly, since the lenses are spatially multiplexed rather than phase multiplexed as in the hybrid lens configuration. In the compound refractive bifocal lens, the CD lens is in the center and the DVD lens (with a larger NA) is fabricated over a doughnut aperture around the DC lens. In a hybrid refractive/diffractive dual-focus lens, the entire lens aperture can be used to implement the CD and the DVD lens.

Figure 7.11 shows such a hybrid refractive/diffractive OPU objective lens and the associated numeric reconstruction of the two focal spots in the near field (the general Fresnel numeric propagator presented in Chapter 11 has been used to perform this numeric reconstruction).

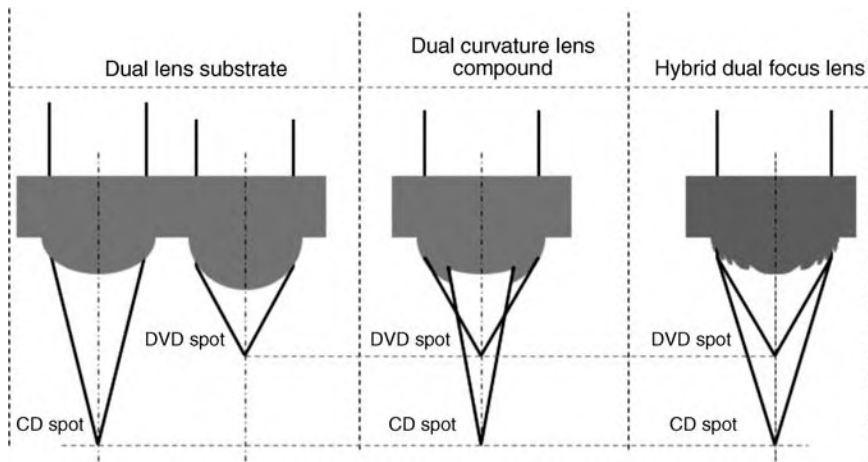


**Figure 7.10** Dual-focus lenses for CD/DVD OPUs

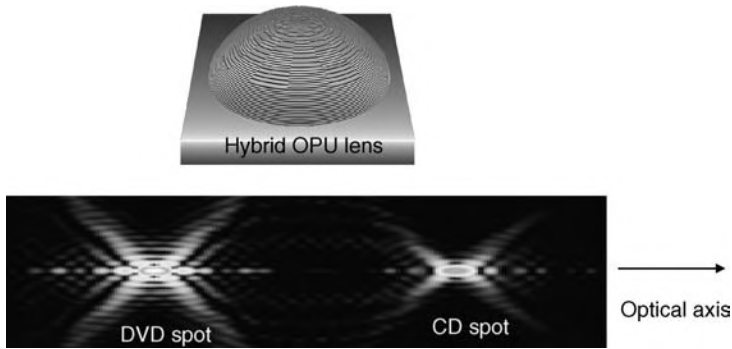


**Figure 7.11** A numeric reconstruction of a hybrid dual-focus OPU objective lens

Hybrid optics and hybrid optical compound lenses are also extremely suitable for the design of triple-focus lenses to be used in the next-generation Blu-ray disks (BD OPU), which feature three different wavelengths (780 nm, 650 nm and 405 nm), three different focal lengths and three different spherical aberrations to compensate for three different disk media overcoat thicknesses. For an example of a tri-focus OPU lens, see Chapter 16.

### 7.4.3 Hybrid GRIN/Free-Space Optics

Chapter 3 has reviewed various implementations of GRIN lenses. Since such GRIN lenses have a planar surface, as opposed to conventional refractive lenses, which produce a surface profile sag, the lithographic fabrication of diffractive elements on such GRINs (especially in 2D arrays) is thus possible. Such a hybrid GRIN/diffractive array is shown in Section 4.4.1.2.

One popular implementation of GRIN lenses with arrays of Thin Film Filters (TFFs) is presented in Figure 7.12. This is a typical example of a waveguide/free-space optics application for the telecom industry. The application in Figure 7.12 is a CWDM Demux/Mux device for telecom applications.

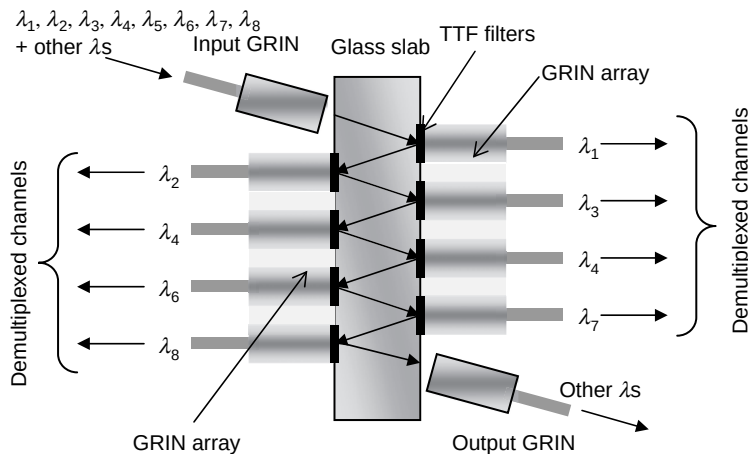


**Figure 7.12** A 1D array of GRINs with TFF arrays for CWDM Demux

Chapter 9 shows more applications of GRIN arrays used in free-space applications (arrays of 2D and 3D micromirrors etc.).

Other hybrid implementations of GRINs/diffractives have been used in endoscopic devices, in which the imaging quality in the end tip of the endoscope had to be improved from the initial GRIN fixed parameters (e.g. for endoscopic zoom applications).

## 7.5 Waveguide-based Hybrid Optics

Digital waveguide optics, and especially waveguide gratings, have been extensively reviewed in Chapter 3. There are mainly two different ways in which waveguide optics are hybridized with either free-space optics or gratings for industrial applications.

The first uses MultiMode Interference (MMI) cavities, in 2D or 3D, after and before waveguide structures, in the same substrate. These cavities are produced to allow the waves out-coupled by the waveguide to diffract freely in free space (free space means not guided – free space can be inside the materials, and for most of the time is actually inside the material).

The second uses gratings (sub-wavelength or scalar regime gratings) to produce waveguide gratings or Fiber Bragg Gratings and integrated Bragg reflectors, as used in DFB and DBR lasers or wavelength lockers in PLC implementations (see Chapter 3).

### 7.5.1 *Multimode Interference Cavities in PLCs*

In Chapter 3, the various implementations of Arrayed Waveguide Gratings (AWG) and Waveguide Grating Routers (WGR), which both use partial waveguide sections and partial free-space sections in the PLC in order to produce the desired functionality, were discussed. Usually, this functionality is spectral dispersion combined with coupling into arrays of waveguides.

The cascading of such WGR or AWGs with other elements such as directional couplers and Mach–Zehnder interferometers is used to produce complex optical functionalities such as integrated router modules, integrated add–drop modules, integrated interleaver modules and so on.

### 7.5.2 *Waveguides and Free-space Gratings*

#### 7.5.2.1 Integrated Bragg Couplers

Figure 7.13 shows an integration of a waveguide coupler with a free-space lensing out-coupling functionality. Such a device can be used as an integrated waveguide Optical Pick-up Unit (OPU).

Waveguide Bragg grating couplers, in which the grating is etched on top of the waveguide as a binary grating, can be used as sensors for biotechnology applications (see Chapter 16) and as VOAs for telecom applications.

#### 7.5.2.2 Integrated Bragg Reflectors

Integrated Bragg reflectors combine waveguide and Bragg grating reflection (as in traditional reflection holography) to produce numerous integrated devices, such as DBR and DFB lasers (see Chapter 3).

#### 7.5.2.3 An Example of a Dynamic Gain Equalizer

When using holographic materials as the cladding of a ridge waveguide, several interesting effects can be triggered. The integration of an effective index cladding material or the integration of a waveguide Bragg coupler could be effected.

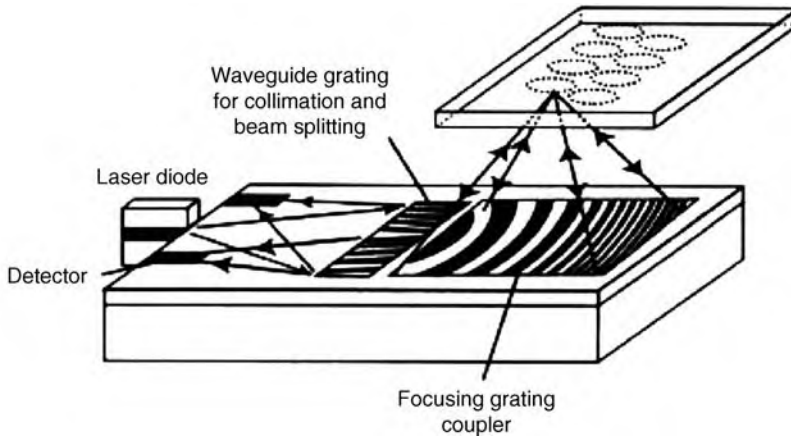


**Figure 7.13** An integrated waveguide lensing coupler for an OPU module

If this holographic grating is a volume grating (composed of Bragg planes) and is chirped (i.e. the grating period increases monotonously), spectral shaping effects can be implemented.

Now, if such a holographic grating can be tuned (see also Chapter 10) from 0% efficiency to 100% efficiency, one can implement a dynamic spectral shaper (much like an audio frequency tuner in a hi-fi device).

Figure 7.14 shows the implementation of such a dynamic hybrid waveguide grating coupler, to be used as a Dynamic Gain Equalizer (DGE) in a DWDM line after gain amplification through an Erbium Doped Fiber Amplifier (EDFA).

The integration of such a device uses here a Silicon OxyNitride (SiON) ridge waveguide for its high index, and an H-PDLC (Holographic Polymer Dispersed Liquid Crystal) holographically originated volume grating around the waveguide.

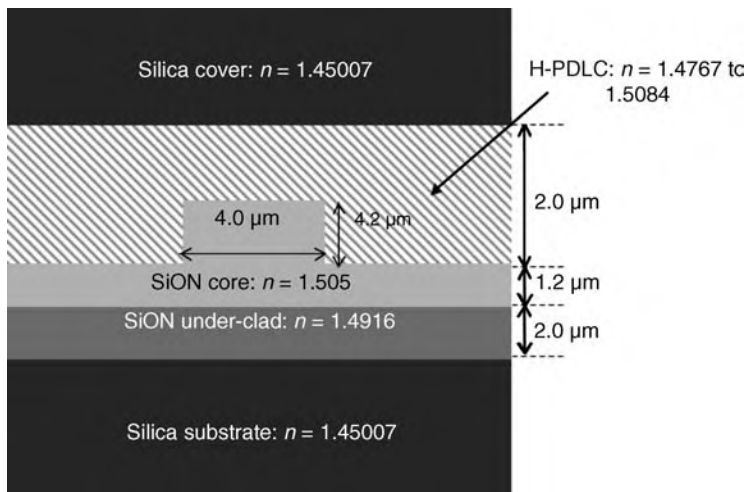


**Figure 7.14** The integration of a hybrid waveguide/hologram system for spectral shaping

7.6 Reducing Weight, Size and Cost

7.6.1 Planar Integrated Optical Benches

Planar integrated optical benches, also referred to as ‘planar optics’, are a substrate on which several optical elements have been patterned lithographically [18–22]. Such optical elements can be positioned with sub-micron accuracy directly on the tooling (photomask or reticle) during the e-beam mask patterning, and lithographically projected onto the final substrate.

The substrate works either in total internal reflection (TIR) mode or in simple reflection mode, with an added reflective coating on all or some of the patterned elements.

The elements can be of various types: surface-relief diffractive, surface-relief refractive, reflective, GRIN and even waveguide. There is also varying functionality: lenses, gratings, analytic DOEs or numeric CGHs, sub-wavelength gratings and so on.

Figure 7.15 shows how such elements can be integrated in a planar optical bench.

The integration of such elements can be done on one side of the substrate or on both sides, with appropriate front/back side alignment techniques (for more information on such alignment techniques, see Chapter 12).

In this instance, the free space is the internal substrate itself. Usually, the beam is coupled to the slab by one of the techniques described in Chapter 3.

All elements are working in off-axis mode: therefore, the aberration controls over refractive and diffractive lenses have to take care of the strong off-axis mode (compensation for coma etc.). Therefore, such systems are unlikely to be used as standard imaging systems, but can find nice applications as special image sensors or for nonimaging tasks such as optical pick-up units (OPUs) for disk drive read-out. The planar optical bench is very desirable for the latter, since it can be replicated in one single process, in glass or plastic, without having to align any of the numerous hybrid optical elements constituting the system (as is done today in the OPU industry). Moreover, such planar optical benches can be used in medical diagnostic applications, as low-cost, mass-replicable disposable devices. Figure 7.16 shows two potential applications for such planar optical benches (see also Figure 7.13, which shows a similar device but integrated with waveguide gratings and waveguide lensing functionalities).

However, due to the high constraints on DVD or Blu-ray read-out functionality (high NA, perfect Strehl ratio etc.), it is unlikely that such planar devices will be used for such high-content disk applications. However, they could be used for lower-content optical disk applications, such as a credit card sized CD, where the reader could be disposable (not for computing applications, but for security).

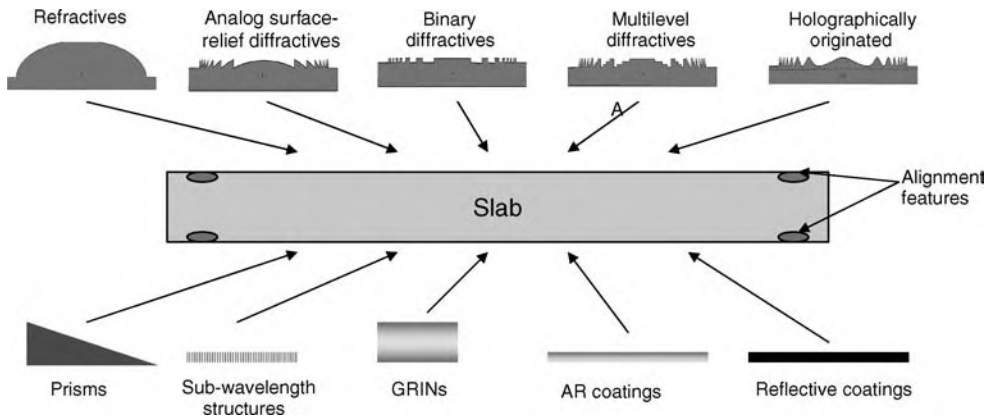


Figure 7.15 The integration of multiple elements in a planar optical bench

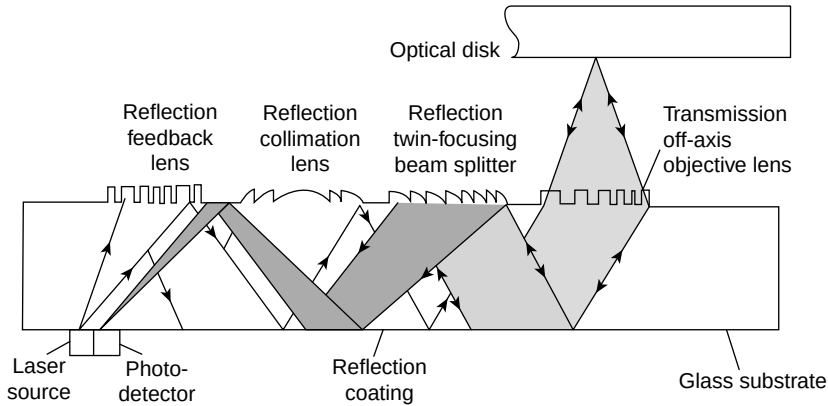


**Figure 7.16** Planar optical benches using hybrid optics

### 7.6.2 Origami Folded Optical Systems

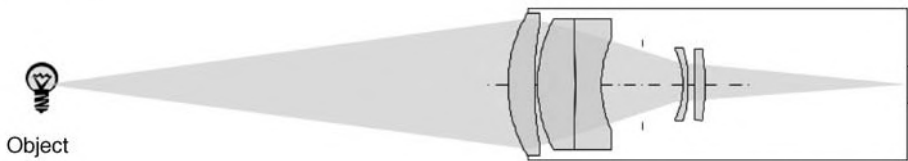
With the same intention of reducing the size, weight and eventually the cost of an optical system, while providing a one-step replication technology, origami space-folded imaging systems have been developed.

Such an imaging system linked to a CMOS image sensor used for a security application is depicted in Figure 7.17.

The origami lens in Figure 7.17 produces a gain in size and weight of more than  $10\times$ .

The main difference between the previous planar optical bench example and an origami optical system is that in a planar optical bench the whole aperture of the optical elements can be used (although in off-axis mode), whereas in an origami lens, due to the on-axis space folding scheme, only a part of the aperture can

#### Refractive lens



#### Folded optics

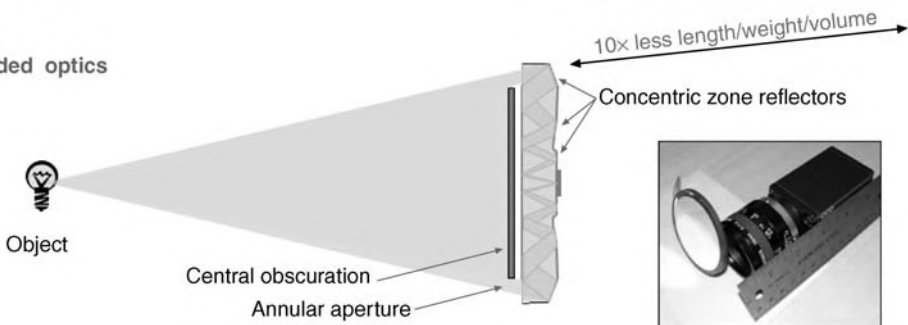


**Figure 7.17** An origami lens developed at the University of California San Diego

be used (usually a toroidal section). However, when an origami lens is used in off-axis mode (as in the CMOS origami imager presented in Chapter 16), the element strongly resembles a planar optical bench.

Such origami optical systems can comprise diffractive elements on curved or planar reflective surfaces, which can – for example – implement wavefront coding functionalities as longitudinal chromatic aberrations (see also Chapter 16), as a pure reflective system does not yield any chromatic aberrations. Also, it has been demonstrated that holographic optical elements can be integrated in such origami optical systems.

Furthermore, such a system can be dynamic, changing its focus by moving one surface with regard to the other (see also Chapter 9).

### 7.6.3 *Hybrid Telefocus Objective Lenses*

Chapter 16 shows a hybrid refractive/diffractive telephoto zoom lens, which has been developed not to increase the performance or produce new functionality, but to reduce the size, weight and cost of the objective.

The telephoto zoom lens is depicted in Figure 16.35 (see Chapter 16). The compound lens uses sandwiched Fresnel lenses with different indices. Actually, the Fresnel lenses used in the Canon example are much more micro-refractive Fresnel lenses than diffractive lenses. Nevertheless, this is a typical example of hybridization for size, weight and cost issues.

## 7.7 Specifying Hybrid Optics in Optical CAD/CAM

There are many ways to specify hybrid refractive/diffractive elements in classical optical CAD/CAM tools. Some of these description techniques are presented below.

Figure 7.18 shows how the optical CAD software Zemax™ specifies the hybrid refractive/diffractive lens shown on the upper right part of the figure.

The lens shown in Figure 7.18 is a typical hybrid refractive/diffractive lens for digital imaging tasks. The first optical surface is usually refractive and spherical, and bears most of the raw focusing, whereas the second optical surface is usually aspheric and also bears the diffractive surface (which is usually also aspheric).

The aspheric coefficients of the refractives and diffractive lenses are listed in Figure 7.18. For more details about such aspheric coefficients, see also Chapter 5.

Great care has to be taken when specifying aspheric coefficients, since the values that are generated by optical CAD on the market can differ markedly from each other (e.g. such coefficients can be expressed in waves, wavenumbers, radians, milliradians, microns, centimeters, millimeters etc.).

In Figure 7.19, it is shown how the FRED™ software defines diffractive surfaces (as would Zemax™ or CodeV™), by specifying aspheric coefficients of an infinitely thin refractive (Sweatt model) with special spectral properties on top of a base surface.

Such CAD tools are also capable of optimizing and modeling such hybrid optical elements, usually by one of the techniques described in Chapter 11. The need to use numeric propagator algorithms (physical optics) rather than simulation techniques based on ray tracing, in order to model effects of multiple diffraction order interference, effects of fabrication errors and so on, should be emphasized.

That said, if the desired outcome is to fabricate such a hybrid lens, by either diamond turning or lithographic techniques (see Chapters 12 and 13), it is essential to make sure that the file generated by such a CAD tool can be read by the foundry – which is usually not the case, especially when lithographic fabrication technologies are to be used. Thus it is essential to provide a dedicated fringe fracture algorithm in order to produce the GDSII data (see Section 12.4.3.3).

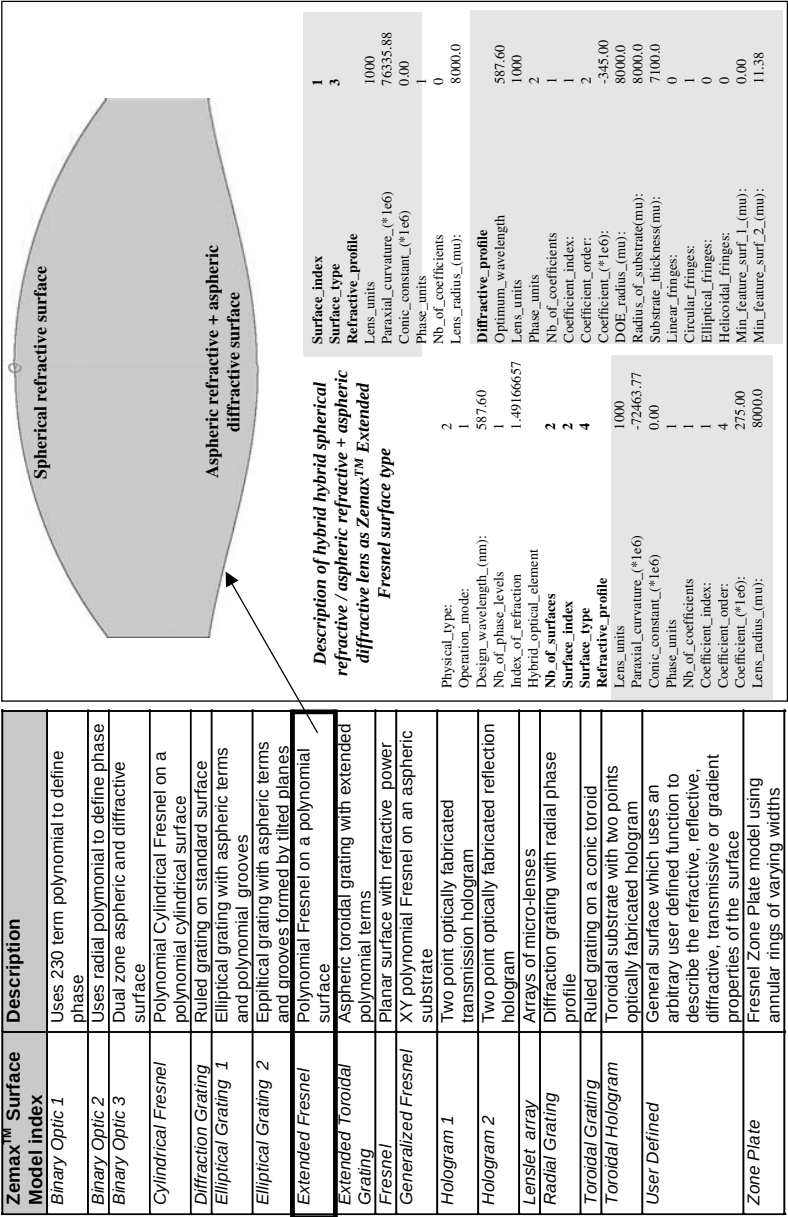


Figure 7.18 Hybrid spherical/aspheric lens description in the Zemax™ CAD tool



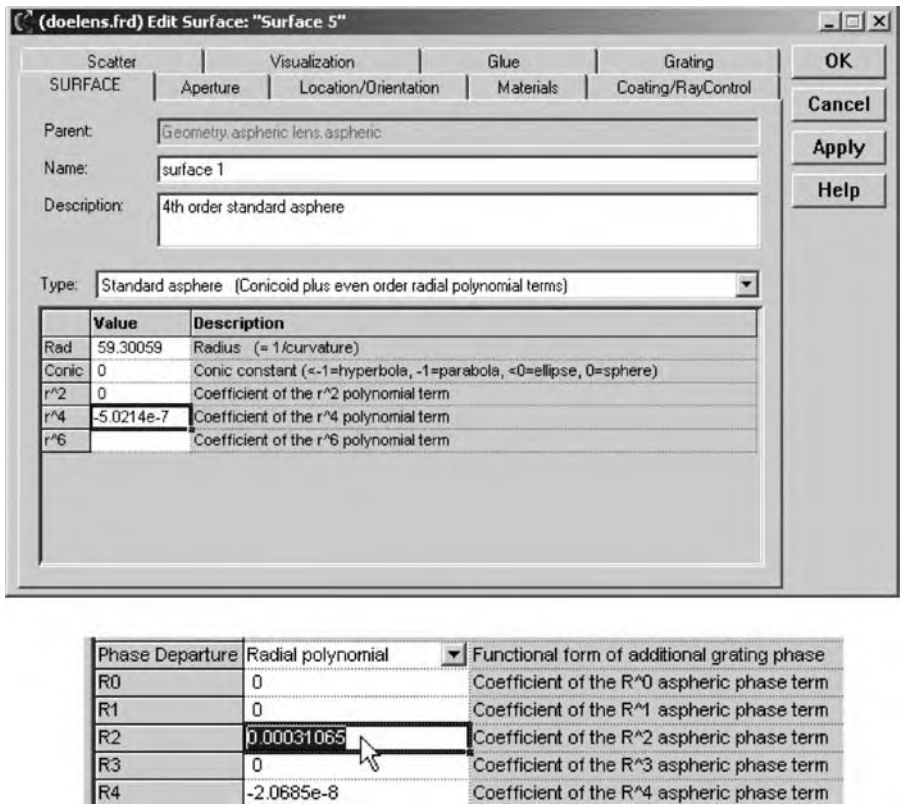


Figure 7.19 Specifying hybrid refractive/diffractive surfaces in optical CAD tools

7.8 A Parametric Design Example of Hybrid Optics via Ray-tracing Techniques

In order to illustrate hybrid diffractive/refractive optics design through ray-tracing techniques with a conventional optical CAD tool and to demonstrate the relative merits of these designs, several representative examples will be shown. The specifications for the design example are as follows:

- Entrance pupil: Diameter 25 mm
- Field of view: On-axis only
- Wavelengths: C (656.3 nm), D (587.6 nm), F (486.1 nm)
- f #: f/10, f/5, f/2.5, f/1.25

Figure 7.20 shows the transverse ray aberration curves for an f/10 hybrid singlet per the above specifications. The substrate material is BK7 glass, and very similar results would arise for acrylic, which would be a fine choice of material if the element were to be mass-produced.

The diffractive surface is located on the second surface of the lens. The spacing or separation between adjacent fringes was allowed to vary with respect to the square as well as the fourth power of the aperture radius, or  $y^2$  and  $y^4$ , where  $y$  is the vertical distance from the vertex of the surface perpendicular to the

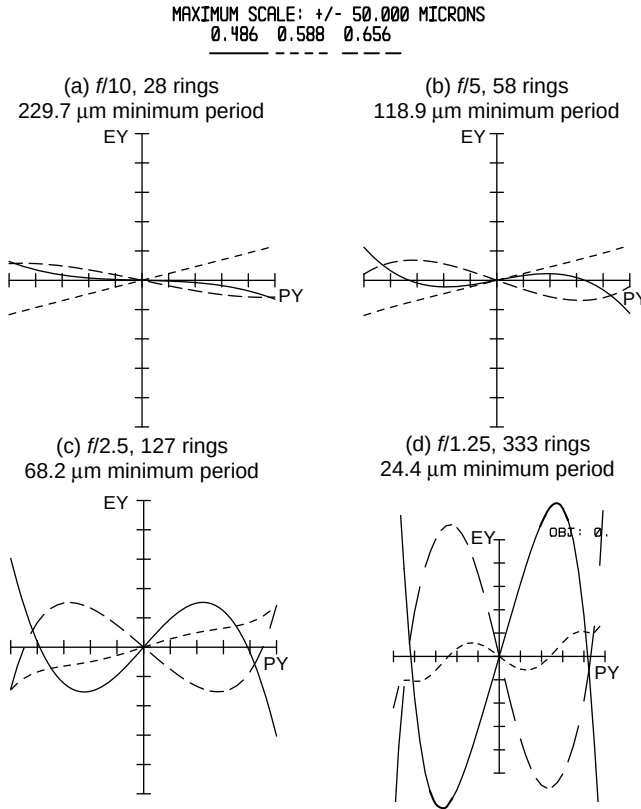


**Figure 7.20** A hybrid refractive/diffractive achromatic singlet as a function of the  $f\#$

optical axis. The quadratic term allows for the correction of the primary axial color, whereby the F and C light (blue and red) are brought to a common focus. The fourth-order term allows correction of the third-order spherical aberration as well. The resulting ray-tracing curves show the classic performance typical of an achromatic doublet. Since the lens is of a relatively high  $f\#$ , the spherical aberration at the central wavelength is fully corrected. There is a residual of spherochromatism, which is the variation of a spherical aberration with wavelength. This residual aberration is due to the fact that the dispersion of the diffractive is linear with wavelength, whereas the material dispersion of BK7 glass is nonlinear. The resulting surface has 28 rings, with a minimum period of 229.7  $\mu\text{m}$ . The insert for injection molding this surface could easily be diamond turned. Note that as the  $f\#$  gets lower and lower, the higher-order spherical aberration increases, and the spherochromatism increases as well, to a point where the spherochromatism is the predominant aberration. The number of rings and the minimum fringe period in the diffractive are listed in Figure 7.21.

Note that these data do not scale directly, as is possible with conventional optical designs. As a diffractive optical element is scaled down in focal length while maintaining its  $f\#$ , a true linear scaling of all parameters (except, of course, the refractive index, which is unitless and thus does not scale) is not correct. This is because it is necessary to maintain the fringe depths to create a total of  $2\pi$  phase shifts for the target wavelength, and a linear scaling would result in only one  $2\pi$  phase shift. Thus, for a 0.5 scaling, this results in one half of the number of fringes with approximately the same minimum fringe period. For example, if the aim is a 12.5 mm diameter  $f/2.5$  hybrid, it is necessary to scale the radii, thickness and

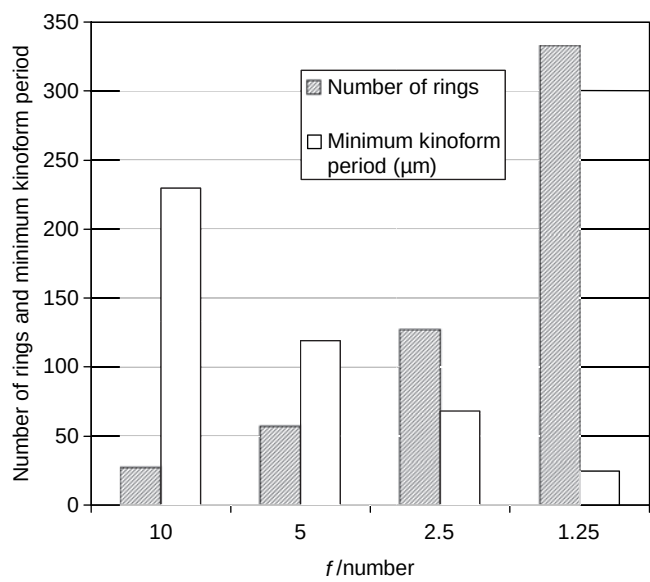


**Figure 7.21** The number of fringes and the minimum period as a function of  $f \#$  for a hybrid singlet

diameter by  $0.5\times$  from the 25 mm starting design. However, the number of fringes will decrease by a factor of 2, with essentially the same minimum fringe period. It is highly recommended to re-optimize any lens or lens system containing one or more diffractive surfaces after scaling, in order to ensure that the surface prescription is correct.

It is feasible to manufacture diffractive surfaces as well as binary surfaces with minimum periods of several microns or less; however, it is best to discuss specific requirements with the foundry prior to finalizing the design. Figure 7.22 shows, for comparison, the performance of classic  $f/10$  and  $f/2.5$  achromatic doublets using BK7 glass. The results are somewhat improved over the hybrid solution. Note

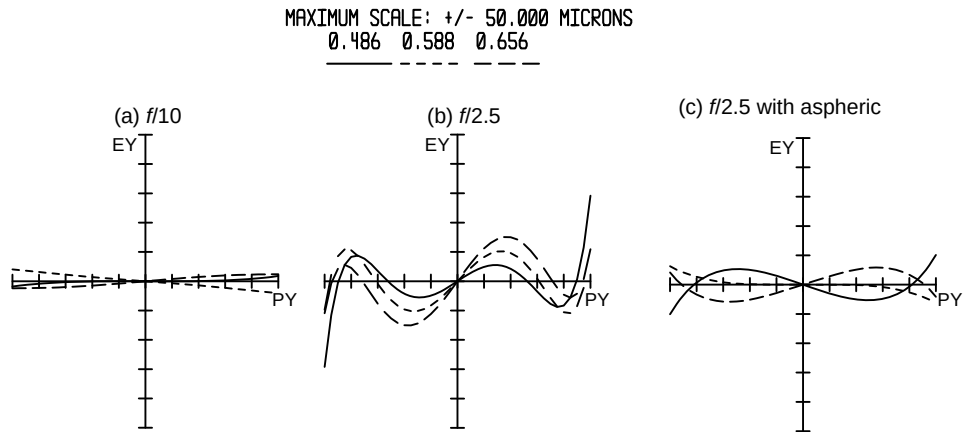


**Figure 7.22** Classic  $f/10$  and  $f/2.5$  achromatic doublets

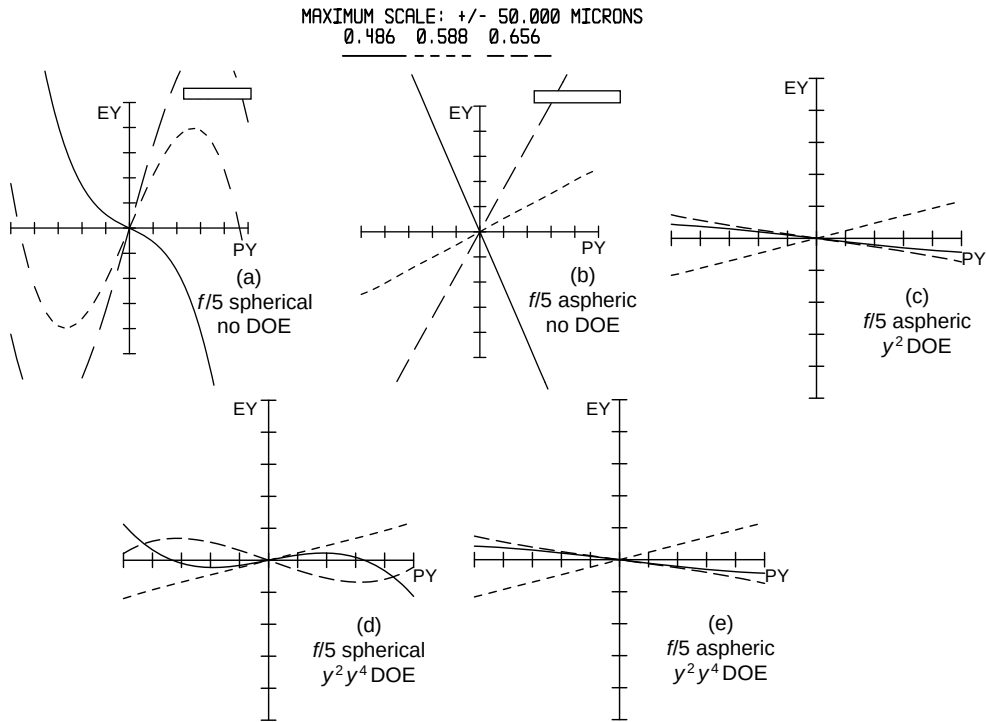


**Figure 7.23** The performance of  $f/5$  hybrid singlets with different surface descriptions

that at the lower  $f\#$  of  $f/2.5$  (Figure 7.22(b)), the spherical aberration of the achromatic doublet is becoming a problem, and an aspheric design is shown in Figure 7.22(c).

Figure 7.23 shows several design scenarios, all for a constant  $f/5$  single-element lens. For reference, Figures 7.23(a) and 7.23(b) both have no diffractive surface; however, Figure 7.23(b) does have an aspheric surface for the correction of spherical aberration. The primary axial color is the same in both lenses, and is quite large as expected. Figure 7.23(c) has an aspheric surface for spherical aberration correction and a quadratic diffractive surface for correction of the primary axial color. Figure 7.23(d) is all spherical, with a quadratic and a fourth-order diffractive fringe width variation. It is interesting that this solution is very similar to Figure 7.23(c), except that this solution has more spherochromatism than the solution with the aspheric surface. Finally, Figure 7.23(e) allows both an asphere as well as a quadratic and fourth-order diffractive fringe width variation. The aspheric along with the quadratic diffractive fringe width variation of Figure 7.23(c) was so well corrected that no further improvement is possible here.

One of the more interesting observations is that the aspheric surface, along with the diffractive surface, allow for the correction of both the spherical aberration as well as the spherochromatism. The all-diffractive surface with the quadratic and the fourth-order fringe width variation has a residual of spherochromatism. The reason for this subtle difference is that in the aspheric case the spherical aberration correction and the chromatic aberration correction are totally separate from one another, thereby allowing better performance. The diffractive designs are more constrained, and do not have sufficient variables to eliminate the spherochromatism as well as the spherical aberration and the primary axial color.

## References

- [1] Y. Sheng, 'Diffractive Optics', short course, Center for Optics, Photonics and Lasers, Department of Physics, Physical Engineering and Optics, Laval University.
- [2] T. Stone and N. George, 'Hybrid diffractive/refractive lenses and achromats', *Applied Optics*, **27**(14), 1988, 2960–2971.
- [3] M.D. Missig and G.M. Morris, 'Diffractive optics applied to eyepiece design', *Applied Optics*, **34**(14), 1995, 2452–2461.
- [4] E. Ibragimov, 'Focusing of ultrashort laser pulses by the combination of diffractive and refractive elements', *Applied Optics*, **34**(31), 1995, 7280–7285.
- [5] R.H. Czichy, D.B. Doyle and J.M. Mayor, 'Hybrid optics for space applications – design, manufacture and testing', in '*Lens and Optical System Design*' H. Zuegge (ed.), Proc. SPIE Vol. **1780**, 1992, 333–344.
- [6] R.L. Roncone and D.W. Sweeney, 'Cancellation of material dispersion in harmonic diffractive lenses', in '*Diffractive and Holographic Optics Technology II*', SPIE Proceedings, I. Cindrich and S.H. Lee (eds), SPIE Press, Bellingham, WA, 1995, 81–88.
- [7] T.D. Milster and R.E. Gerber, 'Compensation of chromatic errors in high N.A. molded objective lenses', *Applied Optics*, **34**(34), 1995, 8079–8080.
- [8] T. Stone and N. George, 'Hybrid diffractive–refractive lenses and achromats', *Applied Optics*, **27**(14), 1988, 2960–2971.
- [9] M.W. Fam and J.W. Goodman, 'Diffractive doublets corrected at two wavelengths', *Journal of the Optical Society of America A*, **8**, 1991, 860.
- [10] W. Li, 'Hybrid diffractive refractive broadband design in visible wavelength region', *Proceedings of SPIE*, **2689**, 1996, 101–110.
- [11] M. Schwab, N. Lindlein, J. Schwider *et al.*, 'Compensation of the wavelength dependence in diffractive star couplers', *Journal of the Optical Society of America A*, **12**(6), 1995, 1290–1297.
- [12] S.M. Ebstein, 'Achromatic diffractive optical elements', in '*Diffractive and Holographic Optics Technology II*', I. Cindrich and S.H. Lee (eds), SPIE Press, Bellingham, WA, 1995, 211–216.
- [13] M.J. Riedl, 'Design example for the use of hybrid optical elements in the infrared', Engineering Laboratory Notes, Optics and Photonics News, May 1996.
- [14] G.P. Behrmann and J. Bowen, 'Color correction in athermalized hybrid lenses', *Optical Society of America Technical Digest*, **9**, 1993, 67–70.
- [15] C. Londoño, W.T. Plummer and P.P. Clark, 'Athermalization of a single-component lens with diffractive optics', *Applied Optics*, **32**(13), 1993, 2295–2302.
- [16] G.P. Behrmann and J. Bowen, 'Influence of temperature on diffractive lens performance', *Applied Optics*, **32**(14), 1993, 2483–2487.
- [17] G.P. Behrmann and J. Bowen, 'Influence of temperature on diffractive lens performance', in *Applied Optics*, **32**, 1993, 8–11.
- [18] J.M. Battiato, R.K. Kostuk and J. Yeh, 'Fabrication of hybrid diffractive optics for fiber interconnects', *IEEE Photonics Technology Letters*, **5**(5), 1993, 563–565.
- [19] M. Gruber, R. Kerssenfischer and J. Jahns, 'Planar-integrated free-space optical fan-out module for MT-connected fiber ribbons', *Journal of Lightwave Technology*, **22**(9), 2004, 2218–2222.
- [20] J. Jahns and B.A. Brumback, 'Advances in the computer aided design of planarized free-space optical circuits: system simulation', in '*Computer and Optically Generated Holographics Optics*', SPIE Vol. 1555, 1991, 2–7.
- [21] J. Jahns, R.A. Morgan, H.N. Nguyen *et al.*, 'Hybrid integration of surface-emitting microlaser chip and planar optics substrate for interconnection applications', *IEEE Photonics Technology Letters*, **4**, 1992, 1369–1372.
- [22] M. Jarczyński, T. Seiler and J. Jahns, 'Integrated three-dimensional optical multilayer using free-space optics', *Applied Optics*, **45**(25), 6335–6341.



# 8

## Digital Holographic Optics

In the realm of optics, a hologram seems at first glance to be the most analog type of element, after the conventional smooth surface refractive lens.

Holography, in its original form, is truly an analog phenomenon and an analog optical element, in its design (the use of analytic techniques) and in its recording (the use of conventional optics) as well as in its internal form (an analogous modulation of refractive indices).

This said, the holograms used in industry today (including HOEs) can be classified as digital elements because of the many improvements to their original nature. These improvements are digital in nature and are as follows:

- the wavefront to be recorded in the hologram calculated by a digital computer rather than produced by real analog optical elements;
- the recording of holograms (Bragg gratings) by the use of digital phase masks;
- the recording of HOEs by the use of digital CGHs (beam splitting, beam shaping etc.);
- the recording of pixelated holograms (for display or optical processing applications); and
- the recording of HOEs into photoresist (rather than analog holographic emulsions) to produce surface-relief elements, and then replicating them on wafers, similar to digital lithography.

Other than for traditional 3D display holograms, which were every kid's (and also most grown-ups') favorite eye candy, today industrial holography is increasingly becoming a digital process, for the integration of complex optical functionality in numerous products (consumer electronics and others), and has slowly taken the place of the more traditional analog processes.

The optical functionality of digital diffractive elements (DOEs or CGHs), phase masks and other micro-optical elements designed and fabricated by microlithography can be recorded in HOEs. Today, even 3D display holograms are much more complex than the original holograms. In this chapter, it is shown that holograms are becoming more pixelated and synthetic and, therefore, the object no longer needs to exist physically (only as a 3D object in digital form).

### 8.1 Conventional Holography

It was Dennis Gabor, in 1948 [1], who first introduced the term 'holography' to the scientific community as part of his work in electron microscopy, well before the invention of the laser. His in-line holographic technique was able to record mainly phase objects, and reproduce them in an on-axis geometry, therefore limiting its application to a small number of very particular cases.

In 1962, Emmett Leith and Juris Upatnieks, at the University of Michigan, used laser light (an early HeNe laser) in holography for the first time, and introduced the concept of off-axis recording, which actually enabled modern holography. Leith and Upatnieks' off-axis recording consisted of splitting the laser beam into two beams and making them interfere after hitting the opaque object. This had a tremendous impact on modern holography, by fixing the problems linked to the Gabor hologram, for the following reasons:

- spatial differentiation of multiple images (or diffraction orders);
- the possibility of recording amplitude objects (reflective objects); and
- the possibility of locating the on-axis beam (the zero order) outside the image areas.

However, off-axis illumination also had some stringent requirements, such as very high plate resolution and therefore very good vibration stability during the recording process (this is why the optical table is usually the most expensive component in a holography lab).

### 8.1.1 *Photography and Holography*

At first glance, a hologram could be compared to a high-resolution photograph. However, there are important differences between these two techniques:

- A photograph is a 2D image of a 3D scene and lacks depth perception or parallax.
- A photographic emulsion is two-dimensional, a holographic emulsion is three-dimensional (note that a diffractive surface can also be two dimensional).
- A hologram is a 3D picture of an amplitude or 'transparent' phase object. Photographs are only pictures of real amplitude objects.
- A photographic film is not sensitive to phase, but only to radiant energy; therefore, the phase relation is lost, and most of the information coming from the object is not recorded in the photograph. The phase relation is recorded in the hologram as an interference pattern.
- A photographic film has a typical grain resolution of a few microns, whereas a hologram requires resolution grains of the order of 30 nm (a grain size that is about 30 000 times smaller in volume than that of the photographic grain!).

So, since holograms, as opposed to photographs, record most of the information coming from the object, their name – which was derived from the Greek *holos*, meaning 'the whole message' – is well deserved.

Now, when reading the previously listed unique features of a hologram (and, for that matter, diffractives – see the previous chapters), it is stunning to see graduate students and novices in the field of holography thinking about it as something that required a quantum leap in technology and processing compared with photography. To prove them wrong and show that there is a clear smooth transition between a photograph and a hologram, a simple crude hologram in a photographic film (preferably on a conventional  $24 \times 36$  mm slide), or a printout of simple diffractive optics (gratings, CGHs or DOEs) using a \$30 black and white 600 dpi inkjet printer on a transparency film, can be recorded, or rather printed in this case. Although such elements are truly holograms and diffractives in their nature, they yield very low diffraction angles, very low efficiency and so on.

### 8.1.2 *Recording a Hologram and Playing it Back*

A traditional (Leith and Upatnieks) holographic recording consists of creating an interference pattern (in other words, a 3D fringe distribution) between two beams in a photosensitive material: one of the beams is called the 'reference beam' and the other the 'object' beam [2, 3]. The photosensitive material can be planar or 3D (thick planar), phase, amplitude or a combination thereof (usually, a highly efficient hologram is recorded in a thick 3D phase material – i.e. a material the refractive index of which varies as a function of the intensity of light); in essence, a hologram takes a 3D phase picture of the fringe distribution.



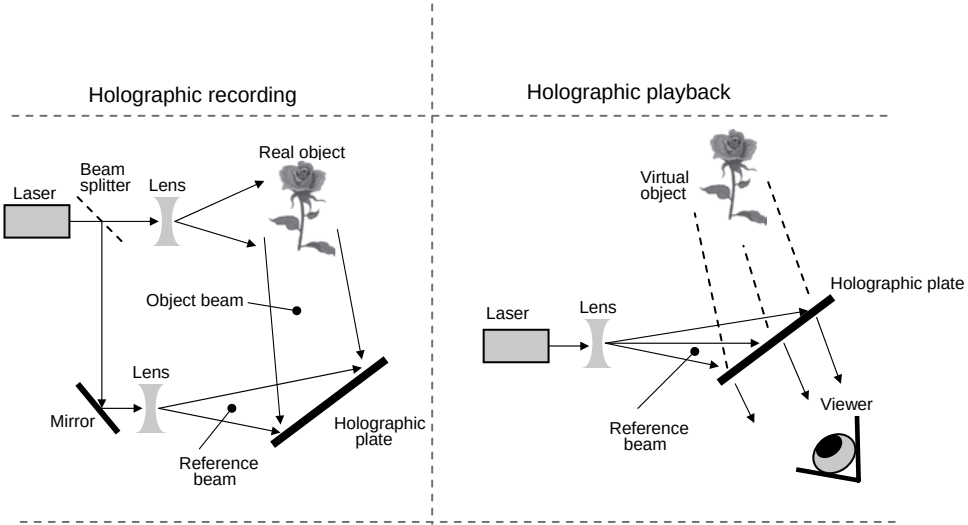


**Figure 8.1** Conventional transmission hologram recording and playback

As its name indicates, the object beam (or wavefront) carries all the information about the object (or signal) that is recorded in the hologram (e.g. a 3D object). The reference beam (or wavefront) is used to create the interference pattern with the object beam, and perhaps give some other effect (e.g. a lensing effect, an off-axis effect etc.). Note that the same (or a similar) reference wavefront has to be used when playing back the hologram after recording.

In order to create an interference pattern, it is always good to have coherent beams of light in the first place. An easy way to produce two coherent beams of light in phase with each other is to use a laser and a beam splitter, as described in Figure 8.1, which also shows that the holographic playback process uses the same reference beam as used for the recording process.

Note that there can be multiple reconstruction orders (multiple images) and stray light (zero order), which are not depicted in Figure 8.1. The off-axis illumination parameters can be tuned so that the viewer only sees one single image, and no zero order.

The previous section showed that holography is a two-step process:

1. Recording through the interference phenomenon – an interference pattern is created by two or more wavefronts.
2. Playback through the diffraction phenomenon – a diffraction pattern is created by light passing through the recorded interference pattern (to recreate the original wavefront).

If the reference and object beams are produced by a point source as depicted in Figure 8.1, the complex amplitudes for both beams can be written as

$$\begin{cases} R(x, y, z) = r(x, y, z)e^{i\phi(x, y, z)} = A_R \cdot e^{ikz} \\ O(x, y, z) = o(x, y, z)e^{i\phi(x, y, z)} = A_0 \cdot e^{ikr} \end{cases} \quad \text{where } r = \sqrt{x^2 + y^2 + z^2} \quad (8.1)$$

Thus, the intensity pattern arising from the interference between these two beams is

$$I(x, y) = |O + R|^2 = OO^* + RR^* + OR^* + O^*R \quad (8.2)$$

When this intensity pattern is recorded on a photographic (or holographic) plate, it gives rise to a transmittance pattern  $T$  that is dependent of the intensity pattern on the plate:

$$T = C + \zeta(|O|^2 + O^*R + OR^*) \quad (8.3)$$

where  $C$  is related to the background amplitude term  $R$ , and  $\zeta$  is a parameter of the holographic recording and development process ( $\zeta$  can be a complex number).

In the holographic plate, this intensity pattern can yield a real gray-scale amplitude pattern (a photograph or an amplitude hologram), or a pure phase pattern (a pure refractive index change or phase hologram) or a combination of both. It is obvious that a pure-phase hologram (with no absorption effect) will give rise to the highest efficiency. This is why diffractives are usually fabricated in pure-phase substrates (quartz, glass, plastic etc.) with the notable exception of the amplitude Fresnel Zone Plate (which remains an academic example, since its efficiency is very low: see Chapter 6).

Chapter 6 describes the interferogram-type diffractive element, which is in essence a synthetic hologram, since it is calculated and fabricated according to same Equation (8.3) as derived here.

When illuminated by wavefront  $R'$ , which is similar to the reference wavefront  $R$  used in the recording process, the reconstructed wavefront gives rise to the following equation, which is a superposition of four wavefronts:

$$R' \cdot T = T \cdot R' + \zeta(OO^*R' + O^*RR' + OR^*R') \quad (8.4)$$

If the illumination wavefront is identical to the initial reference wavefront, Equation (8.4) can be rewritten as

$$R \cdot T = R \cdot (T + \zeta OO^*) + \zeta \cdot R^2 O^* + \zeta \cdot |R|^2 O \quad (8.5)$$

There are three terms in the reconstructed wavefront described in Equation (8.5):

- the left-hand term, the direct wave – identical to the reference wave with an intensity change (zero order);
- the middle term, the conjugate wave – the complex conjugate of the object wave  $O$  (virtual object); and
- right-hand term, the object wave – identical to the object wave with an intensity change (real object).

Figure 8.2 depicts the three different terms described in Equation (8.5), in the simple case of a spherical object wave recording (i.e. a holographic on-axis lens).

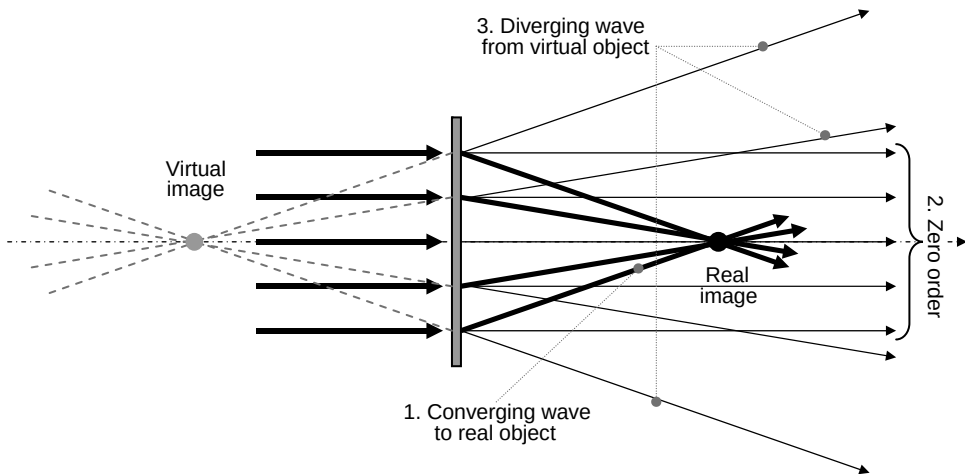


**Figure 8.2** The holographic reconstruction process

The direct wave is the zero order (light that is not affected by the hologram), usually also called the *DC light*. The fundamental negative and positive orders are conjugate diffractive orders (carrying the signal information, and therefore sometimes called the *AC light*), defining the real object and the virtual objects, respectively.

## 8.2 Different Types of Holograms

As for diffractives (see Chapters 5 and 6), the hologram geometry can be optimized in order to push the maximum amount of light into one of these orders, usually the fundamental negative or positive order (virtual or real image), therefore creating a highly efficient hologram. In many cases, an optimal hologram can also consist of a multitude of orders, and even a well-behaved zero order: it all depends on the target application (which is very seldom a display hologram in industrial applications).

### 8.2.1 On- and Off-axis Holograms

In many applications, if there is no way to reduce the energy in the conjugate or zero order, the recording process can be configured in an off-axis mode, therefore spatially demultiplexing the different resulting beams, as shown in Figure 8.3 (two viewers, each seeing a different object, and neither of them bothered by DC light).

The early holograms of Denis Gabor were on-axis holograms (and mainly of phase objects); modern holograms are very often recorded in off-axis configuration.

### 8.2.2 Fraunhofer and Fresnel Holograms

In the previous section, the use of point source holograms to generate lens functionality was discussed. Now consider a hologram that represents an image (2D or 3D), recorded as depicted in Figure 8.1. There are two ways to record such a hologram: as a Fraunhofer (or Fourier) hologram (where only plane waves are used in both beams) or as a Fresnel hologram, where diverging or converging beams are used. The simplest Fraunhofer hologram is a linear grating. The simplest Fresnel hologram is a (Gabor) holographic lens. Figure 8.4 shows the reconstruction geometry for both configurations.

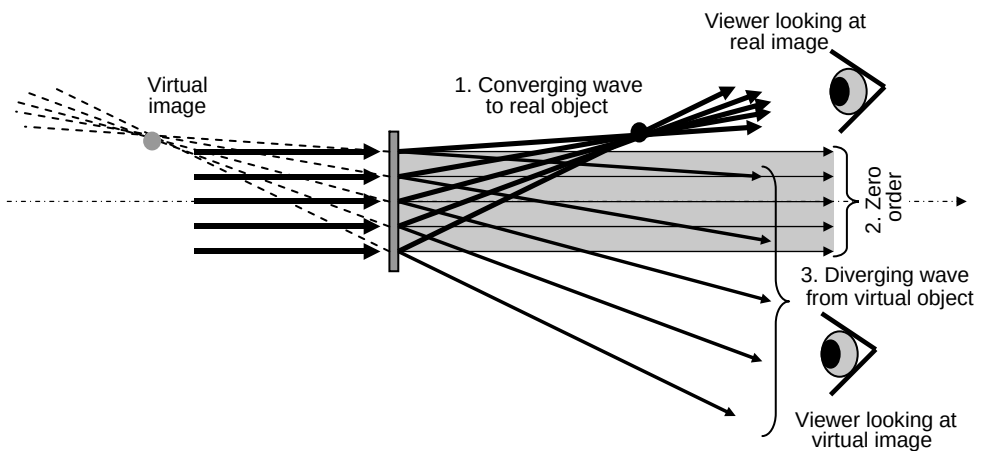


**Figure 8.3** Off-axis holographic configuration

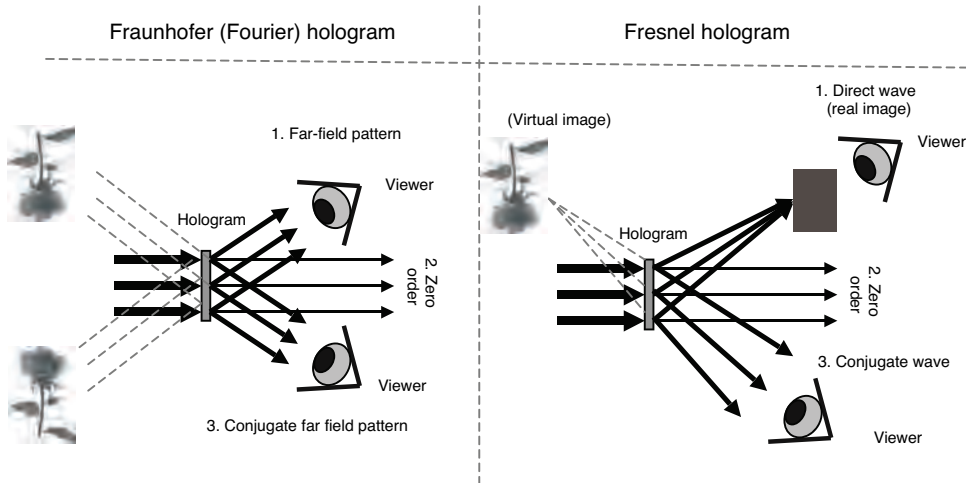


**Figure 8.4** Reconstructions of Fraunhofer and Fresnel holograms

The Fraunhofer hologram reconstructs the pattern in the far field, so the viewer looks at an image that is located at infinity, either through the element, or after diffused reflection off a plane (a wall or paper). The image in a Fraunhofer hologram is mainly two-dimensional, or 2.5D (in the case of holographic stereograms). Since both images are at infinity, both are at the same time real and virtual. Note that the conjugate image is flipped  $180^\circ$  with regard to the direct image.

In a Fresnel hologram, the viewer can look at either the real image formed after the hologram (the direct wave) or the virtual image that floats behind the hologram. These images can be truly 3D. Usually, in a traditional display hologram, the aim is to push all of the energy into the virtual image (the floating object behind the plate).

In both cases, the off-axis configuration helps to spatially demultiplex the three beams, in order to make clean images appear. In most cases, there are more images appearing (higher orders, or ghosts, which are not discussed here – for more insight for higher orders for both Fraunhofer and Fresnel synthetic holograms, see Chapters 5 and 6).

### 8.2.3 Thin and Thick Holograms

Before considering HOE technology in more detail, clarity is needed about the use of the terms ‘diffractive’ and ‘holographic’, since these two words seem to point to the same optical effect (namely the diffraction of light through phase microstructures). Diffractive elements are usually considered as ‘thin holograms’, whereas HOEs are considered as ‘thick holograms’.

A holographic grating is considered as ‘thick’ or ‘thin’ in Bragg incidence when its quality factor  $Q$  is larger than 10 or lower than unity, respectively (a definition that is usually agreed on in the literature). For values of  $Q$  in between 1 and 10, the grating behaves in an intermediate state.  $Q$  is defined as follows:

$$Q = \frac{2\pi\lambda d}{n\Lambda^2 \cos\alpha} \quad (8.6)$$

where  $d$  is the thickness of the grating,  $\Lambda$  is its period,  $n$  is its average refractive index and  $\alpha$  is the incident angle.

Volume holograms, or ‘thick gratings’, have their own advantages and specifications, as do ‘thin holograms’.

A thin hologram is usually a surface-relief element, fabricated by diamond turning or lithography (see Chapters 5 and 6), with an aspect ratio of the structures that can vary from 0.1 to 5. Usually, lateral structures (minimal local periods in the thin hologram) are large when compared to volume holograms (from several tens of microns to about twice the size of the reconstruction wavelength). Therefore, thin holograms have diffraction angles that are seldom larger than 20°. Such elements can work either in transmission (when fabricated in a transparent material) or in reflection (when fabricated in a reflective material).

A thick hologram is usually a volume hologram; that is, a holographic emulsion in which a refractive index modulation is created (silver halide, dichromated gelatin or DCG, photopolymers, photo-refractive materials – Acousto-Optical (AO) Bragg gratings, Holographic Polymer Dispersed Crystals (H-PDLCs) – and other more exotic materials, as discussed at the end of this chapter). The typical minimal periods are much smaller than for thin holograms, typically on the order of the wavelength. Each can work either in transmission mode or reflection mode, although the material itself is always transparent (as opposed to thin holograms).

8.2.4 Transmission and Reflection Holograms

As the reference and object beam sources can be located on either side of the holographic plate, there are numerous potential recording architectures, which yield different types of holograms (reflection or transmission). Figure 8.5 summarizes these various recording architectures.

In a reflection hologram, the Bragg planes are almost parallel to the substrate, whereas in a transmission hologram the Bragg planes are slightly tilted, and can be almost orthogonal to the substrate. This is one reason why it is nearly impossible to fabricate a reflective hologram as a surface-relief hologram (a thin hologram).

Figure 8.6 shows the internal Bragg layers of typical reflection and transmission holograms.

| Point sources |         | Optical recording set-up | Type         | Optical reconstruction set-up |
|---------------|---------|--------------------------|--------------|-------------------------------|
| s1            | Real    |                          | Reflective   |                               |
| s2            | Real    |                          |              |                               |
| s1            | Real    |                          | Transmissive |                               |
| s2            | Real    |                          |              |                               |
| s1            | Real    |                          | Transmissive |                               |
| s2            | Virtual |                          |              |                               |
| s1            | Virtual |                          | Transmissive |                               |
| s2            | Virtual |                          |              |                               |
| s1            | Virtual |                          | Reflective   |                               |
| s2            | Virtual |                          |              |                               |

Figure 8.5 Recording geometries for transmission or reflection holograms

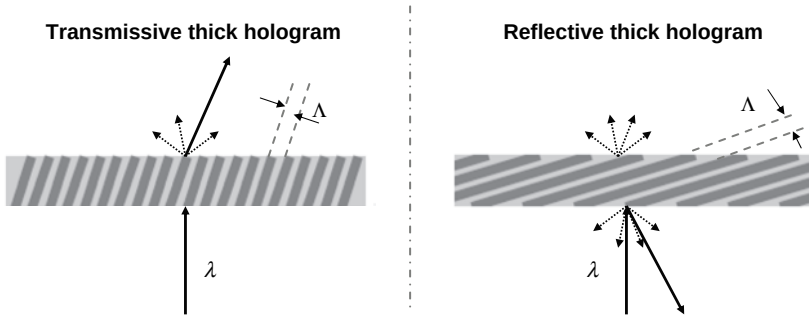


**Figure 8.6** Bragg planes in reflection and transmission holograms

### 8.2.5 Lippmann Holograms

Gabriel Lippmann (1845–1921) was able to record the first color photographs way before the invention of color photography, by recording photographic interferograms (also called interference color photography). The photographic plate produces standing waves that record the different colors. Basically, a Lippmann photograph (or hologram) can be understood as a volume hologram in which the Bragg planes are perfectly parallel to the plate, therefore creating strong interference effects (see Figure 8.7). Lippmann holograms are therefore reflection holograms.

Mercury is used as a reflector in Figure 8.7. A Lippmann photograph can be described as a reflection hologram that has been recorded with very short temporal coherence (white light).

## 8.3 Unique Features of Holograms

In the previous section, the various type of holograms used in industry today (thin or thick holograms, Fraunhofer or Fresnel holograms, and transmission or reflection mode holograms) were discussed. Now, the unique optical specifications and features of such holograms, and why they can provide optical solutions where no other optical element can, will be reviewed.

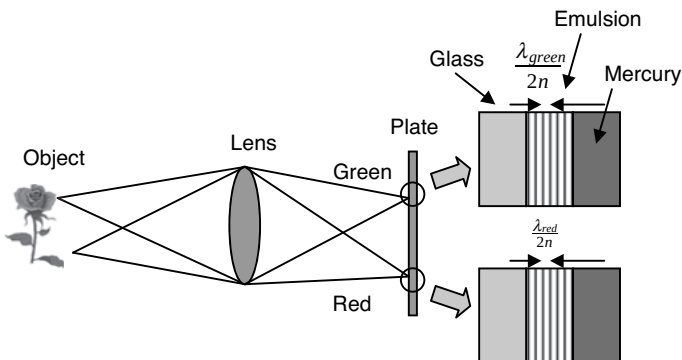


**Figure 8.7** Lippmann photography

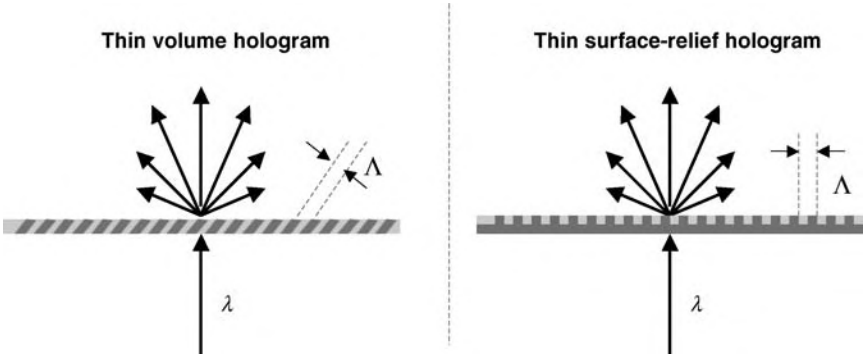


**Figure 8.8** Diffraction orders from thin and thick holograms

8.3.1 Diffraction Orders

Thick holograms (either in reflection or transmission modes) have large  $\lambda/\Lambda$  ratios, in which the incident beam passes through many Bragg planes, thus yielding strong constructive or destructive interference. In thin holograms, there is only one phase shift that is imprinted on the incoming wave, and no Bragg planes. The optimum phase shift for a thin binary hologram is  $\pi$ , and for an thin analog surface-relief hologram this phase shift is  $2\pi$ ). Thin holograms, either synthetic or recorded (especially in their binary form), tend to diffract into numerous orders (see Chapter 5), whereas thick holograms yield mainly one single diffraction order. Figure 8.8 shows the diffraction orders for a thin binary element and for a thick hologram in transmission.

In order to derive the various angles at which the higher orders will appear, consider a one-dimensional thin transmission Fraunhofer hologram recorded as a linear sinusoidal grating of period  $\Lambda$ . An incident plane wave hits this grating at an angle  $\alpha$  (see Figure 8.9).

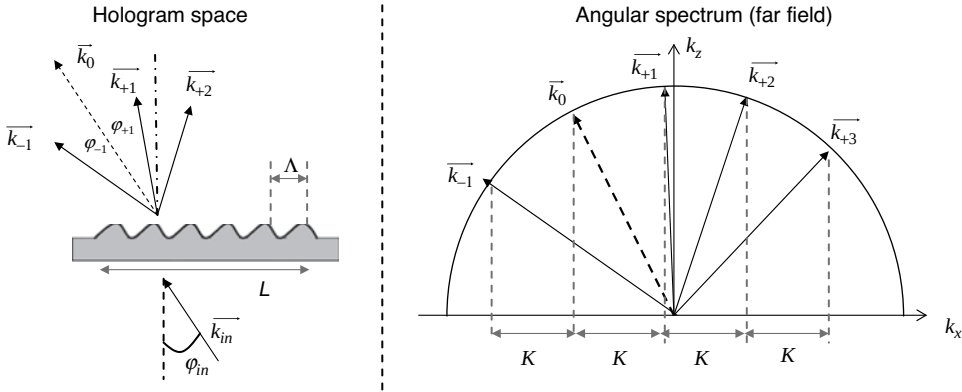


**Figure 8.9** A sinusoidal grating and the resulting Fourier space projection of propagation vectors

The incoming field over a grating length  $L$  just before exiting the linear sinusoidal grating can be written as follows (see also Section 5.3):

$$U(x) = U_{in}(x) \cdot H(x) = U_{in}(x) \cdot \text{rect}\left(\frac{x}{L}\right) \cdot e^{-jk_0 n \frac{d}{2} \sin(Kx)} = U_{in}(x) \cdot \text{rect}\left(\frac{x}{L}\right) \cdot \sum_{m=-\infty}^{+\infty} J_m\left(k_0 n \frac{d}{2}\right) \cdot e^{-jmKx} \quad (8.7)$$

The far-field pattern of such a Fraunhofer hologram is the Fourier transform of the incoming field, and can be expressed as follows:

$$\begin{aligned} U'(k_x) &= FT[U_{in}(x) \cdot H(x)] = U_{in}(k_x) \cdot A_{in} \cdot \delta(k_x - k_{in}) * \sum_{m=-\infty}^{+\infty} J_m\left(k_0 n \frac{d}{2}\right) \cdot \delta(k_x - mK) \\ &= A_{in} \sum_{m=-\infty}^{+\infty} J_m\left(k_0 n \frac{d}{2}\right) \cdot \text{sinc}\left(\frac{L}{2}(k_x - k_{in} - mK)\right) \end{aligned} \quad (8.8)$$

where  $K$  is the grating vector  $2\pi/\Lambda$ . The conservation of the transverse component of the propagation vector  $k$  (see the Fourier space projection of the propagation vectors in Figure 8.8) gives rise to the following relation, which gives the angles of the various diffracted orders:

$$\begin{aligned} k_m &= \frac{2\pi}{\lambda} \sin(\varphi_m) = k_{in} + mK = \frac{2\pi}{\lambda} \sin(\varphi_{in}) + \frac{2\pi m}{\Lambda} \\ \Rightarrow \sin(\varphi_m) &= \sin(\varphi_{in}) + m \frac{\lambda}{\Lambda} \end{aligned} \quad (8.9)$$

where  $\varphi_m$  is the diffraction angle for order  $m$ .

### 8.3.2 Spectral Dispersion

Spectral dispersion, along with angular selectivity (see the next section) is one of the key features of holograms, and the main reason why holograms are used today in numerous applications other than display ( $\lambda$  Mux/Demux, optical page data storage, etc.).

The resolving power of thin grating holograms has been derived in Section 5.3. However, a volume grating can have a much higher resolving power than thin gratings, since the line spacing (the grating pitch) is usually much greater and the efficiency can also be quite high in the fundamental order. It is, however, interesting to note that in most practical spectral dispersion cases, rather than choosing to use a thick volume grating, a thin reflective grating is often used (e.g. a ruled grating). Thin reflective gratings with groove depths that are optimized to diffract in higher orders produce high dispersion and have strong resolving powers. Besides, they have longer lifetimes than volume holograms, especially in reflective modes (see DWDM Demux gratings, spectroscopic gratings etc.).

Figure 8.10 summarizes the Rayleigh resolvability criterion (the zero of the spectral channel has to at least coincide with the maximum of the next spectral channel).

#### 8.3.2.1 Free Spectral Range

Another interesting feature of a hologram (thin or thick) is that two different spectral channels (two different colors) can be diffracted in the same direction. This is made possible by tuning a higher diffraction order for a shorter wavelength to overlap in an angular fashion with a lower diffraction order for a larger wavelength (see Figure 8.11). In other words,  $\lambda$  is diffracted in the fundamental order in the same direction as  $\lambda/2$  in the second order and  $\lambda/3$  in the third order.

The free spectral range of a grating is the largest wavelength interval in a given order that does not overlap the same interval in an adjacent order. If  $\lambda_1$  is the shortest wavelength and  $\lambda_2$  is the longest



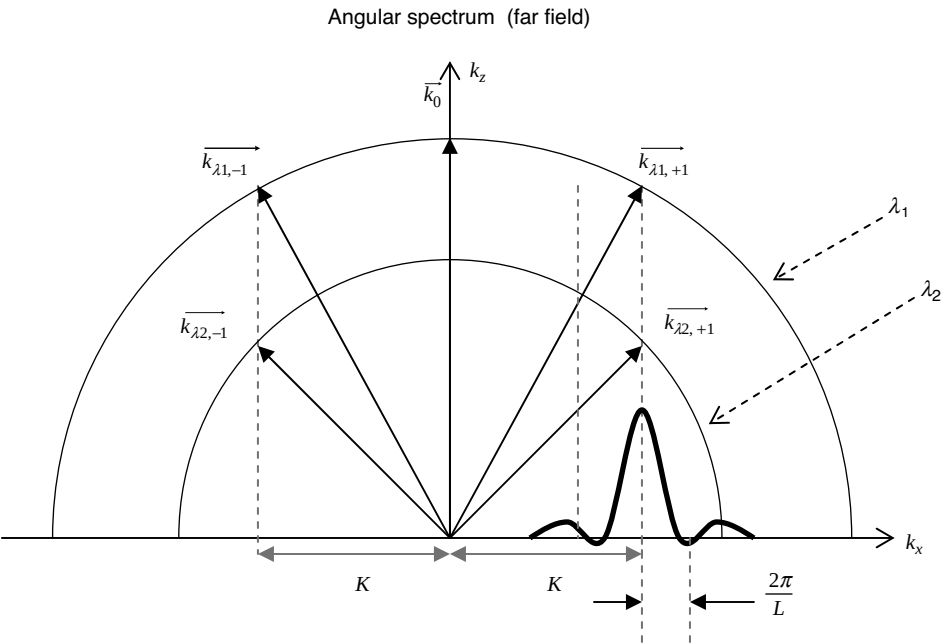


**Figure 8.10** The Rayleigh spectral resolvability criterion

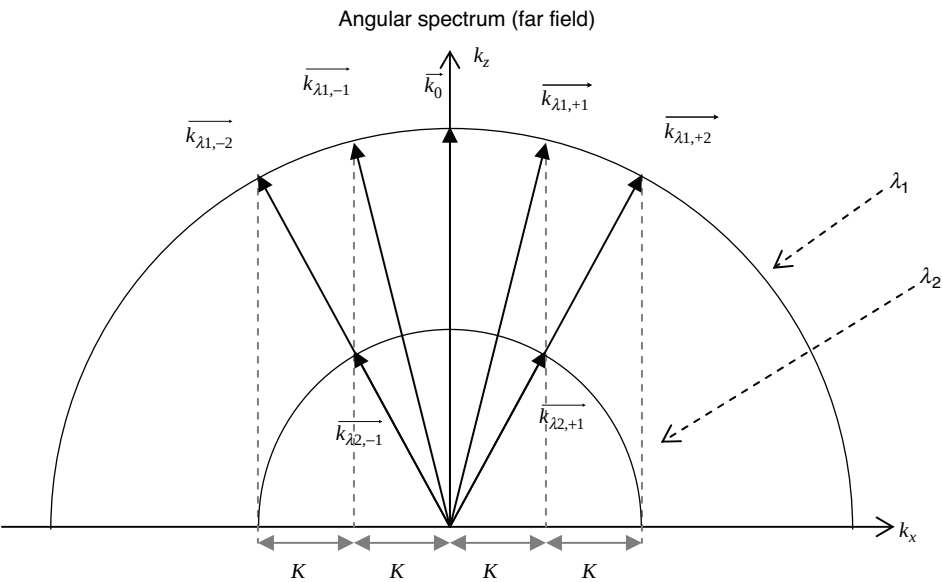


**Figure 8.11** The superposition of different spectral channels by using multiple orders

wavelength in this wavelength interval, then the free spectral range may be expressed as

$$\text{free spectral range} = \lambda_2 - \lambda_1 = \frac{\lambda_2}{m+1} \quad (8.10)$$

### 8.3.3 Diffraction Efficiency and Angular Selectivity

Thick holograms in Bragg incidence can routinely achieve almost 100% diffraction efficiency. However, this is not an easy task. The efficiency depends on many variables, such as:

- the holographic material specifications ( $\Delta n$ , thickness, etc.);
- the exposure (vibrations);
- the smallest and largest fringe widths;
- the exposure beam ratio;
- the coherence of the source;
- how well the Bragg condition is met by the reconstruction wave; and
- temperature, humidity and so on.

The maximum diffraction efficiency (which means reducing the power in the DC light and the higher orders to nearly zero) is very difficult to achieve for a thin hologram [4], especially when fabricated by microlithography, owing to the limitations of these technologies, as seen in the fabrication chapters. As seen previously, a high diffraction efficiency in thick holograms also means a high angular selectivity (the efficiency decreases rapidly when moving away from the Bragg angle). Therefore, the diffraction efficiency is strongly related to the angular selectivity in volume holograms (thick holograms).

In the next section, a modeling technique is derived that allows the quantification of these two aspects: diffraction efficiency as a function of the incoming wavelength and angle.

## 8.4 Modeling the Behavior of Volume Holograms

In order to derive an expression for the diffraction efficiency as a function of the hologram parameters, for thick transmission and reflection holograms, consider the modal theory approach, and more specifically coupled wave theory (see also Chapter 10).

### 8.4.1 Coupled Wave Theory

The modal approach has given rise to Rigorous Coupled Wave Analysis (RCWA), which is ideal for volume holograms with simple fringe geometries. A simplification of coupled wave theory is two-wave coupled wave theory, which only considers the coupling effects between the zero and the fundamental orders. Kogelnik's theory is based on the two-wave coupled wave theory.

### 8.4.2 Kogelnik's Model

Kogelnik's two-wave coupled theory [5] was developed in 1969 at Bell Laboratories, and supposes an incident plane wave that is s polarized. This model gives the best results when in Bragg incidence mode and for a single diffraction order. Any order other than 0 or 1 in reflection or transmission will be considered as evanescent.

Kogelnik's model is a simple solution that is valid for sinusoidal index modulations and slant Bragg planes (perfect for volume holograms) and works either for the transmission or the reflection mode. The limitations of Kogelnik's model are that it is only valid for small index variations, for near Bragg

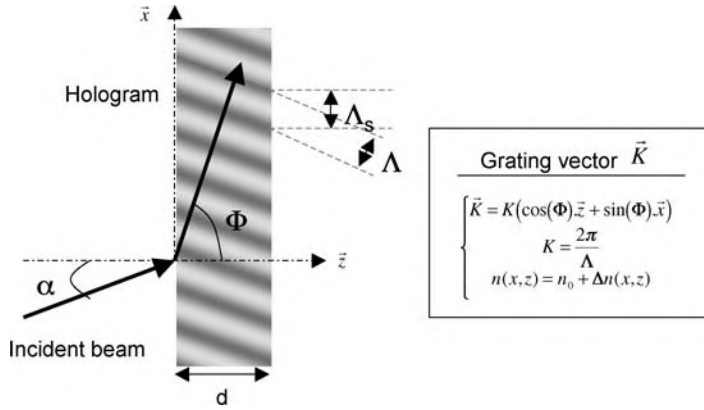


**Figure 8.12** A volume Bragg grating hologram and the grating vector

incidence, and that boundary reflections are not considered. Figure 8.12 shows the grating to which we are considering applying Kogelnik's model.

Based on Figure 8.1, let us derive first the expressions for the fringe slant angle  $\Phi$  and the fringe period  $\Lambda$ , as well as the surface fringe period spacing  $\Lambda_s$ , for a volume hologram:

$$\begin{cases} \Phi = \pi/2 + (\theta_r - \theta_o)/2 \\ \Lambda = \lambda_0 / (2n_0 |\cos(\Phi - \theta_r)|) \\ \Lambda_s = \Lambda / \sin(\Phi) \end{cases} \quad (8.11)$$

where  $\theta_r$  and  $\theta_o$  are, respectively, the object beam free-space angle and the reference beam free-space angle in the recording set-up (see Figure 8.1).

The field expression for both the object ( $O(x)$ ) and reference beams ( $R(x)$ ) can be expressed in the coupled form as follows (see Appendix A and Figure 8.1):

$$\begin{cases} \frac{\partial R(x)}{\partial x} = -j\kappa \cdot e^{-j\zeta \cdot x} \cdot O(x) \\ \frac{\partial O(x)}{\partial x} = -j\kappa \cdot e^{j\zeta \cdot x} \cdot R(x) \end{cases} \quad (8.12)$$

The grating strength  $\kappa$  and the detuning parameter  $\zeta$  of the hologram are defined in the next section. Note that the object beam is actually the diffracted beam when the reference beam is in Bragg incidence.

### 8.4.3 Grating Strength and Detuning Parameter

Before deriving any diffraction expression from the previous coupled equations, a few important hologram parameters need to be defined. The first parameter is the grating strength parameter  $\nu_s$ :

$$\nu_s = \frac{\pi \cdot \Delta n \cdot d}{\lambda \sqrt{c_r c_s}} \quad (8.13)$$

The grating strength parameter is a function of the refractive index modulation amplitude  $\Delta n$ . It is therefore a parameter linked to the holographic material itself. For more insight into  $\Delta n$ , see the next

section. The other two parameters used to define the grating strength are  $c_r$  and  $c_s$ :

$$\begin{cases} c_s = \cos(\alpha_i) - \left(\frac{\lambda}{n_0\Lambda}\right)\cos(\Phi) \\ c_r = \cos(\alpha_i) \end{cases} \quad (8.14)$$

Parameters  $c_r$  and  $c_s$  are both functions of the incident angle (obliquity factors).  $c_s$  is also a function of the recording parameters.

Another parameter that is used in Kogelnik's theory (see also Equation (8.12)) is the hologram detuning parameter  $\zeta$ :

$$\zeta = \left(\frac{Kd}{2c_s}\right) \left( |\cos(\Phi - \alpha_i)| - \frac{K\lambda}{4\pi n_0} \right) \quad (8.15)$$

The detuning parameter (sometimes also called the dephasing parameter) is a function of the recording geometry as well as the illumination geometry. When this detuning parameter is null (i.e. as in Equation (8.12)), the Bragg regime is in operation (i.e. creating maximum local constructive interference and efficiency, maximized for a given grating strength):

$$\cos|\Phi - \alpha_{Bragg}| = \frac{\lambda_{Bragg}}{2n_0\Lambda} \quad (8.16)$$

When plotting the amplitudes of the object beam  $O(x)$  and reference beam  $R(x)$  as a function of the grating strength factor  $k$ , for a detuning parameter nearing zero, a set of sinusoidal curves in anti-phase operation can be obtained, as depicted in Figure (8.13).

When in the Bragg condition (i.e. detuning parameter  $\zeta = 0$ ), the coupling between the two beams is very strong, as one can see from Figure 8.12, and the diffraction efficiency can theoretically reach 100%.

When outside the Bragg condition ( $\zeta \neq 0$ ), the coupling is weaker and the efficiency is reduced. The grating strength  $\kappa$ , which in other words describes the amplitude of the index modulation in the holographic media, provides maximum efficiency for large values, as shown in Figure 8.13.

The Bragg configuration (Floquet's theorem condition satisfied or null detuning parameter) and Bragg detuning are depicted graphically in Figure 8.14.

where  $\vec{O}$  is the object vector,  $\vec{R}$  is the reference vector and  $\vec{K}$  is the grating vector.

Kogelnik's theory does not consider any other derivatives of the coupled wave equations (Equation (8.12)), so no other orders are present in the analysis shown in Figure (8.12).

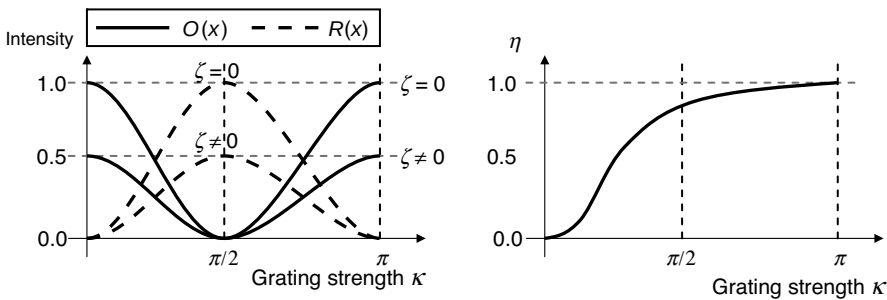


**Figure 8.13** The intensities of  $O(x)$  and  $R(x)$  and the diffraction efficiency  $\eta$  as a function of the grating strength in the Bragg condition

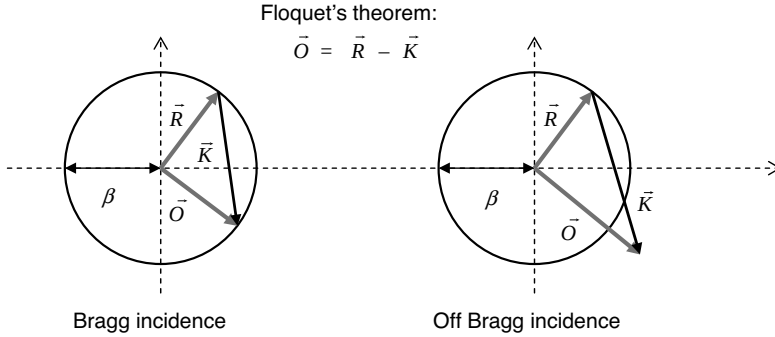


**Figure 8.14** The Bragg condition and Bragg detuning

#### 8.4.4 Diffraction Efficiency

By solving the coupled wave equations in Equation (8.12), the following two expressions for the diffraction efficiency can be derived, for transmission ( $\eta_T$ ) and reflection ( $\eta_R$ ) volume holograms, for s polarization of the incoming beam:

$$\eta_T = \frac{\sin\left(\sqrt{\nu_s^2 + \xi^2}\right)^2}{1 + \frac{\xi^2}{\nu_s^2}} \quad (8.17a)$$

$$\eta_R = \frac{1}{1 + \frac{1 - \left(\frac{\xi}{\nu_s}\right)^2}{\sinh\left(\sqrt{\nu_s^2 - \xi^2}\right)^2}} \quad (8.17b)$$

The incident angle  $\alpha_i$  used is actually the incident angle in the material, which can be related to the incident angle in air by Snell's law:

$$\alpha_i = \arcsin\left(\frac{\sin(\alpha)}{n_0}\right) \quad (8.18)$$

Many variations of this basic theory have been proposed in the literature in order to model more complex volume holograms; for example, multiplexed holograms or complex Bragg plane geometry holograms (see Figure 8.15).

One variation consists of incoherently superimposing multiple Bragg gratings in a similar hologram (e.g. for angular multiplexing in optical page data storage). Kogelnik's theory can thus be applied by considering a series of multiple obliquity parameters  $c_{si}$  in the holographic media (where  $i$  is the index of the  $i$ th Bragg plane in the hologram), and integrating them into the final efficiency expression.

Another variation consists of stacking multiple Bragg gratings one on top of the other (longitudinal spatial multiplexing rather than phase multiplexing) when the Bragg planes are no longer parallel, which can be useful for pseudo-thick surface-relief holograms with tapered slanted gratings, for example. In this case, a series of different grating strengths  $\kappa_i$  and a series of obliquity parameters  $c_{si}$  are considered, for the  $i$ th layer in the hologram.

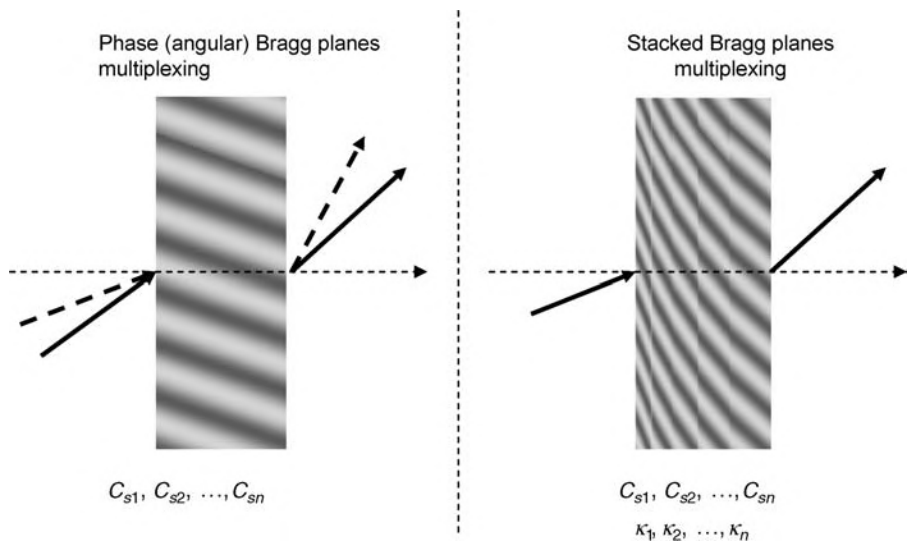


**Figure 8.15** The extension of Kogelnik's theory to complex holograms

8.4.5 Angular Selectivity

Plotting the Kogelnik efficiency results as a function of the incident angle or the reconstruction wavelength can give rise to the graphs presented in Figure 8.16.

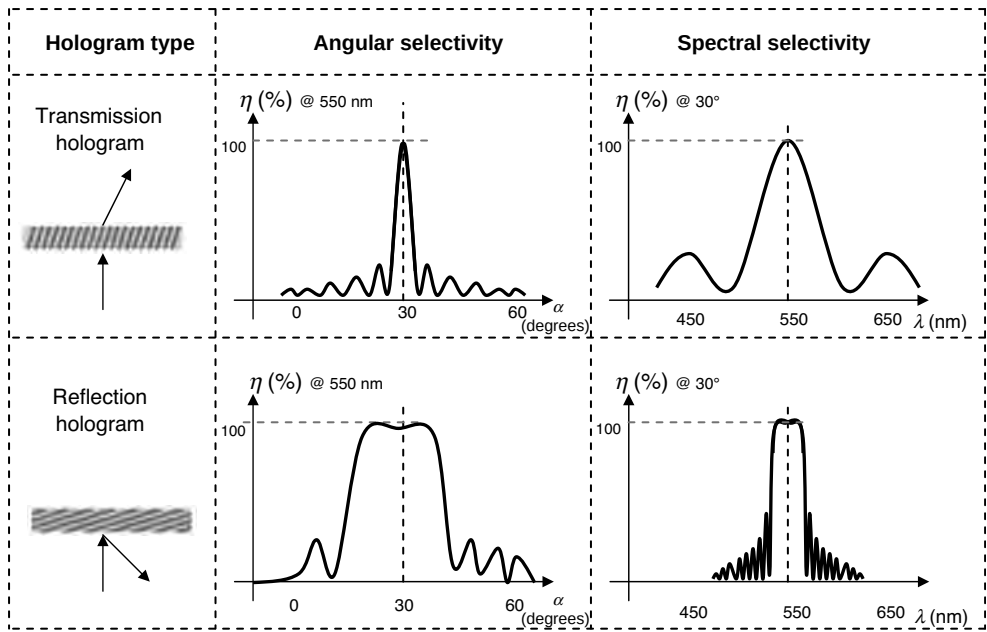


**Figure 8.16** Diffraction efficiency of volume holograms as a function of incident angle and wavelength

The plots on the left-hand side of Figure 8.16 represent the angular selectivity of transmission and reflection holograms (efficiency as a function of the incoming beam angle), and the plots on the right-hand side show the diffraction efficiency as a function of the incoming wavelength.

Figure 8.16 shows that reflection holograms have much narrower angular and spectral selectivity. Moreover, their central bandwidth filter shape is much flatter than for transmission holograms at these very precise locations (in this case, for a Bragg incidence angle of  $30^\circ$  for  $\lambda = 550$  nm).

The narrow spectral filter shape of reflective gratings is used extensively in applications requiring precise spectral dispersion and filtering, such as spectroscopic gratings and DWDM Mux/Demux gratings. The narrow angular selectivity of reflective holograms makes them ideal candidates for angular multiplexing, as in holographic page data storage. In effect, the very narrow angular selectivity of reflection (and, to a lesser extent, transmission) holograms makes it possible to angularly multiplex several holograms in the same hologram area, with the same recording and read-out wavelength.

Figure 8.17 shows the angular and spectral bandwidths of a holographic transmission grating. The efficiency has been calculated over the visible spectrum as a function of the wavelength in air for various incident angles (left) and the efficiency over incident angle for various wavelengths (right).

These graphs show that for a single hologram many combinations of angle and wavelength (Bragg conditions) are possible to achieve a high efficiency (which is the basis for many applications that we will review in this chapter). Figure 8.17 also shows the efficiency when the mean index is increasing (higher indices are more desirable).

Figure 8.18 sketches the various Bragg conditions that can be satisfied for a single wavelength, on both sides of the hologram. The figure shows that symmetrical high-efficiency conditions exist on each side of a volume hologram.

The efficiency of a volume hologram is closely linked to the number and strength of the Bragg planes (or grating strength parameter  $\nu_s$ ) through which the incident beam propagates (and diffracts). The strength of such Bragg planes is a function of the amplitude of the index modulation ( $\Delta n$ ) and the emulsion thickness. The thinner the emulsion or the weaker the amplitude of the index modulation, the higher is the number of Bragg planes required to achieve a specific diffraction efficiency. It is therefore very desirable to use a holographic material with the maximum index amplitude. The available holographic materials (see the next section) have index modulations up to 0.05.

In Section 8.5, the maximum amplitude modulation for conventional holographic materials (H-PDLC) is shown to be about 0.05. However, if the aim is to create Bragg planes as surface-relief elements in a clear material such as glass or plastic, thus going from air to plastic, or from index 1.0 to index 1.5 (therefore creating a  $\Delta n$  of 0.5, one order of magnitude higher than the highest  $\Delta n$  in traditional holography), it is possible to have a huge grating strength parameter and thus not need many Bragg planes to introduce a decent diffraction efficiency. Such elements can be recorded, for example, in thick photoresist (SU-8, for example), and then replicated by UV curing or nano-imprint techniques, by using a negative Nickel master shim (see also Chapter 12).

#### 8.4.6 Polarization Effects in Volume Holograms

Strong polarization effects occur in volume holograms [6, 7]: basically, volume Bragg gratings are polarization-selective elements. Such effects can reduce the spectral and angular bandwidths in  $p$  polarized light, and can trigger different peak efficiencies for  $s$  and  $p$  polarization. In a general way, the polarization effects increase with the angle.

The grating strengths for both polarizations are different, and is described as follows:

$$\nu_p = \nu_s \cos(\alpha_i - \theta_d) \quad (8.19)$$

where  $\alpha_i$  is the incidence angle and  $\theta_d$  is the diffraction angle.

In order to use volume holograms efficiently with both polarizations (i.e. without losing light when using a nonpolarized light source), it is necessary to include a polarization recycling scheme, which is very

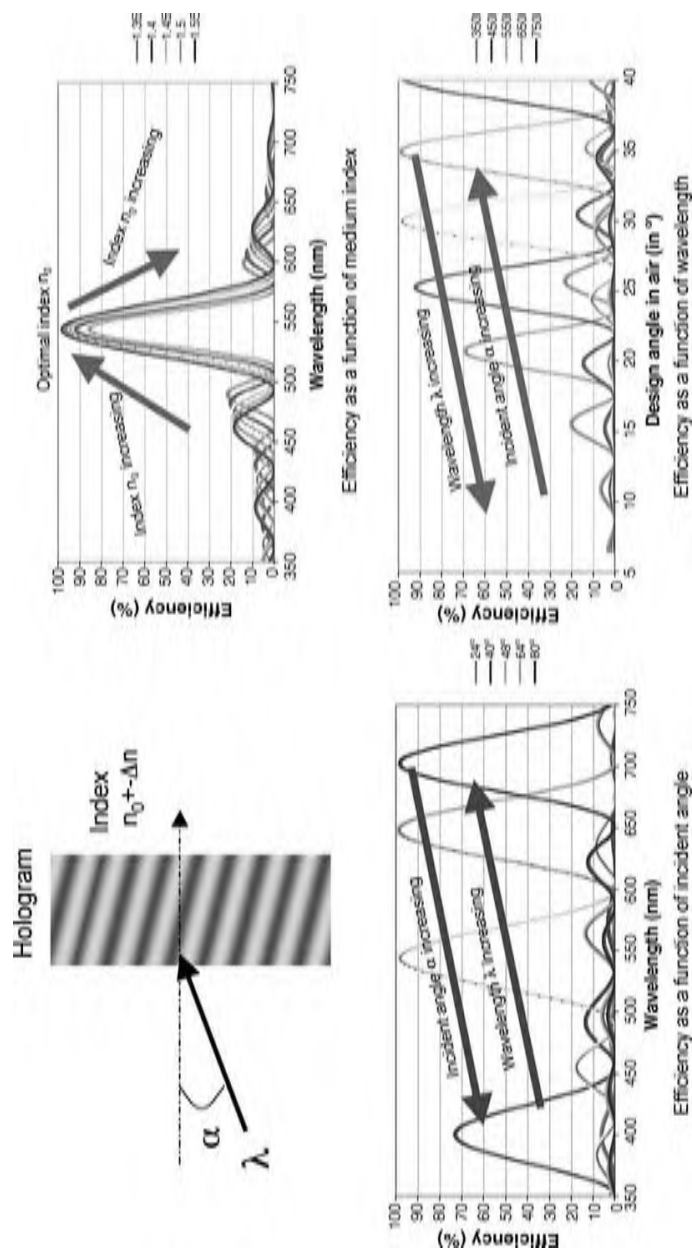


Figure 8.17 Wavelength and angular bandwidths of transmission holograms



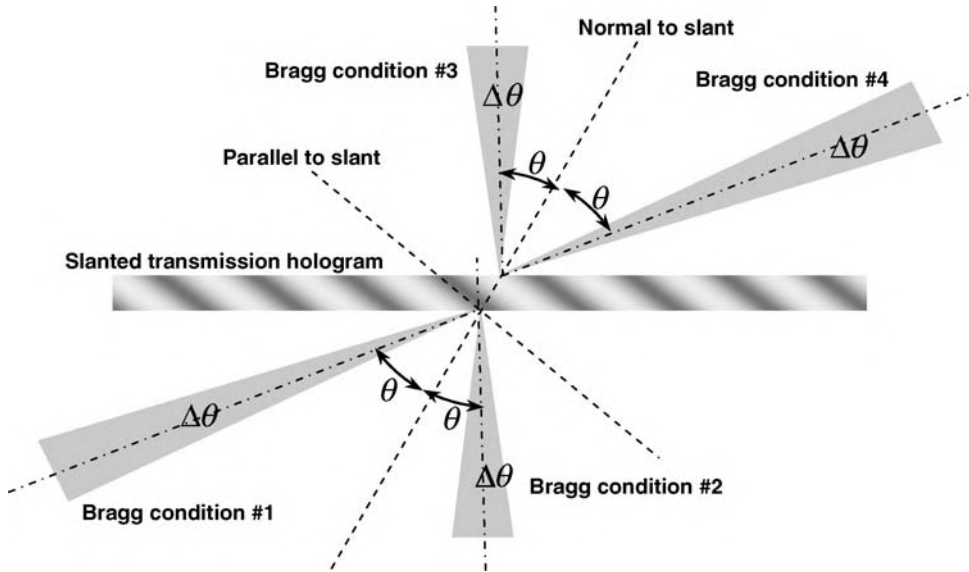


**Figure 8.18** Successive Bragg conditions for transmission holograms

desirable in applications where every photon counts (video projection), or in applications where polarization effects can be a definite show-stopper (Polarization Dependent Loss – PDL – in telecom applications).

## 8.5 HOE Lenses

In Chapter 2, Holographic Optical Elements (HOEs) are listed as Type 1 elements in the classification of digital optics shown.

The previous sections have described the various types of holograms and their optical specifications. Now, the focus will be on particular optical functionalities (other than conventional 3D display holograms) that are recorded in holograms, and associated with other optical elements such as diffractives, waveguides, refractives and catadioptric elements, in order to use them in industrial applications.

As seen in the introductory section of this chapter, HOEs are considered as digital optical (diffractive) elements in the sense that the recording set-up is usually designed by an optical CAD tool on a digital computer, and in many cases, the fabrication often includes a master digital diffractive element fabricated by binary microlithographic techniques.

### 8.5.1 In-line Gabor Holographic Lenses

When Gabor introduced the in-line holographic concept with point sources [1], the most straightforward hologram to produce was not a grating, but actually a lens. The object to be recorded was not an amplitude object, but a phase object (namely, a quadratic phase object). There are two basic parameters involved in an in-line hologram recording (see Figure 8.19):

- the radius of curvature of the wavefront; and
- the inclination angle of the hologram plate.

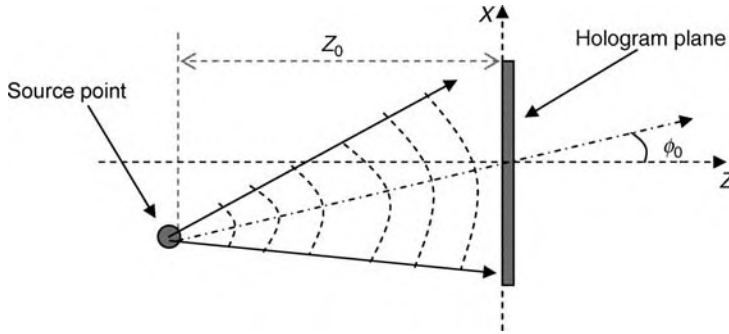


**Figure 8.19** Recording geometry of an in-line Gabor hologram

The phase profile of a spherical wavefront and an inclined plane wavefront (inclined only in the  $x$  direction) along the hologram surface is given by

$$\begin{cases} \varphi_{\text{spherical}} = \frac{2\pi}{\lambda} \left( \frac{(x \cdot \cos\phi_0)^2}{2R_x} + \frac{y^2}{2R_y} \right) \\ \varphi_{\text{linear}} = \varphi_0 + \frac{2\pi}{\lambda} \sin(\phi_0) \end{cases} \quad (8.20)$$

where  $R_x$  and  $R_y$  are, respectively, the radii of curvature of the diverging wavefront in the  $x$  and  $y$  directions.

In a general off-axis configuration, the expressions in Equation (8.20) describing the wavefront can be combined as follows:

$$\varphi_{\text{off axis spherical}} = \varphi_0 + \frac{2\pi}{\lambda} \sin(\phi_0) + \frac{\pi}{\lambda} \left( \frac{\cos^2\phi_0}{R_x} x^2 + \frac{1}{R_y} y^2 \right) \quad (8.21)$$

When the source is on-axis, this expression reduces to

$$\varphi_{\text{on-axis spherical}} = \varphi_0 + \frac{\pi}{\lambda} \left( \frac{x^2}{R_x} + \frac{y^2}{R_y} \right) \quad (8.22)$$

Now consider the interference pattern between two such wavefronts originating from two different on-axis source locations, as depicted in Figure 8.20.

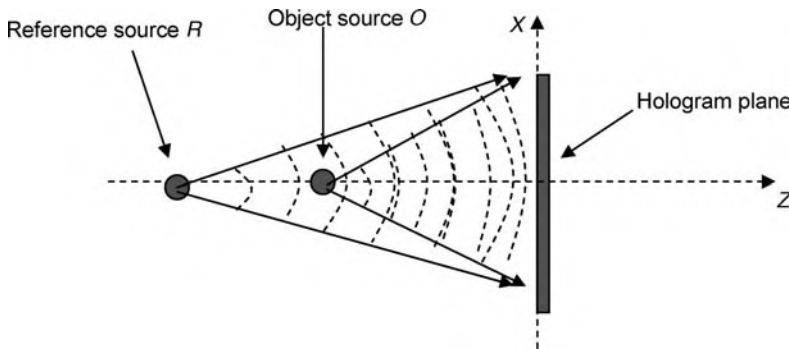


**Figure 8.20** A Gabor in-line hologram recording of a holographic lens

In the introductory section of this chapter, it was shown that the resulting interference pattern between a reference wave and an object wave can be described as follows (see also Figure 8.1 and Equation 8.2):

$$I = I_R + I_O + 2\sqrt{I_R \cdot I_O} \cdot \cos(\theta_R - \theta_O)$$

$$\begin{cases} \theta_R(x, y) = \theta_R + \frac{\pi}{\lambda \cdot R_R} (x^2 + y^2) \\ \theta_O(x, y) = \theta_O + \frac{\pi}{\lambda \cdot R_O} (x^2 + y^2) \end{cases} \quad (8.23)$$

which results in the Gabor Zone Plate (see also Chapter 6). The Gabor Zone Plate equation is as follows:

$$I(x, y) = \frac{1}{2} \left( 1 + \cos \left( \frac{\pi}{\lambda} \left( \frac{1}{R_R} - \frac{1}{R_O} \right) (x^2 + y^2) \right) \right) \quad (8.24)$$

where  $R_O$  and  $R_R$  are, respectively, the radii of curvature of the object and reference beams (note that these radii can be positive (i.e. a diverging wave) or negative (i.e. a converging wave)).

If the hologram is recorded in the configuration shown in Figure 8.21, and if a third wavefront is hitting the hologram (an illumination wavefront of radius of curvature  $R_i$  and wavelength  $\lambda_i$ ), the expression of the radius of curvature of the resulting wavefront (the diffracted wavefront) in the  $m$ th order can be derived by such a hologram in the following way:

$$\frac{1}{R_m} = m \frac{\lambda}{\lambda_i} \cdot \left( \frac{1}{R_O} - \frac{1}{R_R} \right) + \frac{1}{R_i} \quad (8.25)$$

The location of the virtual image ( $m = +1$ ) or the location of the real image ( $m = -1$ ) can thus be computed easily. The locations of the higher orders, for real images ( $m = +2, +3, \dots$ ) or virtual images ( $m = -2, -3, \dots$ ) can also be derived.

Similar relations can be derived in the off-axis architectures (although the equations tend to become complex). Figure 8.21 shows the recording of a transmission off-axis HOE lens with a red laser light and playback with a green laser light. Note the shifts.

Note that the reconstruction wavelength  $\lambda_i$  and the recording wavelength  $\lambda$  in Equation (8.24) do not need to be the same (actually, they are very rarely the same). This is not due to the fact that holographers are funny people who like to confuse ordinary people with complex variations between exposure and

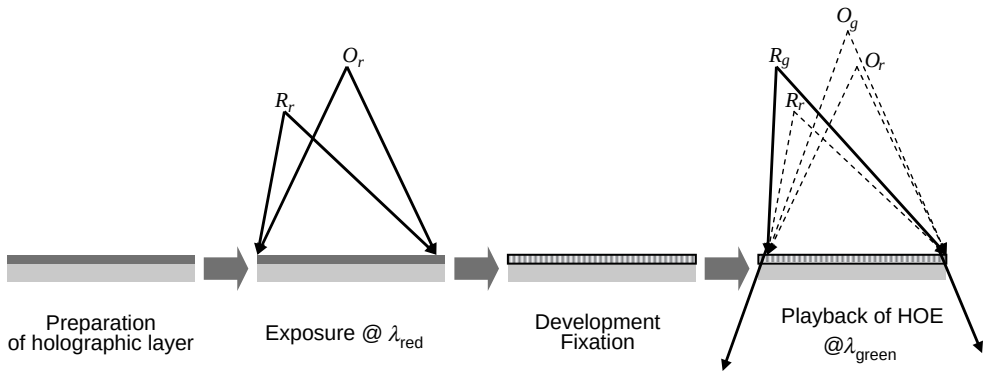


**Figure 8.21** Recording and playback of transmission HOE lens with different wavelengths

playback: this is simply due to material and final application requirements, which might require different wavelengths (e.g. the holographic material could be only sensitive to UV as for exposure, but the hologram has to work on IR light).

Figure 8.21 shows a simple example of the recording of an off-axis lens in red light and playback in green light. In Figure 8.21, note the longitudinal and lateral shifts of the position of the object and reference beams (sources) when one changes the wavelength.

### 8.5.2 *Imaging with HOEs*

Imaging with HOEs is not a simple task, especially if broadband illumination is used. Actually, if the HOE is used without refractive lens for chromatic aberration compensation, this is an impossible task. However, in many cases, imaging tasks may use only a narrow spectral width, for which HOEs provide adequate solutions in terms of size, weight, planarity, ease of off-axis operation and ease of arbitrary aspheric compensation introduction via a CGH recording process (see Section 8.7).

One of the issues to be solved when using HOEs as imaging elements is that the recording wavelength is very rarely even close to the reconstruction wavelength.

The aberrations of holographic lenses are derived and compared to the aberrations of diffractive and refractive lenses in Chapter 7.

### 8.5.3 *Nonspherical HOE Lenses*

In the previous section, simple recording of spherical HOE lenses through the use of two basic building blocks was reviewed:

- the on-axis spherical wavefront (spherical lens profile); and
- the linear ramp phase profile (linear grating, or grating carrier offset for off-axis lenses).

In many cases, the application requires more complex optical functionality than a spherical lens can provide (in either the on- or off-axis modes). Such elements include the following:

- anamorphic HOE lenses;
- cylindrical HOE lenses;
- conical HOE lenses;
- toroidal HOE lenses;
- helicoidal HOE lenses;
- vortex HOE lenses;
- multifocal length HOE lenses (phase multiplexing of on-axis HOE lenses);
- multiple-image HOE lenses (phase multiplexing of off-axis HOE lenses);
- HOE ‘focusators’ (lenses that focus in geometrical shapes other than spots); and
- beam-shaping HOE lenses.

For more insight on such complex imaging functionalities incorporating specific aberration corrections or imaging properties (and nonimaging properties such as beam shaping) – which can be difficult or nearly impossible to record as a conventional hologram, since the wavefronts have to exist physically – see Chapter 5, Section 5.4, in the section on diffractive lenses.

For diffractive elements, the expression of the aspheric wavefront need only be expressed mathematically as a polynomial; for example, by the use of dedicated CAD software.

This leads us to the next section, which deals with design tools (CAD software) for the development of more complex HOEs.

## 8.6 HOE Design Tools

When the optical design engineer decides to incorporate a hologram, a HOE or even a diffractive element in an optical system, he or she has to be convinced that:

- the introduction or replacement of a refractive by a hologram will yield sufficient added value in terms of functionality, footprint, weight, packaging and overall pricing;
- the HOE can be fabricated with state-of-the-art fabrication techniques;
- the HOE will have a MTBF at least as long as that of the lowest-MTBF element in the final system. HOEs have relatively short MTBFs when compared to diffractive, refractive or reflective optics, especially when used in rough environments with high temperature swings, humidity or UV exposure (direct sunlight); and, finally,
- the introduction of a HOE in a commercial system is not related to some odd commercial or marketing reason (which is often the case – holograms are often hyped!).

For more insight on designing a hybrid refractive/holographic or a hybrid refractive/diffractive element, see Chapter 7.

The main issues when considering the use of an HOE in conventional optical systems are as follows:

- chromatic aberrations (spectral bandwidth) – these can be of no importance if the system uses a narrowband source, such as an LED or a laser;
- the angular acceptance bandwidth – the angular acceptance angle can be kept under control if the system has a very low field of view, for example;
- zero-order leakage – this can be addressed by blocking the zero order or setting the reconstruction off-axis;
- higher diffraction orders – these can be addressed by blocking them, or using a CGH master to record a single-order HOE;
- The largest diffraction order achievable – this is linked mainly to materials issues;
- the smallest diffraction angle achievable – this is also linked to materials issues (many holographic media can only record fringes below a maximal allowed width);
- Hologram imaging aberrations – these can be controlled by using a master CGH as an object wavefront (see also aberrations control analysis in the previous section – imaging with HOEs); and
- materials issues (temperature, humidity etc.) – can be addressed through consideration of the packaging (hermetically sealed, temperature controlled, shielded from UV light etc.).

## 8.7 Holographic Origination Techniques

Holographic exposure of HOEs can yield a wide variety of optical functionalities, either as thin surface-relief HOEs (through the use of photoresist spun over a substrate) or as more complex Bragg volume holograms. In this case, the object does not usually need to exist physically. The object beam can be generated by the means of a CGH master fabricated in quartz or fused silica by microlithography (see Chapters 5 and 6), or the HOE can be recorded in a pixelated form, one pixel after the other, with some means of controlling the phase shift between pixels (see the discussion in the next section on fringe locking and fringe writers).

### 8.7.1 Two-step Holography and Phase Conjugation

As described in Chapters 5 and 6, binary diffractives and holograms have a unique property, one that is unique in the realm of optics: they can generate a wavefront that is the conjugate of the object wavefront,

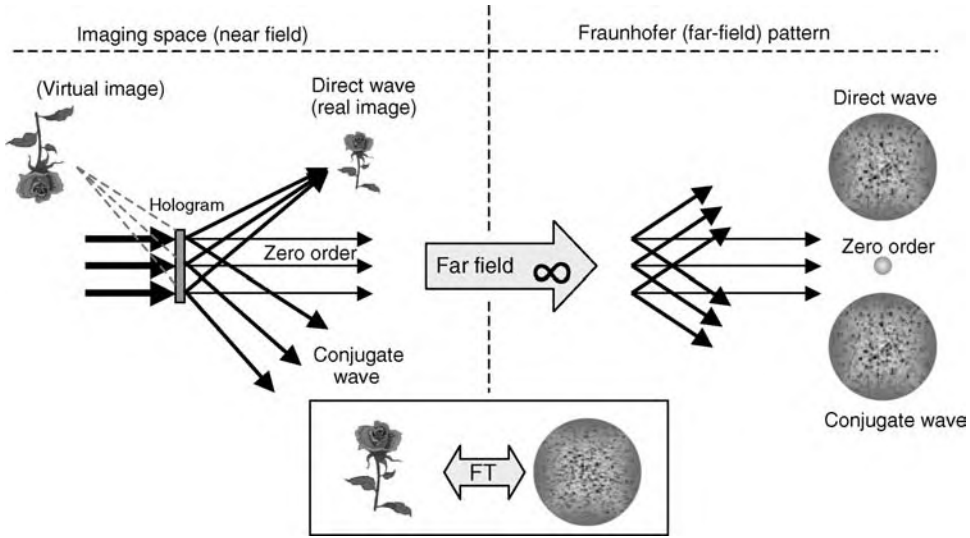


**Figure 8.22** The Fraunhofer pattern of a Fresnel hologram

and therefore produce a second image. The phase of this wavefront is the opposite of the real image, and therefore produces a virtual image. In the case where the hologram is a Fraunhofer hologram, with its reconstruction in the far field, the image is simply the central image that is symmetric to the optical axis. In the case of a Fresnel hologram (the majority of holograms), the positive fundamental order is the real image and the negative fundamental order (the conjugate order) is the virtual image. In an earlier section, Figure 8.4 shows the reconstruction geometry for Fraunhofer holograms, which generate a far-field pattern, and Fresnel holograms, which generate a near-field pattern.

Note that a Fresnel hologram also has a Fraunhofer pattern, and that although the conjugates in a Fresnel hologram are real and virtual images (converging and diverging waves), the far-field pattern of such a hologram (angular spectrum) produces two identical images on a plane, one being the central symmetric of the other. This is shown in Figure 8.22.

### 8.7.2 Rainbow Hologram Recording

Rainbow holography is a technique developed by Steve Benton at Polaroid. The aim is to be able to view color holograms with white light. Benton proposed a two-step recording process (master hologram H1 and transfer hologram H2). In the secondary exposure process, the master hologram H1 is illuminated in phase conjugation and H2 is recorded through a horizontal slit and an inclined reference beam (from below), therefore canceling the parallax in the vertical direction and creating additional spectral dispersion in the vertical direction.

### 8.7.3 Recording from a CGH Master

Complex optical functionalities such as aspherical or nonsymmetric lens profiles can be recorded by using a master diffractive element, such as a DOE or a binary or multilevel CGH. There are two ways to record a CGH functionality in a HOE.

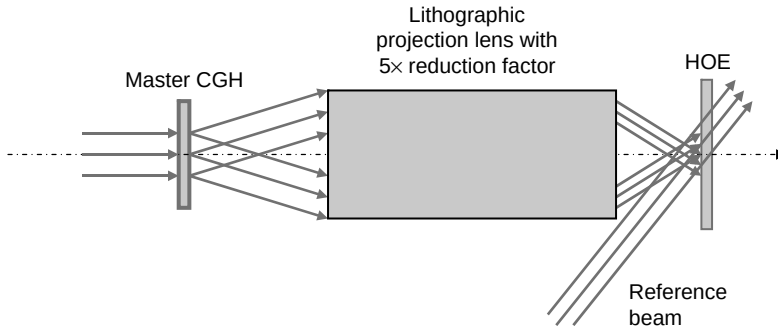


**Figure 8.23** Recording a CGH functionality in a HOE by wavefront imaging and the introduction of a spatial frequency carrier

### 8.7.3.1 CGH to HOE Recording by Imaging

The first recording technique (and also the most straightforward one) is to image the microstructures from the CGH directly into the HOE by means of a high numeric aperture lens that can resolve very small features down to the micron or even sub-micron level. Such a lens is usually an optical lithography lens, such as a stepper lens or a projection aligner lens. In addition, a reduction factor can be used to produce structures that are smaller than the fabricated structures in the original CGH, thus alleviating the fabrication burden in the master CGH, but gaining the final high diffraction angles in the recorded HOE.

This is not a classical imaging task; rather, it is a phase-imaging task, where the wavefront is imaged rather than an intensity map over the wavefront (see Figure 8.23).

For example, in a stepper, the task is to image a binary amplitude function from the reticle (binary chrome mask) onto the wafer. In some cases, the mask is a complex phase-shifting mask that has amplitude and phase information (in order to produce smaller features – see also Chapter 13). Here, only phase information is imaged (the mask or reticle is replaced by the phase CGH, and the wafer is replaced by the holographic material).

In some cases, a spatial frequency carrier (off-axis illumination) is desired in order to produce a grating carrier on top of the CGH profile in order to set the reconstruction off-axis by a considerable amount, which is usually not possible in a CGH due to limitations in the smallest features to be fabricated (and thus the largest diffraction angle that can be generated in that same CGH).

However, in this first technique, all the limitations of the CGH will also appear in the HOE, namely:

- multiple diffraction orders;
- effects due to quantization of the phase profile (quantization noise);
- effects due to the use of square pixels when fabricating the CGH; and
- limited diffraction efficiency in the HOE due to limited diffraction efficiency in the CGH.

### 8.7.3.2 CGH to HOE Recording by Diffraction Order Selection

The second method for recording a CGH into a HOE does not use any imaging task but, rather, uses a Fourier transform lens. In such a recording process, a single diffraction order is used from the CGH (which diffracts multiple propagating orders) to generate the object beam, and all other diffraction orders present are blanked out. Figure 8.24 shows such a CGH/HOE recording process.

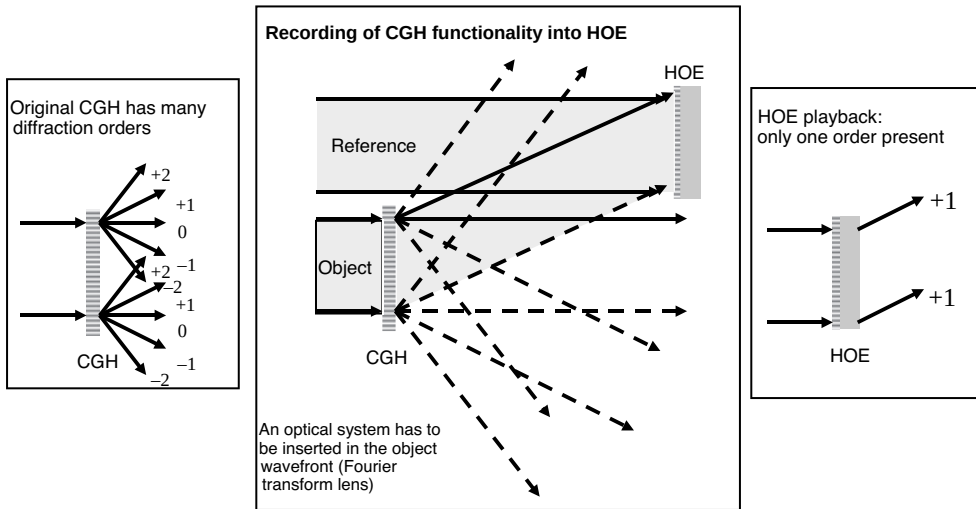


**Figure 8.24** The process for recording a surface-relief CGH into a volume HOE

Although the CGH generates many orders, the resulting HOE will generate only a single diffraction order without altering the initial optical functionality (if the recording has been done properly at the Bragg regime angles).

This second method does not reproduce the limitations in the CGH, namely:

- that the CGH has multiple orders, while the HOE has only one order; and
- that the CGH has limited diffraction efficiency, while the HOE can have maximum efficiency.

However, it cannot get rid of the imperfections of the CGH due to lateral quantization of the phase profile (e.g. into square pixels).

#### 8.7.4 Nonsinusoidal HOE Fringe Profiles

In more and more applications, in order to be effectively used in a product, the HOE has to be mass-replicated, and thus have a low overall price tag. One solution is to record the HOE as a surface-relief element in photoresist and replicate it using injection molding, UV curing or embossing techniques (see Section 8.8 and Chapter 14).

It is particularly important to shape the surface profile of the HOE in order to get the desired effect (see also Section 5.3). For example, a blazed, sawtooth or *échelette* surface profile is a desirable feature in many spectroscopic and wavelength demultiplexing applications. Such nonsinusoidal profiles can be obtained by multiple HOE exposure, as explained in Chapter 12.

#### 8.7.5 Digital Holographic Fringe Writers

The problem with traditional holography is that, as depicted in Figure 8.1, the object has to exist in order for a hologram to be recorded.

In many cases, this is not possible since:

- the object does not, or cannot, exist physically;
- the object beam cannot be produced by a CGH master diffractive;



- the object cannot be fabricated with tight enough tolerances;
- the required tolerances on the optical elements used to record the hologram are prohibitive;
- the exposure time would be too long and the vibration requirements too high to be sustained; and
- the energy of the potential laser needed to illuminate the object (a very large object) is prohibitive.

In order to fabricate volume holograms in the above prohibitive cases, complex fringe writers have been developed that use fringe locking and other compensation techniques to overcome the drawbacks mentioned here. Chapter 12 reviews such fringe-locked fringe writers in detail.

## 8.8 Holographic Materials for HOEs

When it comes to choosing a holographic material for a specific application, there are a variety of points to consider apart from the targeted diffraction efficiency, namely:

1. Availability of materials. The fact that a material is well known does not mean that it is available on the market (e.g. Dupont Photopolymers).
2. If it is available, are there secondary sources for the material?
3. Is a mass-replicable technique already developed for this material?
4. What is the exposure wavelength range?
5. What efficiency can be achieved (as a function of material thickness,  $\Delta n$  etc.)?
6. How does it stand with regard to humidity and temperature swings?
7. Is fancy sealing of the material needed?
8. How does it stand with regard to direct sunlight (UV)?
9. What is the shelf life of the material?

Various holographic materials have been proposed over the past four decades, many of which are used today in industry for consumer products. The different holographic materials that are available can be grouped into four main categories:

- emulsions;
- photopolymers;
- crystals; and
- photoresist-based materials.

Holographic emulsions, photopolymers and crystals are index modulation elements, whereas photoresist-based materials are surface-relief elements.

### 8.8.1 HOEs as Index Modulation Elements

Index modulation holograms are volume holograms, and can yield very high efficiency, over a thick emulsion (from less than a micron to several tens of microns). Transparent index modulation holograms can act as reflection holograms. Bragg planes are well defined in such elements. Surface-relief holographic elements do not have Bragg planes, and therefore cannot act as reflection holograms.

Most of the holograms used in industry today are layered index modulation materials. The recording process of a hologram, either in index modulation or surface modulation form, can be summarized as follows:

- preparation of the holographic layer (DCG layering and sealing, photopolymer peeling and gluing etc.);
- the exposure set-up, which can be optimized by CAD tools;
- the laying out of the exposure set-up on a vibration-free table;
- exposure in H1 (the master hologram) and optional second-step exposure in H2 (the transfer hologram);

- development and post-bake;
- optional beaching (for silver halides and other index modulation elements with partial amplitude variations);
- optional etching (for surface-relief modulation);
- optional replication (holographic for photopolymers, and casting/stamping/injection molding for etched elements); and
- integration/playback in the final application.

### 8.8.1.1 Holographic Emulsions

#### Silver Halides

Silver halides were the first mass-produced holographic media. Silver halide grains (AgBr, AgI etc.) are suspended in a gelatin emulsion mounted on a glass substrate (see Figure 8.25).

The grain sizes are around 5–10 nm. Dyes are added to sensitize the AgH in the visible (normally AgH is only sensitive in the UV range). The silver halide photographic process is shown in Figure 8.32 and can be described as follows:



Approximately four Ag ions must be reduced to silver to make the grain developable by wet processing.

Examples of silver halide emulsions are Agfa's 8E75 and Kodak's 649F. Ultra fine grain panchromatic silver halide light-sensitive material for RGB recording of reflection holograms has been recently developed.

#### Dichromated Gelatin (DCG)

It is well known that dichromated gelatin is one of the best materials for phase hologram recording. It produces a very high efficiency while reducing scattering noise, and can yield up to 5000 lines per millimeter. Several competing hypotheses for the nature of the photo-induced response in DCG have been discussed in the literature, and there is still no generally accepted point of view. However, one of the main

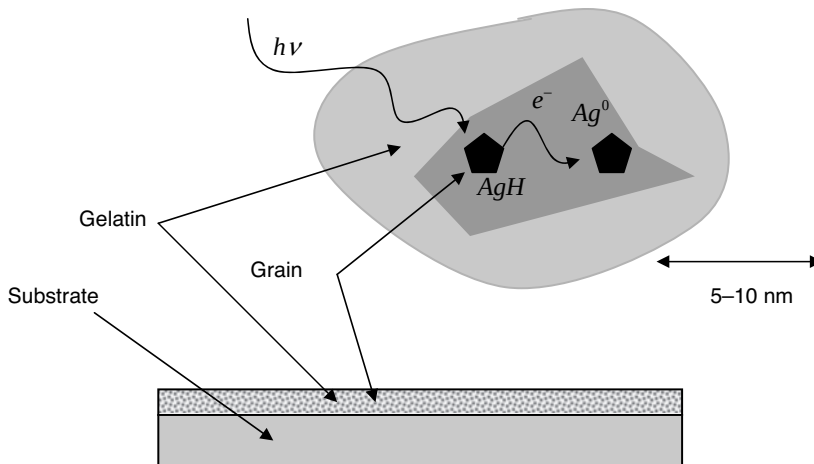


**Figure 8.25** Silver halide holographic exposure

disadvantages of DCG is its short shelf life. The thickness of the DCG layer can be increased or decreased by controlling the exposure and processing conditions.

A widely used method of preparing a DCG film is to dissolve out the silver halide in a silver halide photographic plate by soaking the unexposed plate in fixer. It is also possible to coat glass plates with gelatin films. Such a DCG layer can be made by mixing 1 g of ammonium dichromate with 3 g of gelatin and 25 g of water. Popular gelatin production methods include pork skin and chicken legs gelatin.

### 8.8.1.2 Photopolymers

Today, photopolymers are the workhorse of holography. Many of the consumer applications of holography are based on photopolymers.

#### *Passive Photopolymers*

Some of the most used photopolymers are Aprilis, Polaroid's DMP128, Dupont's Photopolymer 600, 705 and 150 series, and PQ-MMA. One of the problems with photopolymers is that they tend to shrink during exposure and development. Techniques to reduce shrinkage have been developed (PQ-MMA). Photopolymers can be mass-replicated by techniques developed by companies such as Dupont. One example is the credit card hologram. However, due to the very strict regulation in the optical security market, such photopolymers are very hard to obtain for companies that are not in the right field, or that do not have the right connections.

Dual photopolymers have been developed for applications in optical data storage, and one company has actually put such a device on the market recently (see Section 16.8.2.4).

#### *Active Photopolymers*

Chapter 10 reviews in detail some of the active photopolymer materials used in industry today, including the Holographic-Polymer Dispersed Liquid Crystal (H-PDLC).

### 8.8.1.3 Photorefractive Crystals

In a photorefractive material, the holographic exposure process forms an interference pattern that creates a distribution of free carriers through ionization. These free carriers diffuse, leaving a distribution of fixed charges. The separated fixed charges form an electric field distribution, which induces a change in the refractive index through the electro-optic effect of refractive index modulation  $\Delta n$  described by

$$\Delta n = \frac{1}{2} n^3 r E \quad (8.27)$$

Typical photorefractive materials include doped  $\text{Bi}_{12}\text{TiO}_{20}$ ,  $\text{Bi}_4\text{Ge}_3\text{O}_{12}$ , Te-doped  $\text{Sn}_2\text{P}_2\text{S}_6$ ,  $\text{LiTaO}_3$ ,  $\text{Pb}_2\text{ScTaO}_6$  crystals and many more!

### 8.8.1.4 Other Phase Materials

Exotic index modulation holographic materials have been developed alongside the various types described in this section. These include the nonlinear effects of bacterial rhodopsin, which can implement erasable holographic media.

Another exotic material has recently been reported by the optical science center of the University of Arizona: an updatable photorefractive polymer that can implement an updatable holographic media for 3D display or other applications. A material with such properties is capable of recording and displaying new images every few minutes. Such a hologram can be erased by flood exposure and then re-recorded ad infinitum.

8.8.2 HOEs as Surface Modulation Elements

8.8.2.1 Photoresists

Although a very versatile and efficient method, holographic recording of volume HOEs cannot be used in many practical cases, where materials with long lifetimes are necessary (due to temperature swings, vibration, humidity or daylight/UV exposure), or simply when mass replication of cheap diffractive elements is just a *de facto* requirement (the automotive industry, consumer electronics etc.).

In many cases, it is desirable to produce a surface-relief hologram (a thin hologram) via a holographic recording process, and create a master negative mold from this surface-relief element to inject, cast or emboss thousands of replicas (see also Chapter 14). Such surface-relief holograms can be recorded in a material such as photoresist, both in positive or negative form (see Figure 8.26).

8.8.2.2 Etched Substrates

As a photoresist is a polymer, it is not usually the best material to be used in highly demanding applications. Photoresist profiles can be transferred into the underlying substrate as a sinusoidal profile (by proportional RIE etching – see Chapter 12) or as binary gratings with a simple RIE etch.

For mass production, the photoresist profile can be used to produce by electroplating a nickel shim that can act as a negative mold for injection molding, UV casting or embossing in plastic (see Chapter 14). In some cases, the etched profile is used to produce the shim.

Table 8.1 summarizes the various holographic materials used in industry today and their specifications. Recently, rewritable holographic materials have been demonstrated.

Technically, there is another material for holography, though basically it is nonexistent: in this case, we are referring to electronic holography. The next section will review two of these techniques.

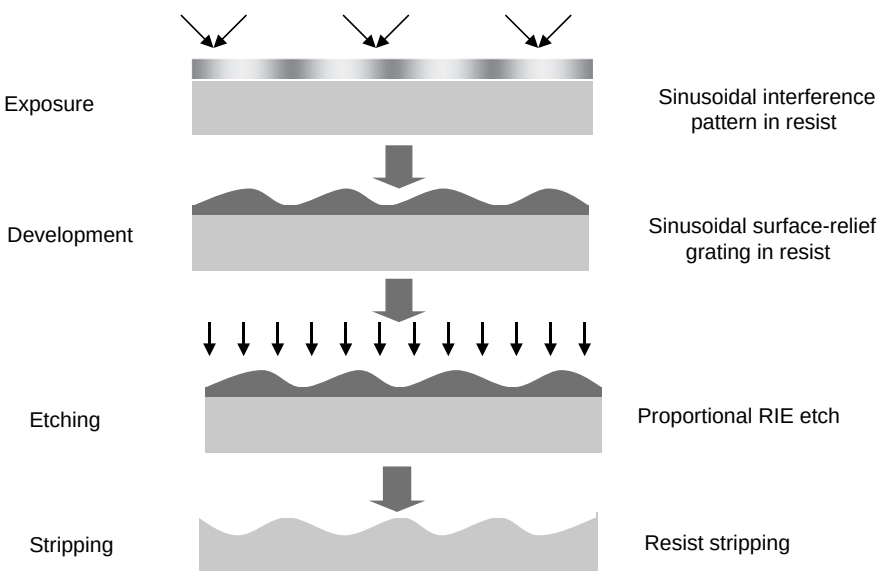


Figure 8.26 Surface-relief HOE recording in a photoresist material

**Table 8.1** Holographic materials used for the recording of HOEs

| Materials     | Type                                     | Hologram type                                             | Application                                           | Advantage                                                          | Inconvenient                                              |
|---------------|------------------------------------------|-----------------------------------------------------------|-------------------------------------------------------|--------------------------------------------------------------------|-----------------------------------------------------------|
| Photoresist   | Thin resists (positive or negative)      | Thin hologram (structures normal to substrate surface)    | Diffraction optics/lithographic patterning            | Simple process; resist pattern can be etched by RIE in substrate   | Low efficiency due to resist modulation                   |
|               | SU-8 resist                              | Slanted structures, replicated by UV casting/nano-imprint | HUD, solar, LCD diffusers, phase masks, . . .         | Bragg planes from air to plastic (highest $\Delta n$ of 0.5!)      | Difficult to replicate; use of complex fringe writers     |
| Films         | Silver halides                           | Volume                                                    | First historical holograms                            | Cheap process                                                      | Thin, low selectivity, low efficiency                     |
|               | Dichromated Gelatin (DCG)                | Thick volume hologram                                     | High selectivity applications (optical storage, DWDM) | Cheap, strong angular and spectral selectivity                     | Long-term stability (from animal gelatin–pork or chicken) |
| Photopolymers | Passive (Dupont type)                    | Thin hologram                                             | Anticounterfeiting holograms                          | Replicable in mass                                                 | Low resolution; unavailable outside DuPont                |
|               | Polymer Dispersed Liquid Crystals (PDLC) | Volume hologram                                           | Display, telecom, datacom, storage                    | High index, high $\Delta n$ , switchable through gray levels, fast | Long-term stability–polarization issues                   |
| Crystals      | Photorefractive                          | Very thick volume hologram                                | Various                                               | Erasable/rewritable                                                | Expensive–polarization issues                             |
|               | Acousto-Optic (AO) modules               | Volume Bragg grating (Raman–Nath regime)                  | Beam intensity modulation and beam steering           | Fast reconfigurable grating angle                                  | Only linear gratings, need piezo-transducers              |
| Bacteria      | Rhodopsin                                | Thin (surface relief)                                     | Gratings                                              | High efficiency                                                    | Complex formulation                                       |

## 8.9 Other Holographic Techniques

In the previous sections the various methodologies and technologies used to produce holograms in industry today have been discussed. In this last section, the focus will be on additional techniques used to implement specific applications, by either multiple exposures or material-less holograms.

### 8.9.1 Holographic Interferometry

Holographic interferometry is an interferometric technique for small or fast displacement analysis (stress, vibration modes, distortion, deformation etc.), which can replace a traditional interferometric device as depicted in Figure 8.27.

In holographic interferometry, the hologram creates one or both wavefronts. These wavefronts interfere to produce the desired interference patterns. There are three main holographic interferometry techniques, as detailed below.

#### 8.9.1.1 Double-exposure Holography

This technique was the first holographic technique developed, and is also the easiest to implement. Two holograms are recorded in a single plate, one after the other, developed and then played back. Two objects beam are thus created, interfere and produce fringes if the object has moved between the holographic exposures. This is a static process in which the movement is analyzed between two exposure sets. The recording time between each shot can be very small, thus enabling fast interferometric measurements via fast pulsed laser exposure (flying bullets etc.), fast vibration modes and so on.

#### 8.9.1.2 Real-time Holographic Interferometry

Here, a single hologram is recorded, and then placed back in its exact original location after development. In this case, the reference beam becomes the object beam. The interference pattern is observed on the object as the object moves. The problem here is to align the object and hologram to sub-wavelength accuracy. Such a technique is ideal with a thermoplastic holographic camera, since the exposure is done in

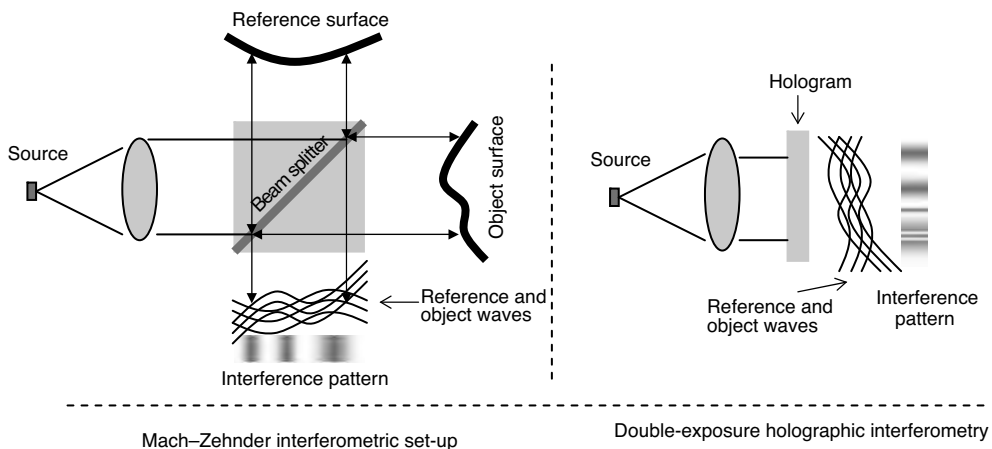


**Figure 8.27** Interferometer and holographic interferometry

place (without moving the hologram). Applications include slow-motion analysis such as shrinkage control and so on.

### 8.9.1.3 Time-averaged Holographic Interferometry

Here, a single exposure is taken, but over a long time period, much longer than the vibration period of the movement under analysis. The result is fuzzy fringes rather than sharp fringes as with the previous techniques, but bright fringes and dark regions can relate to the total amplitude of the vibration mode.

## 8.9.2 Holographic Angular Multiplexing

Previously, it was shown with Kogelnik's theory that volume holograms have relatively narrow angular bandwidths. These angular bandwidths narrow when the material thickness and  $\Delta n$  increase. The idea in angular multiplexing is to record many holograms in the same location by varying the reference (or object) wave over an angle that is large enough not to result in overlap between adjacent images [8]. Applications include fast imaging of moving objects (varying object angles) or fast variation of reference angles with a moving object, and of course optical data storage (Figure 8.28).

In optical data storage, digital binary images (or gray-scale images) are stored in a crystal or emulsion, using up to 30 or more different angles. Although, theoretically, several terabytes of digital content can be stored in a square centimeter crystal, the main challenges for this technique are as follows:

- reducing the overlap between images, so that a reasonable Bit Error Rate (BER) can be achieved that is compatible with current data storage devices (a physical challenge);
- having a read-out capability able to point at the right location on the hologram at the right angle (an opto-mechanical challenge); and
- producing a material with a long shelf life (material challenge).

Chapter 16 shows, as an example, the first holographic optical data storage device on the market (InPhase, Inc.)

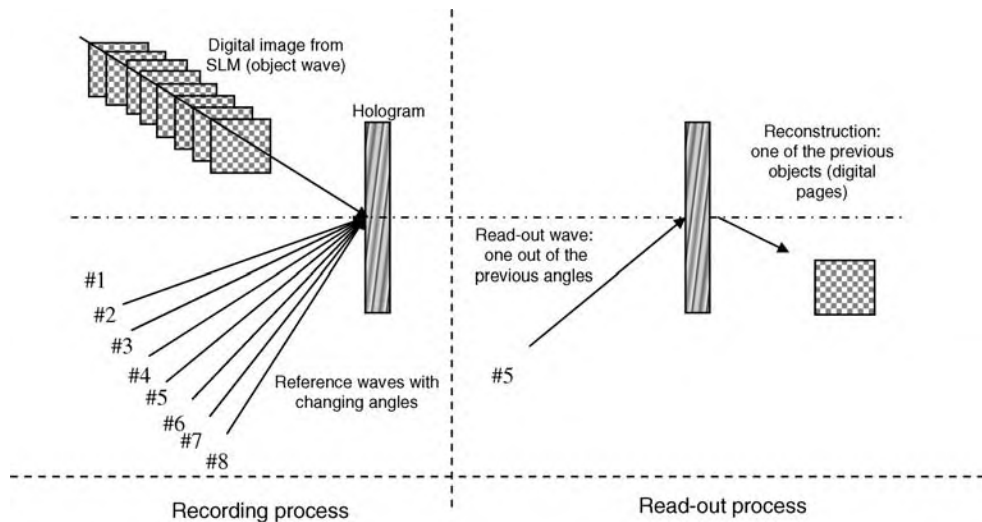


**Figure 8.28** Angular multiplexing in holographic page data storage

### 8.9.2.1 An Important Note about Efficiency and Reverse Operation as an Angular Beam Combiner

One important feature to remember when angularly multiplexing holograms is that (as nothing comes free in this world) the overall diffraction efficiency drops as the number of multiplexed holograms increases. If the efficiency is not reduced, the ability to produce (in the reverse operation) a perfect beam combiner (combining similar wavelengths of light into a single beam, the holy grail of laser fusion, laser weapons, fiber amplifiers and so on) is not possible. But it is good enough for optical data storage applications.

Note that it is possible to implement beam combiners when one of the beam parameter changes (e.g. the wavelength, the polarization and so on).

### 8.9.3 Digital Holography

Digital holography is a new technique that gets rid of one of the main problems in holography: the holographic material. Instead, a digital camera is used to record the interference patterns. The interference patterns are then analyzed by an algorithm that can back-propagate the wavefront to the object from which it was reflected (or transmitted). Therefore, an object can be reconstructed from just the information of its hologram (interference between the object wavefront and a reference wavefront). Figure 8.29 summarizes the digital holography process. The diffraction models used by the algorithm are those described in Chapter 11 and Appendix B.

The challenges in digital holography reside in the hardware and software:

- having an individual sensor pixel size small enough to be capable of resolving the interference fringes, and large enough to grasp the required space bandwidth product of the interference pattern in order to perform the inverse problem; and
- retrieving the complex amplitude (phase information) of the waves from only the intensity interference pattern created on the digital sensor array.

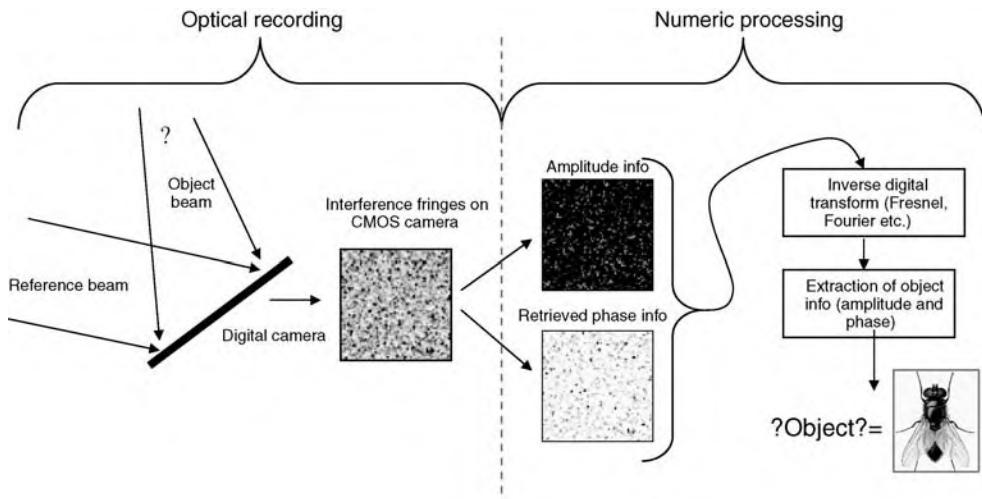


**Figure 8.29** Digital holography



Chapter 16 reviews two applications using digital holography, namely industrial 3D shape acquisition via off-axis holographic exposure of amplitude objects and holographic confocal imaging (or optical diffraction tomography) via in-line Gabor holographic exposure of phase objects.

In this chapter, the various holographic techniques used today in industry have been reviewed. The various characteristics of volume holograms and how they can be applied to specific products have also been reviewed.

The next chapter will focus on dynamic digital optical elements, and especially on dynamical holograms.

## References

- [1] D. Gabor, 'A new microscope principle', *Nature*, **161**, 1948, 777–778.
- [2] P. Hariharan, '*Optical Holography*', Cambridge University Press, Cambridge, 1984.
- [3] W.J. Dallas, 'Holography in a Nutshell', Optical Sciences course 627 on Computer Holography, January 12, 2005.
- [4] M.G. Moharam and T.K. Gaylord, 'Rigorous coupled-wave analysis of planar grating diffraction', *Journal of the Optical Society of America*, **71**, 1981, 811–818.
- [5] H. Kogelnik, 'Coupled wave theory for thick hologram gratings', *Bell Systems Technical Journal*, **48**, 1969, 2909–2947.
- [6] I.K. Baldry, J. Bland-Hawthorn and J.G. Robertson, 'Volume phase holographic gratings: polarization properties and diffraction efficiency', *Publications of the Astronomical Society of the Pacific*, **116**, 2004, 403–414.
- [7] S. Pau, Intermediate Optics Lab 8, College of Optical Sciences, University of Arizona, Fall 2007.
- [8] J. Ma, T. Chang, J. Hong *et al.*, 'Electrical fixing of 1000 angle-multiplexed holograms in SBN:75', *Optics Letters*, **22**(14), 1997, 1116–1118.



# 9

## Dynamic Digital Optics

The focus in Chapters 3–8 has been set on static digital optics, including waveguides, micro-refractives, diffractives, holographic, hybrid and sub-wavelength optical elements fabricated in hard materials such as glass and fused silica, or recorded in holographic emulsions. In Chapters 4, 6 and 8, the third dimension of space was introduced (through 3D displays). In this chapter, the fourth dimension (i.e. time) will be introduced. This chapter reviews dynamic digital optics that can be implemented on technological platforms similar to the ones described previously, and on new platforms such as LCs, MEMS and MOEMS.

### 9.1 An Introduction to Dynamic Digital Optics

Dynamic optics and dynamic optical micro-devices are an emerging class of optical elements that have a high potential to provide solutions for demanding applications [1], especially in the consumer electronics, transportation, optical telecoms and biotechnology market segments (for product descriptions, see also Chapter 16). Dynamic optics opens a whole new window for the use of digital optics. New applications are being proposed every day and this is becoming the fastest-growing segment of the digital optics realm.

#### 9.1.1 Definitions

Dynamic digital optics can be classified into three groups (see Figure 9.1):

- switchable digital optics;
- tunable digital optics; and
- reconfigurable digital optics.

Note that there is an additional sub-field, similar to tunable optics, called software optics. This is a class of optics that cannot work properly without associated image-processing software, and it will be reviewed at the end of this chapter.

- A *switchable optical element* has a preconfigured optical functionality, which can be switched ON or OFF, either in a binary ON/OFF mode or in a continuous OFF to ON mode. Examples include:
  - A diffractive optical lens (DOE) with a fixed NA in which the diffraction efficiency can vary from a minimum level to a maximum level.

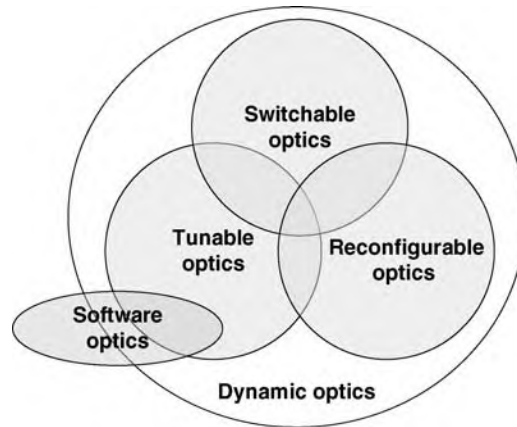


**Figure 9.1** The three dynamic optics element types

- A grating with a diffraction angle that would switch instantly between one set angle and another set angle.
- A *tunable optical element* has a preconfigured optical functionality, that specifications of which can be tuned from one value to the other.
  - Example: a refractive optical element the focal length of which could be tuned from one set value to another set value, or in a continuous way.
- A *reconfigurable optical element* has neither a preconfigured optical functionality nor a preconfigured efficiency. The functionality, as well as the efficiency, can be reconfigured in real time.
  - Example: a laser-based diffractive pico-projector would produce many different sets of distant field patterns through the generation of Fraunhofer diffractive elements that could be calculated and stored in real time in a pixelated phase element.

### 9.1.2 Key Requirements for Dynamic Optics

The complexity, hardware requirements and final price increase linearly or even exponentially from simple switchable optics to tunable optics [2], and to more complex and functional reconfigurable optics.

The key requirements for dynamic optical elements are common to all three types, and include the following:

- the speed of the transition from one state to the other (switching, tuning, reconfiguring speed rate);
- remnant functionality in the OFF position (an absolutely clear window?);
- the maximum reachable efficiency in the ON position (100% efficiency achieved?);
- the appearance of parasitic effects (higher orders, aberrations, noise etc.);
- the level of complexity of the technological platform (single technology or hybrid technology);
- ease of replication and mass-production; and
- the Mean Time Before Failure (MTBF).

However, the most desirable features in dynamic optical elements are high switching speeds and low levels of remnant functionality in the OFF position (also known as the extinction ratio).

In the sections that follow, the various technological platforms and optical implementations for each of these three types of dynamical digital optical elements will be reviewed.

### 9.1.3 Cross-technological Platforms

A variety of platforms can be used to implement dynamic optics. The following will be discussed:

- liquid crystal optics;
- MEMS and MOEMS; and
- microdisplays.

#### 9.1.3.1 Liquid Crystals (LCs)

A *liquid crystal* is a substance that has intermediate properties between those of a crystalline solid and those of a liquid. Its molecules are more orderly than those of a liquid, but less orderly than those in a pure crystalline solid. One of the main features of an LC is that it can rotate linear polarized light beam when no voltage is applied, and return to an inert configuration when a voltage is applied (see Figure 9.2). Such electrodes are usually formed using Indium Tin Oxide (ITO), which is conductive and transparent. Liquid crystals are used in an impressive number of applications, especially in display-related applications (flat-screen televisions, wristwatches, laptop screens, digital clocks etc.).

To create a display, the liquid crystal is placed between two polarizers and glass plates, and electrical contacts are applied to the liquid crystal. In Figure 9.2, light passes through the polarizer on the right and is reflected back to the observer, resulting in a bright segment. The liquid crystal does not rotate the plane of polarization when a voltage is applied. The light is absorbed by the polarizer on the right and none is reflected back to the observer, resulting in a dark segment. Changing the voltage applied to the crystal using a precise electrode pattern at a precise time can create dynamic display patterns, such as a digital watch display, or more complex alphanumeric or logo displays.

The illumination of the LC cells can come from the front (see Figure 9.2) or from the back (see Figure 9.3). When illumination is provided from the front, about the display is a transreflective display, and when the illumination comes from the back, the display is referred to as back-illuminated. Back-illumination can be performed by series of white longitudinal lamps actually located in the back of the display, or by LEDs. In the case of edge illumination of LCD screens by linear LED arrays, the light is propagated through the display panel by total internal reflection (TIR) in a wedge waveguide,

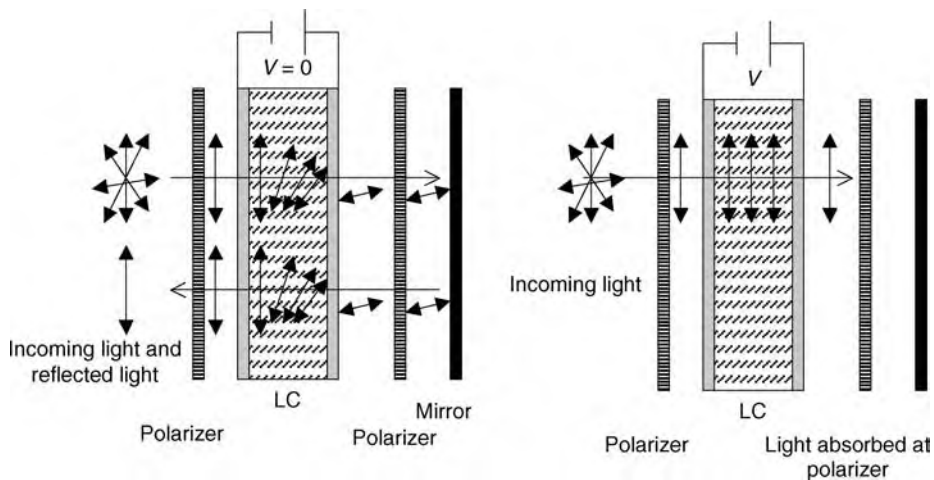


**Figure 9.2** The principle of the Liquid Crystal (LC) display cell

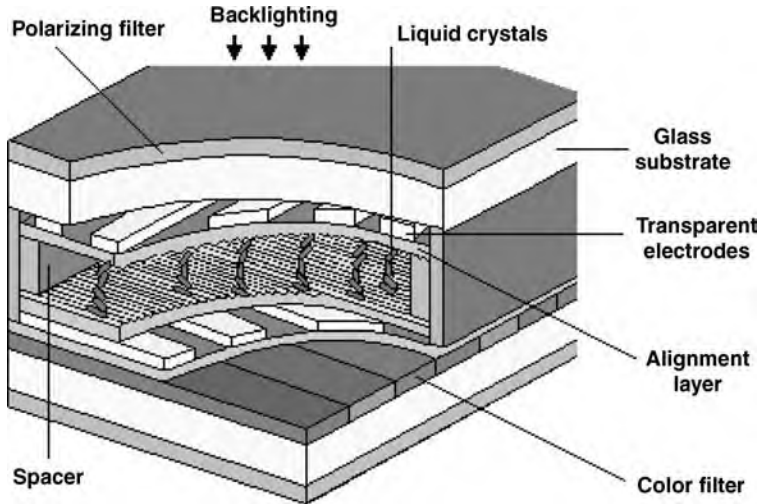


**Figure 9.3** The back-illumination of a Liquid Crystal Display (LCD) pixel

and extracted from that waveguide by either refractive micro-optics (prism arrays) or diffractive gratings (see Chapters 3 and 16).

#### ***Bulk Liquid Crystals***

The phase-shifting effect allows LCs to rotate the polarization state of light. The LC prints the phase shift on only one polarization, leaving the other one unchanged. The phase shift is introduced by changing the refractive index for the specific polarization of interest. This is the most interesting property of LCs, which will be used in numerous examples in this chapter.

Bulk LCs have no patterned electrodes as they have in a display (the electrodes are actually coated on the entire window). One can therefore change the relative index of refraction of the LC lying between two surfaces (these surfaces are not necessarily planar).

#### ***LC-based Spatial Light Modulators (SLM)***

An LC-based Spatial Light Modulator (SLM) is an LC window that has pixelated ITO electrode patterns, which can be addressed individually, and can therefore produce a binary aperture stop for polarized light (such as ferro-electric LCs). Such elements are also referred to as ‘light valves’. A transparent SLM is a phase SLM where the pixels imprint a phase difference rather than an amplitude difference. Therefore, in a phase SLM, an analyzer is not necessary (see the second polarizer in Figure 9.2).

### **9.1.3.2 MEMS and MOEMS**

The emerging field of Micro-Opto-Electro-Mechanical Systems (MOEMS), also called Optical Micro-Electro-Mechanical Systems (MEMS), [3–5] is an impressive example of the efficient integration and combination of optics with electronics and mechanics (see Figure 9.3).

In MEMS technology, classical macroscopic mechanics no longer apply, since weight is almost insignificant [6–8]. However, other forces – especially electrostatic forces – become more important. Therefore, MEMS devices can move very rapidly without the movement creating mechanical wear or cracks, and can yield very high switching rates over a very long life (e.g. operations at several kilohertz over a number of years have been reported).

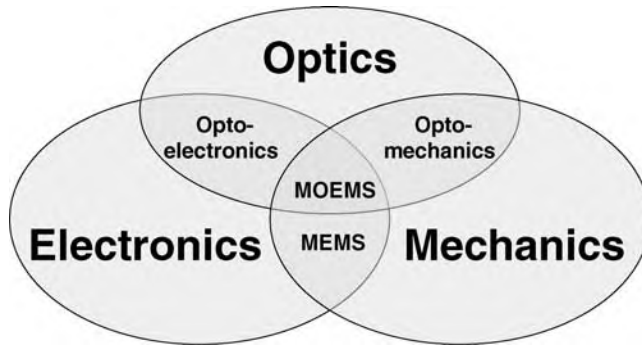


**Figure 9.4** Merging optics, electronics and mechanics

The fabrication of a MEMS or MOEMS device is performed by microlithography [15], by using several layers of polysilicon sandwiched between sacrificial layers on top of a silicon substrate. The sacrificial layers are removed after the successive polysilicon layers have been patterned and etched through. Electronic, mechanical and optical devices can be fabricated by this technology. The most popular MEMS implementations is the micromirror array (see Figure 9.4) [10, 11]. Texas Instrument's Digital Light Processor (DLP<sup>®</sup>) is formed by arrays of such micromirrors, which can be switched individually in a display architecture.

### 9.1.3.3 Microdisplays

In the previous section, a reflective or transmissive display was implemented by the use of patterned ITO electrodes on an LC window. A microdisplay is basically a dynamic display matrix (reflective or transmission), the size of which has been reduced so that it can be used in small projection display devices (digital projectors). The LCD microdisplay is often used in transmission mode (LCD projectors or HTPS microdisplays), whereas the micromirror array is used in reflection mode (DLP<sup>®</sup> projectors) [12, 13]. When a DLP<sup>®</sup> chip is synchronized with a digital video or graphic signal, a light source and a projection lens, its mirrors can reflect a digital image onto a screen or other surface (see Figure 9.6).

A DLP<sup>®</sup> chip's micromirrors are mounted on tiny hinges that enable them to tilt either toward the light source in a DLP<sup>®</sup> projection system (ON) or away from it (OFF), creating a light or dark pixel on the projection surface (see Figure 9.5). The bit-streamed image code entering the semiconductor directs each mirror to switch on and off up to several thousand times per second. When a mirror is frequently switched on more than off, it reflects a light gray pixel, whereas a mirror that is switched off frequently reflects a darker gray pixel. In this application, the mirrors in a DLP<sup>®</sup> projection system can reflect pixels with up to 1024 shades of gray to convert the video or graphic signal entering the DLP<sup>®</sup> chip into a highly detailed gray-scale image.

The leading application for nonoptical MEMS devices is the airbag sensors that are now implemented in nearly every air bag in the automotive industry. This was the first industrial application of the MEMS technology.

### 9.1.3.4 Waveguide and PLC Platforms

Dynamic optical devices can be integrated in waveguide platforms, especially in Planar Lightwave Circuits (PLCs), as described in Chapter 3. Dynamical PLCs are important devices for today's DWDM optical telecom industry.

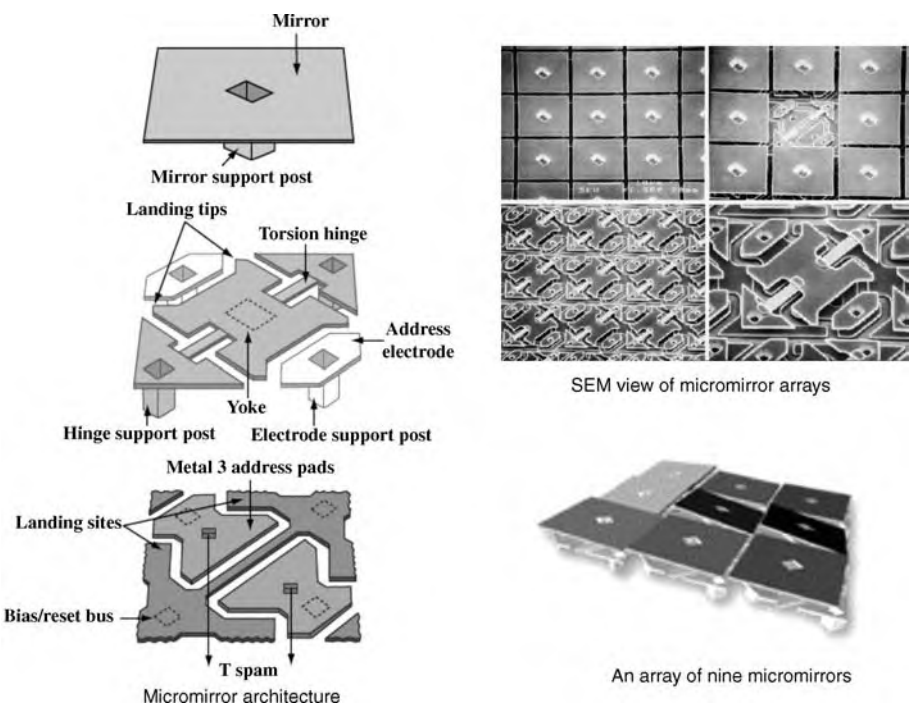


Figure 9.5 The MEMS micromirror architecture

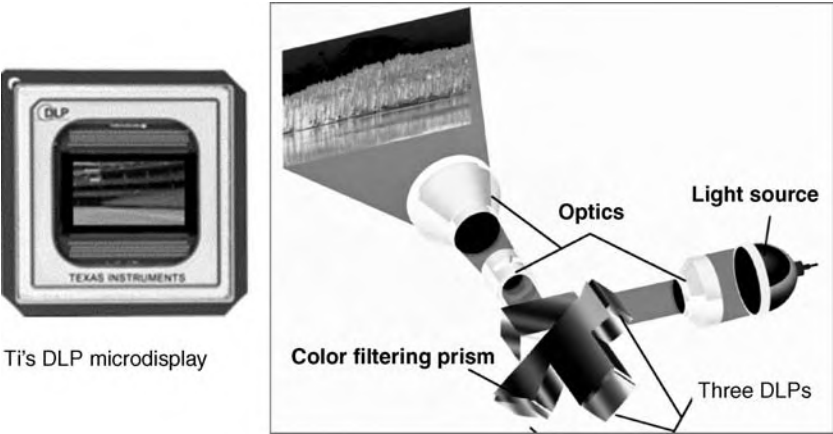


Figure 9.6 A DLP®-based digital projector (a three-chip design)



## 9.2 Switchable Digital Optics

When implementing a switchable digital optics element (e.g. a digital diffractive element [2]), one has to make sure that the Space Bandwidth Product (SBWP: see the definition in Chapter 6) of that device is sufficient for the diffraction task in hand. The SBWP is proportional to the number of individual pixels and to the number of states that such a pixel can implement:

- binary switchable optics can take on only two states, ON and OFF;
- analog switchable optics can have a large series of states between a minimal and a maximal state (e.g. ON, OFF and an entire set of gray levels in between).

Therefore, an analog switchable optical element has a much larger SBWP than a digital one. Digital and analog switchable optics will be reviewed in the following sections.

### 9.2.1 Binary Switchable Optics

#### 9.2.1.1 Polarization-selective Gratings

Polarization-selective gratings are fabricated in form-birefringent materials (Chapter 10), thus producing a specific refractive index for a specific incoming polarization state (see Figure 9.7). Such elements are binary elements when the incoming beam is either pure s or pure p polarized, but can also be analog switchable when the polarization state is set in between these states. A polarization-sensitive grating can implement either a polarization combiner or a polarization splitter (see Figure 9.7).

Form-birefringent CGHs, which integrate more complex optical functionalities, can be designed and fabricated. For example, such elements can integrate a set of fan-out beams, which can be switched from one state to the other (Figure 9.8). Such polarization CGH switches can be used in numerous applications, especially in optical computing and optical interconnections systems.

#### 9.2.1.2 Binary MEMS Digital Optics

Binary MEMS digital optics comes in various forms: by either moving a single digital optical element (mirror or lens) [14] or by moving several elements to produce a digital optical element (arrays of micromirrors) [15].

The most basic example consists of moving a digital optical element on a MEMS system by electrostatic forces. The SEM photograph in Figure 9.9 shows a MOEM device that incorporates a

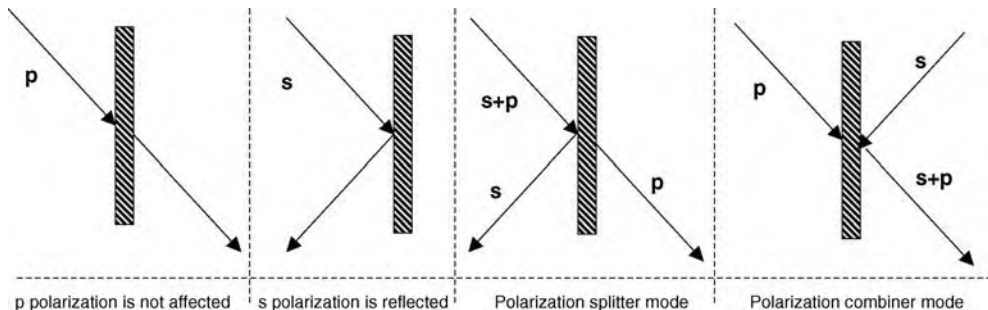


Figure 9.7 Polarization grating switches

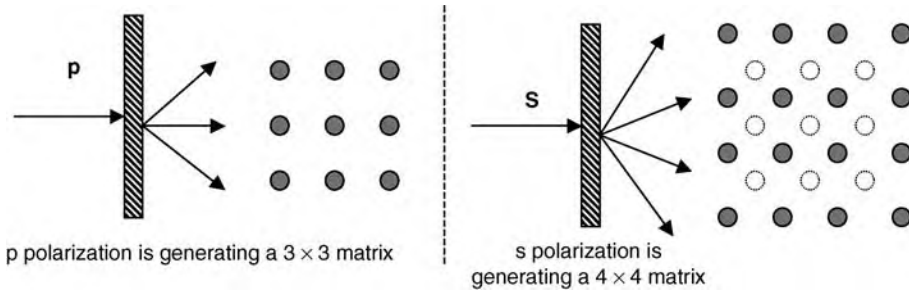


**Figure 9.8** A form-birefringent fan-out CGH

digital diffractive lens that can be moved upward by the use of four electrostatic combs etched into a polysilicon layer.

Another range of applications that fully utilize the potential of integrated optics and mechanics are the optical switches necessary in the deployment of high-density/high-bandwidth DWDM fiber networks. These switches need to be scalable and very rapid. Such switches are addressed by the use of the MOEM devices depicted below (Figure 9.10).

Such arrays can be pigtailed with single-mode fibers and use arrays of refractive lenses, diffractive Fresnel lenses (Figure 9.10) or GRIN lenses (see Chapter 4). These are used to collimate the beams for the free-space switching applications before re-coupling the beams into the fibers.

One of the main issues with free-space MEMS binary optical switches is the natural divergence of the beam, which has to be re-coupled into a fiber after being switched around by the micromirrors. Figure 9.11 shows how one can control the natural divergence of the free-space beamlets. Focusing the waist of the beam at mid-range between the two fibers, rather than collimating the beam, will ultimately reduce the coupling losses due to Gaussian beam divergence. These effects are more severe in 3D optical switches,

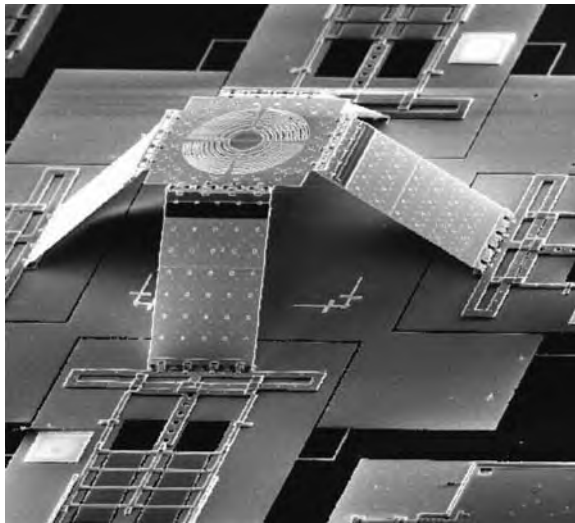


**Figure 9.9** A MOEMS switchable diffractive lens

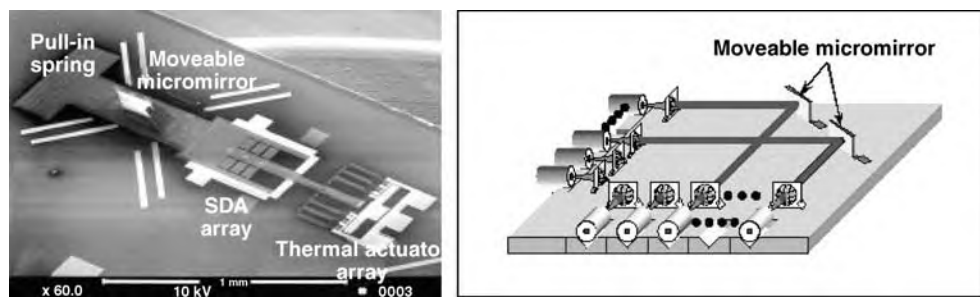


Figure 9.10 Two-dimensional optical switches integrated with MOEM micromirrors

where the mirrors can be oriented in many different angles, and the distance between the two fibers can greatly vary when compared to 2D switches.

9.2.1.3 Aperture-modulated Switchable Elements

The switchable digital diffractive element presented here consists of a compound element composed of an amplitude ferro-electric liquid crystal SLM. Chapter 5 has reviewed interferogram-type diffractive lenses, which can implement sets of diffractive lenses, including cylindrical, conical, helicoidal and doughnut lenses. This produces sets of focal elements (segments, curves etc.) in the near field (focused 3D wire-framed shapes).

As an example, a holographic animation composed of tens of space-multiplexed interferogram DOEs, arranged in a matrix, produces focal structures in a 2D plane (segments, curves and spots), as a pseudo-dynamic diffractive element compound. Figure 9.12 shows this implementation and the optical results. A spatial amplitude light modulator, placed in front of a matrix of interferograms with the same geometry, allows the laser light to be diffracted through specific parts of the elements to reconstruct specific segments and curves, thus resulting in a dynamic animation.

9.2.1.4 Binary MEMS Gratings

In the previous sections, single MEMS binary elements have been reviewed. Arrays of similar MEMS elements that can produce a digital optical configuration such as a linear grating will now be discussed.

Switchable blazed gratings can be implemented by arrays of binary micromirrors, which switch from an OFF position to an ON position as shown in Figure 9.13. Here, the diffraction angle is defined by the micromirror periodicity and not by the micromirror angle. The angle dictates the diffraction efficiency. In this case, only one diffraction angle can be set. If the micromirrors were to move in a vertical motion

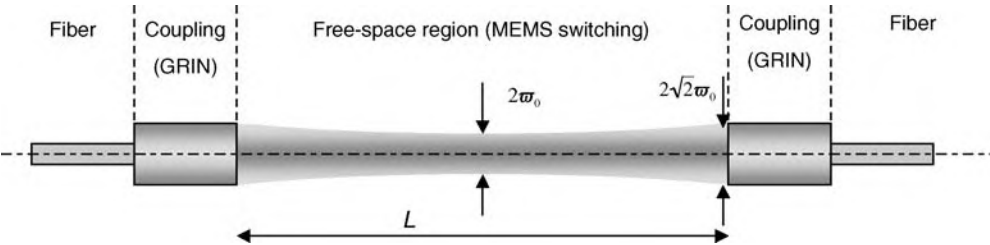


Figure 9.11 Reducing the divergence of the beam in free space

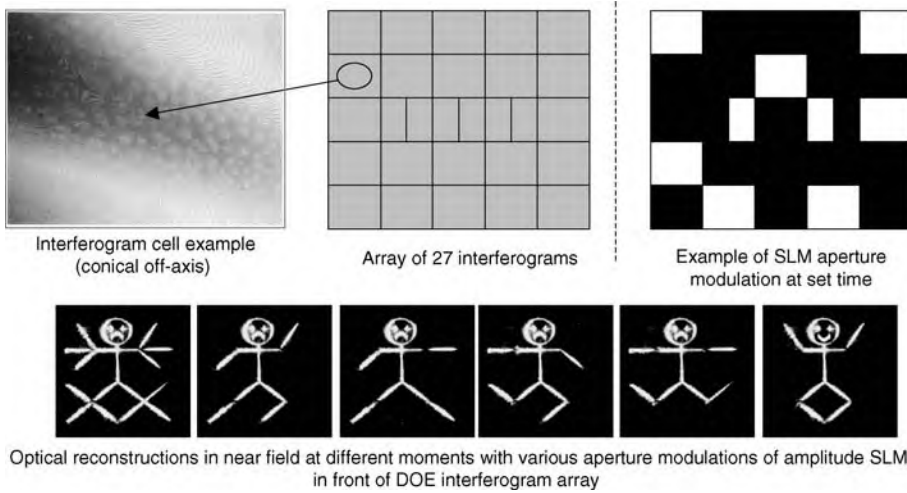


**Figure 9.12** An example of an array of interferogram-type DOEs producing, in a 2D plane, sets of linear and curve focal elements combined with an amplitude SLM aperture stop array to add animation

rather than in a rotational motion, one would get a reconfigurable CGH, each mirror becoming a CGH pixel that could take on various phase values (heights).

### 9.2.2 Analog Switchable Optics

Analog switchable optics can produce a range of effects in between two states (as gray shades between an OFF position and an ON position).

#### 9.2.2.1 MEMS Gratings

As seen previously, the linear arrays of micromirrors can implement switchable blazed gratings, and can act as beam deflectors [16–19]. If the angle of the micromirrors can be controlled in an analog

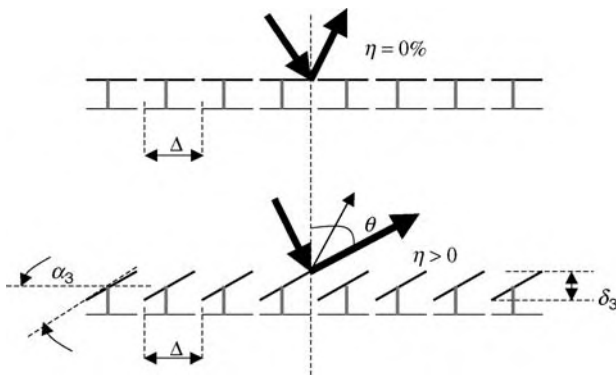


**Figure 9.13** Switchable blazed gratings

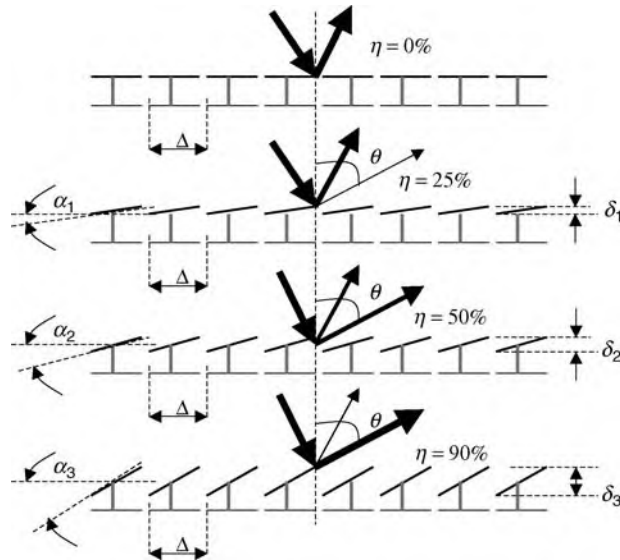


**Figure 9.14** An analog micromirror-based switchable blazed grating

way (e.g. by using torsion sensors in a closed-loop operation), the switchable blazed gratings can implement a desired diffraction efficiency that can be set between 0% and almost 100% (see Figure 9.14).

The angle of the mirror prescribes the amount of phase shift on the incoming beams, and therefore sets the diffraction efficiency (for more details of the efficiency calculation, see Chapter 5). Note that the diffraction angle is always the same, and does not depend on the angle of the micromirrors.

More complex functions, such as super-grating effects, can be implemented in a series of mirrors set to a specific angle, and then another series of mirrors set to another angle, and so forth. In this case, two diffraction effects will occur, one from the individual mirrors and the other one from the repetitive groups of mirrors.

MEMS gratings can also be integrated in another way, by using deformable membranes or actuated linear membranes [20]. These membranes can be set on a vertical movement (see Figure 9.15), thus implementing various binary grating diffraction efficiencies. Additionally, the diffraction angle can also be set by packing membranes together and thus forming super-membranes and allowing diffraction in a lower harmonic (see Figure 9.16).

Such elements are used as variable optical attenuators in many applications. The attenuation resides in the zero order (subtractive mode operation). The stronger the diffraction efficiency, the stronger is the attenuation in the zero order. Finally, due to their binary nature, these deformable membranes diffract in many different orders (see Figure 9.17 and Chapter 5), from the conjugate to the zero reflected order. This is also the reason why such a device is used in subtractive mode rather than in additive mode (where the signal would consist of the fundamental positive order, for example).

If the membrane were to be controlled in an analog way, more optical functionalities could be implemented, as discussed in Figure 9.17, where three examples are shown: a linearly variable efficiency binary grating (for spectral shaping), a blazed linear grating with variable efficiency and quantized variable periods (harmonics), and a reflective refractive lens. One aspect to consider is the presence of the diffraction effects arising from the unavoidable gaps in between the mirrors (as shown in Figure 9.16).

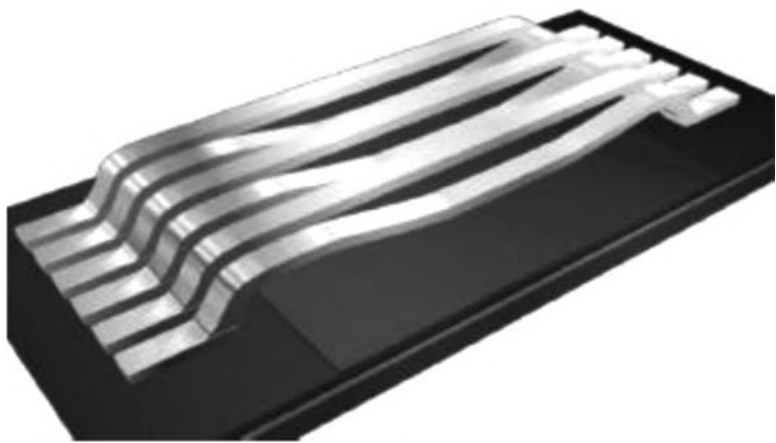


**Figure 9.15** A deformable MEMS membrane dynamic switchable grating

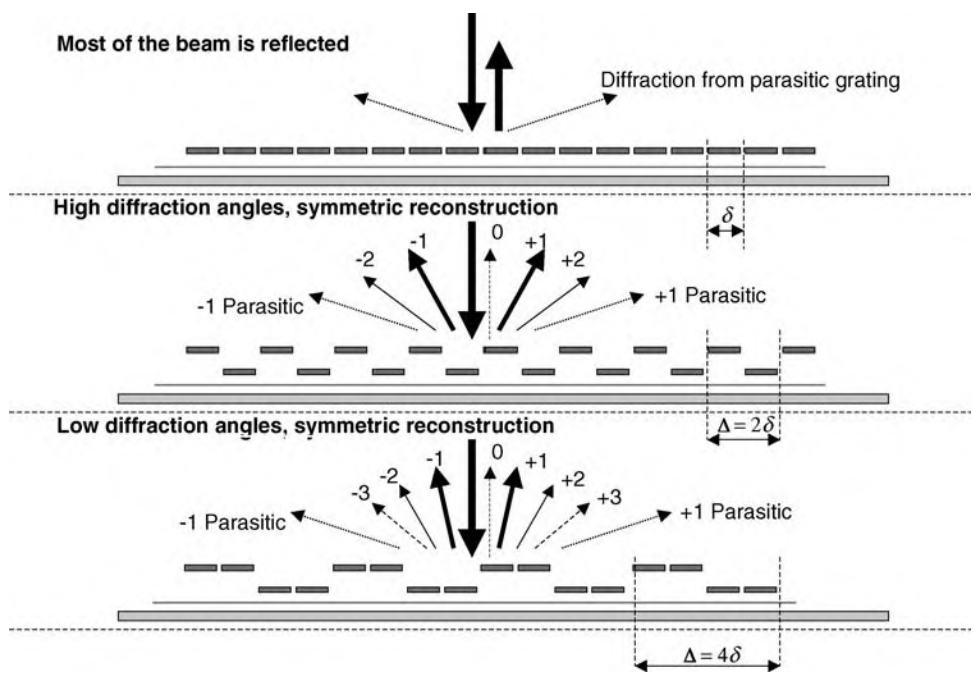


**Figure 9.16** The use of a deformable MEMS membrane device to implement various harmonic angles

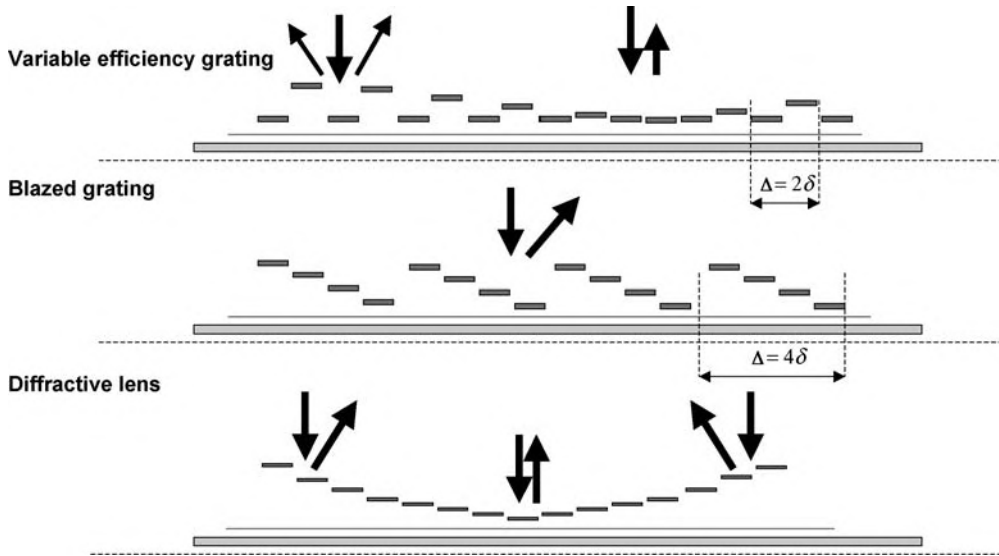


**Figure 9.17** A deformable MEMS membrane with analog height control

Applications using deformable membranes include spectral plane filtering (for dynamic gain equalizers in DWDM devices by using the first example in Figure 9.17), laser displays, laser printing and so on.

### 9.2.2.2 Liquid Crystal Switchable Optics

Switchable optics implemented via liquid crystal techniques will now be reviewed [21–24].

#### *Switchable Refractives*

In the previous section, the bulk LC effect was used to modulate a refractive index between two ITO electrode surfaces. If the ITO electrodes are layered on a plano-convex geometry filled with LCs, a switchable lens can be created (see Figure 9.18). If the voltage is set in a binary way, the lens is a binary switchable lens. If the voltage is set in an analog way, such a lens becomes a tunable-focus lens.

Figure 9.19 shows the example of a switchable prism, where the prism angle varies as a function of the applied voltage. This prism can also be placed within another medium with a similar index; thus in an OFF position, it acts like an inert window.

One has to remember that this only occurs for a specific polarization. For the other polarizations, there is only one lensing effect occurring, whatever voltage is applied. There is only one (refracted) beam in a switchable refractive optical element, as opposed to two or more (diffracted) beams in the switchable gratings presented earlier. With switchable refractive elements, the optical functionality is therefore much ‘cleaner’. There are no numerous diffraction orders present, as with switchable and tunable diffractive elements.

#### *Switchable Diffractives*

Similarly to a refractive switchable element, a switchable diffractive element can be implemented with the bulk LC effect [21–24]. The internal surface of a simple LC cell is etched with a specific diffractive element (e.g. a diffractive lens as depicted in Figure 9.20).

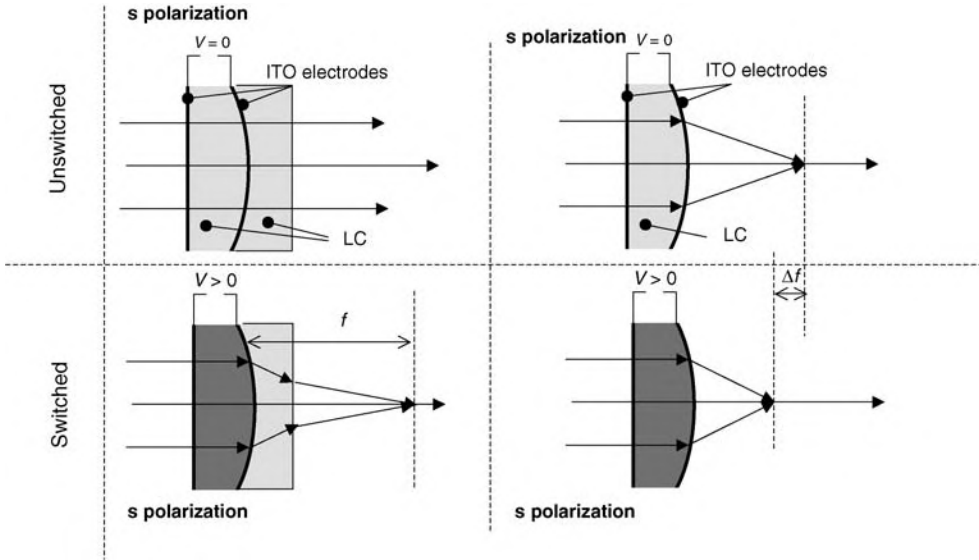


**Figure 9.18** A switchable refractive lens with a bulk LC effect

As the diffractive elements are very thin, a polarization recovery scheme can be implemented to use both polarizations (see Figure 9.21), and thus the device can be used with unpolarized light. Such a polarization recovery scheme is difficult to implement in the previously mentioned switchable refractive elements, due to their size and geometry (even in a plano-convex form).

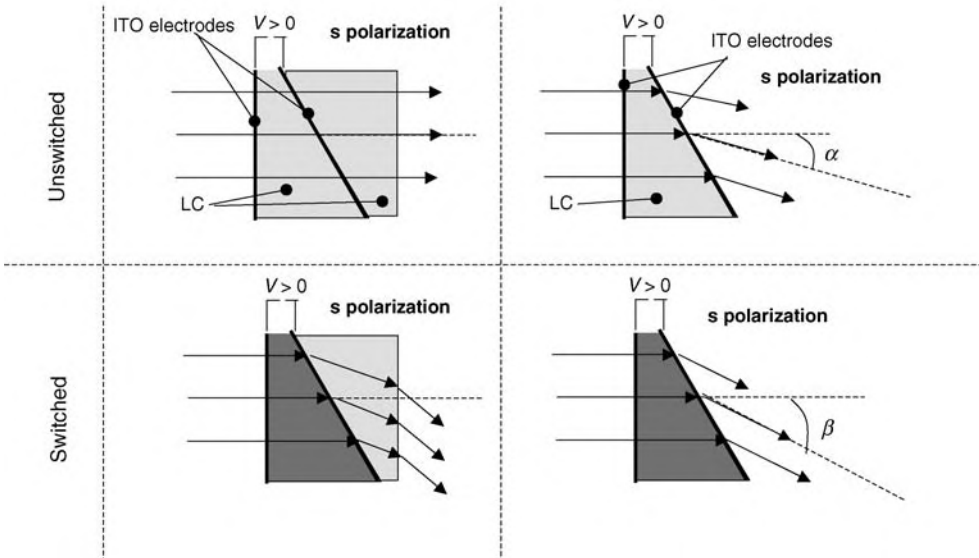


**Figure 9.19** A switchable prism with a bulk LC effect



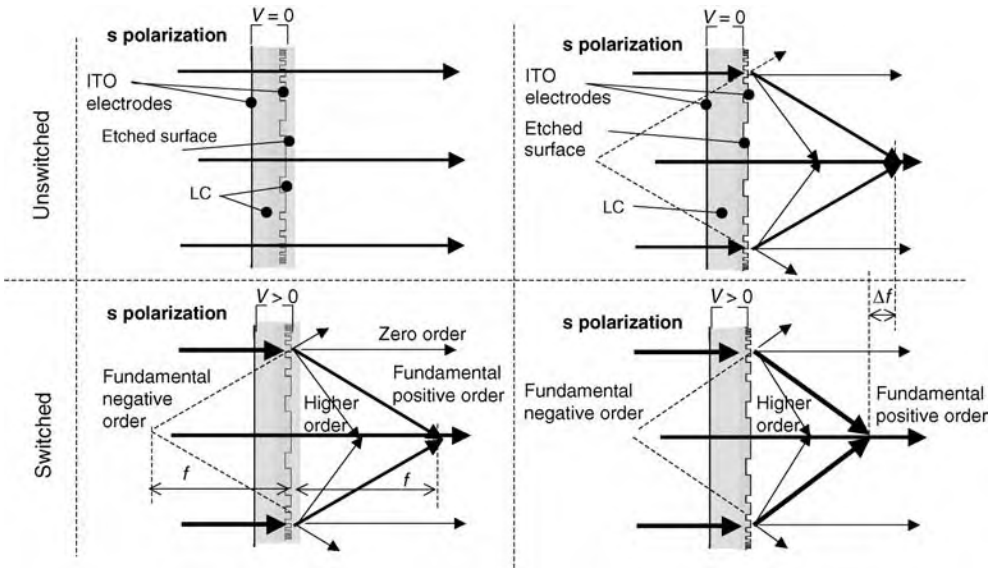


Figure 9.20 A switchable diffractive element with a bulk LC effect

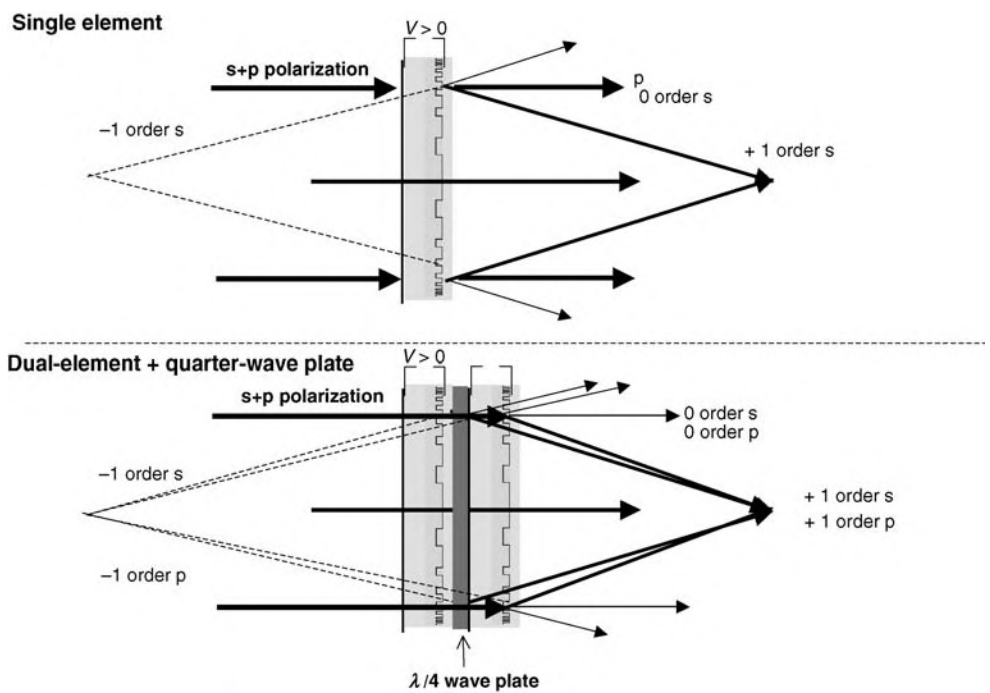


Figure 9.21 A polarization recovery scheme in a dual switchable diffractive element with a bulk LC effect

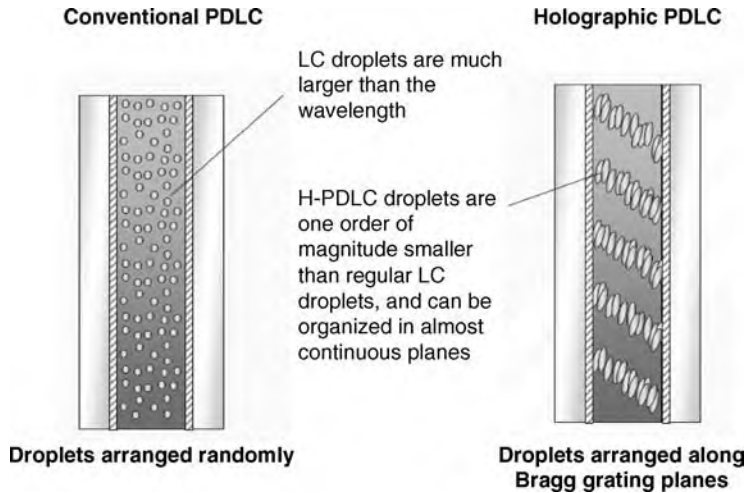


**Figure 9.22** PDLC and LC droplets. Reproduced by permission of Jonathan Waldern, CEO SBG Labs Inc.

### 9.2.2.3 H-PDLC Switchable Holograms

Polymer Dispersed Liquid Crystals (PDLCs) are nanodroplets of liquid crystals, smaller than regular liquid crystals, dispersed in a monomer solution [25]. Holographic-PDLC (H-PDLC) is a technique in which a hologram is recorded in the PDLC material. Figure 9.22 shows the difference between H-PDLC and regular PDLC droplets. Such H-PDLCs can be recorded holographically in the same way that HOEs are recorded (see Chapter 8). Such H-PDLCs can be reflective or transmission HOEs, and they implement a variety of holographic functionalities.

During holographic recording (see Figure 9.23), the initial solution liquid monomer and liquid crystal undergoes a phase separation, creating regions that are densely populated by liquid crystal micro-droplets, interspersed with regions of clear polymer.

The grating created has the same geometry as the interference fringe pattern created by the original recording laser beams. Typical H-PDLC hologram Diffraction Efficiency (DE) can exceed 80%, and because of the small droplet size, switching speeds are typically tens of microseconds. When an electric field is applied to the hologram via a pair of transparent electrodes (ITO), the natural orientation of the LC molecules within the droplets is changed, causing the refractive index modulation of the fringes to reduce and the hologram diffraction efficiency to drop to very low levels, effectively erasing the hologram by switching it into the OFF state. Removal of the voltage restores it to its passively ON state.

H-PDLCs can implement similar optical functionalities in the form of other HOEs, namely spectral filters, beam deflection gratings, diffractive lenses that incorporate specific aberrations or optical correction, diffusers, structured illumination generators, fan-out gratings and so forth. As for HOEs (see Chapter 8), one can use a master CGH in order to record a specific H-PDLC.

SBG Labs, Inc., of Sunnyvale, California ([www.sbg labs.com](http://www.sbg labs.com)), has developed a higher-index, nano-clay enhanced H-PDLC material system, which the company calls a Reactive Monomer Liquid Crystal Mix (RMLCM). This system enables the recording of dynamic holograms at 95% DE and lower voltage (8 V). Using the higher-index material, SBG have pioneered the recording of complex binary optics without reducing DE, and are bringing to the market several new applications, principally in the areas of display and imaging. Their new RMLCM-based optics provide multiple functionalities (beam shaping,

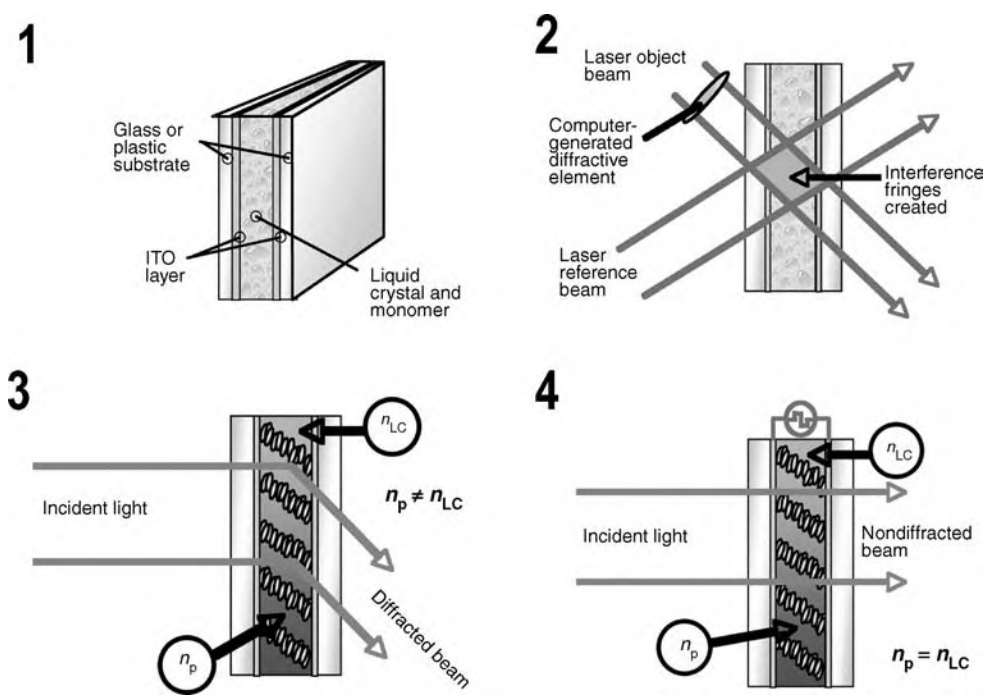


**Figure 9.23** The holographic recording of a H-PDLC. Reproduced by permission of Jonathan Waldern, CEO SBG Labs Inc.

homogenization, laser de-speckling) in a compact form factor, and may be most relevant in low-cost ‘pico’ microdisplay-based cell phone projection modules.

Figure 9.24 shows a light beam (blue light) first being passed unperturbed through a hologram in its OFF state and then, in the lower part of the figure, being deflected and focused by the diffractive optical power of the hologram in the ON state.

In the ON mode, no voltage is applied. The LCs droplets are essentially aligned normal to fringe planes. The diffraction occurs only for the s polarization state. For p polarization, we have

$$\Delta n = n_e - n_p \tag{9.1}$$

and for s polarization

$$\begin{cases} \Delta n = n_p - n_0 \\ \Delta n = 0 & \text{if } n_p = n_0 \end{cases} \tag{9.2}$$

In the ON state, an AC voltage is applied, which orients the optical axis of the LC molecules within the droplets to produce an effective index that matches the polymer refractive index, creating a transparent cell.

The switching voltage and switching time are very important features of the H-PDLC. Figure 9.25 shows the switching time for a typical H-PDLC holographic transmission grating. A typical switching time to achieve 95% of the response is around 30–50 μs.

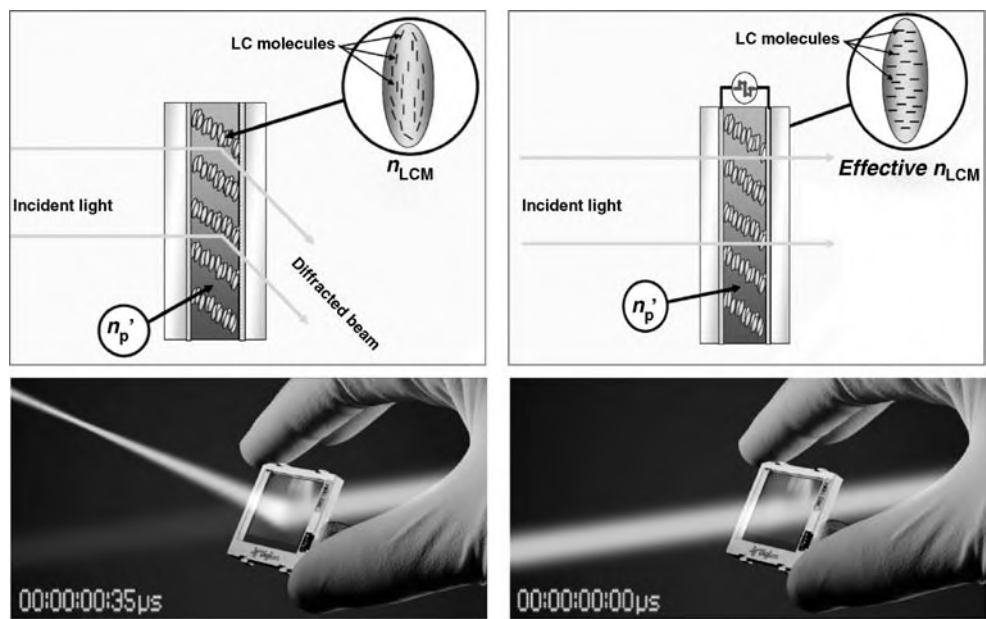


**Figure 9.24** An example of the implementation of an off-axis diffractive lens in an H-PDLC. Reproduced by permission of Jonathan Waldern, CEO SBG Labs Inc.

The replication of such elements is also facilitated thanks to the developments in the LC and LCD replication industry, with which H-PDLC fabrication is compatible. The H-PDLC cell origination is similar to that of LCD cells (performed since the 1970s). Very little material is required (a liter would be enough for millions of 1 inch square elements). The cell filling is also identical to that of an LCD; and for the holographic recording, the technique is identical to conventional holographic reproduction in industry.

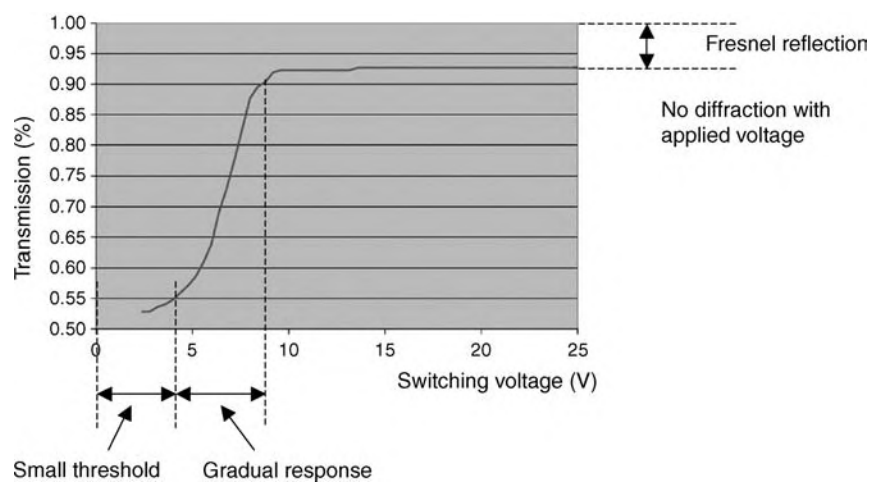


**Figure 9.25** A switching curve for a typical H-PDLC holographic transmission grating

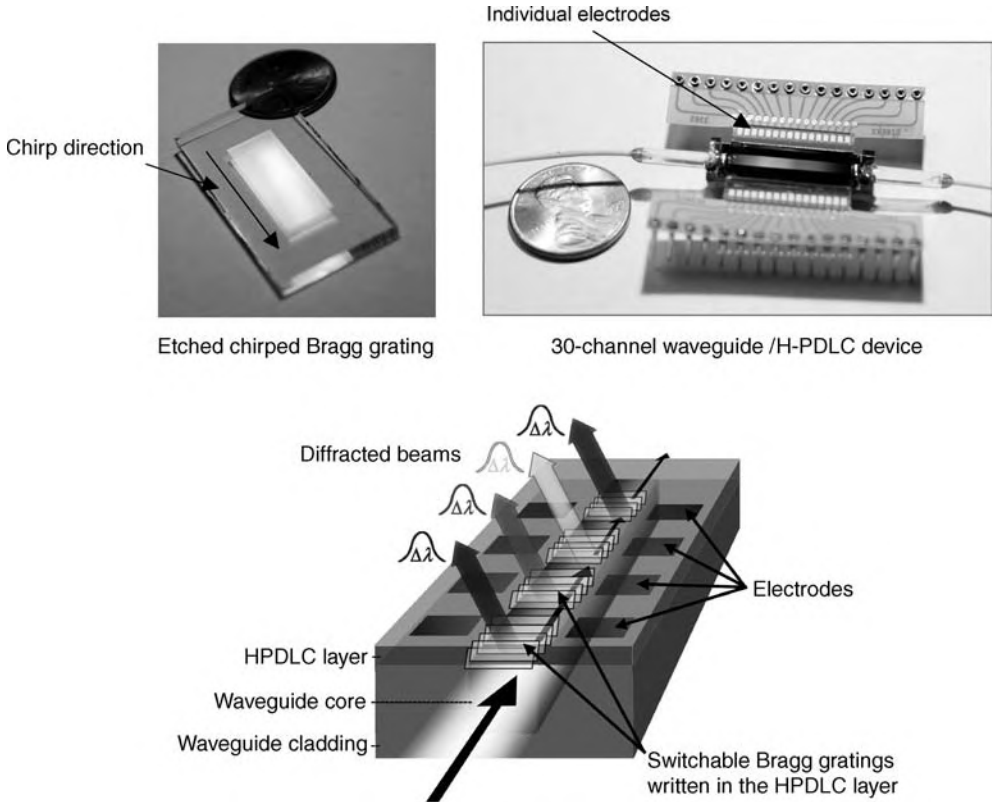


**Figure 9.26** Waveguide H-PDLC implementation in the form of a dynamic gain equalizer

H-PDLCs can be integrated on a waveguide platform, as switchable Bragg grating couplers. Such a device can be used for dynamic gain equalization after an EDFA amplifier (see also Chapter 16). In this application, the H-PDLC is set as the cladding around a diffused or ridge-core waveguide. The H-PDLC is then exposed as a chirped Bragg grating through a phase grating etched into a quartz plate (see Figure 9.26), similar to fiber Bragg exposure.

Various electrodes located along the chirped grating can activate specific grating periods to extract specific wavelengths. By modulating the voltage, the Bragg coupling efficiency can also be modulated; therefore, a spectrum of channels propagating in that waveguide can be equalized. Such spectrum deformations occur when an EDFA amplifier is used in the line.

### 9.3 Tunable Digital Optics

Tunable digital optics are optics in which one of the optical parameters can be tuned continuously from one state to the other (e.g. diffraction efficiency or diffraction angle). The other optical parameters remain unchanged (e.g. diffraction angle or diffraction efficiency).

#### 9.3.1 Acousto-optical Devices

An Acousto-Optical Modulator (AOM) is a form of dynamic linear volume (or Bragg) grating, controlled by an array of piezo-electric transducers. This type of transducer is electrically addressed and is placed

on top of a crystal material ( $\text{TeO}_2$  crystal, for example). The acoustic waves are generated by the RF/piezo-transducers propagating through the crystal, which create local refractive index modulations. The frequency of the driving AC voltage defines the period of the traveling sinusoidal phase gratings inside the crystal. Therefore, AOMs are actually tunable digital elements that can deflect light into a desired direction with a desired efficiency, by controlling the frequency as well as the amplitude of the piezo-electric transducer. By using two AOMs placed orthogonally to one another, separable two-dimensional phase functions can be implemented dynamically at a very high rate (e.g. a 2D scanning function).

Applications include high-speed beam steering and beam modulation (e.g. in laser beam patterning systems), and also full color holographic video, by using three RGB lasers and three sets of orthogonal AOMs (see Hologvideo, the first holographic video, developed by Professor Steve Benton and Pierre St-Hilaire at the Media Lab, Massachusetts Institute of Technology, in the early 1990s). Acousto-optical modulators may actually be the oldest dynamic diffractives yet integrated.

### 9.3.2 *Electro-optical Modulators*

Electro-Optical Modulators (EOMs) are similar to AOMs in the way they modulate local refractive indices by the use of electrically applied transducers. Such modulators can be integrated either in free space or in waveguide platforms. Most of the applications are actually based on PLC architectures.

#### 9.3.2.1 **Free-space EO Modulators**

Here, the transducers are electrodes placed on each side of a substrate or crystal. These electrodes can be patterned on the substrate by lithography; hence the geometry of the structures can be very complex and is not restrained to linear geometries, as is the case in AOMs.

#### 9.3.2.2 **Integrated-waveguide EO Modulators**

Integrated waveguides and PLCs have used EO modulators extensively. Chapter 3 has investigated some these architectures, such as Mach-Zehnder interferometric modulators (switches, directional couplers etc.). Very complex dynamic functionality integration can be performed on these OE waveguides, namely dynamic add-drop devices or dynamic cross-connects for high channel count DWDM networks. More complex routers and switches can also be implemented using approach (for more details of these architectures, see Chapter 3).

### 9.3.3 *Mechanical Action Required*

In the previous section, dynamic digital optics have been implemented using MEMS technology, where mechanical action is conducted by means of tiny electrostatic motors (micromirrors etc.). Dynamic optics can also be implemented in a macroscopic way, by moving one element or by moving another element with regard to the other. This movement can be linear or rotational. Movements that do not alter the volume of the optical element – such as lateral rotation, and vertical and lateral linear movements – are of great interest, rather than traditional longitudinal movements (as in the case of traditional zoom lenses).

#### 9.3.3.1 **Pulling and Stretching**

A conventional approach to implementing macroscopic dynamic optics is to deform the optical surface in real time with electro-mechanical transducers such as pistons. Adaptive telescopes have primary mirrors that can be deformed by arrays of pistons placed underneath various sectioned parts of the mirror, which are controlled in real time to compensate measured wavefront aberrations. Shack-Hartmann wavefront sensors are used to measure these aberrations in real time, as shown in Chapter 4 (based on

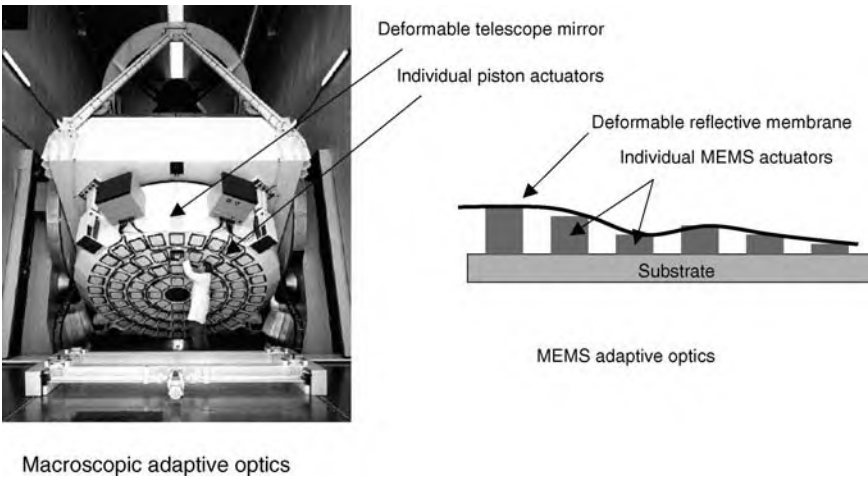


**Figure 9.27** The implementation of macroscopic and microscopic adaptive optics

microlens arrays). Based on this architecture, adaptive optics have also been implemented at the microscopic scale, on digital optics (see Figure 9.27).

Bragg waveguide grating couplers, which become dynamic under stress, can act as Variable Optical Attenuators (VOAs) or Dynamic Gain Equalizers (DGEs) in DWDM optical telecom networks.

An etched Bragg grating on a piezo-electric actuator can be placed close to a fiber core, in order to implement a variable Bragg coupler (or a weakly diffused core) (Figure 9.28). As the grating approaches the core, the Bragg coupling effect gets stronger, and the signal in the core is attenuated.  $N$  Bragg gratings with different periods (acting on  $N$  different wavelengths) placed on  $N$  individual actuators can produce a DGE instead of a simple VOA.

Other examples include strain sensors, which are basically fiber Bragg gratings, that are stretched and have variable periods. The variation in the periods is related to the amount of Bragg reflected signal for a specific wavelength, or the amount of Bragg coupling (losses) for a specific wavelength. Such sensors are implemented in large infrastructures; for instance, in bridges to sense small movements.

Stress in the form of pressure and movement has been mentioned in the previous applications. Stress in fiber Bragg gratings can be implemented by increasing the temperature, therefore dilating the gratings

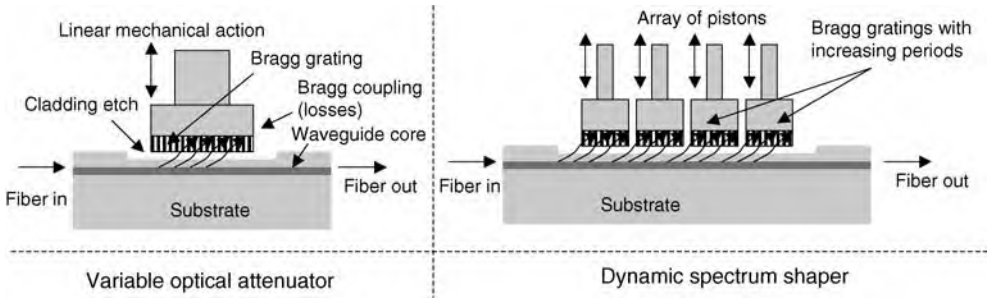


**Figure 9.28** Dynamic VOA or DGE by moving a Bragg grating with regard to a waveguide core

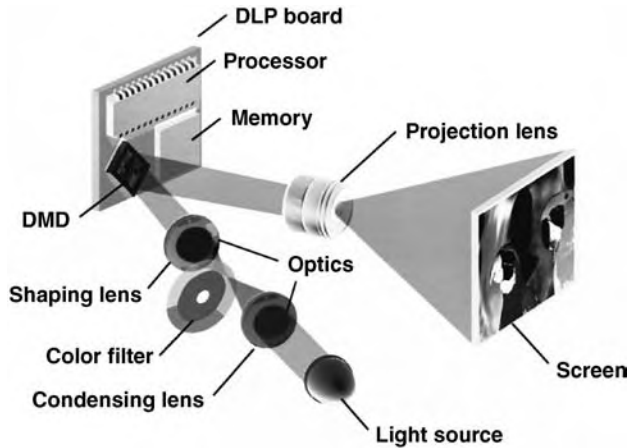


**Figure 9.29** A rotating color filter in a single-chip digital projector

and increasing the grating period. This exemplifies a method of implementing wavelength Bragg selectors in laser diodes, where the output wavelength can thus be tuned over a specific range.

### 9.3.3.2 Translation and Rotation

The potential to tune an optical functionality by using a movement that does not alter the overall size of the optical system, namely lateral translation or lateral rotation, without using any longitudinal translation (as is done in traditional zoom lenses, for example) will be reviewed in this section.

Rotating an optical element without a nonrotationally symmetric pattern is a straightforward approach to producing a dynamic tunable optical element. Such rotation or translation can be performed as a single element or as a differential movement from one element to another.

#### *Single Digital Element*

##### *Color Filter Wheels*

Color filters in digital projectors are a simple example of a dynamic optical element. Figure 9.29 shows the color wheel implemented in a single-chip DLP projector. A wheel can be made of dichroic filters or holographic gratings.

H-PDLC switchable holographic elements can also implement a color filter wheel, by using a sandwiched stack of three H-PDLC elements that have different Bragg angles for different wavelengths (for the theoretical background, see Chapter 8). Such an element is a mechanically static but nevertheless dynamic color filter. Such a filter works in subtractive mode, diffracting the red and blue parts of the spectrum. There will be always two elements ON at each time. All three elements ON would blank the beam.

##### *Speckle-reducing Diffusers*

Speckle can be reduced in laser applications (displays, sensing etc.) by using a rotating diffuser wheel. The diffuser wheel produces the same diffusing cone but creates a multitude of different speckle patterns, which are smoothed out temporally during the integration time of the detector (or the eye).

The diffuser in Figure 9.30 has been designed using an IFTA algorithm presented in Chapter 6, and fabricated by lithography (see Chapter 12). The replication has been done by plastic embossing



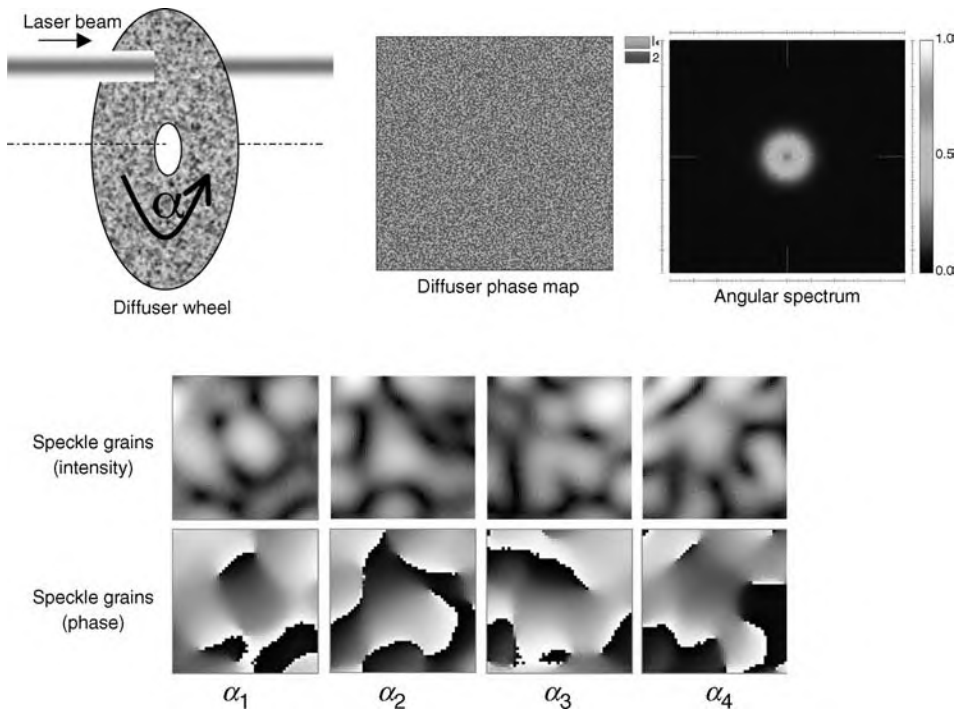


**Figure 9.30** A rotating diffuser wheel to reduce objective speckle

(see Chapter 14). Typically, such diffusers have a very low diffusion angle ( $\pm 0.25^\circ$ ) and are isotropic. When used in display applications, care must be taken to ensure that the speckle reduction scheme only works on the objective speckle. The subjective speckle needs to be tackled by some other means.

*Rotational Encoders*

Incremental and absolute rotational encoders can be implemented as a succession of different CGHs, producing different binary codes in the far field (or on a detector array). For example, this diffraction pattern code can be a binary Gray code. Chapter 16 describes such encoders and compares them to classical optical encoders. Figure 9.31 illustrates an example of an encoder etched into a quartz substrate. An element in rotation can be considered to be a dynamical digital optical element, since the CGH presented to the laser beam changes as the disk turns.

Chapter 16 shows an example of a dynamically structured illumination projector (which projects sets of fringes), used to remotely acquire 3D shapes and 3D contours. Arbitrary repetitive optical patterns can also be implemented in order to produce a repetitive small diffractive video for metrology applications, or even for entertainment applications.

*The Helicoidal Lens*

A helicoidal diffractive lens is a lens that has a continuously circularly variable focal length (see Chapter 5). When a laser beam is launched off-axis to the lens and the lens turns on its optical axis, a prescribed and repetitive off-axis imaging functionality can be implemented on that beam (the beam focuses back and forth between a minimal and a maximal focus value, in off-axis mode) – see Figure 9.32.

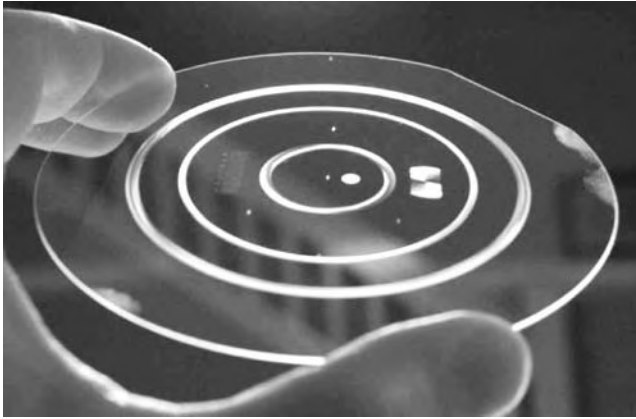


**Figure 9.31** An absolute rotational diffractive encoder

### *The Daisy Lens*

A Daisy lens is very similar to a helicoidal lens; however, the variations of the focus have angular periods narrower than  $360^\circ$ . Several examples of Daisy diffractive lenses are given in Chapter 5.

### *Sandwiched Digital Elements*

#### *Beam Steering with Microlens Arrays*

Chapter 4 described a beam-steering example that used two arrays of microlenses in linear motion. Such beam steerers are now implemented in many applications.

### *Moiré Diffractive Optical Elements (M-DOEs)*

A desirable feature in an imaging system is to vary its focal length without having to vary the overall size of the compound lens system. This approach has been reviewed with the helicoidal lens in the previous section. However, only a component of the system could be used. It is very desirable to use the entire lens aperture in imaging applications.

Recently, an elegant solution involving stacked phase plates has been introduced by using the moiré effect over sandwiched DOE lenses. The focus changes as a function of the angle of rotation,  $\alpha$ , of one plate to the (see Figure 9.33). These elements are called Moiré DOE (M-DOE), and can be applied to

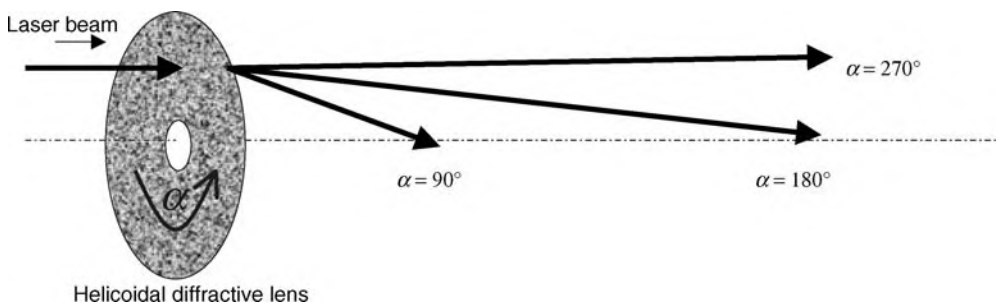


**Figure 9.32** Dynamic focus in a helicoidal diffractive lens

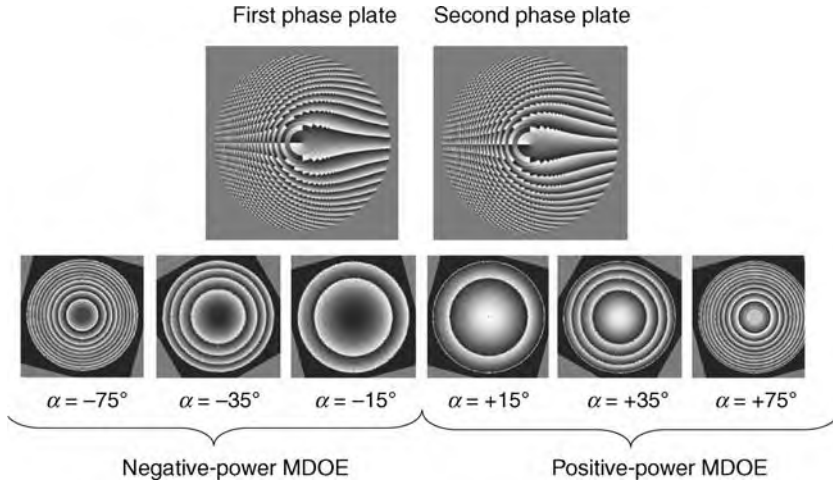


**Figure 9.33** Moiré (M-DOE) tunable-focus diffractive lenses. Reproduced by permission of Prof Stefan Bernet

functionalities other than on-axis spherical phase profiles (beam-shape modulation, spherical aberration modulation etc.).

### 9.3.3.3 The Hybrid Fresnel/Fourier Tunable Lens

Here, the task is to produce a tunable-focus lens by integrating a Fourier CGH and a synthetic aperture diffractive lens. The Fourier element is used in reflection mode and produces a circular array of  $N$  beams. These  $N$  beams are collimated and fall onto a set of  $N$  synthetic apertures on a second reflective substrate, each encoding diffractive helicoidal lenses. When the second substrate (synthetic aperture lenses) turns, the  $N$  beams hit the helicoidal lenses and diffract the beam into a wavefront with a specific convergence. The focus position varies with the angle. The angular range is therefore limited to  $360^\circ/N$ . This example is described in Figure 9.34, where  $N = 4$ . The apertures of the four helicoidal lenses are shaped in a sausage geometry.

The  $N$  beams are diffracted by the set of  $N$  helicoidal lenses, and recombine to interfere at a single focus. A sharp spot (however highly aberrated) is thus produced. This example was used in space-folded mode (both elements reflective), with a high-power 2 kW CW YAG laser, for laser material processing, where the focus had to be continuously modulated to adapt to the 3D contour of the workpiece to be cut. Experimental results are shown on photographic paper using  $N = 8$ . These eight beams interfere at the focus.

### 9.3.4 Tunable Objective Lenses

Tunable objective lenses have recently gained a lot of interest, mainly due to two consumer electronic products that can benefit greatly from these features:

- camera cell phones (zoom lenses); and
- backward compatibility in Blu-ray Optical Pick-up Units (OPUs).

It should be noted that the term ‘tunable’ is often mistaken with ‘switchable’. In a switchable lens, the focal length cannot be tuned – only the efficiency of that lens can be tuned (as in H-PDLC switchable lenses). In

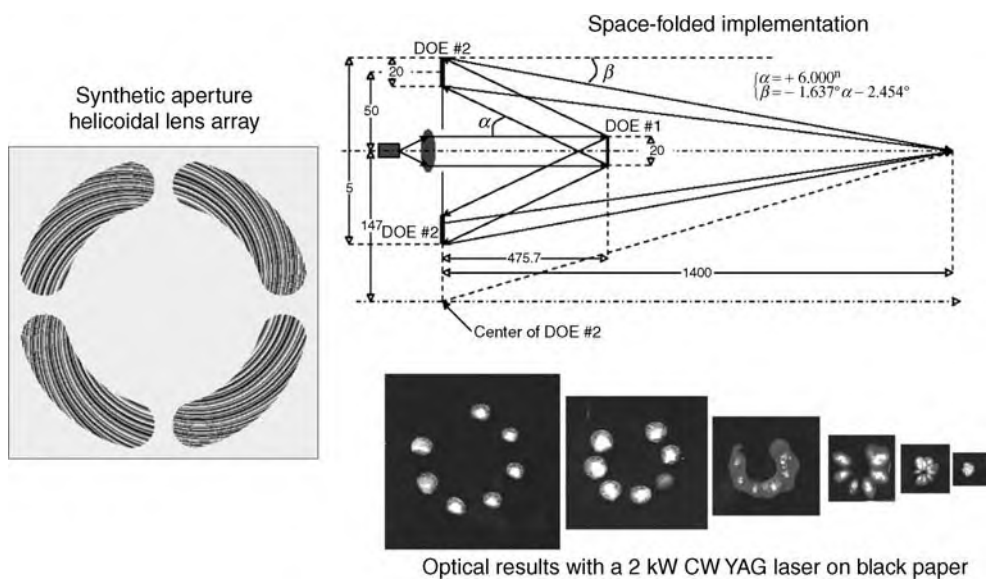


**Figure 9.34** A dynamic focus diffractive compound lens

a tunable lens, the focal length can be varied smoothly between a minimal and a maximal value. Zoom lenses in camera phones are potential applications of such lenses. Tunable lenses to be used in backward-compatible disk drive OPUs are defined by the various wavelengths to be used, the various spherical aberrations to be compensated (variable-thickness media) as well as the various NAs to be produced. Chapter 16 reviews OPU applications.

The following section will discuss some of the techniques used today to implement such lenses, especially for camera phone zoom lens objectives.

**9.3.4.1 MEMS-based Lenses**

Moving the lens on top of the CMOS display is the most straightforward technique for implementing a zoom functionality with a fixed plastic lens. The vertical movement can be implemented with piezo-electric actuators or MEMS actuators in numerous ways. Figure 9.35 shows three different ways in which such a movement can be implemented.

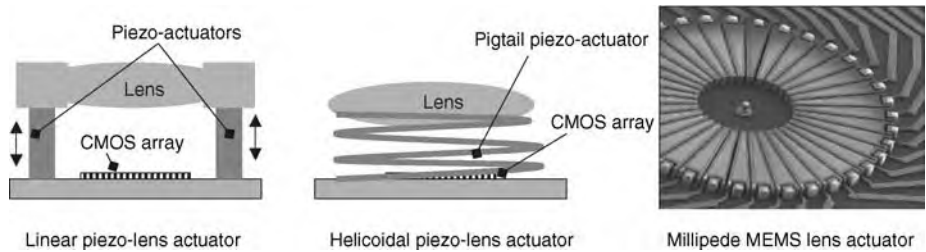


**Figure 9.35** Mechanically actuated zoom lenses in camera phone objectives

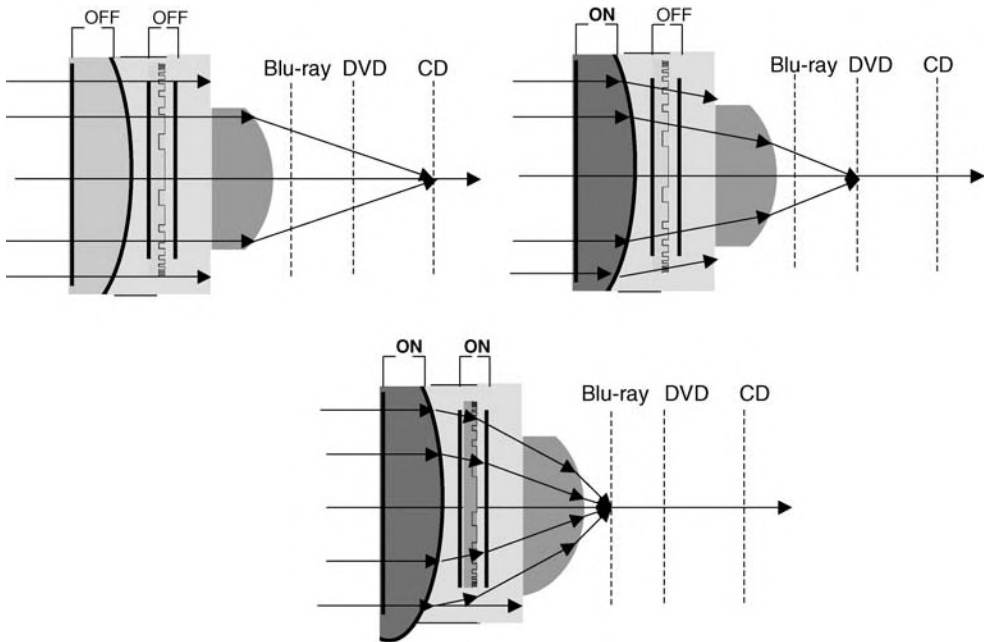


**Figure 9.36** A switchable dual- or triple-focus objective lens for a Blu-ray/DVD/CD OPU

The first example in Figure 9.35 shows a linear piezo-actuator moving the lens up and down. The amplitude range of such an actuator is limited. The second example shows a piezo-actuator built as a pigtail (helicoidal structure). This multiplies the dynamic range. The third example is a MEMS actuator, constituted by numerous small latches, which can move a lens up and down (also called the millipede lens actuator).

#### 9.3.4.2 LC-based Lenses

In the previous section, a mechanical method for implementing lens movement for zoom applications has been reviewed. A static solid state solution (no moving parts) is, however, almost always preferred to a moving-part solution, particularly for consumer electronics applications (smaller, cheaper, enhanced functionality). A method for implementing a switchable lens has been discussed earlier in this chapter, in the form of the bulk LC index change effect. Figure 9.36 shows another way to implement a switchable dual-focus, or even triple-focus, lens. These can switch in a binary manner between two or three different lens configurations, for different wavelengths.

In Figure 9.37, two electrically addressable LC elements are integrated with a static lens. The first electrically addressable lens is refractive, and the second one is diffractive while the static lens is refractive. The diffractive lens is not used here for chromatic aberration control, as discussed in Chapter 7, but for an additional aspheric lensing effect. When no voltage is applied, the OPU produces the CD spot with the target NA and for the target wavelength. When the first voltage is applied, the focal length reduces, the NA increases and the spherical aberrations are corrected for a DVD medium half the size of the CD ( $600\text{ }\mu\text{m}$  instead of  $1200\text{ }\mu\text{m}$ ). More light is coupled into the DVD spot. When both voltages are ON, the OPU produces the Blu-ray spot for  $405\text{ nm}$ . The additional lens not only reduces the focus but also corrects the aberrations for the reduced Blu-ray disk thickness of  $100\text{ }\mu\text{m}$ .

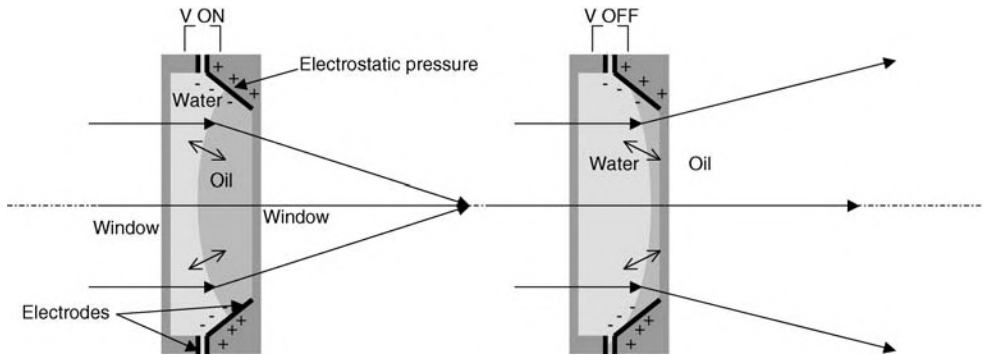


**Figure 9.37** The electro-wetting process in a tunable liquid objective lens for a camera phone

### 9.3.4.3 Electro-wetting in Liquid Lenses

Recently, a novel concept for a switchable refractive lens has been introduced based on the electro-wetting effect in liquid lenses. The concept of using liquid lenses is quite old. Transparent plastic or glass containers filled with fluids have been used as lenses for over 100 years. Cylindrical liquid lenses have been used in large display applications, and large turning liquid-filled basins with a reflective liquid (liquid mercury!) today implement very large telescope mirrors. The shape of the lens can be changed, thus changing the focal length of the lens (see Figure 9.37).

When oil and water are placed in a hermetically sealed tank, a liquid lens is formed at the interface (oil has a larger refractive index than water). Electrodes are then placed on each side of the oil droplet. Depending on the voltage applied, the electrostatic pressure increases in the region of the electrodes. This changes the shape of the oil droplet due to the electro-wetting process. The surface tension is reduced, and the focal length of the resulting compound lens is changed. When a continuous voltage is applied, the change can be continuous. The changes in the lens shape can be quite large, from a converging lens effect to a diverging lens effect.

## 9.4 Reconfigurable Digital Optics

Completely reconfigurable digital optics appears to be the '*nec plus ultra*' in terms of dynamic digital optics. Such a totally reconfigurable element could implement a lens at  $T = T_0$  and a Fourier CGH at  $T = T_1$ . However, this does not imply that these are the best-suited techniques for all the applications listed previously. A totally reconfigurable digital optical element is basically a reconfigurable diffractive element, where all the pixels can be reconfigured in real time. Microdisplays are some of the technologies used for these tasks. The various microdisplay technologies are reviewed below.

### 9.4.1 Microdisplay Technology

The various technologies that can be implemented as reconfigurable digital elements are as follows:

- MEMS microdisplays;
- LC-based microdisplays;
- LCoS microdisplays; and
- H-PDLC microdisplays.

### 9.4.1.1 MEMS Grating Light Valves (GLV)

MEMS micromirror or micro-membrane elements that can implement dynamic diffractive elements in reflection mode were mentioned in Section 9.2. In order to develop reconfigurable diffractive elements, such MEMS devices need to have a sufficiently large 2D SBWP, to produce not only a large number of individually addressable pixels, but also large sets of phase shifts between 0 and  $2\pi$  for the wavelength under consideration. A binary version would be very limiting in terms of efficiency and image geometry (creating multiple diffraction orders). An array of digital micromirrors as implemented in DLP<sup>®</sup>-based digital projectors is not very appropriate for such a task.

#### *The Grating Light Valve (GLV)*

The most interesting technology resides in the deformable membrane technology presented earlier – set out here in a 2D version – implementing a Grating Light Valve (GLV). The various pillars of the GLV can be actuated over a wide range of levels in a large 2D array, and a quasi-continuous phase function can be imprinted onto the incoming beam. The result is an entirely reconfigurable digital diffractive element in reflection mode. Strong research efforts are geared toward GLV technology. In order to produce a superior diffraction effect, the SBWP of such a 2D deformable MEMS membrane should be much larger when compared to existing deformable membrane devices. Dynamic GLV devices are already used in maskless lithography, as described in Chapter 13.

Other microdisplays can also be used, perhaps more efficiently than MEMS, in order to produce large SBWP digital diffractive elements. Three of these techniques are discussed below.

### 9.4.1.2 Liquid Crystal Spatial Light Modulators

LC Spatial Light Modulators (LC SLMs) are a unique and well-known category of reconfigurable diffractive optics (see Figure 9.38).

The simplest approach is to implement a binary amplitude ferro-electric SLM. However, this will not create a high efficiency (it is limited to about 8%) and it will not produce a decent diffracted image, mainly due to the high number of orders diffracted, the symmetry of these orders and the very large zero order (for details of the numbers involved, see Chapter 5).

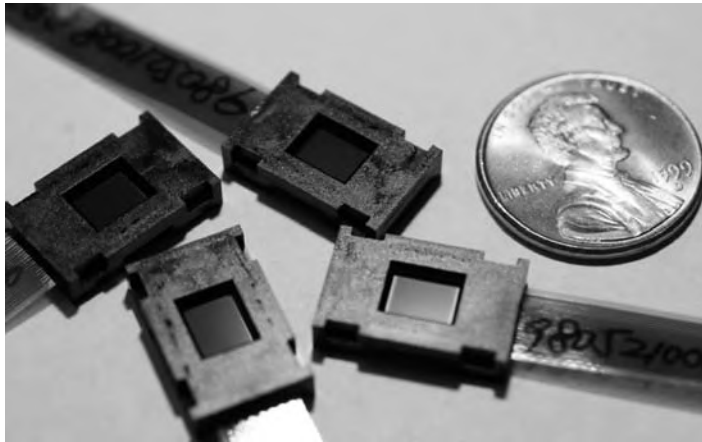


**Figure 9.38** LC SLMs as microdisplays

A phase SLM is a much better choice. However, a phase SLM is different from a traditional LC-based microdisplay, since the SLM is completely transparent (there is no analyzer sheet, only an initial polarizer – or nothing at all in the case of polarized laser illumination). The phase imprinted on the desired polarization is not intended to turn the polarization  $90^\circ$  as in an LC-based microdisplay but, rather, to produce a specific phase shift. This phase shift can be set between 0 and  $2\pi$  for the desired wavelength. Therefore, the width of the LC layer needs to be optimized and calibrated accurately, and the LC phase response stored in a Look-Up Table (LUT). For example, Hamamatsu's PAL-SLM is an electrically addressed phase SLM that can yield a phase shift in excess of  $2\pi$  for an incoming red laser beam.

However, one main inconvenience of these SMLs is the large size of the cells (the smallest cells available are still around  $8\text{--}10\text{ }\mu\text{m}$ ), and also the relatively large cell inter-spacing needed in order to locate the driving electronics (about one quarter of the cell size). The large cell sizes cannot yield high diffraction angles and the inter-pixel spacing absorbs much of the incoming light and creates parasitic super-grating effects. Another disadvantage is the binary values (switched on or off, no multi-value or analog modulation possible in most cases), producing numerous orders. This explains why such phase SLMs are used only in research areas; for example, to optimize in real time a CGH with an IFTA iterative algorithm (as presented in Chapter 6), by producing the Fourier or Fresnel transforms optically rather than by the use of an algorithm.

### 9.4.1.3 LCoS Microdisplays

Liquid Crystal on Silicon (LCoS) is a technology that brings together traditional IC and LC technologies. LCoS devices can operate either in transmission or reflective modes. Reflective LCoS have been optimized to produce high-quality microdisplay with conventional twisted nematic LC layers, which have virtually no pixelization, no dead spaces between pixels and very high electronics integration (see Figure 9.39). Thus, LCoS technology seems to be the preferred method to implement a reconfigurable display, rather than the previous DLP or phase SLM technologies.

Several companies are actively pursuing this technology to produce diffractive pico-projectors (see the next section).

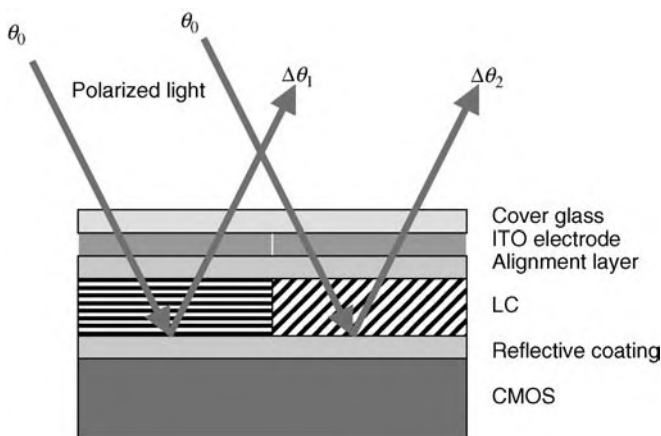


**Figure 9.39** Reflective LCoS microdisplay operation architecture



#### 9.4.1.4 Dynamic Grating Array Microdisplays

Another technique used to implement high-quality reconfigurable phase digital diffractives with high SBWP is to use dynamic gratings arrays to form microdisplays [26, 27]. We have described the principle of the Grating Light Valve (GLV) in Section 9.4.1.1. Dynamic grating microdisplays are similar, but are set in an array that forms the microdisplay.

Dynamic grating arrays can work in three different modes in order to implement microdisplays: such dynamic gratings can be implemented as MEMS gratings, H-PCLD gratings or etched gratings in a bulk LC layer.

##### *Amplitude Light Valves*

Such gratings can implement an array of amplitude light valves, similar to conventional DLP or HTPS LCD (High Temperature Poly Silicon) microdisplays, either in additive or subtractive modes depending on whether the image to be used is the diffracted field or the nondiffracted field (see Figure 9.40). The gray scales in the image can be generated by modulating the driving voltages, thus modulating the overall diffraction efficiency of the microdisplay.

An improvement of this technique is presented in Figure 9.41, with the edge illumination grating array microdisplay architecture. The sources (lasers or LEDs) are located at the edges of the display and launched in TIR within the slab. Each pixel is incorporating a dynamic Bragg coupler grating, which can out-couple more or less light depending on the voltage level at that pixel. Such gratings can be fabricated as MEMS, H-PDLC or etched gratings in bulk LC. For optimal Bragg coupling, the gratings should be tilted.

See-through displays (Figure 9.42) are perfect candidates to implement Head-Up Displays (HUD) in which the combiner and the image-generation functionalities are located in the same element: a clear window. They are also perfect candidates for other applications, such as near-eye displays or Helmet-Mounted Displays (HMDs) for augmented reality, which are especially useful for military and industrial applications. If the combination can be operated in two directions of space, a stereographic head-up display can be designed (e.g. by the use of a conjugate Bragg regime in volume gratings).

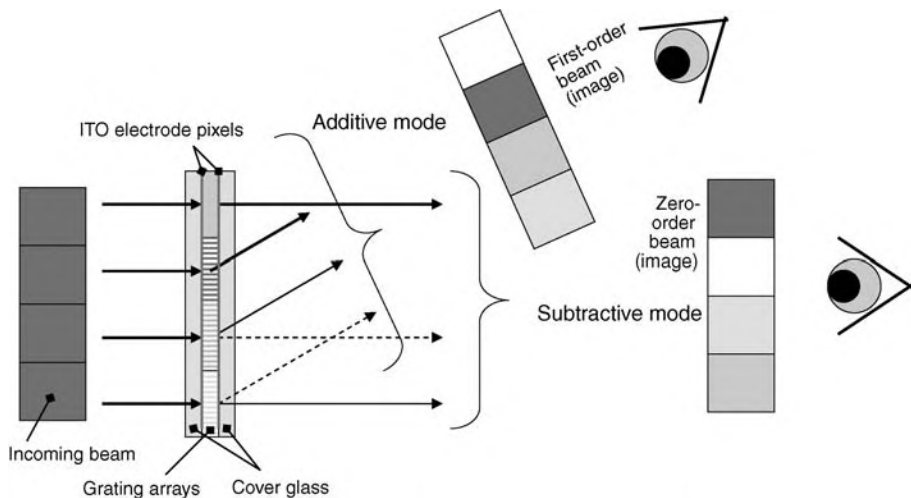


**Figure 9.40** Dynamic grating array microdisplay architectures

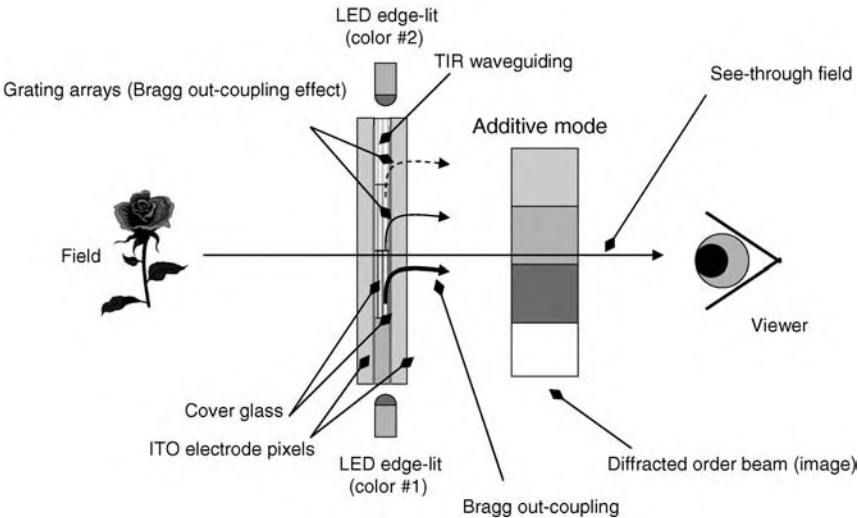


**Figure 9.41** An edge-illuminated grating array microdisplay

**Pixelated Phase Microdisplays**

A pixelated phase display is transparent: it does not produce an intensity image but, rather, a desired wavefront. The phase shift is introduced here by the sub-wavelength grating Effective Medium Theory effect (EMT – see Chapter 10). The display plane is recorded holographically or fabricated lithographically as a large grating with sub-wavelength gratings, which are nondiffracting for visible light but transfer a phase shift proportional to the voltage associated to each pixel (see Figure 9.42).

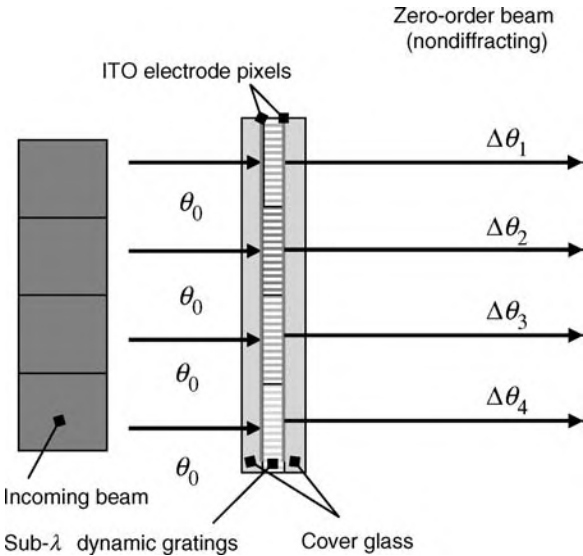


**Figure 9.42** A phase microdisplay based on sub-wavelength dynamic gratings

Such phase microdisplays are best suited (with the previous reflective LCoS) to implement reconfigurable digital diffractive optics or, in real time, generate wavefronts for correcting atmospheric aberrations sensed by diffractive Shack–Hartmann sensors, as discussed in Chapter 16. The next section shows some current applications of these architectures in pico-projectors.

### 9.4.2 *Diffractive Pico-projectors*

Diffractive projectors are entirely reconfigurable digital-phase diffractives, based on some of the microdisplay technologies reviewed in the previous section. Pico-projectors are tiny projectors, small enough to fit in a cell phone, but powerful enough to project a decent-sized, WVGA ( $854 \times 480$  pixels) image onto a nearby surface (approximately 300–500 lumens for a 2 foot high image at 3 feet). Chapter 16 shows some of the first prototype diffractive projectors.

Such laser diffractive projectors are very desirable, since they can be miniaturized more than common LED-based projectors. They consume less power than LEDs or high-pressure bulbs. Most interestingly, objective lenses are not required to produce an image, since the diffractive pattern is created in free space in the far field (Fourier CGH). Therefore, the image can be projected onto nearly any surface and will still be in focus (provided that the surface is located further away than the Rayleigh distance – see Chapter 11).

These projectors also suffer from numerous problems, which is perhaps the reason why they are not yet available on the market, although the phase microdisplay is already available (LCoS especially). The main disadvantages of diffractive projectors are as follows:

- The image set on the microdisplay is the Fourier transform of the projected image; therefore strong computing power has to be available in the tiny projector (hard-wired Fourier transform, for example – see Appendix C). The Fourier CGH has to be optimized with, for example, an IFTA algorithm (see Chapter 6).
- Multiple orders are diffracted, since the SBWP is limited by the number and size of the pixels, although the pixels can take on a great variety of phase values. Such orders have to be blanked out.
- Multiple colors need complex time sequencing architecture and three times more addressable phase levels in the microdisplay (for the three RGB lasers).
- The smallest pixels are between 5 and  $10\text{ }\mu\text{m}$ , which produces a smallest grating period of  $10\text{--}20\text{ }\mu\text{m}$ . Such a grating can only diffract in a very small angular cone. Therefore, the image remains in that small cone unless a complex angle enlarger (afocal inverted telescope system) is inserted as an objective (which kills off one of the main advantages of such a system).
- Uniform illumination from frame to frame is difficult, since the diffraction efficiency is the same if one pixel or 1000 pixels are set on in the image (the laser power remains constant). Therefore, either a complex algorithm has to be included to reduce the diffraction efficiency in proportion to the total number of pixels generated, or the power of the laser has to be modulated in real time.
- Laser illumination creates speckle, which needs to be reduced. It is not possible to use a spinning diffuser here, since the beam has to interfere in order to produce the desired image at infinity. One solution is to produce many different microdisplay frames (CGHs) within a single image frame, which correspond to the same intensity but with different phase maps and therefore different speckle patterns. However, this can only be achieved by sacrificing speed (see Figure 9.43).

For these various reasons, diffractive pico-projectors might actually be more suitable for applications with limited projection content, such as Head-Up Display (HUD), Helmet-Mounted Display (HMD) or metrology applications using structured laser illumination, rather than for pure video content.

In this section, various diffractive digital projectors have been discussed, which have been implemented through the following architectures:

- amplitude SLM microdisplay;
- phase SLM microdisplay;

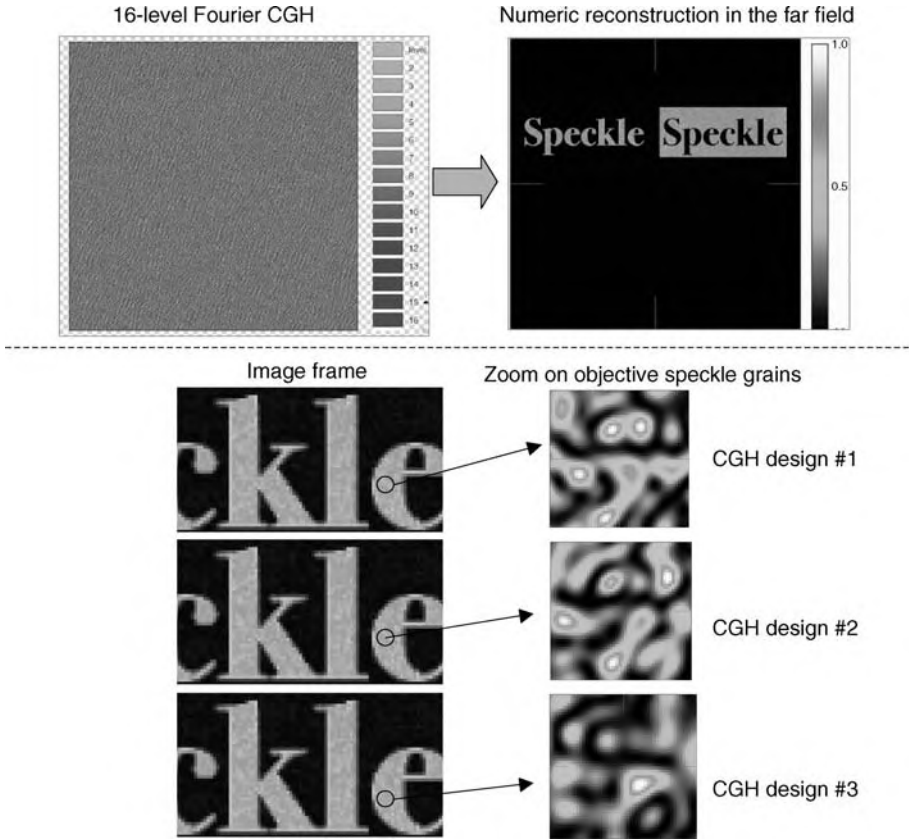


**Figure 9.43** The speckle reduction method in diffractive projectors

- deformable membrane microdisplay;
- Grating Light Valve (GLV) displays;
- LCoS microdisplays; and
- dynamic grating microdisplays.

## 9.5 Digital Software Lenses: Wavefront Coding

A digital software lens is a lens that cannot work correctly without its specific image-processing software. Such a technique is called wavefront coding.

The paradigm here is to combine software with a (complex) lens – a lens that is optimized not to produce a good image at a given focus, but a given controlled aberration over a wide range of focal lengths – and associate to that lens a numeric algorithm that can compute an aberration-free image at any depth plane. This is achieved by using the fact that digital cameras usually have much larger pixel resolutions than the camera lens itself allows (as CMOS image sensors get cheaper, quality optical objectives remain expensive, an over-dimensioned CMOS sensors can be purchased for a given objective lens). There is plenty of room at the bottom, which can be used precisely to compute a focused image from a well-controlled and well-aberrated (or blurred) raw image.

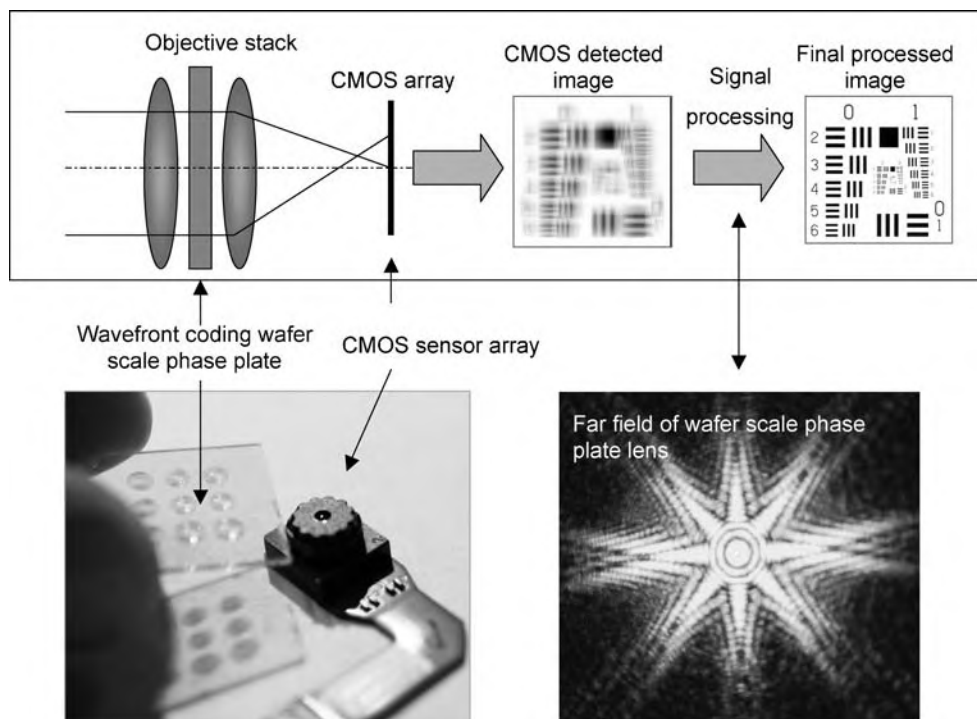


**Figure 9.44** A diffractive wavefront-coding software lens for a CMOS camera objective

Wavefront coding consists of introducing carefully crafted aberrations into a camera objective lens in order to produce an image in which the aberrations are carefully controlled. For example, large longitudinal chromatic aberrations can be implemented. One of the tasks of such lenses is to be able to focus at different distances without the lens having to be moved using one of the available techniques (piezo, MEMS etc.) or without the lens shape having to be changed (bulk LC, liquid lens, reconfigurable lens etc.), to simplify the objective lens design, reduce the size and lower the price.

Longitudinal aberrations can give hints about a focus closer to or further away from the central wavelength. Therefore, a carefully crafted algorithm, which can be integrated in a DSP on the back side of the CMOS sensors array, calculates the resulting unaberrated image, for any object location, without any mechanical focusing device. The wavefront-coding element can be a lens or – better – it can be a digital phase plate, or even a digital diffractive element. Figure 9.44 shows such a device on top of a conventional CMOS sensor. In this case, the lens comprises both the hardware lens and the software algorithm, as a hybrid hardware/software lens compound.

Other applications of such ‘digital software lenses’ are athermalization of imaging elements (for example, in IR imaging apparatus in missile heads).

## References

- [1] H. Zappe, ‘Novel components for tunable micro-optics’, *Optoelectronics Letters*, **4**(2), March 1, 2008.
- [2] L.G. Commander, S.E. Day and D.R. Selviah, ‘Variable focal length microlenses’, *Optics Communications*, **177**, 2000, 157–170.

- [3] J.J. Sniegowski, 'Multi-level polysilicon surface micromachining technology: applications and issues' (Invited Paper), ASME 1996 International Mechanical Engineering Congress and Exposition, *Proceedings of the ASME Aerospace Division, Atlanta, GA*, Vol. 52, November 1996, 751–759.
- [4] J.J. Sniegowski, 'Moving the world with surface micromachining', *Solid State Technology*, February 1996, 83–90.
- [5] Sandia National Laboratories, Introductory MEMS Short Course, June 29–July 1, 1998.
- [6] K.E. Petersen, 'Silicon as a mechanical material', *Proceedings of the IEEE*, **70**(5), 1982, 470–457.
- [7] C. Livermore, course materials for 6.777J/2.372J Design and Fabrication of Microelectromechanical Devices, Spring 2007, MIT Open Course Ware.
- [8] L.-V. Starman, 'Micro-Electro-Mechanical Systems (MEMS)', EE480/680, Department of Electrical and Computer Engineering, Wright State University, Summer 2006.
- [9] D.A. Koester *et al.*, SmartMUMPs Design Handbook, Revision 5.0; [www.memsrus.com/cronos/svcs/mumps.html](http://www.memsrus.com/cronos/svcs/mumps.html)
- [10] D. Bishop, V. Aksyuk, C. Bolle *et al.*, 'MEMS/MOEMS for lightwave networks: can little machines make it big?' In '*Analog Fabrication Methods*', SPIE Vol. 4179, September 2000, 2–5.
- [11] J.J. Sniegowski and E.J. Garcia, 'Microfabricated actuators and their application to optics', in '*Proceedings of Photonics West '95*', SPIE Vol. 2383, 1996, 46–64.
- [12] L.J. Hornbeck, 'From cathode rays to digital micromirrors: a history of electronic projection display technology', *TI Technical Journal*, **15**(3), 1998, 7–46.
- [13] Texas Instruments, Digital Light Processing® technology; [www.dlp.com](http://www.dlp.com)
- [14] M.K. Katayama, T. Okuyama, H. Sano and T. Koyama, 'Micromachined curling optical switch array for PLC-based integrated programmable add/drop multiplexer', Technical Digest OFC 2002, paper WX4.
- [15] L.H. Domash, T. Chen, B.N. Gomatam *et al.*, 'Switchable-focus lenses in holographic polymer-dispersed liquid crystal', in '*Diffraction and Holographic Optics Technology III*', I. Cindrich and S.H. Lee (eds), Proc. SPIE No. 2689, SPIE Press, Bellingham, WA, 1995, 188–194.
- [16] J.E. Ford, 'Micromachines for wavelength multiplexed telecommunications', in '*Proceedings of MOEMS '99*', Mainz, Germany, August 30, 1999.
- [17] T. Bifano, J. Perreault, R. Mali and M. Horenstein, 'Microelectromechanical deformable mirrors', *IEEE Journal of Selected Topics in Quantum Electronics*, **5**(1), 1999, 83–89.
- [18] W.D. Cowan, V.M. Bright, M.K. Lee and J.H. Comtois, 'Design and testing of polysilicon surface-micromachined piston mirror arrays', *Proceedings of SPIE*, **3292**, 1998, 60–70.
- [19] M.C. Roggemann, V.M. Bright, B.M. Welsh, W.D. Cowan and M. Lee, 'Use of micro-electro-mechanical deformable mirrors to control aberrations in optical systems: theoretical and experimental results', *Optical Engineering*, **36**(5), 1997, 1326–1338.
- [20] J.E. Ford and J.A. Walker, 'Dynamic spectral power equalization using micro-opto-mechanics', *Photonics Technology Letters*, **10**, 1998, 1440–1442.
- [21] S. Masuda, S. Fujioka, M. Honma, T. Nose and S. Sato, 'Dependence of optical properties on the device and material parameters in liquid crystal microlenses', *Japanese Journal of Applied Physics*, **35**, 1996, 4668–4672.
- [22] S. Masuda, T. Nose and S. Sato, 'Optical properties of a polymer-stabilized liquid crystal microlens', *Japanese Journal of Applied Physics*, **37**, 1998, L1251–L1253.
- [23] R.G. Lindquist, J.H. Kulick, G.P. Nordin *et al.*, 'High-resolution liquid-crystal phase grating formed by fringing fields from interdigitated electrodes', *Optics Letters*, **19**, 1994, 670–672.
- [24] M. Kulishov, 'Adjustable electro-optic microlens with two concentric ring electrodes', *Optics Letters*, **23**, 1998, 1936–1938.
- [25] A.R. Nelson, T. Chen, L.A. Jauniskis and L.H. Domash, 'Computer-generated electrically switchable holographic composites', in '*Diffraction and Holographic Optics Technology III*', I. Cindrich and S.H. Lee (eds), Proc. SPIE No. 2689, SPIE Press, Bellingham, WA, 1995, 132–143.
- [26] C. Altman, E. Bassous, C.M. Osburn *et al.*, 'Mirror array light valve', US Patent No. 4,592,628, 1986.
- [27] K.E. Petersen, 'Micromechanical light modulator array fabricated on silicon', *Applied Physics Letters*, **31**(8), 1977, 521–523.

# 10

## Digital Nano-optics

The previous chapters (from Chapter 3 to Chapter 9, with the exception of Chapter 8, which focused on holographic optics) have described micro-optical elements that have smallest lateral feature sizes on the order of several hundred down to about 3–4 times the reconstruction wavelength. Based on Appendices B and C, Chapter 11 will show how such elements can be modeled via scalar diffraction theory. When the smallest lateral feature sizes constituting the elements are nearing the wavelength of light (reconstruction wavelength), down to a fraction of that wavelength, scalar diffraction design and modeling tools can no longer be used, and one has to develop appropriate design and modeling tools based on rigorous diffraction theory (see also Appendix A and Chapter 11).

This chapter reviews the various sub-wavelength digital elements and nano-optical elements used in industry today, and describes for each of them the appropriate modeling techniques that are used today in industry.

### 10.1 The Concept of ‘Nano’ in Optics

In the semiconductor fabrication and Integrated Circuit (IC) industries, the term ‘nano’ usually refers to printed structures that are smaller than 100 nm, thus defining the realm of structures below that of traditional microstructures (microtechnology versus nanotechnology).

However, in optics (and especially in digital optics), the term ‘nano’ is not so much related to an absolute dimension, as is the case in the IC industry, but rather to the ratio between the wavelength  $\lambda$  and the smallest structure period  $\Lambda$  present in the optical element [1]. This is roughly the limit of validity of the scalar theory of diffraction. When the ratio  $\Lambda/\lambda$  nears unity or even goes below unity, it is commonly agreed in optics that one is in the nano-optics realm. Figure 10.1 depicts where the notion of ‘nano-optics’ can be used in optics when considering the  $\Lambda/\lambda$  ratio.

Therefore, depending on the wavelength of light used, the ‘nano’ realm in optics can refer to structures that are larger or smaller than the structures considered as ‘nano’ in the IC industry. For example, a nano-optical element engineered for CO<sub>2</sub> laser light would be a micro-optical element for visible light.

### 10.2 Sub-wavelength Gratings

Sub-wavelength (SW) diffractives (Type 4 elements in the digital optics classification of Chapter 2) are gratings (linear, circular, chirped etc.) in which the smallest grating period is smaller than the reconstruction wavelength ( $\Lambda/\lambda < 1$ ). Such gratings can operate in either the transmission or the

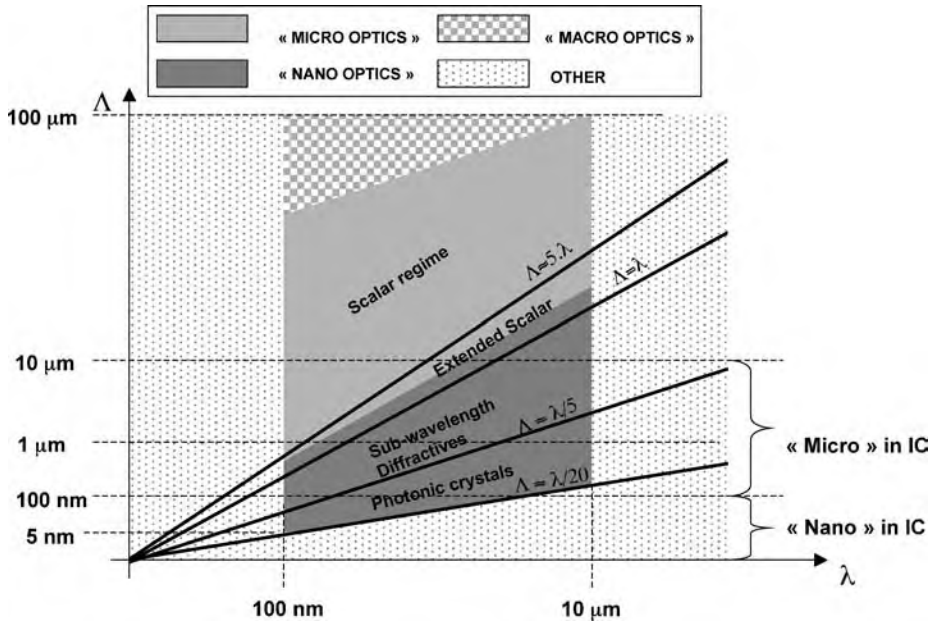


Figure 10.1 The macro-, micro- and nano-optics realms

reflection regime, and have only the zero – backward and forward – orders propagating, all other higher diffraction orders being evanescent [2, 3]. That is why these gratings are also referred to as ‘zero-order gratings’.

Figure 10.2 shows the differences in diffracted and propagating orders between a standard multi-order diffractive (grating periods much larger than the wavelength) and a SW period grating.

Note that SW diffractives can be implemented either as thin surface-relief phase elements or index-modulated holographic elements (see Figure 10.3). SW gratings, as volume holograms described in Chapter 8, are elements that also show very strong polarization dependence, unlike standard diffractive with much larger feature sizes.

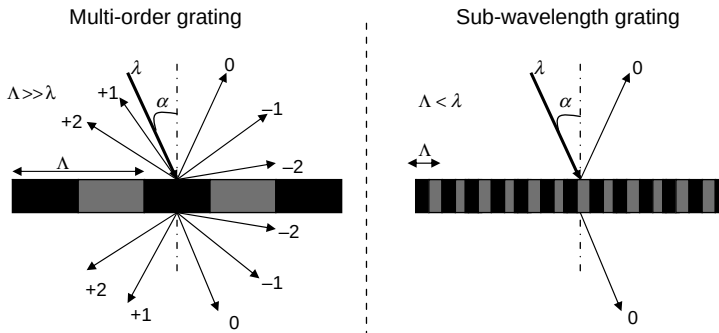


Figure 10.2 Multi-order and sub-wavelength gratings



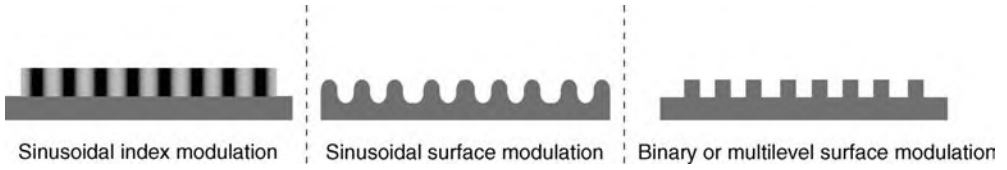


**Figure 10.3** The various physical implementations of sub-wavelength gratings

Applications requiring SW diffractives include mainly:

- polarization-sensitive elements (polarization splitting, polarization combining);
- anti-reflection surfaces (AR);
- high-resolution resonant filters (in reflection or transmission);
- integrated waveguide gratings (with Bragg reflection or coupling effects); and
- phase plates (in reflection and transmission).

SW gratings can be fabricated in various geometries, which are more complex than the grating example shown in Figure 10.3. More complex SW gratings can be fabricated with sawtooth, échelette or arbitrary 2D repetitive surface-relief elements.

### 10.3 Modeling Sub-wavelength Gratings

Solving Maxwell's time-harmonic equations (see below and Appendix A) at the boundaries [4] of the sub-wavelength grating structures gives an exact solution of the diffraction problem:

$$\begin{cases} \text{curl}(\vec{E}) = -\mu \frac{\partial \vec{H}}{\partial t} \\ \text{curl}(\vec{H}) = \epsilon \frac{\partial \vec{E}}{\partial t} \\ \text{div}(\epsilon \vec{E}) = 0 \\ \text{div}(\mu \vec{H}) = 0 \end{cases} \quad (10.1)$$

Unfortunately, analytic solutions to Maxwell's equations are very hard to derive, and in most practical cases the equations have to be solved numerically. Various approaches have been used to solve this diffraction problem.

#### 10.3.1 Rigorous Diffraction Models

It was probably G.W. Stokes, in defending his thesis in 1960, who mentioned the need to use Maxwell's equations to predict the efficiency of sub-wavelength gratings in various operating configurations. Lord Rayleigh, in the early 20th century, proposed the same idea. Unfortunately, due to the lack of analytic solutions and due to the total lack of computing power in those days, any serious exploitation of the idea was put aside.

Half a century after Lord Rayleigh's propositions, as computing power became available, his theory regained popularity in the scientific community. Since 1964, several thousand papers have been published in this field, proposing new numeric solution techniques and, especially, to derive new applications (see Chapter 16).

Several calculation techniques have thus been proposed: the integral method; the differential method and its derivatives; the modal approach, which includes the Rigorous Coupled Wave Analysis (RCWA) approach [5]; and Kogelnik's two-wave coupled wave theory (see Chapter 8), the Raman–Nath method and Rytov's method.

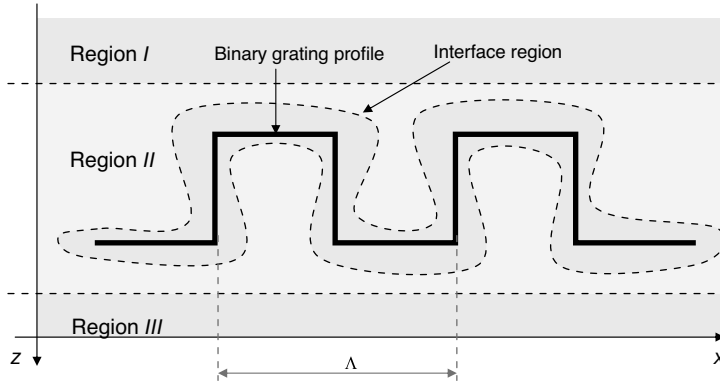


**Figure 10.4** Boundary regions in a binary surface-relief grating

### 10.3.1.1 Rayleigh's Early Proposition

As a common description of a 1D grating, let us consider three different regions of interest in a binary surface-relief grating (see Figure 10.4). An incident wavefront is propagating in region I, a medium of homogeneous permittivity  $\epsilon_1$ . Within this medium, incident and reflected wavefronts (diffracted or not) will propagate. The diffracted wavefronts will propagate in region II, a medium of homogeneous permittivity  $\epsilon_2$ . In between these two media, a region III is defined by a permittivity  $\epsilon(z)$ , which varies from  $\epsilon_1$  to  $\epsilon_2$ , and describes the diffractive sub-wavelength phase-relief structures.

Let  $U_i$  be the incident field over the grating and  $U_d$  the diffracted field; the total field  $U$  can then be described as follows:

$$U = U_i + U_d \Leftrightarrow U_d = U - U_i \quad (10.2)$$

The periodicity of the grating in one direction allows us to express  $U_d$  as a Fourier series:

$$U_d = \sum_{n=-\infty}^{+\infty} U_n(z) \cdot e^{j(\sin(\varphi)k + nK)x}, \quad \text{where} \quad \begin{cases} K = \frac{2\pi}{\Lambda} \\ k = \frac{2\pi}{\lambda} \end{cases} \quad (10.3)$$

Then, by using Helmholtz's equation (see Appendix A), we can write

$$\nabla^2 U_d + k^2 U_d = 0 \Leftrightarrow \frac{\partial^2 U_n(z)}{\partial z^2} + (k^2 - (\sin(\varphi) \cdot k + nK)) U_n(z) = 0 \quad (10.4)$$

Hence, the diffracted field  $U_d$  can be expressed as

$$U_d = \sum_{n=-\infty}^{+\infty} \alpha_n \cdot e^{j((\sin(\varphi) \cdot k + nK)x + (k^2 - (\sin(\varphi) \cdot k + nK)^2) \cdot z)} \quad (10.5)$$

Note that there are some restrictions to this method. If the Helmholtz's equation is valid in regions I and II, it is verified only partially in region III. To be rigorous (and thus for the equation to be valid everywhere in region III), it is necessary to consider the first derivative of  $U_d$  as a continuous function, in order to be able to calculate the term  $\Delta U_d$  in Helmholtz's equation. Miller demonstrated in 1966 that Rayleigh's approximation was valid in the case of a sinusoidal grating where  $d/\Lambda < 0.072$ .

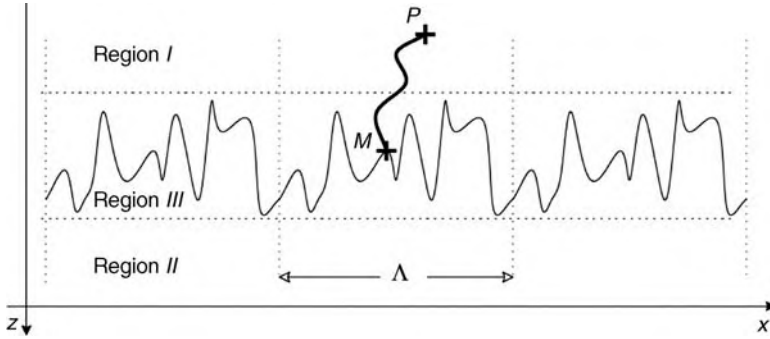


**Figure 10.5** The integral resolution method

### 10.3.1.2 The Integral Method

In the integral method, the electric field is described as an unknown function  $\Psi(M)$  located at the point  $M$  on top of the grating profile (see Figure 10.5).

If  $U(P)$  is the expression of the field at a point  $P$  in space,  $U(P)$  can be expressed as follows:

$$U(P) = \int_{\Lambda} G(P, M) \cdot \Psi(M) \cdot dS \quad (10.6)$$

where  $G$  is a specific kernel function. The function  $\Psi$  can be determined in an iterative numeric way. It gives the value of the fields on both sides of the grating profile. As the iterative algorithm converges, both fields converge to the grating profile and eventually become equal one to each other. This method usually gives poor results when compared to experimental data.

### 10.3.1.3 The Differential Method

This method is often used today and several contiguous methods have been derived from it. Also, it is the method that accommodates the best theory and experimental results (which is always good). It was in 1969 when, for the first time, Cerutti-Maori [6] developed the idea of layering the grating profile into several regions, giving the grating profile a layered representation, each layer (referred to by the index  $n$ ) having its own constant dielectric permittivity  $\epsilon_n(x)$  ( $\epsilon_n$  is either equal to  $\epsilon_1$  or  $\epsilon_2$ ). Being periodical,  $\epsilon_n(x)$  can be decomposed as a Fourier series:

$$\epsilon_n(x) = \sum_{l=-\infty}^{+\infty} \tilde{\epsilon}_{n,l}(K) \cdot e^{jKx} \quad (10.7)$$

Figure 10.6 depicts the layered grating structure.

The coupling equations between the electric field TE and magnetic field TM are ruled by Maxwell's equations (Equation (10.1)). The combination of these equations gives the wave equation, which is easily expressed in a homogeneous medium, as it occurs within each layer  $n$ . Within each of these layers, the wave equation for the TE mode can then be expressed as

$$\nabla^2 E_y + k^2 \epsilon_n(x) \cdot E_y = 0 \quad (10.8)$$

The resolution of such an equation (which links  $E_y(x, z)$  to its derivatives – hence the name of this method), is performed by substituting in the Fourier expansion of  $\epsilon_n(x)$  for each layer. In order to simplify the calculation, let us assume the grating profile is sinusoidal, and thus that the Fourier expansion of the

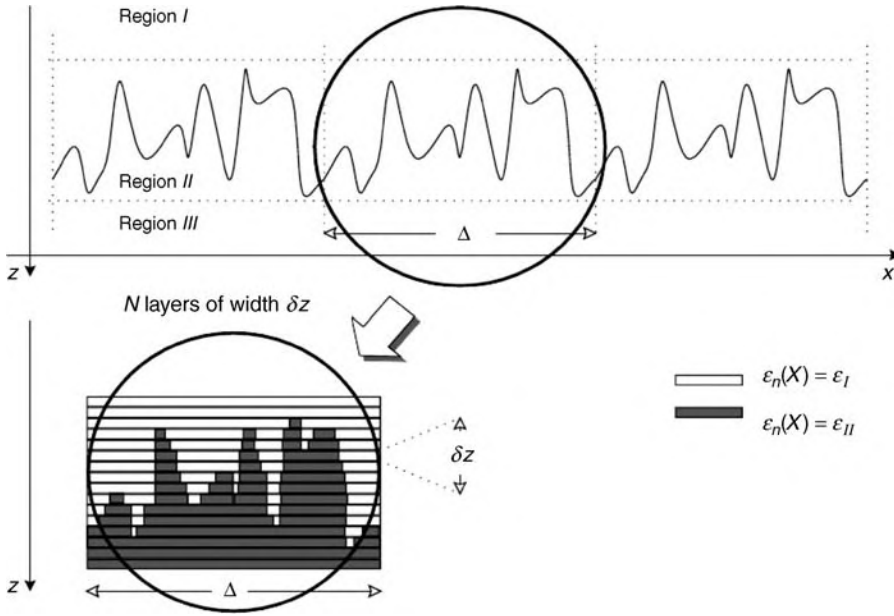


**Figure 10.6** The layered grating structure

permittivity reduces dramatically (to unity). The resolution of the wave equation can be done quite effectively, since  $E_y(x, z)$  can be decomposed in an infinite sum of reflected or transmitted modes, either propagating or evanescent. The resulting expression for  $E_y(x, z)$  then becomes

$$E_y(x, z) = \sum_{p=-\infty}^{+\infty} A_p \sum_{q=-\infty}^{+\infty} B_{p,q} \cdot e^{j(\beta_p - qK)x} \cdot e^{j\zeta_p z} \quad (10.9)$$

where  $\zeta_p$  corresponds to the projection of wave vector  $k_p$  of mode  $p$  in the layer  $n$  along the  $x$ - and  $z$ -axes respectively, and  $A_p$  and  $B_{p,q}$  are constants to be defined by solving the wave equation and applying the conditions at the limits. The summation over  $p$  describes the modal decomposition of the field, whereas the summation over  $q$  reveals the coupling between each mode.

The values  $A_p$  and  $B_{p,q}$  give the value of the electric field  $E_y(x, z)$  in the corresponding layer  $n$ . By iteratively calculating the field within each layer, and finally within regions I and II, and by taking into account the conditions at the limits, the diffracted field can be computed, and thus can predict the diffraction efficiencies for each order. However, this is easier said than done. Below are presented two different approaches for the representation of  $E_y(x, z)$ : the general modal approach and the rigorous coupled wave analysis (RCWA).

### 10.3.1.4 The Modal Approach

In the general modal approach, the summation describing the coupling is replaced by a function  $\phi$ , which depends on vector  $\vec{r}$ . The expression for  $E_y(x, z)$  then becomes

$$E_y(x, z) = \sum_{p=-\infty}^{+\infty} A_p \cdot \phi(\vec{r}) \cdot e^{-j\beta_p x} \cdot e^{j\zeta_p z}, \quad \text{where} \quad \phi(\vec{r}) = \sum_{q=-\infty}^{+\infty} B_{p,q} \cdot e^{jqKx} \quad (10.10)$$

This approach considers the grating as a waveguide. However, it is very seldom used for diffractive optics, since the function  $\phi(r)$  has to be evaluated.

### 10.3.1.5 The RCWA Model

The Rigorous Coupled Wave Analysis (RCWA) of grating diffraction has been applied to single planar gratings [7], surface-relief gratings [8], anisotropic and cascaded gratings and volume multiplexed gratings. It is today one of the most widely used techniques for the modeling of sub-wavelength diffractives. As opposed to the modal approach, the coupling effects between the different diffracted orders are considered. By using Floquet's formalism, one can express  $E_y(x, z)$  in the form

$$E_y(x, z) = \sum_{q=-\infty}^{+\infty} S_q(z) \cdot e^{-j\sigma_q x}, \quad \text{where} \quad S_q(z) = \sum_{p=-\infty}^{+\infty} A_q \cdot B_{p,q} \cdot e^{ipKz} \quad (10.11)$$

where the  $S(z)$  are the complex amplitudes of the space harmonics and  $\sigma_q = \beta_q - qK$  is the Floquet wave vector.

By replacing the expression for  $E_y(x, z)$  in the wave equation, one obtains a set of relations between space harmonics and their primary and secondary derivatives. The complete calculation of the diffraction efficiencies boils down to the resolution of a system of equations. This system of equations can be solved using a matrix expression (i.e. by solving for the eigenvalues and eigenvectors of the matrix). The complete set of equations can be solved by applying the boundary conditions for each layer  $n$ . To do so, and in order to keep the required CPU time relatively small, only a finite number of orders on both sides of the zero order will be considered. This is one of the limitations of this method, as well as a limitation of all of the rigorous numeric methods.

The accuracy of the RCWA method and the convergence rate is closely linked to the number of harmonics included in the analysis. In order to solve practical problems for perfect dielectric gratings, several approximations and propositions have been made to the generic coupled wave theory.

### 10.3.1.6 The Raman–Nath Coupled Wave Model

The technique proposed by C.C. Raman and N.S. Nagendra Nath back in 1935–6 [9] neglected the secondary derivative of  $S_n(z)$ , and considered all the diffracted orders. Also, the phase shift from any selected order to the next one was neglected. The complex amplitudes of the space harmonics  $S_n(z)$  can then be considered as a Bessel function of order  $n$ .

### 10.3.1.7 Kogelnik's Coupled Wave Model

The *two-wave coupled wave theory* considers only the zero and fundamental positive orders, in transmission and reflection. This severe approximation is possible only when the grating is illuminated at the first Bragg angle. However, this approximation has provided quite accurate solutions for several applications. Based on the Raman–Nath theory, Kogelnik developed this technique In 1969, considering an incident plane wave, in TE mode, launched onto a thick grating. The incident waves are in Bragg incidence to the  $n$ th order (incidence angle and wavelength); thus, any orders different than 0 and  $n$  in reflection or transmission are considered as evanescent. Kogelnik's approximations resemble the previous ones, except that the electric field is considered to be varying very slowly in  $y$ , and hence the second derivative of  $S_n(z)$  can be considered to be null. Thus, an important simplification occurs in the wave equation, and gives rise to immediate solutions for a large number of applications.

Note that this is only valid for a thick grating. For further analysis of Kogelnik's theory, and for practical examples, as applied to holographic optical elements, see Chapter 8.

### 10.3.1.8 Analytic Methods

Analytic methods have also been proposed to solve the problem for particular gratings. In 1959, Marechal and Stokes adapted this idea to a multilevel binary grating operating in TM mode.

### 10.3.2 Numeric Implementations

Numeric resolution methods have been flourishing recently in order to provide design tools for applications stemming from sub-wavelength gratings and sub-wavelength structures, the number of which is increasing every day. Some of these applications are listed below:

- SWL blazed grating and fan-out gratings;
- SWL diffractive lenses;
- anti-reflection surfaces (ARS);
- polarization components (splitting, combining);
- resonant filters (reflection and transmission);
- resonant waveguide gratings; and
- custom phase plates (reflection and transmission).

The following sections will review two of the main numeric techniques used today to implement the modal and integral resolution models (RCWA and FDTD).

#### 10.3.2.1 Numeric Implementation of the RCWA Model

The RCWA model was first proposed by Moharam and Gaylor in 1981. The arbitrary surface-relief profile of a grating (within a single period in region III) can be expressed as a linear combination of several different periodic binaries [5], as described in the previous section. The permittivity  $\epsilon_n(x)$  within region III and layer  $n$  can be expressed as the modulation of the two extreme permittivities  $\epsilon_I$  and  $\epsilon_{II}$  by a function  $f_n(x)$  as follows:

$$\epsilon_n(x) = \epsilon_I + (\epsilon_{II} - \epsilon_I) \cdot f_n(x) \quad \text{where} \quad f_n(x) = f_n^0(x) \sum_{p=-\infty}^{+\infty} \delta(x - p\Lambda) \quad (10.12)$$

As  $\epsilon_n(x)$  is periodic, it can be decomposed into the following Fourier expansion:

$$\begin{cases} \epsilon_n(x) = \epsilon_I + (\epsilon_{II} - \epsilon_I) \sum_{h=-\infty}^{+\infty} \tilde{\epsilon}_{h,n} \cdot e^{jhKx} \\ \tilde{\epsilon}_{h,n} = \frac{1}{\Lambda} \int_0^\Lambda f_n^0(x) \cdot e^{-jhKx} dx \end{cases} \quad (10.13)$$

#### Global Wavefront Coupling

As seen earlier, the electric field  $E_y^{(n)}(x, z)$  can be described by considering the global coupling effects between all of the diffracted orders, propagating in either transmission mode or reflection mode. This global coupling occurs effectively everywhere in space, and the projection of the wave vector  $\vec{k}$  on the  $x$ -axis is the same for any diffraction order in any medium:

$$\vec{k}_I^{(p)} \cdot \vec{x} = \vec{k}_{III}^{(p,n)} \cdot \vec{x} = \vec{k}_{II}^{(p)} \cdot \vec{x} = k_{0,x} - pK \quad \text{where} \quad \begin{cases} K = 2\pi/\Lambda \\ k_{0,x} = (2\pi/\lambda) \cdot \sin(\theta_I) \\ k_{0,y} = 0 \\ k_{0,z} = (2\pi/\lambda) \cdot \cos(\theta_I) \end{cases} \quad (10.14)$$

On the other hand, the projections of the wave vector on the  $z$ -axis tell us about the evanescent or propagation nature of the field considered:

$$\begin{cases} \vec{k}_I^p \cdot \vec{z} = \sqrt{|k_0|^2 \epsilon_I - (k_{0,x} - pK)^2} \\ \vec{k}_{II}^p \cdot \vec{z} = \sqrt{|k_0|^2 \epsilon_{II} - (k_{0,x} - pK)^2} \end{cases} \quad (10.15)$$

$E_y^{(n)}(x, z)$  can be written as a linear combination of space harmonics:

$$E_y^{(n)}(x, z) = \sum_{p=-\infty}^{+\infty} S_p^{(n)}(z) \cdot e^{\left(-j(k_{0,x} - pK)x + j\sqrt{(k_{III}^{(n)})^2 - (k_{0,x} - pK)^2}z\right)} \quad (10.16)$$

If  $k_{III}^{(n)}$  remains unchanged whichever order is considered, it can be expressed as follows:

$$k_{III}^{(n)} = \frac{2\pi}{\lambda} \sqrt{\epsilon_n^{(0)}} \quad \text{where} \quad \epsilon_n^{(0)} = \epsilon_1 + (\epsilon_{II} - \epsilon_I) \frac{1}{\Lambda} \int_0^\Lambda f_n^0(x) dx \quad (10.17)$$

As the exponential in the expression for  $E_y^{(n)}(x, z)$  is a function of both  $x$  and  $z$ , and thus is quite difficult to manipulate, Floquet proposed combining the expression of the space harmonic  $S_p^n(z)$  and the term that varies in the  $z$ -direction in the exponential in a single form,  $\widehat{S}_p^n(z)$ . The field expression then becomes

$$E_y^{(n)} = \sum_{p=-\infty}^{+\infty} \widehat{S}_p^n(z) \cdot e^{-j\sigma_p x} \quad (10.18)$$

which gives rise to a conveniently separable equation in  $x$  and  $z$ , where  $\widehat{S}_p^n(z)$  are the complex space harmonics and  $\sigma_p = k_{0,x} - pK$  the Floquet wave vectors.

### Field Description outside the Diffractive Structures Region

The electric fields  $E_y^I(x, z)$  and  $E_y^{II}(x, z)$ , in the two homogeneous regions I and II, can be described as follows:

$$\begin{cases} E_y^I(x, z) = e^{-j(k_{0,x}x + k_{0,z}z)} + \sum_{p=-\infty}^{+\infty} R_p e^{-j(k_{0,x} - pK)x + j\sqrt{\epsilon_I |k_0|^2 - (k_{0,x} - pK)^2}z} \\ E_y^{II}(x, z) = \sum_{p=-\infty}^{+\infty} T_p e^{-j(k_{0,x} - pK)x - j\sqrt{\epsilon_{II} |k_0|^2 - (k_{0,x} - pK)^2}z} \end{cases} \quad (10.19)$$

### Resolution of the Wave Equation

Once the field in the three regions and the diffractive medium in region II are described, the wave equation can be resolved. For each layer  $n$ , this differential equation becomes

$$\Delta E_y^{(n)}(x, z) + \epsilon_y^{(n)}(x, z) |k_0|^2 E_y^{(n)}(x, z) = 0 \quad (10.20)$$

By substituting the expressions for  $E_y^{(n)}(x, z)$  and  $\epsilon_y^{(n)}(x, z)$  into the wave equation, one obtains

$$\begin{aligned} & \frac{\partial^2 S_p^{(n)}(z)}{\partial z^2} - 2j\sqrt{(k_{III}^{(n)})^2 - (k_{0,x})^2} \cdot \frac{\partial^2 S_p^{(n)}(z)}{\partial z^2} + pK^2(m^{(n)} - p)S_p^{(n)}(z) \\ & + |k_0|^2(\epsilon_{II} - \epsilon_I) \cdot \sum_{h=1}^{+\infty} (\widehat{\epsilon}_{h,n} \cdot S_{p-h}^{(n)}(z) + \widehat{\epsilon}_{h,n}^* \cdot S_{p+h}^{(n)}(z)) = 0 \end{aligned} \quad (10.21)$$

where  $m^{(n)}$  describes the Bragg condition. One then writes

$$m^{(n)} = \frac{2\Lambda\sqrt{\epsilon_n^{(0)}}}{\lambda} \sin(\theta_I) \quad (10.22)$$





The conditions can be described as follows: ‘The electric and magnetic fields are equal in the same neighborhood of each layer, as well as in the neighborhood of regions I and II.’ The following system of equations has to be solved:

- $2s(N + 1)$  relations;
- $2sN$  unknowns  $C_q$ ;
- $s$  unknowns  $T_p$ ;
- $s$  unknowns  $R_p$ .

This system is homogeneous, and can thus be solved completely.

### ***Discussion on the Coupled Wave Approximations***

As seen previously, the coupled wave theory requires truncation of the number of diffracted orders considered in the resolution process, to only some lower orders adjacent to the zero order. Apart from the CPU time, which is very high for such a matrix resolution (see the following section), the errors that accumulate during the simplified matrix resolution can give rise to high fluctuations between the exact ‘virtual’ solution and the actual eigenvalues of the simplified matrix. Furthermore, there are some aberrations that occur when, for example, a highly energetic order propagates near the evanescence limit (Wood’s aberration).

However, this theory can be applied to a wide range of gratings, because the ‘real’ permittivity can be set for a perfect dielectric medium or an absorbing medium. It is also possible to solve the wave equation with the TM polarization. In this case, the wave equation becomes vectorial, and the difficulty of resolving it increases dramatically. This is the case when  $\vec{K}$  is no longer in the incidence plane; thus, the TE and TM modes are coupled, and cannot be solved independently (the conical diffraction domain).

### ***CPU Time Considerations***

When solving a grating problem using the RCWA or any other rigorous method, the CPU time consumption is a major drawback. Here, the CPU time grows linearly with the number of layers considered within region III, but grows as the cube of the number of propagating orders considered in the numeric calculations.

Note that the number of propagating orders is strongly dependent on the  $\Lambda/\lambda$  ratio of the grating considered. Therefore, the CPU time is strongly affected by that ratio. Thus, one can estimate the CPU time consumption to solve the problem for  $N$  layers and  $M$  desired diffracted orders as follows:

$$CPU_{time} \propto NM^3 \quad (10.27)$$

Since the number of propagating orders increases with the ratio  $\Lambda/\lambda$ , rigorous theories are realistic modeling tools for solving the grating problem only for very low  $\Lambda/\lambda$  ratios; that is, when the evanescent orders are located near the zero order. The perfect case is the zero-order grating, where the only propagating order remains the zero order.

### **10.3.2.2 Numeric Implementation of the FDTD Method**

The Finite Distance Time Domain (FDTD) method is an efficient method for solving Maxwell’s equations directly for surface-relief sub-wavelength diffractives through heavy CPU processing. Maxwell’s equations, as seen earlier, are formed using Gauss’s law, Faraday’s law of induction and Ampère’s circuital law. Faraday’s and Ampère’s laws can be expressed as follows:

$$\begin{cases} \nabla \times \vec{H} = \sigma \vec{E} + \epsilon \frac{\partial \vec{E}}{\partial t} \\ \nabla \times \vec{E} = -\mu \frac{\partial \vec{H}}{\partial t} \end{cases} \quad (10.28)$$

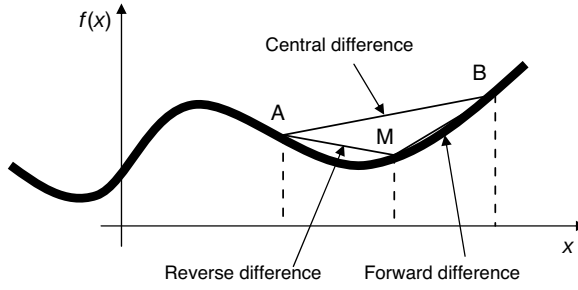


**Figure 10.7** The simple finite difference method

For a TE polarized wave in 2D, one can rewrite this as

$$\frac{\partial E_z}{\partial t} = -\frac{1}{\epsilon} \left( \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} \right) \quad \text{and} \quad \begin{cases} \frac{\partial H_y}{\partial t} = -\frac{1}{\mu} \frac{\partial E_z}{\partial x} \\ \frac{\partial H_x}{\partial t} = -\frac{1}{\mu} \frac{\partial E_z}{\partial y} \end{cases} \quad (10.29)$$

Thus, these partial differential equations have to be solved. In order to find the derivative  $f'(x)$  of a function  $f(x)$ , the simple finite difference method

$$\frac{\partial f(M)}{\partial x} = \lim_{\Delta x \rightarrow 0} \left( \frac{f(B) - f(A)}{\Delta x} \right) \approx \frac{f(B) - f(A)}{\Delta x} \quad (10.30)$$

will be used, as depicted in Figure 10.7.

In order to implement the finite difference over the electric and magnetic fields  $E$  and  $H$ , the space dimension as well as the time dimension will be sampled and interleaved. Figure 10.8 shows a 1D example of the  $E$ - $H$  sampled-interleaved space/time fields used in FDTD methods.

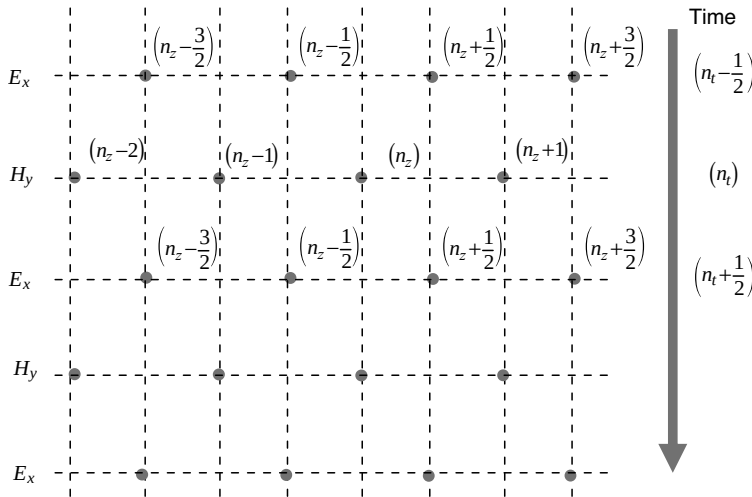


**Figure 10.8** Sampled interleaved  $E$ - $H$  fields in space/time

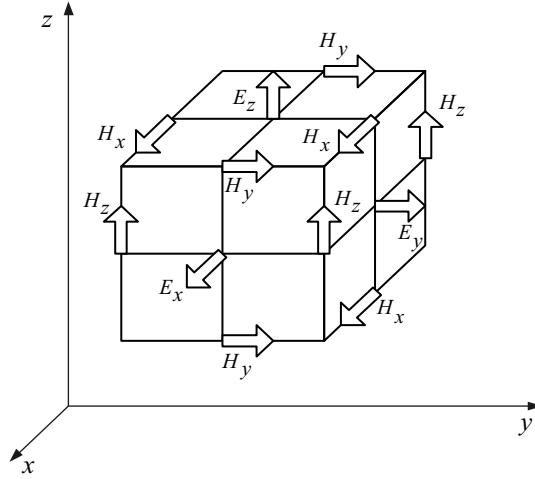


**Figure 10.9** Three-dimensional space sampling along the Yee cell

Such interleaving results in the following sets of equations:

$$E_z^{(n+1)}(i,j) = E_z^{(n)}(i,j) - \frac{\Delta t}{\epsilon \Delta x} \left( H_y^{(n+1/2)}(i+1/2,j) - H_y^{(n+1/2)}(i-1/2,j) \right) \\ + \frac{\Delta t}{\epsilon \Delta y} \left( H_x^{(n+1/2)}(i,j+1/2) - H_x^{(n+1/2)}(i,j-1/2) \right)$$

and

$$\begin{cases} H_x^{(n+1/2)}(i,j+1/2) = H_x^{(n-1/2)}(i,j+1/2) - \frac{\Delta t}{\mu \Delta y} \left( E_z^{(n)}(i,j+1) - E_z^{(n)}(i,j) \right) \\ H_y^{(n+1/2)}(i+1/2,j) = H_y^{(n-1/2)}(i+1/2,j) - \frac{\Delta t}{\mu \Delta x} \left( E_z^{(n)}(i+1,j) - E_z^{(n)}(i,j) \right) \end{cases} \quad (10.31)$$

For computation in a 3D space, the Yee cell is usually taken into consideration. The 3D Yee cell is depicted in Figure 10.9. The space lattice built on the Yee cell architecture is the basis of the 3D FDTD computational grid. The locations of the  $E$  and  $H$  fields are interlocked in space and the solution is ‘time stepped’.

The FDTD algorithm can be approximated as follows:

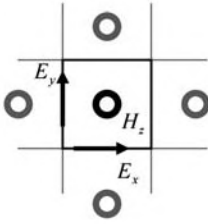
- compute Faraday’s equation for all nodes inside the simulation region;
- compute the 1D Ampère–Maxwell equation for all nodes inside the simulation region;
- compute the electric current density excitation for all excitation nodes  $\hat{E}_y^{(i,j+1/2)}$ ; and
- compute the boundary conditions for all PEC boundary nodes:  $\hat{E}_y^{(i,j+1/2)} = 0$ .

When focusing on a single cell, with the  $E$  field locations at the edges of a square and  $H$  in the middle (see the previous sampling/interleaving condition), the computational update of the cell according to the previous steps is depicted in Figure 10.10.

Following Figure 10.10, at time  $t$ , the  $E$  fields are updated everywhere using spatial derivatives of  $H$ . Then at time  $t + 0.5$ , the  $H$  fields are updated everywhere using spatial derivatives of  $E$ . Every cell must be updated. The boundary condition  $\hat{E}_y^{(i,j+1/2)} = 0$  is applied at the cell boundary for each cell in the array.

At time  $t$ : update  $E_x$  and  $E_y$  using  $H_z$

$$\begin{cases} E_x^{t+\Delta t} = E_x^t + \frac{\Delta t}{\Delta y \epsilon_r \epsilon_0} (H_z^{y+\Delta y} - H_z^y) \\ E_y^{t+\Delta t} = E_y^t + \frac{\Delta t}{\Delta x \epsilon_r \epsilon_0} (H_z^{x+\Delta x} - H_z^x) \end{cases}$$



At time  $t + \Delta t$ : update  $H_z$  using  $E_x$  and  $E_y$

$$H_z^{t+\Delta t} = H_z^t + \frac{\Delta t}{\mu} \left( \frac{E_x^{y+\Delta y} - E_x^y}{\Delta y} + \frac{E_y^{x+\Delta x} - E_y^x}{\Delta x} \right)$$

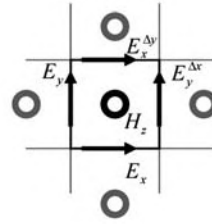


**Figure 10.10** The FDTD square cell update for  $E$  and  $H$

In order for the FDTD algorithm to be able to converge to a solution, an absorbing boundary has to be inserted around the device (see Figure 10.11). Figure 10.11 also shows how the boundary conditions of single cells are applied, especially for periodic structures such as the blazed grating depicted in the figure.

The advantage of the FDTD method is that it is relatively straightforward to implement. It can model inhomogeneous and anisotropic media, and it can be applied to periodic and nonperiodic structures. The disadvantages of the FDTD method are that it needs to sample at very fine grids ( $\lambda/20$ ) for acceptable accuracy, and that it is also difficult to implement it in nonorthogonal grids.

The FDTD implementation discussed here is one of the most widely used FDTD implementations today. The FDTD method can be implemented in numerous ways, which have been widely discussed in the literature. The implementation methods include the Finite Element Method (FEM), the Boundary

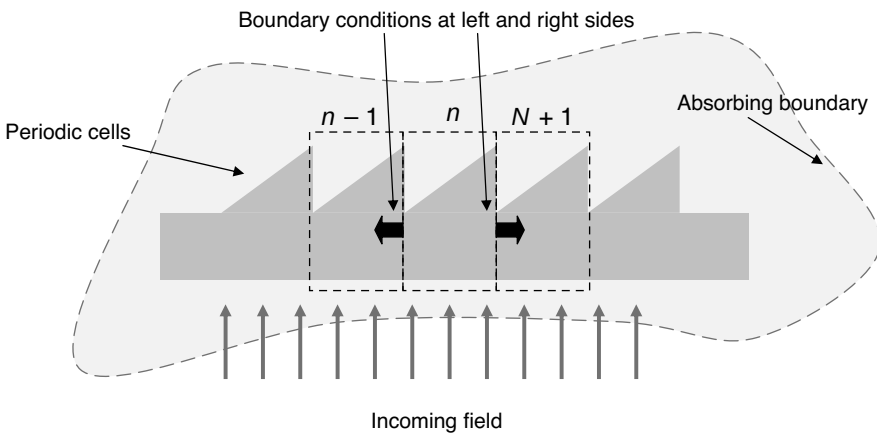


**Figure 10.11** Repetitive cell structures (blazed grating) and the absorbing boundary

Element Method (BEM), the Boundary Integral Method (BIM), the hybrid FEM–BEM method and other finite difference methods.

## 10.4 Engineering Effective Medium Optical Elements

When light of wavelength  $\lambda$  interacts with a micro- or nanostructured optical surface, there are mainly three scenarios that can take place:

- *The traveling photon*: the light sees the macroscopic structures on which it bounces or traverses, and does not see any lateral or longitudinal replications of such structures, but only the individual structure (since they do not exist, or they are too far away). Waves can thus be considered as being composed of individual rays, which are reflected or refracted through these individual structures.
- *The hawk-eye wave*: when this same light falls onto a structured surface (index modulation or surface modulation) that lateral dimensions of which are larger than the wavelength, but not too large, the light might be able to see the replicated structures laterally or longitudinally, and thus diffracts according to the geometry and replication patterns of these structures.
- *The blind wave*: when this same light falls onto a microstructured surface, the structures of which are smaller than its wavelength, the light no longer sees these structures (the light becomes ‘blind’), and it is affected by an effective medium the refractive index of which is defined locally by the micro- or nanostructures composing the structure.

This section shows how a blind wave gets diffracted by sub-wavelength structures that it can no longer ‘see’ as individual structures, but that form an effective index distribution. Such an effective index distribution will then act on the incoming wave and diffract it according to its modulation function.

As the incident wave cannot resolve the sub-wavelength structures, it sees only the spatial average of its material properties (the effective refractive index). In such a case, binary (two-level) structures can yield effective indices that are not binary, and can change in a quasi-continuous way along the substrate. This is a very desirable feature in industry, since multilevel or quasi-analog surface-profile fabrication of microstructures (which is required for high diffraction efficiency) is very costly and difficult (see Chapters 12 and 13), whereas binary fabrication is a relatively cheap way of producing microstructured optical elements, even if these structures are sub-wavelength.

For example, the design and fabrication of sub-wavelength blazed Fresnel lenses or sawtooth gratings have been reported in the literature. The incident wave considers the material as a smoothly varying phase element. The Effective Medium Theory (EMT) is applied to such elements in order to model them.

### 10.4.1 Rytov’s Method

In 1956, S.M. Rytov [10] developed several theories that allow the calculation of the effective permittivity for each layer, considering that the grating can act as a thin film multi-layer element, in order to ‘layer’ the grating profile into several distinctive homogeneous layers. Figure 10.12 depicts Rytov’s technique for the effective permittivity of layered media.

For more details on the implementation of modal and other resolution techniques for Maxwell’s equations, see Chapter 8 for the application of Kogelnik’s theory, and Chapter 11 for a numeric implementation of the Finite Distance Time Domain (FDTD) algorithm. These methods permits a dramatic reduction in CPU time and are used especially for applications involving Anti-Reflection Surfaces (ARS) and diffractive lenses.

### 10.4.2 Effective Index Calculations

The approximate EMT method (the zero-order EPT method) yields good approximations and ease of calculation of effective indices [11–13]. For example, for a binary surface-relief grating with a duty

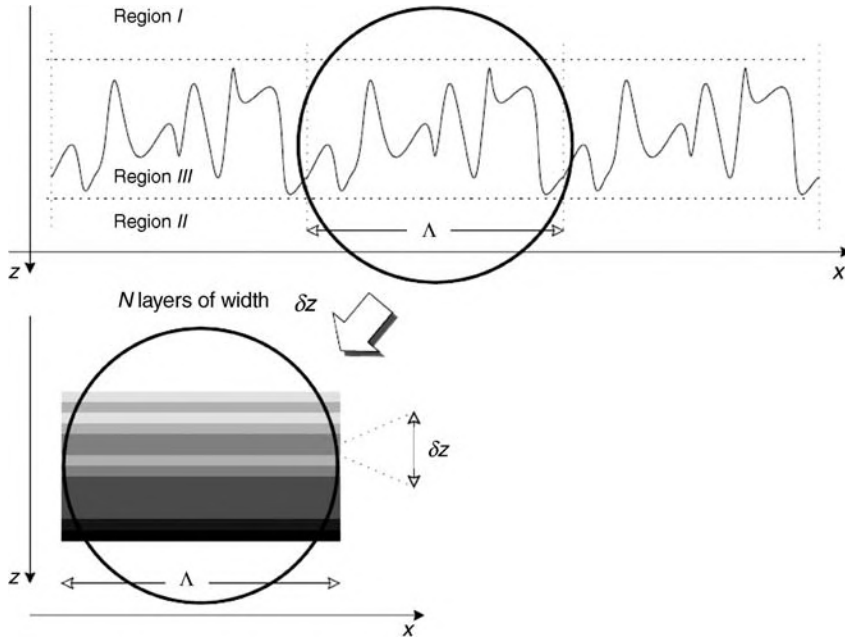


**Figure 10.12** Rytov's technique for the effective permittivity of layered media

cycle of  $c$  ( $0 < c < 1.0$ ), the effective index  $n_{eff}$  can be calculated as follows:

$$n_{eff} = \begin{cases} n_0^{(0)} = \sqrt{n_1^2(1-c) + n_3^2c} & \text{(TE)} \\ n_E^{(0)} = 1/\sqrt{(1-c)/n_1^2 + c/n_3^2} & \text{(TE)} \end{cases} \quad (10.32)$$

Engineering an effective index boils down to calculating the optimal duty cycle  $c$ , for a given wavelength and material index  $n_3$  and a surrounding medium  $n_1$  (in the case of air,  $n_1 = 1$ ). The duty cycle is thus given by the following equation (a duty cycle of 1.0 means that all grating grooves are filled with  $n_3$  material):

$$c = \begin{cases} n_1/(n_1 + n_3) & \text{(TE)} \\ n_3/(n_1 + n_3) & \text{(TM)} \end{cases} \quad (10.33)$$

This calculation is valid for zero-order refractive indices. The expressions for second-order and higher-order refractive indices can be used to yield greater accuracy in the effective index calculations.

The following equations link the coupled effective indices  $n_O$  and  $n_E$ , respectively, for the TE and TM polarizations, for higher-order EMT theories, for a normal incidence angle. For a second-order refractive index expression, the coupled effective index relations become

$$\begin{cases} n_O = \sqrt{n_O^2 + \frac{1}{3} \left( \pi \frac{\Lambda}{\lambda_0} c \cdot (1-c) \right)^2 (n_2^2 - n_1^2)^2} \\ n_E = \sqrt{n_E^2 + \frac{1}{3} \left( \pi \frac{\Lambda}{\lambda_0} c \cdot (1-c) \right)^2 \left( \frac{1}{n_2^2} - \frac{1}{n_1^2} \right)^2 \cdot n_E^6 \cdot n_O^2} \end{cases} \quad (10.34)$$

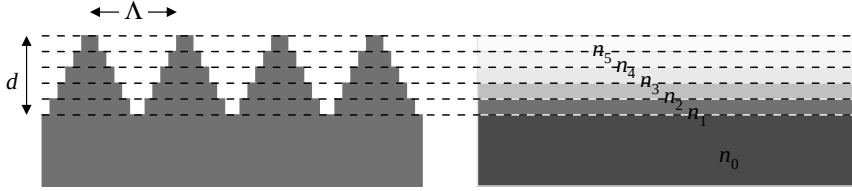


**Figure 10.13** Anti-reflection sub-wavelength structures

For higher-order refractive index expressions, the following coupled equations have to be solved for  $n_O$  and  $n_E$ :

$$\begin{cases} \sqrt{n_1^2 - n_O^2} \cdot \tan\left(\pi \frac{\Lambda}{\lambda_0} (1 - c) \sqrt{n_1^2 - n_O^2}\right) = -\sqrt{n_2^2 - n_O^2} \cdot \tan\left(\pi \frac{\Lambda}{\lambda_0} c \sqrt{n_2^2 - n_O^2}\right) \\ \frac{\sqrt{n_1^2 - n_E^2}}{n_1^2} \cdot \tan\left(\pi \frac{\Lambda}{\lambda_0} (1 - c) \sqrt{n_1^2 - n_E^2}\right) = -\frac{\sqrt{n_2^2 - n_O^2}}{n_2^2} \cdot \tan\left(\pi \frac{\Lambda}{\lambda_0} c \sqrt{n_2^2 - n_E^2}\right) \end{cases} \quad (10.35)$$

When  $\Lambda/\lambda$  approaches zero, the second- and higher-order refractive index expressions converge to the zero-order expression.

#### 10.4.3 The Longitudinal Effective Index (ARS Surfaces)

Anti-reflection surfaces (ARS structures) can usually be fabricated as pyramidal structures with periods smaller than the wavelength [14–16]. The resulting effective index is thus formed in the longitudinal way (normal to the surface). Figure 10.13 shows how a one-dimensional pyramidal structure can produce a continuous slowly varying effective medium, which introduces a smooth transition from a substrate index into air (or to any other index in which this surface structure's substrate is inserted).

There are many advantages of using such pyramidal etched ARS structures, as compared to more traditional multilayer thin film technology (either multiple index layers or binary Fabry–Perot type filter thin film stacks). These advantages are as follows:

- the structures can be mass-produced by etching, embossing, UV curing or injection molding;
- there is a monolithic substrate with single optical characteristics; and
- there is a very high laser damage threshold if etched into a hard substrate (no multi-layer effects and potential lift-off).

#### 10.4.4 Lateral Effective Index Engineering

More recently, in 1987, M. Farn at Lincoln Laboratory developed an analogous theory [12] to the one described in the previous section. Although he also considered an effective medium, he considered the effective medium to appear to the incident wave in the lateral direction, parallel to the surface rather than normal to that surface (see Figure 10.14).

As in Rytov's method (see Figure 10.12), where the effective medium was considered to appear in the normal direction, here also only the *effective index of refraction* is of interest; hence, the grating can be considered as a graded-index grating. Here, the periods of the grating can be greater, or much greater, than the actual reconstruction wavelength, but the structures that compose one period of that same grating are of sub-wavelength dimensions. In fact, this graded-index profile acts like a nonlinear phase shift.

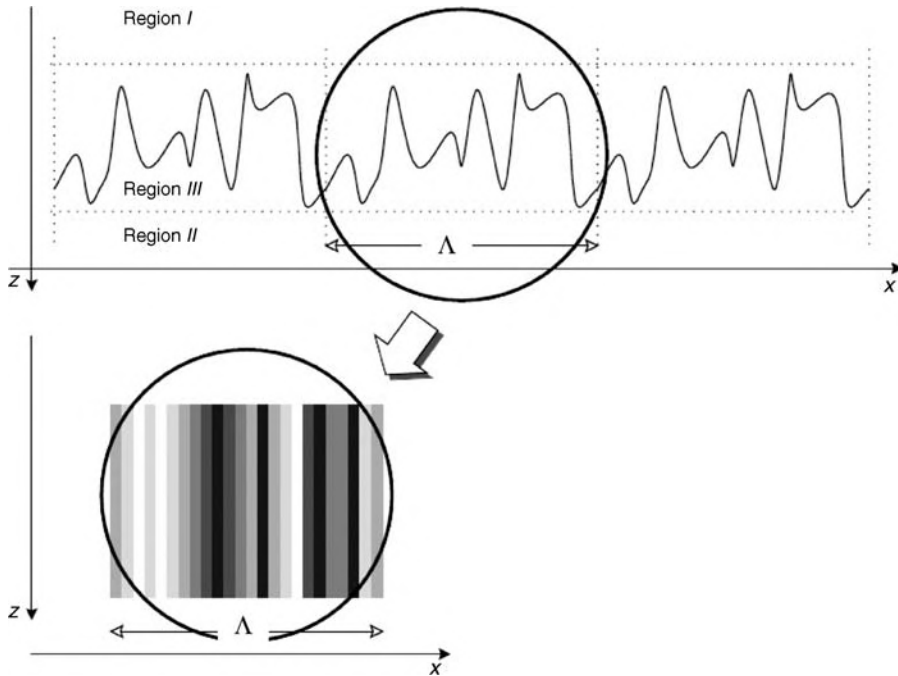


**Figure 10.14** The effective medium in the lateral dimension

This theory is mainly used for the analysis of the behavior of sub-wavelength chirped gratings. In particular cases, a binary chirped grating (with two phase levels) can be modeled efficiently using this theory as a smooth blazed grating that has the same period (Figure 10.15).

Figure 10.15 regroups the four different physical grating implementations reviewed up to now:

- A. The analog surface-relief element (equivalent to a refractive – or prism – element).
- B. The multilevel surface-relief grating (fabricated by multiple masking – see Chapter 12).
- C. The graded-index grating, fabricated by continuously varying refractive indices.
- D. The binary sub-wavelength grating, fabricated by digital lithography.

Note that in both the multilevel and binary versions (B and D in Figure 10.15), the individual binary features are much smaller than the wavelength, but the resulting period is larger than the wavelength, thus producing a regular diffractive element with sub-wavelength structures.

#### 10.4.5 Example: the Blazed Binary Fresnel Lens

The chirp of the local binary sub-wavelength grating can be modulated in order to imprint the successive zone widths of a Fresnel lens (see Chapter 5). If these gratings are made circular, one can implement a circular blazed high-efficiency diffractive Fresnel lens with only binary structures (see Figure 10.16).

There are, however, several differences between the specifications of a blazed Fresnel lens and its binary sub-wavelength counterpart. The diffraction efficiency can be in some cases very different, this being due to strong polarization effects in D that are not present in A, to changes in efficiency versus wavelength (stronger in D) and finally to the angular bandwidth (also stronger in D). In a general way,



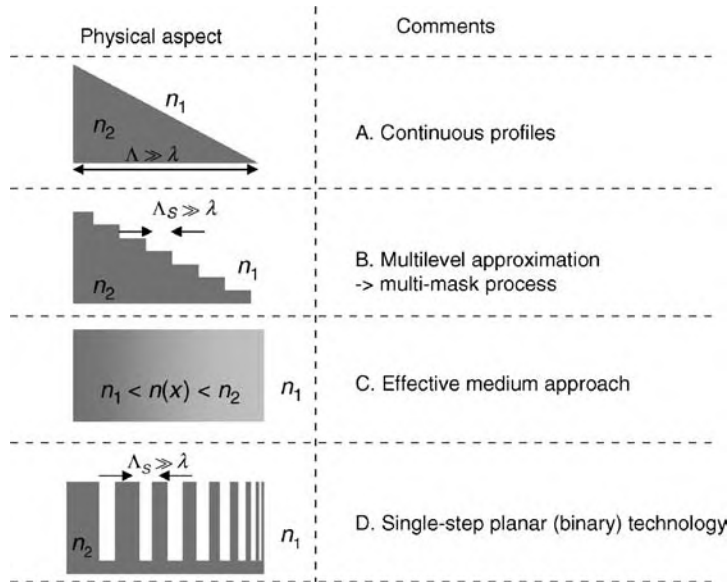


**Figure 10.15** A chirped grating, the corresponding effective index medium and the corresponding surface-relief element (in multilevel or analog surface-relief profile)

version A is more forgiving and has a greater wavelength and larger angular bandwidths than the D versions.

*10.4.6 The Pulse Density and Pulse Modulation for an Effective Medium*

There are mainly two ways to produce effective analog phase profiles with binary structures: pulse width modulation and pulse density modulation. These techniques are well known in the publishing industry,

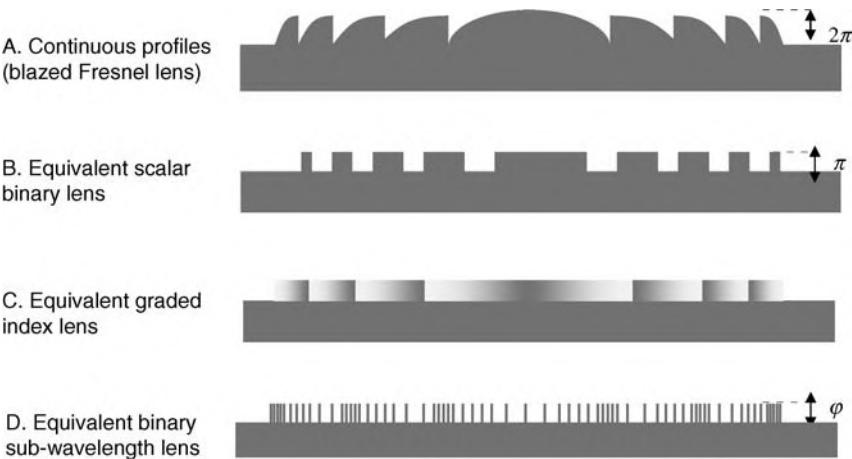


**Figure 10.16** A scalar blazed Fresnel lens fabricated by sub-wavelength chirped binary gratings

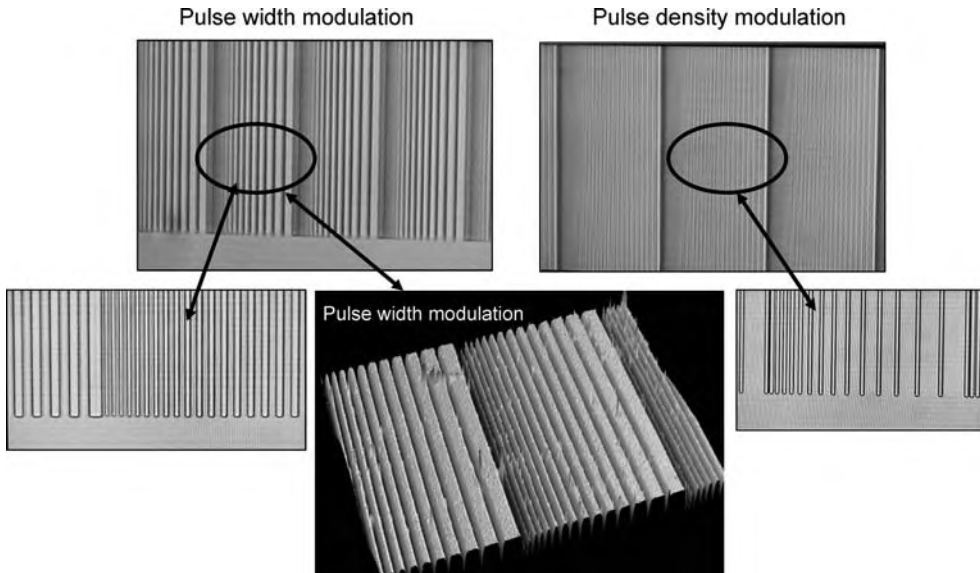


**Figure 10.17** The pulse density and pulse width modulation of sub-wavelength binary structures etched in quartz

and have been developed as part of traditional printing techniques (magazines, newspapers) and also especially the laser printers, to produce gray-scale patterns on paper, with patterns of black dots. Note that gray-scale photomasks have actually been produced by high-resolution PostScript printers using such binary pulse modulation techniques.

Figure 10.17 shows some optical microscope photographs and profilometry plots of the pulse density modulation and pulse width modulation of sub-wavelength gratings to produce effective analog phase profiles for grating structures (plots generated by a confocal interferometric white light microscope). The period of the local effective medium fringe shown in Figure 10.17 is much larger than the wavelength used, whereas the smallest binary features are well below the wavelength used (this element is to be used in reflection mode with a  $\text{CO}_2$  laser).

It is interesting to note the similarity between the pulse density and pulse width modulation techniques used to produce binary gray-scale effects in lithography (see Chapters 12 and 13) and the pulse density and pulse width modulation of sub-wavelength binary structures to produce gray-scale phase profiles (Figure 10.18).

In order to overcome polarization effects, one method is to produce two-dimensional grating lines rather than one-dimensional grating lines, even though the index in one dimension is constant, as depicted in Figure 10.17.

Figure 10.17 shows pulse density and pulse width modulations over 1D linear structures (which are highly polarization dependent), whereas Figure 10.18 shows the same type of grating with a 2D distributed pulse density or pulse width modulation, which greatly reduces the polarization effects.

## 10.5 Form Birefringence Materials

Diffraction gratings with periods much larger than the wavelength do not produce any polarization effects, and thus scalar theory is applicable for either s or p polarized light (see Chapters 5 and 6). One of the characteristics of sub-wavelength gratings is their high polarization dependency [17]. Such polarization

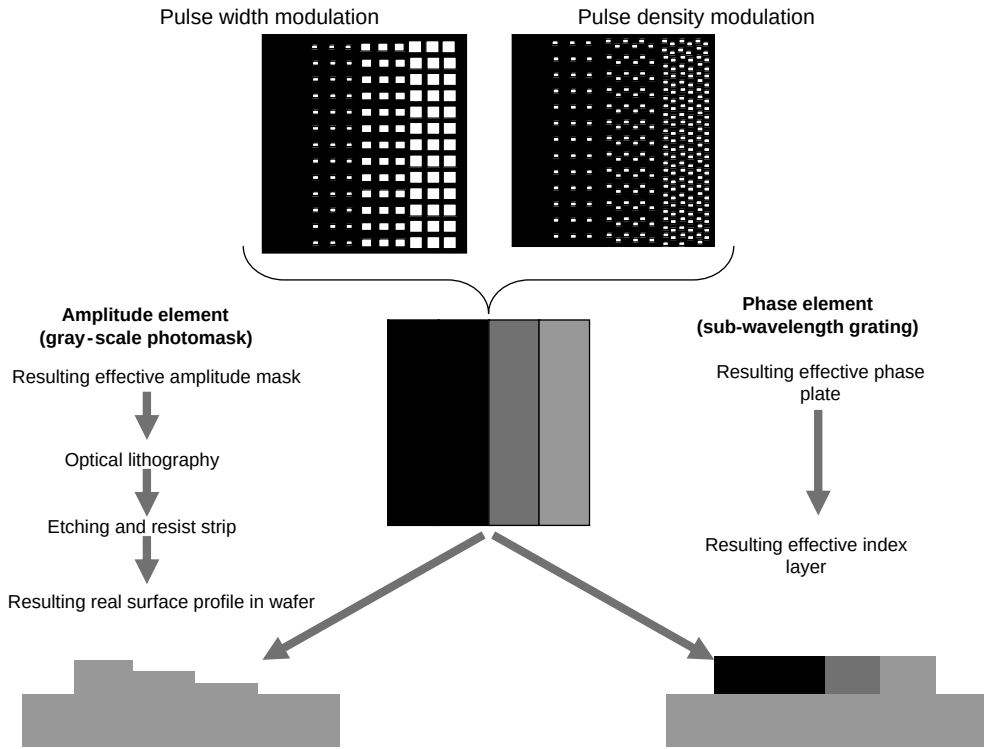


**Figure 10.18** The similarity between the pulse density modulation of amplitude gray-scale lithography and pulse sub-wavelength structures

effects on anisotropic materials are defined as form birefringence. Traditional polarization functionality generally requires highly birefringent materials or complex thin film polarization filters.

Sub-wavelength gratings can induce artificial birefringence at any wavelength in anisotropic materials (there is no need for the material to be birefringent). The amount of birefringence (or index difference  $\Delta n$ ) can be controlled accurately, and can be modulated across the substrate with the grating pattern modulation (period, duty cycle, depth, chirp etc.). Table 10.1 shows some of the  $\Delta n$  values achieved with form birefringence in 50% duty cycle gratings in zero-order EMT theory [18].

As can be seen in Figure 10.15, the form birefringence effect can be much stronger than the highest natural birefringence found in natural materials such as calcite or proustite. Form birefringence induced by such sub-wavelength grating surface profiles thus seems to be the perfect technology with which to engineer custom birefringence in materials that are already widely used in lithography (silicon or fused silica wafers).

### 10.5.1 Example 1: Polarization Beam Splitters/Combiners

Polarization beam splitters and polarization beam combiners are very desirable elements in a wide range of applications, especially in telecoms. Such applications include pump laser polarization combiners and polarization splitters for polarization diversity techniques in order to reduce PDL in active or passive optical PLCs (see also Chapter 4). Laser cavity polarization mirrors can also be implemented by using this

**Table 10.1** Natural and form birefringence for various media

| Form birefringence    |                        |       |                        |
|-----------------------|------------------------|-------|------------------------|
| Material              | $\lambda_0(\text{nm})$ | $n$   | $\Delta n = n_E - n_0$ |
| Fused Silica          | 630                    | 1.46  | -0.084                 |
| Photoresist           | 630                    | 1.64  | -0.151                 |
| ZnSe                  | 1.5                    | 2.46  | -0.568                 |
| GaAs                  | 1.0                    | 3.40  | -1.149                 |
| Si                    | 1.0                    | 3.50  | -1.214                 |
| Ge                    | 10.0                   | 4.00  | -1.543                 |
| Natural birefringence |                        |       |                        |
| Material              | $n_0$                  | $n_E$ | $\Delta n = n_E - n_0$ |
| Quartz                | 1.544                  | 1.533 | 0.009                  |
| Rutile                | 2.616                  | 2.903 | 0.287                  |
| ADP                   | 1.522                  | 1.478 | -0.044                 |
| KDP                   | 1.507                  | 1.467 | -0.040                 |
| Calcite               | 1.658                  | 1.486 | -0.172                 |
| Proustite             | 3.019                  | 2.739 | -0.280                 |

method, in order to produce a highly polarized beam. Gratings can be designed in a circular configuration in order to produce more complex polarization states in laser cavities.

Highly polarization dependent reflection gratings can be designed with the effective medium theory and modeled with modal analysis as seen previously (RCWA). Such a polarization splitting grating is depicted in Figure 10.19.

The reflectance curve of the polarization splitter in shown in Figure 10.19. The optimal operation for this grating lies in the C band (around 1.5  $\mu\text{m}$ ), where the extinction ratio between TE and TM

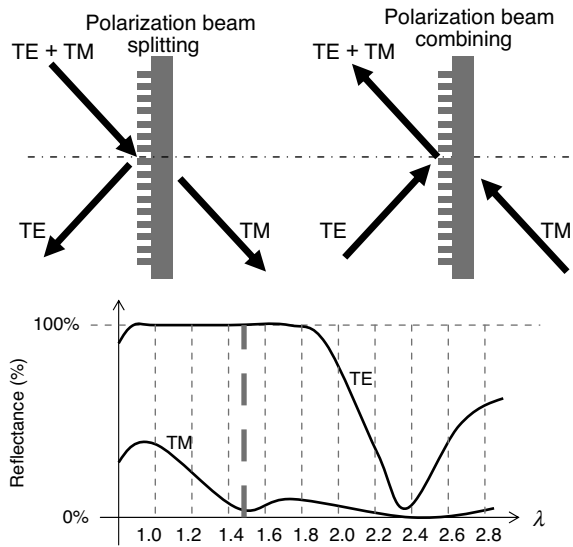


**Figure 10.19** Polarization splitter and polarization combiner gratings

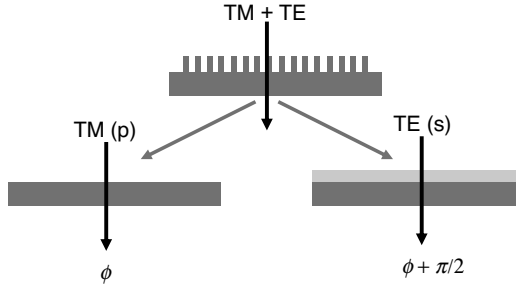


**Figure 10.20** The implementation of a quarter wave plate with sub-wavelength gratings

polarizations is maximal. Typically, such gratings are fabricated by alternating layers with high and low indices; for example, Si ( $n = 3.48$ ) and  $\text{SiO}_2$  ( $n = 1.44$ ).

### 10.5.2 Example 2: Wave Plates

Wave plates are used to slow down one polarization state in regard to the other in order to rotate the polarization state. Effective medium gratings can produce such wave plates. Figure 10.20 shows a binary grating producing an effective index required for a  $\lambda/4$  phase shift for s polarized light, thus implementing a quarter wave plate for a specific wavelength and a specific incoming angle.

Bear in mind that although form birefringence polarization beam splitters and wave plates seem to be very desirable, the constraints linked to such devices are mainly that:

- as the wavelength decreases, it becomes increasingly difficult to fabricate such sub-wavelength structures; and
- the extinction ratio is a maximum only for a specific spectral bandwidth (small).

Therefore, natural birefringence materials still have good prospects, when they are used in preference to broadband spectral and angular spreads, especially for shorter wavelengths.

## 10.6 Guided Mode Resonance Gratings

Based on the effective medium theory developed in the previous section, guided mode resonant waveguide gratings can be designed and used as precise spectral filters, for a variety of applications, including narrowband spectral filters for DWDM telecom devices (see also Chapter 3). Such gratings are usually fabricated in optical materials with high indices. Metallic resonant gratings will be investigated in Section 10.7.

A guided mode resonance filter structure is a zero-order grating (surface-relief or index modulation) that transmits and/or reflects a single order. This type of grating is depicted in Figure 10.21.

The grating here is composed of materials with alternating high and low indices, which forms an effective index  $\epsilon_{II}$ . The condition for zero-order grating operation is given by

$$\frac{\Lambda}{\lambda} < \frac{1}{\text{Max}\{\sqrt{\epsilon_I}, \sqrt{\epsilon_{III}}\} + \sqrt{\epsilon_{II}} \sin \theta_{\max}} \quad (10.36)$$

where  $\epsilon_I$ ,  $\epsilon_{II}$  and  $\epsilon_{III}$  are defined in Figure 10.21. A guided mode resonant grating will reflect and transmit zero-order beams, with respect to the wavelength. Such filters can be very narrow notch rejection filters.

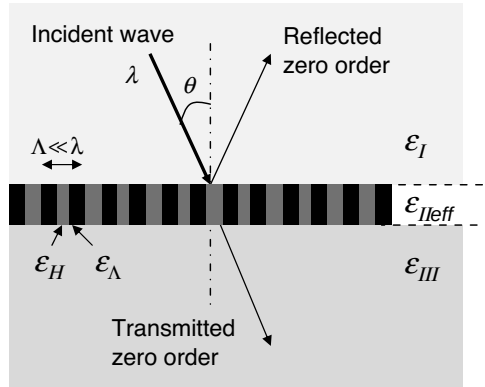


**Figure 10.21** A zero-order grating

The thickness of the dielectric grating can be considered as a waveguide structure with an effective refractive index. This effective refractive index is a function of the wavelength (see Section 10.4). The effective index as well as the waveguide thickness is important in determining the parameters of the filter response, such as the sideband reflectance, line shape and free spectral range (see also Chapter 8). A symmetrical filter response is usually desired in most applications, and can be implemented by choosing the waveguide thickness to induce a  $\pi$  phase shift (i.e. a multiple of half-wavelengths). The resonance wavelength is set by the grating period.

There are numerous ways to implement a guided-wave resonance grating filter, with one or many additional thin film layers over or under the grating, and with one or more gratings. Figure 10.22 shows some of the implementations used in industry today and their typical spectral responses.

When many thin film layers are used in the filter design, the filter response is dictated by a combination of the waveguide grating and thin film interference effects. Close to the resonance wavelength, the strong

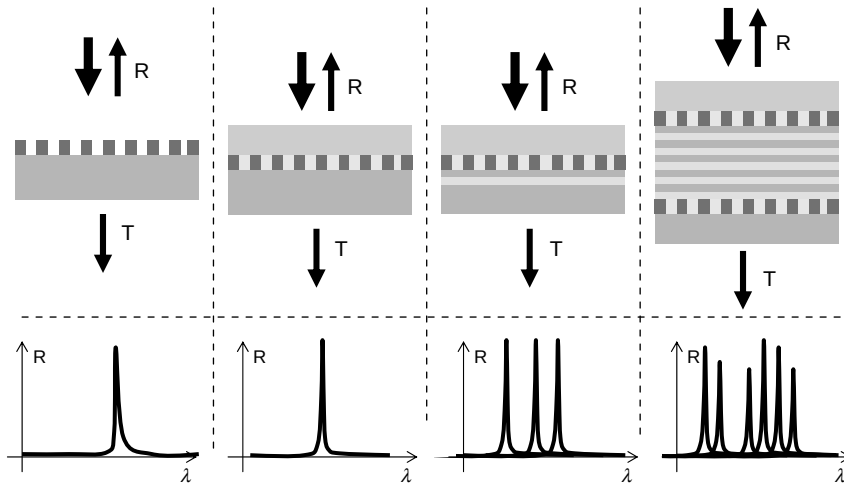


**Figure 10.22** Physical implementations of guided wave resonance grating filters

coupling between the external propagating waves and the adjacent evanescent waves produces a rapid variation in reflectance. Outside the resonance wavelength, the conventional thin film layer interference effect is predominant.

Such gratings have strong polarization effects, and behave very differently depending on the injected polarization state (which can be a very desirable effect). Therefore, resonant gratings can also be applied to laser mode selections, as resonant laser mirrors, and for sharper wavelength selection, polarization stabilization and transverse mode filtering.

## 10.7 Surface Plasmonics

The previous section reviewed resonant guided mode gratings fabricated in optical dielectric materials. Another type of resonant grating will now be reviewed, this time fabricated as thin metallic structures on top of a waveguide structure.

Recently, strong interest has been focused on sub-wavelength three-dimensional nano-metallic structures that trigger the Surface Plasmon (SP). Such effects were already well known and used in photonic sensors based on metallic sub-wavelength linear gratings patterned on a waveguide structure (see Section 16.6).

### 10.7.1 Surface Plasmon Polaritons (SPPs)

Surface Plasmon Polaritons (SPPs) are electromagnetic surface waves that are confined to the interface between materials with dielectric constants of opposite sign (e.g. at a metal/glass interface). Figure 10.23 shows such an interface.

This surface confinement of the SPP field to the interface opens up the possibility of overcoming the diffraction limit encountered in classical optics and enables the fabrication of new digital waveguides with sub-wavelength structures (grating) to produce highly integrated optical devices.

### 10.7.2 In- and Out-coupling

One of the main problems with surface plasmon devices is to couple light into the metallic grating region and to detect the light (which cannot be detected by classical optics). Figure 10.24 shows how light can be coupled onto an SPP grating by a surface-relief binary grating coupler etched into the slab (see also Section 3.4).

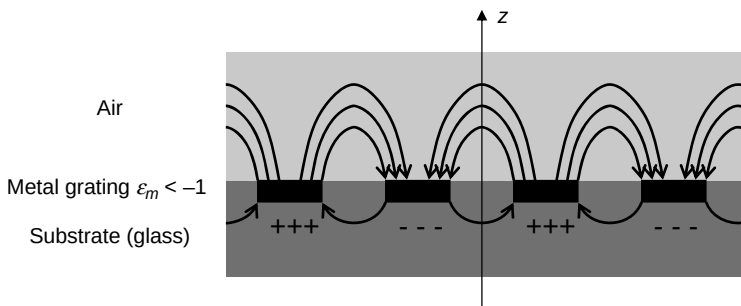


**Figure 10.23** A metallic interface producing surface plasmon polaritons

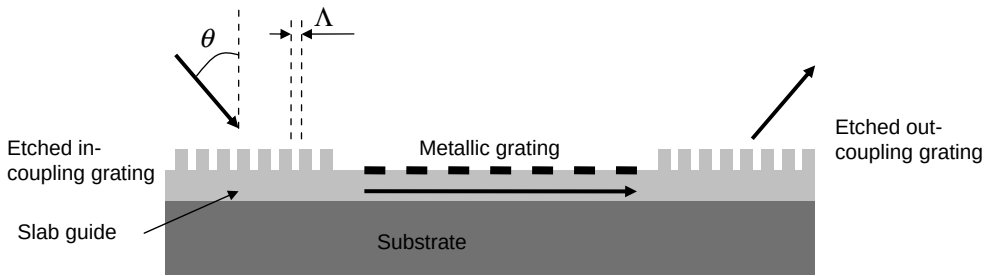


**Figure 10.24** SPP light in- and out-coupling

The coupled SPP propagation vector  $k_{SPP}$  can thus be written as

$$k_{SPP} = \frac{2\pi}{\lambda} \sin \theta + n \frac{2\pi}{\Lambda} \quad (10.37)$$

### 10.7.3 Implementing Planar Optical Plasmon Functionality

Simple holes (of nanoscale dimension) etched into thin metallic films create SPP effects and act as point sources. This is a unique feature that can be used to implement various optical functionalities in planar waveguides, namely lensing effects, grating effects or even beam-splitting effects. Figure 10.25 shows a hybrid implementation of such SPP-induced optical functionality on a planar substrate.

Such optical functionalities can be coupled with conventional waveguide devices or with linear strips of metallic materials, which can produce the lightning rod effect (energy focused by the SPP lens and building up around the nano-rod).

### 10.7.4 SPP Sensors

The surface plasmon effect extends only a few hundred nanometers over the surface of the metallic grating. The propagation of this surface plasmon is strongly affected by the local refractive index in this region (air, solid, gas, liquid etc.) Therefore, if the medium in which the grating is in contact changes (i.e. its refractive index changes), the amount of light coupled into that medium also changes, and

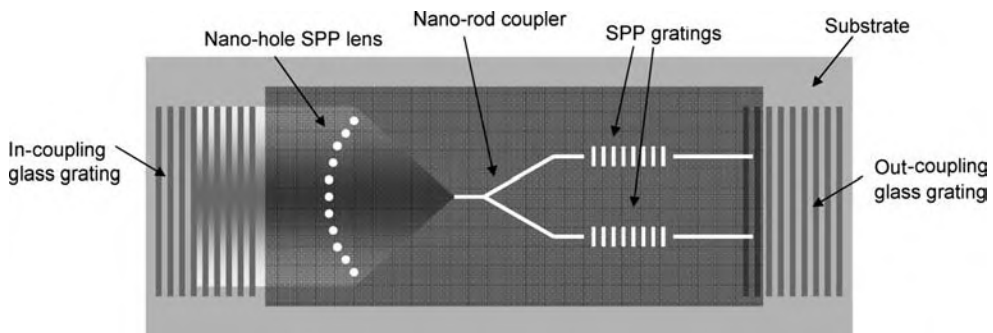


**Figure 10.25** The implementation of SPP optical functionality on a slab with holes in a thin metallic film



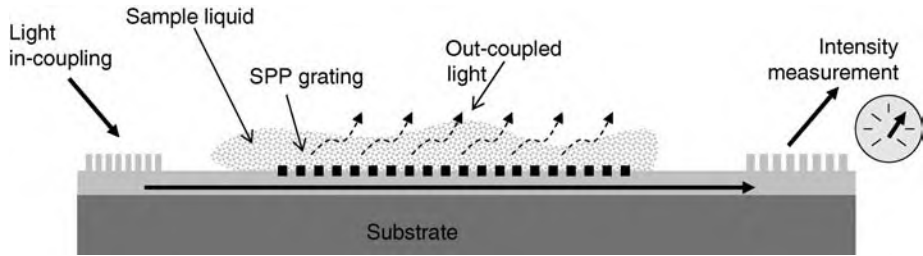


**Figure 10.26** The implementation of a surface plasmon sensor based on a sub-wavelength metallic grating

therefore yields a modulation of the light that can be out-coupled by a grating to be measured. Such devices can measure refractive index variations in a solution for the implementation of, for example, chemical sensors (see Figure 10.26).

SPP sensors can measure the light reflected (with light lost due to the SPP effect) or, on the contrary, the amount of light propagated due to the SPP (see also Chapter 16).

The resonant SPP effect can build up in 1D or 2D metallic sub-wavelength gratings and increase the intensity of the field at the edges of these nanoscopic metallic elements, thus creating nonlinear optical effects, which can be used to develop new optical functionalities. Many efforts have been recently directed to surface plasmon effects in metallic nano-wires and metallic nano-rods (see Section 10.9).

## 10.8 Photonic Crystals

The term ‘Photonic Crystal’ (PC) and its basic effects have been established relatively recently, but are based on predictions formulated back in 1987 by Eli Yablonovitch [19, 20] (at Bell Communications Research in New Jersey) and Sajeev John (at the University of Toronto).

### 10.8.1 From Holograms to Photonic Crystals

Photonic crystals are basically semiconductor materials for light waves, since they behave in the same way. The term ‘photonic band gap’ has thus been proposed to point out such similarity with traditional electronic band gap structures. PCs are based on nanostructures with high index differences,  $\Delta n$ , of over 1, 2 and even 3! Chapter 8 has reviewed holographic elements with  $\Delta n$  values ranging from 0.005 to 0.05, and in the previous sections surface-relief elements with a  $\Delta n$  of 0.5 (from air to etched glass or embossed plastic) up to larger modulations with higher-index dielectric materials, used in the semiconductor industry, have been considered. Photonic crystals can be implemented as 1D, 2D or 3D structures, depending on the structure of the basic cell used and its replication geometry in the crystal. Figure 10.27 shows the slow migration from holographics, diffractives and sub-wavelength gratings to photonic crystals as a function of  $\Delta/\lambda$  and  $\Delta n$ .

Figure 10.27 shows that there is no discrete technological leap between holographics, diffractives and photonic crystals but, rather, a continuous evolution along smooth conducting lines, such as grating periods (the ratio  $\Delta/\lambda$ ), materials indices ( $n_{eff}$  and  $\Delta n$ ) and spatial geometry (1D, 2D and 3D). Such a continuous evolution has been enabled partially by the availability of new nano-fabrication techniques and strong index contrasts. The overlap between the three fields in Figure 10.23 shows how some holograms with high effective indices and  $\Delta n$  values (such as H-PDLC) can actually implement partial PC effects, and how surface-relief gratings with high  $\Delta n$  values can implement 1D or 2D PCs.

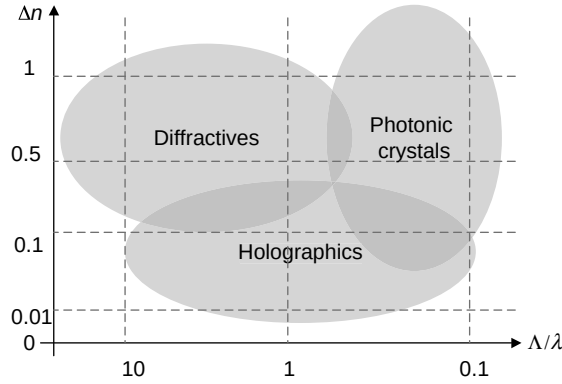


**Figure 10.27** From holograms to photonic crystals

### 10.8.2 Natural Photonic Crystals

Note that PCs have been around for quite some time. Natural photonic crystals include opals, abalone shells, peacock feathers and other morpho-butterfly wings. Opals are perhaps the most interesting natural PC structures, since they yield early perfect photonic band gaps over wide ranges of wavelengths and angles (see Figure 10.28).

### 10.8.3 The Theoretical Background

Maxwell's equations (see Equation (10.1) and Appendix A) yield the following separable sets of equations for the  $E$  and  $H$  fields in the steady state:

$$\begin{cases} \frac{1}{\epsilon(r)} \nabla \times \nabla \times \vec{E}(r) = \left(\frac{\omega}{c}\right)^2 \vec{E}(r) \\ \nabla \times \frac{1}{\epsilon(r)} \nabla \times \vec{H}(r) = \left(\frac{\omega}{c}\right)^2 \vec{H}(r) \end{cases} \quad (10.38)$$

Thus, Maxwell's equations for the steady state can be expressed in terms of an eigenvalue problem, in a direct analogy with quantum mechanics, which governs the behavior of electrons (see Table 10.2).

#### 10.8.3.1 The Master Equation and the Variation Theorem

The solution  $H(r)$  to the master equation (described in Equation (10.38)) can be expressed as the minimization of the functional  $E_f$ :

$$E_f[H] = \frac{1}{2} \frac{(H, \Theta H)}{(H, H)} \quad (10.39)$$

Equation (10.38) can be rewritten, by using the variational theorem, as follows:

$$E_f[H] = \frac{1}{2} \frac{1}{(H, H)} \int \frac{1}{\epsilon(r)} \left| \frac{\omega}{c} D \right|^2 dr \quad (10.40)$$

From this equation, the functional  $E_f$  is minimized when the displacement field  $D$  is concentrated in the regions of high dielectric constant; thus, the lower-order modes tend to concentrate their displacement fields in the region of high dielectric constant.

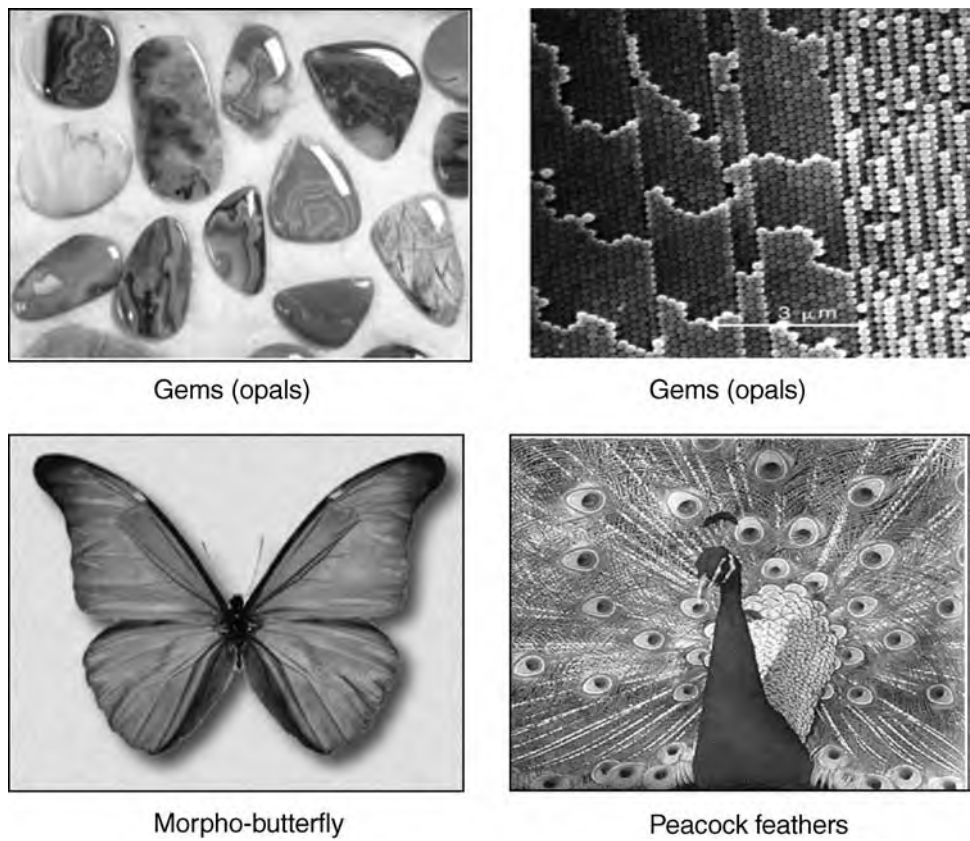


Figure 10.28 Natural photonic crystals

10.8.3.2 The Band Structure

The band structure, or dispersion relation, defines the relation between the frequency  $\omega$  and the wave vector  $\omega = c|k|$ . Figure 10.29 shows a simple band structure for a vacuum, in 1D and 2D. There are many ways in which the band structure can be represented, some of which are shown in Figure 10.29.

10.8.3.3 The Crystal Lattice and Bloch’s Theorem

Photonic crystals are periodic structures, which have singularity geometries where light can propagate. The PC lattice  $U_k(r)$  produces a periodic dielectric distribution, which can be written as  $U_k(r + a) =$

Table 10.2 The analogy between quantum mechanics and electromagnetism

|                   | Quantum mechanics                         | Electromagnetism                                                          |
|-------------------|-------------------------------------------|---------------------------------------------------------------------------|
| Field             | $\Psi(r, t) = \Psi(r)e^{i\omega t}$       | $H(r, t) = H(r) \cdot e^{i\omega t}$                                      |
| Eigenvalue system | $H\Psi = E\Psi$                           | $\Theta H = \left(\frac{\omega}{c}\right)^2 H$                            |
| Operator          | $H = -\frac{\hbar^2 \nabla^2}{2m} + V(r)$ | $\Theta = \nabla \times \left(\frac{1}{\epsilon(r)} \nabla \times\right)$ |

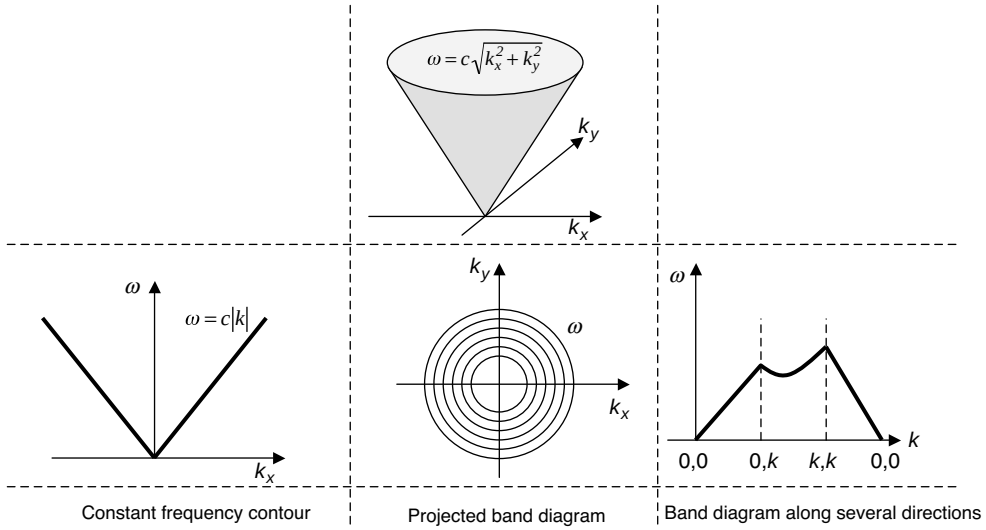


**Figure 10.29** The band structure for a vacuum

$U_k(r)$ . Bragg scattering through this periodic structure provides strong and coherent reflections at particular wavelengths, and is the origin of the photonic band gap. Light can be localized in a PC at defects (singularities), which are due to multiple scattering (interferences).

Figure 10.30 shows a general photonic lattice (here, in 2D), in which a plane wave is propagating with an electric field (Equation (10.41)).

$$\begin{cases} \epsilon(r) \cdot E_\lambda e^{j(\vec{k} \cdot \vec{r})} \\ \epsilon(r) = \epsilon(r + a) \end{cases} \quad (10.41)$$

The permittivity can thus be rewritten with reciprocal lattice vector  $\vec{G}$  as follows:

$$\epsilon(r) = \sum_G \epsilon_G \cdot e^{j(\vec{G} \cdot \vec{r})} \quad (e^{j(\vec{G} \cdot \vec{a})} = 1) \quad (10.42)$$

Therefore, the solution for the electromagnetic wave in a photonic crystal is

$$H(r) = e^{j(\vec{k} \cdot \vec{r})} \sum_G H_G \cdot e^{j(\vec{G} \cdot \vec{r})} \quad (10.43)$$

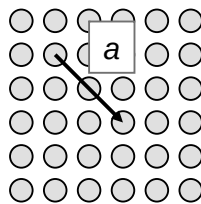


**Figure 10.30** An example of a 2D PC lattice

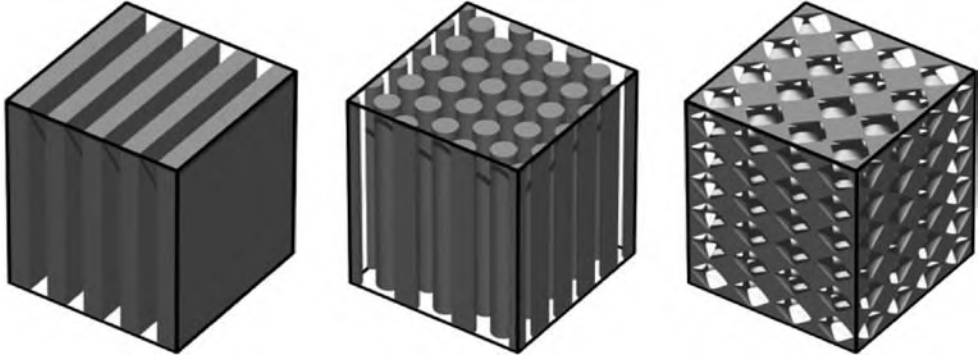


**Figure 10.31** Various photonic crystal configurations

A plane wave will only scatter to those plane waves that have their wave vectors differing by a *reciprocal lattice vector*  $G$ . The reflection is maximal if the Bragg condition is satisfied:

$$e^{2ika} = 1 \Leftrightarrow k = \pi/a \quad (10.44)$$

Photonic crystals can take on many different geometries. Four main geometries are reviewed below: the 1D, 2D and 3D crystal structures and the PC-based fiber (see Figure 10.31).

#### 10.8.4 One-dimensional Photonic Crystals

One-dimensional photonic crystals can be considered as periodic Bragg planes. Strong  $\Delta n$  effects in Bragg holograms had shown PC behavior long before the term PC itself was proposed. For example, a DBR or DFB laser is a prefiguration of a 1D photonic crystal.

Electromagnetic waves cannot propagate to the edges of the Brillouin zones (i.e.  $k = \pi/a$ ), and therefore can only take the form of a standing wave. Figure 10.32 shows the band structure for a simple 1D photonic crystal.

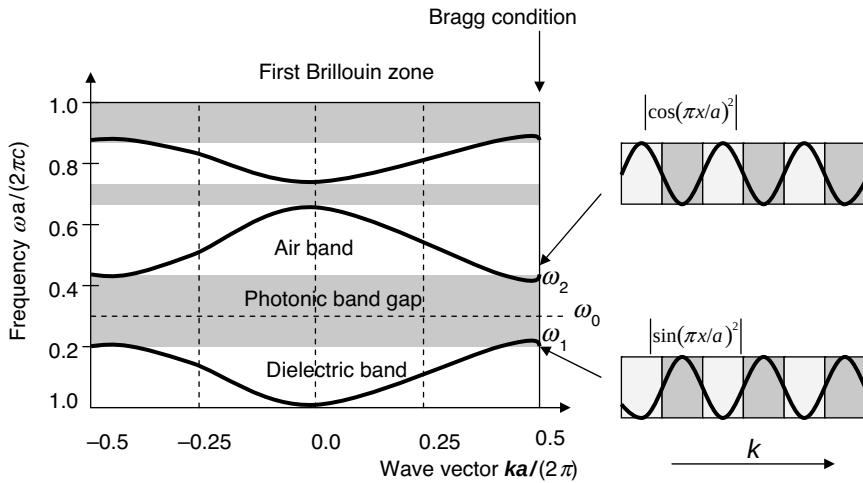


**Figure 10.32** The band structure and Brillouin zones for 1D PCs

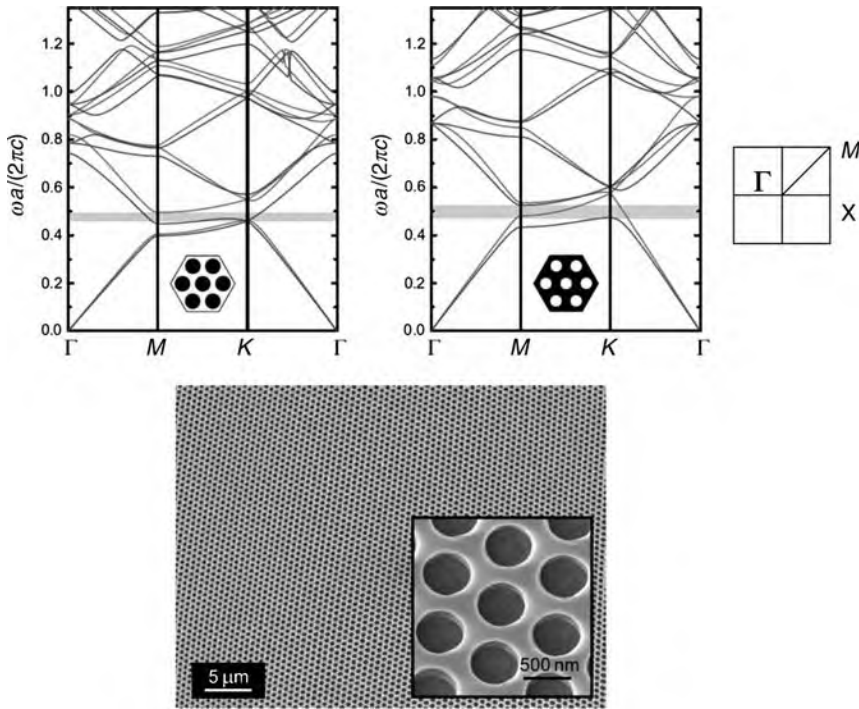


**Figure 10.33** Two-dimensional photonic crystals in a honeycomb lattice (courtesy of Dr Martin Hermatschweiler)

crystal. The Bragg condition coincides with the edges of the first Brieoullin zone. Here, the band structure is depicted in one propagation direction (the  $k$  vector). In multiple dimensions, one can depict several propagation directions on the same band structure. Polarization effects have to be included in the band structure. This makes it difficult to have complete photonic band gaps for each direction and each polarization.

### 10.8.5 Two-dimensional Photonic Crystals

Two-dimensional photonic crystals come in many different forms. Most of them are fabricated via nano-rods or nano-holes in a strong index material. Figure 10.33 shows the band structure of such 2D PCs in a honeycomb lattice, on the left side as arrays of holes and on the right side as arrays of rods. Note that here the band structure is plotted in different directions in 2D space. An SEM photograph of the fabricated structures is also shown in the same figure.

### 10.8.6 Three-dimensional Photonic Crystals

Three-dimensional photonic crystals are perhaps the most impressive, and the wood-pile structure the most picturesque of all. Standard polysilicon MEMS fabrication techniques have been used for the production of 3D wood-pile structures PCs (see Figure 10.34).

These crystals are also the most difficult to fabricate. Figure 10.34 also shows the natural assembly of nanospheres and casting through the assembly of nanospheres, close to opal structures. Natural 3D PCs

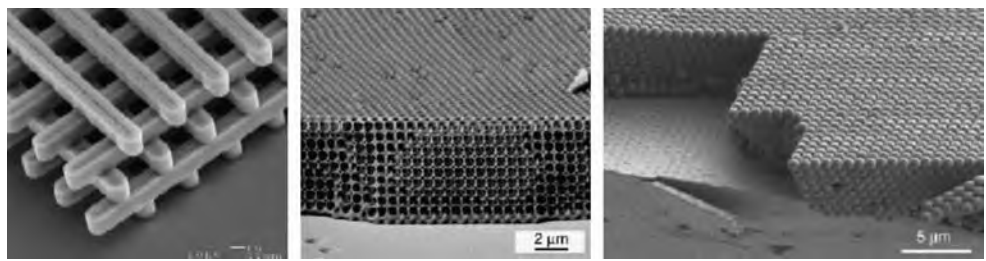


**Figure 10.34** Three-dimensional photonic crystals

include opals, morpho-butterfly wings and peacock feathers. The 3D hexagonal graphite structures provide such a total band gap over all directions and all polarizations.

*10.8.7 Ultrarefraction and Superdispersion*

The group velocity  $v_g$  is given by

$$v_g = \frac{\partial \omega}{\partial k} \tag{10.45}$$

Such a group velocity can be calculated from the band gap diagram. The group velocity in a PC is strongly modified by the highly anisotropic nature of the bandgap structure (anomalous group velocity dispersion). This can give rise to the ‘super-prism phenomenon’ that occurs in prisms based on PC materials. Figure 10.35 shows the strong angle-sensitive light propagation and dispersion effect. Only small changes in the angle of incidence can produce large changes in propagation of the refracted beam (up to one order of magnitude). This phenomenon is also called ultrarefraction.

Figure 10.35 also shows how superdispersion can cause a stronger separation of the different wavelength components, spreading over a much wider angle than in conventional prisms. Ultra-refraction can be used for beam steering over a wide angle.

*10.8.8 The ‘Holey’ Fiber*

The holey fiber PC (see Figure 10.36) is one of the first PCs to be adopted by industry (especially the optical telecom industry).

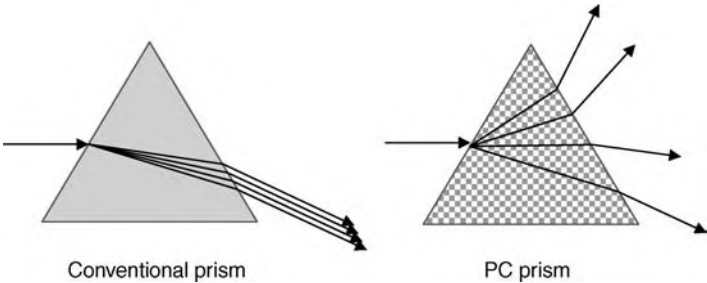


**Figure 10.35** The super-prism effect in a PC

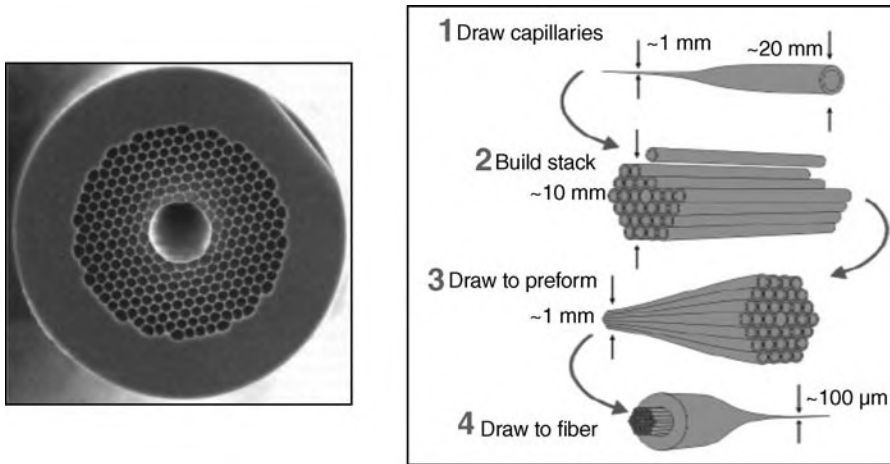


**Figure 10.36** The holey fiber and fabrication process

In a holey fiber, an assembly of micro-holes produce the band gap around the central air gap. Such a fiber is fabricated via a traditional fiber drawing technique, and is thus no more expensive than standard fibers. Such PC effects are optimized for operation with  $1.5\ \mu\text{m}$  and  $1.3\ \mu\text{m}$  laser light. In a PC fiber, there is no need to produce an index modulation in the fiber to create the core and cladding. The injected wave is coupled and guided along the PC band gap.

### 10.8.9 Photonic Crystals and PLCs

Planar Lightwave Circuits (PLCs; see Chapter 3) are a perfect platforms for photonic crystals. In a PC-based waveguide (see Figure 10.37), the light is not guided through total internal reflection but, rather, by the photonic bandgap effect around the nano-hole lattice. This opens the door to a complete new way of designing integrated PLCs, which have traditionally been constrained to design rules including slow bends, materials issues and other conventional waveguide rules.

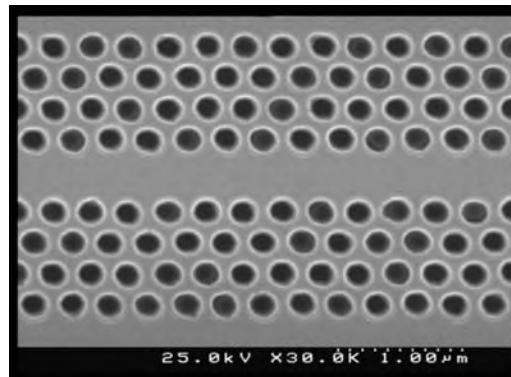


**Figure 10.37** The photonic crystal waveguide structure



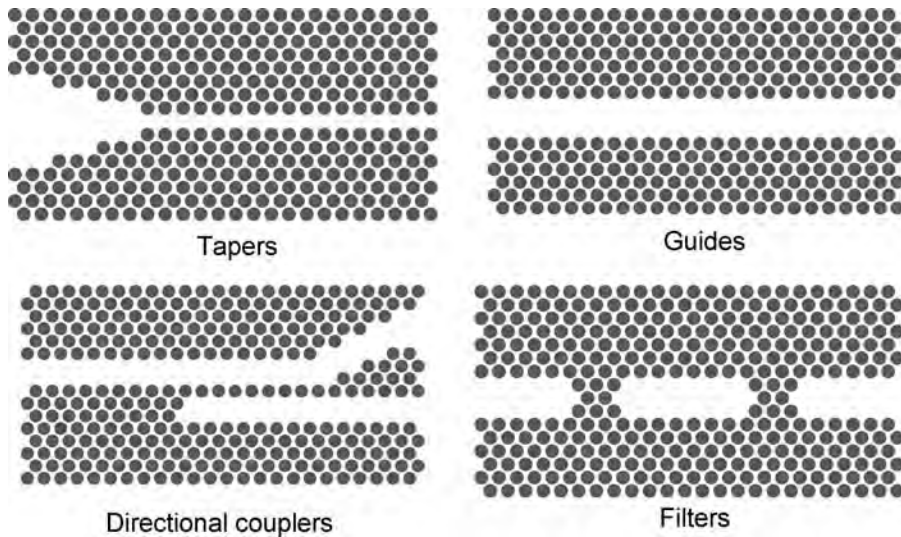


**Figure 10.38** Photonic crystal PLC building block functionality

In a PC-based PLC, the waveguide can produce sharp bends, even larger than  $90^\circ$ , without substantial loss of light, with very strong guidance. Therefore, PLCs can be integrated in smaller surfaces, with more complex functionalities. Some of these functionalities are depicted in Figure 10.38.

The functionality building blocks described in Figure 10.38 are the most basic ones, and can produce smaller and more efficient versions of the traditional PLCs described in Chapter 3. PC-PLCs are monolithic; no index modulation is needed. However, nanoscale lithography is needed along with high aspect ratio structuring (deep nano-holes etc.).

Photonic crystals are attracting considerable interest from industry; however, due to several issues they have been slow to penetrate conventional markets such as optical telecoms, displays and consumer electronics. These issues are mainly the lack of available fabrication techniques and the difficulty of light launching (coupling) into PC-based PLCs (in and out).

It is already difficult to couple efficiently light from a laser diode into a fiber, which has a core section of  $9\text{ }\mu\text{m}$ . It is much more difficult to couple light into a PLC PC with core dimensions in the nanoscale region. Similar techniques to the ones described in Chapter 3 are also used for photonic crystals. PC tapers can also be used on the PLC (see Figure 10.37).

### 10.8.10 Fabrication of Photonic Crystals

The fabrication and replication of the 3D nanoscale PC devices, in a material compound exhibiting a high enough  $\Delta n$  for strong effects, is a difficult task. Standard microlithography has been used to produce 1D and 2D photonic crystals, and MEMS planar polysilicon technology with sacrificial layers to produce more complex 3D structures (wood-piles etc.). Holographic exposure, and especially multiple holographic exposure, has been used to produce 3D periodic structures in thick resist (SU8). Furthermore, fringe-locking fringe writers and other nanoscale direct laser write techniques (such as two-photon lithography) have been developed (see Chapter 13).

Perhaps one of the most straightforward and simple fabrication techniques for photonic crystals is fiber drawing (for PC or ‘holey’ fibers). The PC fiber is fabricated like any regular fiber, from a macroscopic glass preform in which a series of relatively large holes is formed (see Figure 10.35). These holes are

reduced in size down to nanometer scale by simple fiber drawing. This is perhaps why holey fibers became the first PC application to be introduced to industry.

Self-assembly of nano-materials has also been developed in order to assemble crystals in 3D with various lattice geometries. Colloidal self-assembly can be used to produce an inverse opal structure to form close-packed polystyrene spheres. The resulting gaps can be filled with a high index refractive material, such as GaP ( $n \sim 3.5$ ).

In gravity sedimentation, particles in suspension settle to the bottom of a container, while the solvent evaporates. Therefore the proper conditions have to be found so that periodic lattices are formed during the evaporation process. Such a process can take up to several weeks to produce PCs.

In the cell method, an aqueous dispersion of spherical particles is injected into a cell made of two glass substrates. The bottom substrate is coated with a frame of photoresist. One side of the frame contains openings to allow the liquid/solvent to pass through, while the particles are retained. The latter settle to form an ordered structure. The thickness of the PC structure usually does not exceed  $20\text{ }\mu\text{m}$  and lateral extensions are on the order of 1 cm.

PCs are very sensitive to the smallest defects appearing during fabrication and replication.

## 10.9 Optical Metamaterials

Section 10.7 has reviewed surface plasmonics made by nanoscale metallic gratings on optical waveguide slabs and Section 10.8 has reviewed 3D photonic crystals fabricated in high-index materials. Based on these two nanoscale architectures, new materials involving metallic three-dimensional nanostructures have been developed, which include these two technological platforms.

Professor Richard Feynman predicted in the 1960s that ‘there is plenty of space at the bottom’ meaning that there is lots to do in the nanoscale region, as seen previously, especially with photonic crystals. One can now say on top of this that ‘there is plenty of space below the bottom’, when considering metamaterials and their impressive potential.

The exact definition of a metamaterial is a composite or structured material that exhibits properties *not found* in naturally occurring materials or compounds [21, 22]. Left-handed (negative refractive index – or optical antimatter) materials have electromagnetic properties that are distinct from those of any known material, and hence are examples of metamaterials. The metallic nanostructures composing a metamaterial have an active rather than passive effect on the incident light, as is the case in photonic crystals, diffractives and holograms. The oscillation of the field created by the light interacting with the material changes how the light is actually propagated in that same material.

Previous sections have reviewed sub-wavelength gratings, resonant waveguide gratings, photonic crystals, surface plasmon gratings. Figure 10.39 shows how these nanostructured optical elements can be organized.

David Smith and his colleagues at the University of California San Diego (UCSD) have developed a splitting resonator structure, having negative indices at microwave wavelengths (see Figure 10.40).

### 10.9.1 Left-handed Materials

One of most interesting characteristics of metamaterials is that they are left-handed materials, which means that their effective refractive index is negative. Figure 10.41 depicts the reversal of Snell’s law by negative-index metamaterials. It is also interesting to note that in the case of metamaterials, the wave vector is in the opposite direction to the propagation of light (Figure 10.41).

### 10.9.2 What’s in the Pipeline for Metamaterials?

Recently, industry – as well as the media and the academic world – have described with great enthusiasm and excitement the latest developments in optical metamaterials, which could lead to

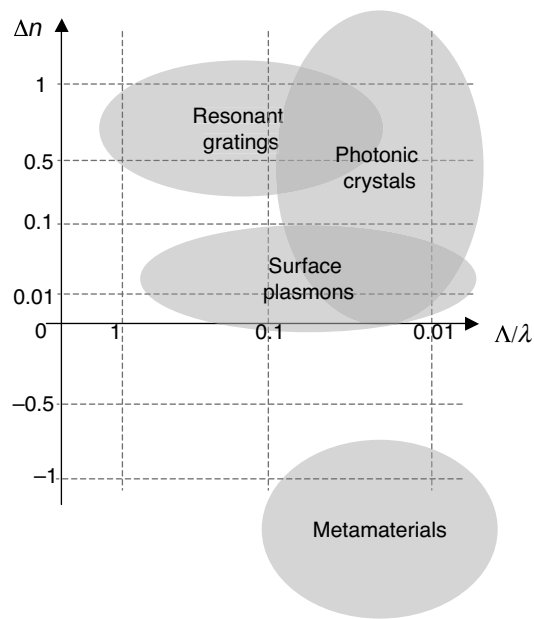


**Figure 10.39** SWL gratings, resonant waveguide gratings, PCs, SPP and metamaterials

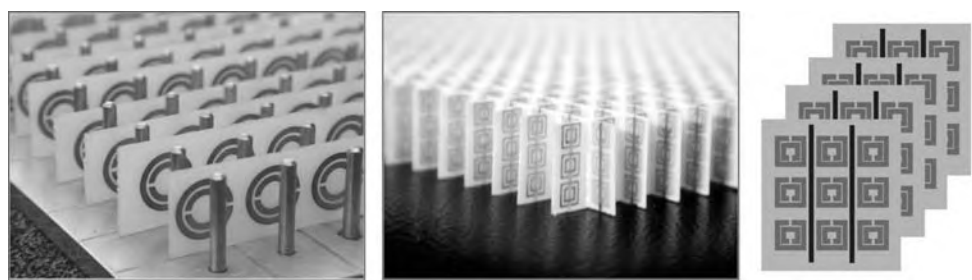


**Figure 10.40** Microwave metamaterials developed by David Smith and his colleagues at UCSD

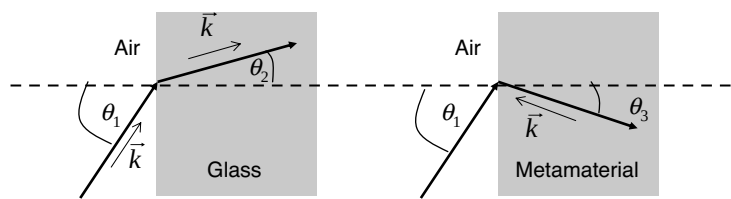


**Figure 10.41** The reversal of Snell's law by metamaterials

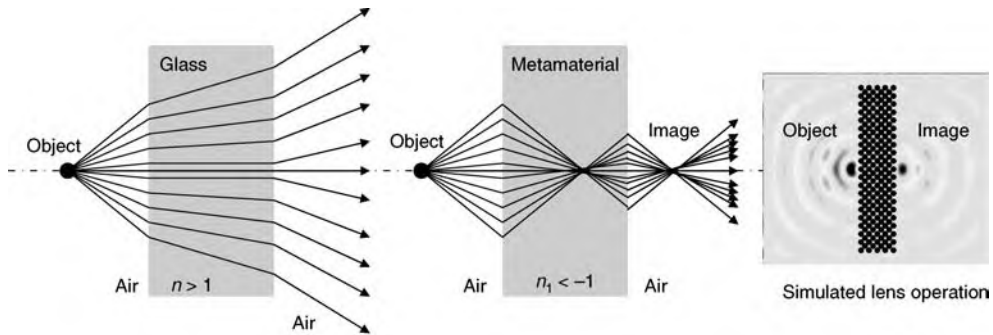


**Figure 10.42** Ray directions in left-handed and right-handed materials

fantastic applications such as Harry Potter's invisible cloak and various highly classified military applications. This is not entirely misleading, since the development of metamaterials in the optical regime has been demonstrated by several research groups to date.

### 10.9.2.1 The Perfect Lens

In optics, a negative index [23, 24] is a very desirable feature. Forty years ago, Victor Veselago showed that if such a material were to exist, it would be possible to implement a perfect lens in a planar substrate. Sir John Pendry from Imperial College demonstrated that not only would such a lens focus light through its planar structure, but it would also focus light beyond the diffraction limit. Based on the reversal of Snell's law (see Figure 10.41), one can draw the ray paths in Figure 10.42 at the interface of conventional materials and at the interface between conventional materials and (left-handed) metamaterials.

Depending on the characteristics of the metamaterial, there can be an additional image in the center of the slab. Such devices could be perfect imaging devices in microlithography for imaging a reticle pattern directly onto a wafer without having to design complex and costly optical projection systems.

### 10.9.2.2 Optical Cloaking

Metamaterials are also perfect candidates to implement cloaking functionality, at microwave and perhaps optical frequencies. Figure 10.43 shows the principle of the optical cloaking operation.

The effect used in optical cloaking is a surface plasmon created by metallic nanostructures on a dielectric substrate (see the previous section on surface plasmon polaritons).

Although the optical cloaking effect has been demonstrated recently, there is still a great deal of work to be done to develop a metamaterial that would not absorb light, and that would work over a broad range of frequencies and a broad range of incident angles, without being perturbed by temperature, pressure or tiny movements.

### 10.9.2.3 Slowing Down Light

Intriguing effects at the interface between a standard material and a metamaterial can slow down light, and could therefore potentially be used as a storage medium, especially for telecom applications.

### 10.9.2.4 Metamaterial Sensors

In addition to the previous effects, metamaterials can also be used as detectors in biological and chemical agents. The nanoscale geometries of metamaterials can be carefully shaped so that they resonate

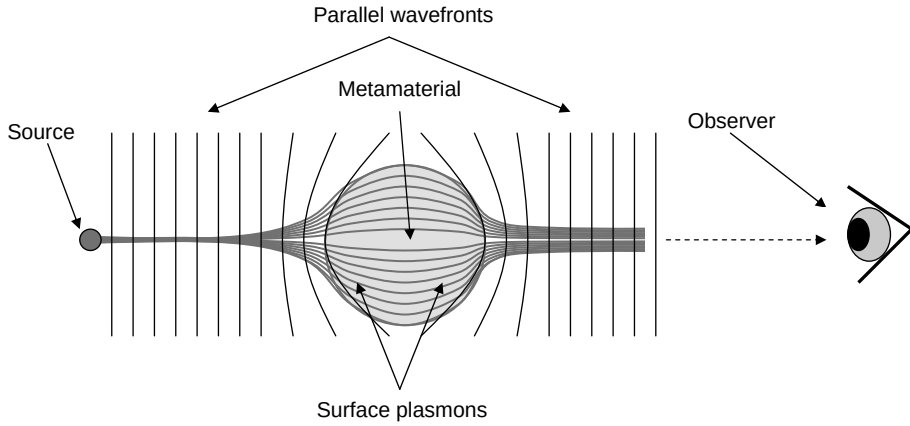


**Figure 10.43** Optical cloaking

at chosen frequencies, such as a molecular vibration frequency. An electromagnetic signature of a chemical agent could therefore be detected by metamaterials.

### 10.9.3 Issues with Metamaterials

Applications of metamaterials seem to be very promising. However, there are many burdens linked to such applications, namely:

- losses in metamaterials;
- fabrication problems;
- the operating range; and
- perturbations.

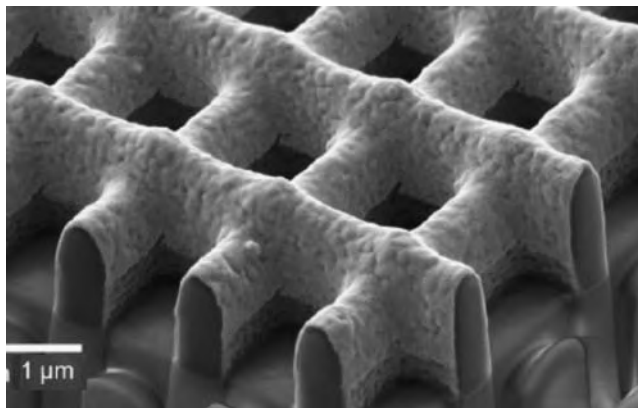


**Figure 10.44** An example of a metamaterial fabricated by direct 3D laser beam exposure (courtesy of M. Hermatschweiler)

Figure 10.44 shows a metamaterial in the optical frequency regime fabricated by direct nanoscale laser write in thick photoresist.

The first fabrication techniques have employed 3D direct laser writing (DLW) and a combination of SiO<sub>2</sub> atomic-layer deposition and silver chemical vapor deposition methods. Such approaches appear to be very promising for the rapid prototyping of truly 3D metamaterials. However, theoretical blueprints for meaningful metamaterial structures compatible with these approaches still need to be developed.

The lack of fabrication techniques and large losses within metamaterials are drawbacks that hinder their use in industrial applications today.

## References

- [1] H.P. Herzig, 'Micro-optics: Toward the Nanoscale', Institute of Microtechnology, University of Neuchâtel, Switzerland, Short Course, Optics Lab, May 2007.
- [2] R. Petit (ed.), '*Electromagnetic Theory of Gratings*', Topics in Current Physics, Vol. 22, Springer-Verlag, Heidelberg, 1980.
- [3] Ph. Lalanne, S. Astilean, P. Chavel, E. Cambril and H. Launois, 'Design and fabrication of blazed-binary diffractive elements with sampling periods smaller than the structural cutoff', *Journal of the Optical Society of America A*, **16**, 1999, 1143–1156.
- [4] E.W. Marchand and E. Wolf, 'Boundary diffraction wave in the domain of the Rayleigh–Kirchhoff diffraction theory', *Journal of the Optical Society of America*, **52**(7), 1962, 761–767.
- [5] T.K. Gaylord and M.G. Moharam, 'Analysis and applications of optical diffraction by gratings', *Proceedings of the IEEE*, **73**, 1985, 894–937.
- [6] G. Cerutti-Maori, R. Petit and M. Cadilhac, 'Etude numérique du champ diffracté par un réseau', *Comptes Rendu de l'Académie des Sciences, Paris*, **268**, 1969, 1060–1063.
- [7] L. Li and C.W. Haggans, 'Convergence of the coupled-wave method for metallic lamellar diffraction gratings', *Journal of the Optical Society of America A*, **10**, 1993, 1184–1191.
- [8] E. Noponen and J. Turunen, 'Complex amplitude modulation by high-carrier-frequency diffractive elements', *Journal of the Optical Society of America A*, **13**(7), 1996, 1422–1428.
- [9] C.C. Raman and N.S. Nagendra Nath, *Proceedings of the Indian Academy of Sciences*, **2A**, 1935, 406–413; **3A**, 1936, **119**, 495.
- [10] S.M. Rytov, *Soviet Physics – JETP*, **2**, 1956, 466.
- [11] C.W. Haggans and R.K. Kostuk, 'Effective medium theory of zeroth-order lamellar gratings in conical mounting', *Journal of the Optical Society of America A*, **10**, 1993, 2217–2225.
- [12] M.W. Farn, 'Binary gratings with increased efficiency', *Applied Optics*, **31**, 1992, 4453–4458.
- [13] H. Haidner, P. Kipfer, W. Stork and N. Streibl, 'Zero-order gratings used as an artificial distributed index medium', *Optik*, **89**, 1992, 107–112.
- [14] D.H. Raguin and G.M. Morris, 'Antireflection structured surfaces for the infrared spectral region', *Applied Optics*, **32**(7), 1993, 1154–1167.
- [15] M.E. Motamedi, W.H. Southwell and W.J. Gunning, 'Antireflection surfaces in silicon using binary optic technology', *Applied Optics*, **31**(22), 1992, 4371–4376.
- [16] S.J. Wilson and M.C. Hutley, 'The optical properties of "moth eye" antireflection surfaces', *Optica Acta*, **29**, 1982, 993–1009.
- [17] C.W. Haggans and R.K. Kostuk, 'Polarization transformation properties of high-spatial frequency surface-relief gratings and their applications', Chapter 12 in '*Micro-optics: Elements, Systems and Applications*', P. Herzig (ed.), CRC Press, Boca Raton, FL, 1997.
- [18] R. Magnusson, SPIE Short Course SC019.
- [19] E. Yablonovitch, 'Inhibited spontaneous emission in solid-state physics and electronics', *Physical Review Letters*, **58**, 1987, 2059–2062.
- [20] E. Yablonovitch, 'Photonic band-gap structures', *Journal of the Optical Society of America B*, **10**, 1993, 283–295.

- [21] D.R. Smith, J.B. Pendry and M.C.K. Wiltshire, 'Metamaterials and negative refractive index', *Science*, **305**, 2004, 788–792.
- [22] F. Zolla, S. Guenneau, A. Nicolet *et al.*, 'Electromagnetic analysis of cylindrical invisibility cloaks and the mirage effect', *Optics Letters*, **32**, 2007, 1069–1071.
- [23] D. Schurig, J.J. Mock, B.J. Justice *et al.*, 'Metamaterial electromagnetic cloak at microwave frequencies', *Science*, **314**, 2006, 977–980.
- [24] C.Y. Luo, S.G. Johnson, J.D. Joannopoulos *et al.*, 'Subwavelength imaging in photonic crystals', *Physical Review B*, **68**, 2003, 153 901–153 916.





# 11

## Digital Optics Modeling Techniques

Chapters 3–10 have described the digital optical elements used in industry today, as well as the various design techniques used today to calculate, design and optimize them.

This chapter will focus on the various numerical tools available to the optical engineer to accurately model the behavior of the digital optics designed with the techniques described in the previous chapters. Such software tools include CAD programs used to model the effects of illumination and opto-mechanical tolerancing as well as the effects of systematic fabrication errors. Chapters 12–15 describe systematic fabrication errors linked to the various fabrication techniques and technologies investigated throughout this book.

The first section will deal with ray-tracing techniques, whereas the second section will deal with more complex numeric propagators specifically adapted to the modeling of digital diffractives in the scalar diffraction regime. When the diffractive element is composed of structures the dimensions of which are in the vicinity of the wavelength, or the reconstruction window lies very close to the diffractive, more complex vector electromagnetic techniques should be used, and are described in Section 11.3.

Figure 11.1 shows the realm of validity of ray tracing, scalar and semi-scalar propagators and vector EM methods for the modeling of diffractives, as a function of the ratio between the smallest grating period in the diffractive ( $\Lambda$ ) and the reconstruction wavelength  $\lambda$ . Note that  $\Lambda/2$  is also the smallest structure in the diffractive element considered, also called the critical dimension (CD) when it comes to fabricating diffractives (see Chapters 12–15).

### 11.1 Tools Based on Ray Tracing

Since diffractives are very often used in conjunction with other optical elements (refractive optics, catadioptric optics, graded-index optics etc.), diffractive optics simulation tools have to be able to interface with standard optical modeling tools, which are mostly based on ray tracing [1, 2]. Therefore, many CAD optical tools available on the market use simple ray-tracing algorithms through diffractives, as if the diffractive was a special refractive element with no thickness. However, we will see that this approach gives results that are often very far from the real behavior of the diffractive.

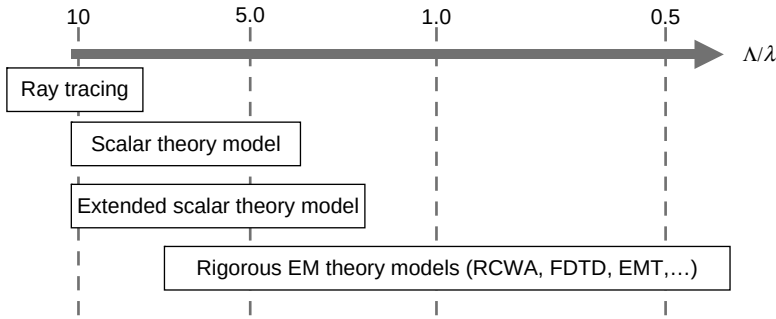


**Figure 11.1** The realm of validity of various numeric modeling techniques

### 11.1.1 The Equivalent Refractive Lens Method (ERL)

The simplest and also the first method used to model diffractives is to consider the refractive counterpart of a diffractive lens, and simply state that the equivalent refractive has zero thickness (see Figure 11.2). The phase profile is then imprinted on the incoming wavefront, as it would be for an atmospheric perturbation of infinite thickness [3, 4]. Of course, this is only possible if the diffractive has a simple refractive counterpart such as a Fresnel lens or a cylindrical lens.

### 11.1.2 The Sweatt Model

The Sweatt model is very similar to the ERL model in the way it compares the diffractive to its refractive counterpart (see Figure 11.3). However, in the Sweatt model [5, 6], the equivalent refractive lens is made of a material with a very high refractive index (e.g.  $n = 1000$ ). Thus, its thickness is reduced down to its underlying surface (planar or any other curvature), resembling a planar diffractive lens, and thus allowing ray tracing through the lens according to the Snell–Descartes law (see Chapter 1).

Here again, the diffractive lens has a simple refractive counterpart model. This method can be easily integrated within standard optical CAD tools that rely on ray tracing, without any major changes. However, it will not provide any information about diffraction efficiency, diffraction orders, quantization noise, interference and so on. Nevertheless, the Sweatt model allows access to paraxial data, and to the third- and fifth-order aberration coefficients derived for diffractive optics (see Chapter 5).

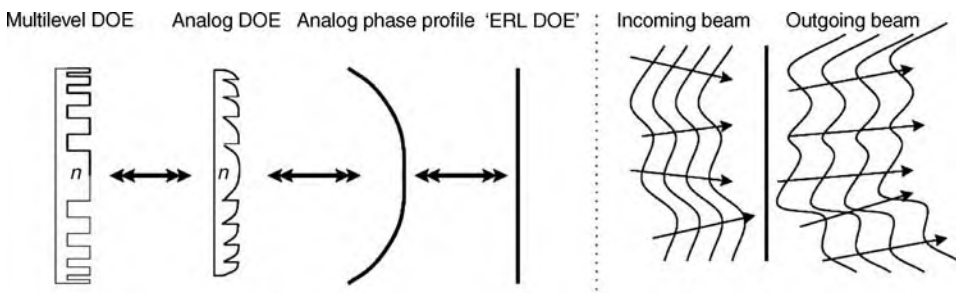


**Figure 11.2** The equivalent refractive lens model (ERL)

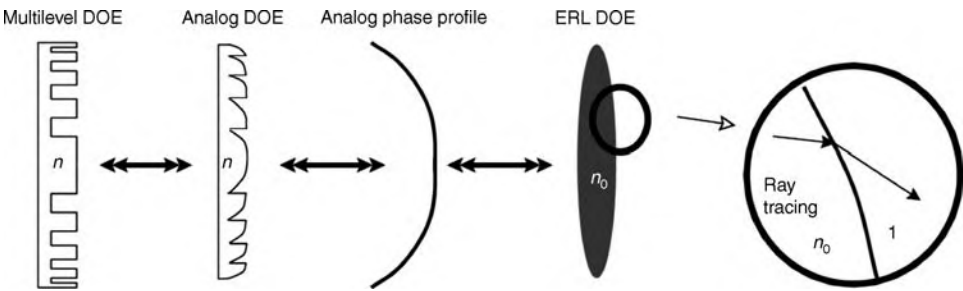


**Figure 11.3** The Sweatt model for modeling diffractives

11.1.3 The Local Grating Approximation Method (LGA)

An intermediate modeling technique between real diffraction effects (see Section 11.2) and ray tracing (the ERL and the Sweatt model) is the Local Grating Approximation (LGA) method [3, 7]. Today, most of the ray-tracing techniques in optical CAD tools use the LGA method to model a wide range of diffractive elements. However, in order to fully implement the LGA method, the diffractive has to be constituted by well-defined fringes.

When taking a closer look at a diffractive composed of well-defined fringes, one can always locally approximate the diffractive function to a linear grating, with a well-defined period and a well-defined orientation, the two specifications used to compute the diffraction angle and the diffraction direction, respectively. Such diffractive elements can include any diffractive lens (spherical, cylindrical, conical, toroidal etc.), any linear or curved grating elements (with the exception of complex 2D gratings, which do not yield fringes), and can also include most of the Fresnel-type CGHs (see Chapter 6). However, they cannot include complex Fourier-type diffractives (such as Fourier CGHs), as these latter elements do not yield fringe-like structures.

Figure 11.4 shows how the local grating approximation is implemented by using the standard grating equation (see Chapters 1 and 5) in order to predict the direction of the ray passing through a particular area of the DOE.

The diffractive lens presented in Figure 11.4 is a binary nonrotationally symmetric aspheric lens that corrects mostly astigmatism aberrations, which yield complex sets of fringes – much more complex than those of conventional Fresnel lenses – but which can nevertheless be locally approximated by a linear grating. The local angle of diffraction is given by the grating equation at the equivalent grating location,

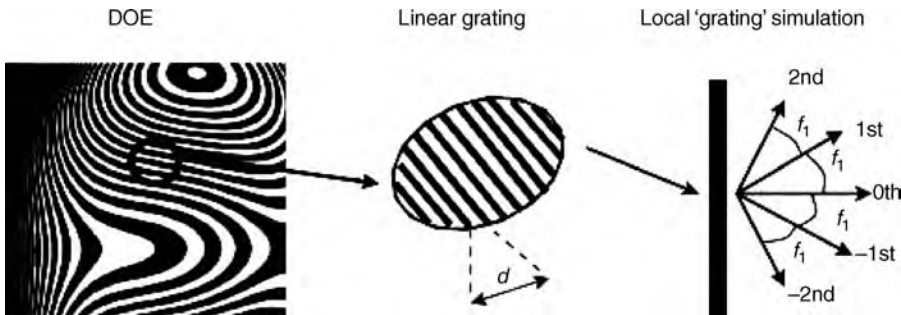


**Figure 11.4** Ray tracing by local grating approximation (LGA) of an aspherical diffractive element

and the efficiency at that location is computed as a function of the diffraction order considered, the groove depth, the wavelength and the number of phase levels.

This technique does not inform the optical designer about crucial aspects such as the real diffraction efficiency and multi-order diffraction (which is present in most cases) and the resulting multi-order interferences. Therefore, this technique can only be used effectively for 100% efficient elements, such as blazed Fresnel lenses or gratings.

The LGA modeling technique is thus best suited for the modeling of hybrid optics, where the diffractive element is usually fabricated by diamond turning on top of a refractive profile, and thus yields very smooth fringes and a high diffraction efficiency. The ray tracing can then be performed directly through the dual-profile refractive/diffractive (see also Chapter 7).

The LGA modeling method constitutes the principal tool for the vast majority of diffractive optics modeling tools available in optical CADs on the market today. Other techniques consider the diffractive element's unwrapped phase as a refractive element that has an infinite refractive index, thus becoming a planar nondiffracting element. This is very similar to the Sweatt model.

## 11.2 Scalar Diffraction Based Propagators

When the optical designer wants to have a precise view of how well (or not) a diffractive element will behave given a set of opto-mechanical constraints (and maybe also fabrication constraints), it is very often best to use numeric propagation tools based on rigorous or scalar diffraction theory (see Appendices A and B).

Rigorous numeric propagators are very seldom used today in industry, owing to the complexity of the numeric implementation and the amount of CPU time required. Therefore, most of the numeric propagators are today based on scalar diffraction theory (see Appendix B).

Such numeric propagators are very powerful in the way they predict the diffraction patterns and thus the optical functionality of the diffractives in the near and/or far fields, by taking the following points into consideration:

- the effects of multi-order diffractions and the interferences between them;
- zero-order light (stray light);
- quantization noise and noise created by cell structures (CGH type);
- effects arising from specific fabrication techniques;
- effects arising from systematic fabrication errors; and
- numeric reconstruction windows not only located in planes parallel to the elements, but in any volume before or after the diffractive.

### 11.2.1 *Scalar Diffraction Propagator Integrals*

By using scalar or semi-scalar diffraction based propagators (i.e. physical optics rather than geometrical optics), one can take into account all diffraction orders propagating through the diffractive, in a nonsequential parallel way [8, 9]. Such scalar diffraction propagators can model any type of diffractive structure, whether composed of fringes or not (diffractive lenses, gratings, DOEs, CGHs etc.), in analog, binary or multilevel surface-relief implementation. This is, however, true only in the realm of scalar diffraction theory; that is, as long as one is in the paraxial regime (low NAs, low angles and smallest structures much larger than the wavelength, without any polarization effects).

Helmholtz's wave equation, with Huygen's principle of secondary sources over the wavefront's envelope, substituted in Green's function, is the major foundation of scalar theory (see Appendix B), and gives rise to the Helmholtz–Kirchhoff integral theorem. The Rayleigh–Sommerfeld diffraction formulation for monochromatic waves follows, and gives rise to the Fresnel and then the Fourier approximations of the diffraction through a thin planar screen in the far and near fields. These two formulations are the basis of most of the physical optics modeling tools used today.

The formulations for Fourier and Fresnel transformation can be summarized as follows:

$$\begin{cases} U(u, v) = \iint_{\infty} U(x, y) \cdot e^{-j2\pi(u \cdot x + v \cdot y)} \cdot dx \cdot dy & \text{(a)} \\ U(x', y') = \left( \frac{e^{jkd}}{j\lambda d} \right) \iint_{\infty} U(x, y) \cdot e^{j\frac{\pi}{\lambda d}((x' - x)^2 + (y' - y)^2)} \cdot dx \cdot dy & \text{(b)} \end{cases} \quad (11.1)$$

where  $(x, y)$  describes the space in the diffractive element plane,  $(u, v)$  the angular spectrum in the far field and  $(x', y')$  the space in the near field at distance  $d$ .

For example, when modeling a diffractive Fresnel lens of focal length  $f$ , it is best to reconstruct the focal plane of that lens, and consider the size and shape of the focal spot (e.g. by using the Strehl ratio method). The following equation expresses the complex amplitude at the focal plane of such a lens:

$$U(x', y') = \left( \frac{e^{jkf}}{j\lambda f} \right) \iint_{\infty} U(x', y') \cdot e^{j\frac{\pi}{\lambda f}((x - x')^2 + (y - y')^2)} \cdot dx \cdot dy \quad (11.2)$$

Note that the Fresnel approximation (Equation (11.1b)) can be described in two different ways, as a direct integral

$$U(x', y') = \left( \frac{e^{jkd}}{j\lambda d} \right) \cdot e^{j\frac{\pi}{\lambda d}(x'^2 + y'^2)} \cdot \iint_{\infty} U(x, y) \cdot e^{j\frac{\pi}{\lambda d}(x^2 + y^2)} \cdot e^{-j\frac{2\pi}{\lambda d}(x'x + y'y)} \cdot dx \cdot dy \quad (11.3)$$

or as a convolution

$$U(x', y') = \left( \frac{e^{jkd}}{j\lambda d} \right) \cdot \iint_{\infty} U(x, y) \cdot e^{j\frac{\pi}{\lambda d}((x' - x)^2 + (y' - y)^2)} \cdot dx \cdot dy \quad (11.4)$$

Although the direct Fresnel integral (Equation (11.3)) and the convolution-based expression of the Fresnel integral (Equation (11.4)) are similar, they differ in the way in which they are implemented numerically, and they yield different near-field reconstruction window sizes.

In effect, while the Fourier or Fresnel transform integrals describe a field of infinite dimension (both in the angular spectrum and in near-field space), the actual numeric implementation of these integrals yields finite reconstruction windows. The optical designer has thus to fully understand the implications of using the direct or convolution-based Fresnel transform when modeling its elements, as we will see in the next section.

### 11.2.2 Numeric Implementation of Scalar Propagators

This section describes how the previous analytic propagator integrals can be implemented numerically in a computer. Although there are many ways to implement the Fourier and Fresnel integrals in numeric algorithms, the most popular implementation techniques wisely use various 2D Fast Fourier Transform (FFT) algorithms (see Appendix C). These algorithms are very effective in terms of speed, and are used in numerous sectors of industry today (physics, biotechnology, microelectronics, economics etc.).

Conventional FFT-based numeric propagators, although being very fast, in many cases suffer from severe drawbacks, which can frustrate the optical designer – or, on the contrary, in other cases, predict very good results for an element which, once fabricated, would yield very poor performance, far removed from that predicted by an FFT-based propagator (a typical example is a sampled Fresnel lens fabricated with square pixels – see Chapter 12). Some of these limitations of FFT-based propagators are linked to the size

of the sampled field or the location of the reconstruction windows, as well as the amount of off-axis of these reconstruction windows.

In order to overcome these constraints, we will see how Discrete Fourier Transform (DFT) algorithms can be used in order to implement the same Fresnel and Fourier approximations of the Rayleigh–Sommerfeld integral (Equations (11.3) and (11.4)). It is worth noting that the exact Rayleigh–Sommerfeld integral can actually be implemented by using DFT-based algorithms, which is not the case with FFT algorithms.

### 11.2.2.1 FFT-based Numeric Scalar Propagators

This section describes the various numeric propagators that have been implemented in the literature through FFT algorithms (complex 2D FFTs).

#### *Fraunhofer Propagation through the FFT Algorithm (Far Field)*

We have seen that the Fraunhofer diffraction formulation can be expressed as a Fourier transform (Appendix B and Equation (11.1a)). Therefore, the far-field diffraction pattern  $U'$  (or the angular spectrum of plane waves) can be expressed in its sampled form as the two-dimensional Complex Fourier Transform (CFT) of the sampled complex amplitude function  $U$  as described at the diffractive element plane:

$$U'(u_n, v_m) = FFT[U(x_n, y_m)] \quad \text{where} \quad \begin{cases} U(x_n, y_m) = A(x_n, y_m) \cdot e^{-j\varphi(x_n, y_m)} \\ U'(u_n, v_m) = A'(u_n, v_m) \cdot e^{-j\varphi'(u_n, v_m)} \end{cases} \quad (11.5)$$

with 
$$\begin{cases} x_n = c_x \left( n - \frac{N}{2} \right) & \forall (-N/2 < n < N/2) \\ y_m = c_y \left( m - \frac{M}{2} \right) & \forall (-M/2 < m < M/2) \end{cases}$$

where  $c_x$  and  $c_y$  are, respectively, the sampling distances in the plane of the diffractive element. The incoming complex amplitude  $U$  is the complex diffractive element plane (usually phase only) modulated by the complex amplitude of the incoming wavefront. For example, the incoming complex amplitude for a generic diffractive illuminated by an on-axis Gaussian diverging astigmatic beam can be described as follows:

$$\begin{cases} A(u_n, v_m) = D_{n,m} \cdot A_0 e^{-2 \left( \frac{x_n^2}{w_{ox}^2} + \frac{y_m^2}{w_{oy}^2} \right)} \\ \varphi(u_n, v_m) = \Phi_{n,m} + e^{-j \frac{\pi}{\lambda} \left( \frac{x_n^2}{f_x} + \frac{y_m^2}{f_y} \right)} \end{cases} \quad (11.6)$$

where  $A_0$  is the amplitude of the Gaussian TEM<sub>00</sub> beam on the optical axis;  $f_x$  and  $f_y$  are, respectively, the focal lengths of a lens describing the divergence of the beam hitting the diffractive (note that the wavefront is astigmatic if  $f_x \neq f_y$ ); and, finally,  $D_{n,m}$  and  $\Phi_{n,m}$  are, respectively, the sampled amplitude and the sampled phase of the diffractive element itself (usually,  $D_{n,m} = 1.0$ ). In Equation (11.6),  $w_{ox}$  and  $w_{oy}$  are, respectively, the beam waists of the incoming laser beam in the  $x$  and  $y$  directions

These sampling rates have to be chosen carefully, since they have to satisfy the Nyquist criteria as described in Appendix C and in the next section. For example, if the diffractive is a CGH sampled by  $N$  cells in the  $x$  direction and  $M$  cells in the  $y$  direction, and is illuminated by a uniform plane wave, the appropriate sampling distance would be the size of the CGH cells in the  $x$  and  $y$  directions (or smaller).

### ***Fresnel Propagation through the FFT Algorithm (Near Field)***

We have seen that the Fresnel diffraction formulation (see Equation (11.1b)) can be defined either as the FT of the incoming complex amplitude multiplied by a quadratic phase, or the convolution of the complex incoming amplitude by a quadratic phase (see Equations (11.2) and (11.3)). We can therefore express these two transforms using the FFT algorithm as follows:

$$\begin{cases} U'(x'_{1n}, y'_{1m}) = e^{\frac{j\pi}{\lambda d}(x_{1n}^2 + y_{1m}^2)} \cdot \text{FFT} \left[ U(x_n, y_m) \cdot e^{\frac{j\pi}{\lambda d}(x_n^2 + y_n^2)} \right] & \text{(a)} \\ U'(x'_{2n}, y'_{2m}) = \text{FFT}^{-1} \left[ \text{FFT}[U(x_n, y_m)] \cdot \text{FFT} \left[ \frac{j\pi}{\lambda d}(x_n^2 + y_n^2) \right] \right] & \text{(b)} \end{cases} \quad (11.7)$$

Here, one can apply the same complex diffractive plane decomposition as has been done for the Fraunhofer far-field propagator in Equation (11.5). However, here we have two more variables, which are the sampling rates in the near field, namely  $c_x$  and  $c_y$ . We will see how to derive these sampling rates in the next section.

### **11.2.2.2 DFT-based Numeric Scalar Propagators**

Here, let us derive similar expressions for the Fraunhofer and Fresnel diffraction patterns by using the DFT computation instead of an FFT algorithm. As seen in Appendix C, the main disadvantage of DFT-based propagators is the CPU time required to compute the numeric reconstruction (the CPU time scales as  $N^2$  for DFT-based propagators and as  $N \log(N)$  for FFT-based propagators). However, the DFT-based propagators have numerous other advantages, which make them suitable choices for most of the applications that require diffractives within the scalar regime.

The main advantages of using a DFT rather than a faster FFT algorithm are as follows:

- arbitrary location of the reconstruction window in the near  $(x, y, z)$  or far field  $(u, v)$  (however, it has to be within the realm of scalar diffraction – that is, not too close to the initial window);
- arbitrary resolution of the reconstruction window; and
- the fact that the number of samples (pixels) in the reconstruction window is not limited to powers of 2.

### ***Fraunhofer Propagation through a DFT (Far Field)***

The implementation of a Fraunhofer propagator, using a simple DFT calculation process, is as follows:

$$U'(u_k, v_l) = \text{DFT}[U(x_n, y_m)] = \sum_{n=-N/2}^{N/2} \sum_{m=-M/2}^{M/2} U(x_n, y_m) \cdot e^{-j2\pi(x_n \cdot (u_n - u_0) + y_n \cdot (v_m - v_0))}$$

with  $\begin{cases} x_n = c_x \left( n - \frac{N}{2} \right) \\ y_n = c_y \left( m - \frac{M}{2} \right) \end{cases}$  and  $\begin{cases} u_k = c_u \left( k - \frac{K}{2} \right) - u_0 \\ v_l = c_v \left( l - \frac{L}{2} \right) - v_0 \end{cases}$  (11.8)

where  $\begin{cases} -N/2 < n < N/2 \\ -M/2 < m < M/2 \end{cases}$  and  $\begin{cases} -K/2 < k < K/2 \\ -L/2 < l < L/2 \end{cases}$

Here,  $N$  and  $M$  can be chosen arbitrarily, as can  $c_u$  and  $c_v$ , and most of all one can define the angular spectrum window off-axis  $u_0$  and  $v_0$  to take arbitrary values (within the realm of scalar theory). In a FFT-based Fraunhofer propagator (see the previous section),  $c_{u0}$  and  $c_{v0}$  were both null,  $c_u$  and  $c_v$  were set in stone, and  $N$  and  $M$  were powers of 2.

### Fresnel Propagation through a DFT (Near Field)

Due to the flexibility of DFT-based propagators (other than the long CPU time required to perform the task), there is no special need to consider a convolution-based DFT propagator (since there would be three FTs to compute (see Equation (11.7b)) – which is fine for a fast FFT-based implementation, but which would be too CPU intensive for DFT implementation). Anyway, in the case of the FFT implementation, we have proposed the use of the convolution-based propagator only for scaling considerations (see also the next section). These scaling limitations no longer hold for DFT implementation. Therefore, we will only derive the simple FT-based Fresnel propagator for DFT implementation:

$$\begin{aligned}
 U'(x_k, x_l) &= e^{j\lambda d} \frac{j\pi}{(x_k^2 + y_l^2)} \cdot \text{DFT} \left[ U(x_n, y_m) \cdot e^{j\lambda f} \frac{j\pi}{(x_n^2 + y_n^2)} \right] \\
 &= e^{j\lambda d} \frac{j\pi}{(x_k^2 + y_l^2)} \sum_{n=-N/2}^{N/2} \sum_{m=-M/2}^{M/2} U(x_n, y_m) \cdot e^{j\lambda f} \frac{j\pi}{(x_n^2 + y_n^2)} \cdot e^{-j2\pi(x_n \cdot (u_n - u_0) + y_n \cdot (v_m - v_0))} \\
 &\text{with } \begin{cases} x_n = c_x \left( n - \frac{N}{2} \right) \\ y_n = c_y \left( m - \frac{M}{2} \right) \end{cases} \quad \text{and} \quad \begin{cases} x'_k = c'_x \left( k - \frac{K}{2} \right) - x_0 \\ x'_l = c'_y \left( l - \frac{L}{2} \right) - y_0 \end{cases}
 \end{aligned} \tag{11.9}$$

Note here that the reconstruction cell sizes  $c'_x$  and  $c'_y$  can be chosen quasi-arbitrarily. For the limitations linked to the DFT sampling of reconstruction windows, see Section 11.2.2.4.

### 11.2.2.3 Sampling Considerations with DFT and FFT Algorithms

We derive in this section the various sampling conditions associated with Gaussian beam propagation, FFT-based propagators and DFT-based propagators, for near- and far-field numeric reconstruction windows.

#### Sampling of Gaussian Beams

Before considering the sampling issues for diffractives and other micro-optical elements (see the next section, and see also Chapters 5 and 6), we will consider here the simple sampling process for a Gaussian laser beam. Although it seems to be a straightforward process, there are a few things to remember when attempting to sample a coherent complex wavefront.

First, the Nyquist criterion has to be satisfied: the sampling rate has to be at least twice the highest spatial frequency of the laser field. The complex amplitude of a circular Gaussian laser beam is given by

$$U(r, z) = A_0(z) \cdot e^{-\frac{r^2}{w(z)^2}} \cdot e^{-j\varphi_x(r)} \cdot e^{-j\varphi_y(r)} \tag{11.10}$$

To accurately sample  $U(r, z)$  in the  $x$  direction, we have to determine the highest-frequency components of  $U(r, z)$ , which is not trivial since the profile is a smooth Gaussian profile. As a Gaussian beam has an infinite extent (i.e. the intensity never reaches zero), we have to define a point beyond which the intensity can be considered as null. We will therefore consider that the Gaussian beam vanishes for  $r > \alpha\omega$ , where  $\alpha > 0$  ( $\omega$  being the waist of the laser beam). It is commonly agreed that a decent value for  $\alpha$  is  $\alpha = 2.3$ . This criterion ensures that the aperture will transmit 99% of the energy and that diffraction ripple oscillations will have less than 1% amplitude. If  $w$  is the Gaussian beam waist, the beam is considered to vanish when  $|r| \geq \sqrt{2}\alpha w$ .

When considering the FT pair described in the following equation:

$$U(x) = e^{-\pi \frac{x^2}{w^2}} \mapsto U'(v) = \frac{1}{w} e^{-\pi v^2 w^2} \tag{11.11}$$

we can similarly assert that the FT  $U'(v)$  should vanish for any frequency  $v$  greater than  $|r| \geq \sqrt{2}\alpha/(\pi w)$ .



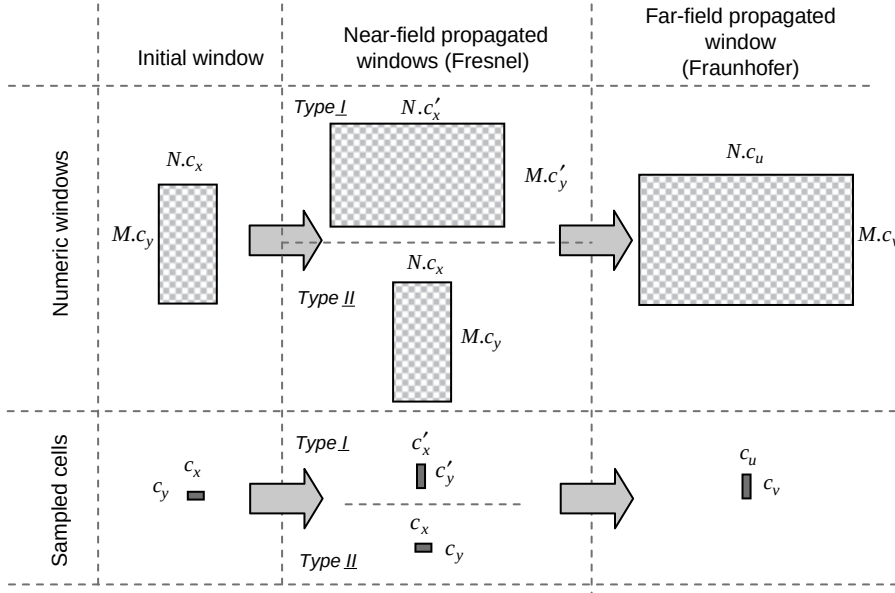


**Figure 11.5** Near- and far-field FFT-based propagation: scaling limitations

We have therefore defined the highest frequency in the Gaussian beam,  $v_{\max} = \sqrt{2}\alpha/(\pi w)$ , and we can therefore accordingly apply the Nyquist criterion to safely sample the Gaussian laser beam, by sampling the laser beam at a rate higher or equal to  $2\sqrt{2}\alpha/(\pi w)$ . Note that if the laser beam has an asymmetric profile, the sampling will be different in  $x$  and  $y$ , according to the beam waists in both directions.

### Field Sampling with FFT-based Propagators

When using an FFT propagator to propagate the complex field into the near or far field, automatic scaling of the reconstruction window occurs. Figure 11.5 shows the various scaling that occurs when one propagates a complex field into the near field (a choice of two different propagators) and into the far field via FFT algorithms.

#### Fraunhofer FFT propagator: Fourier window scaling

When using a FFT-based algorithm to reconstruct the angular spectrum of a complex amplitude defined in a square aperture (e.g. a Fourier-type diffractive element), the Fourier reconstruction window size is scaled so that it includes the largest frequency present in the original element:

$$\begin{cases} U(x_n, y_m) = A(x_n, y_m) \cdot e^{\varphi(x_n, y_m)} \\ U'(u_n, v_m) = \text{FFT}[U(x_n, y_m)] = A'(u_n, v_m) \cdot e^{\varphi'(u_n, v_m)} \end{cases}$$

where  $\begin{cases} x_n = c_x \cdot \left(n - \frac{N}{2}\right), & y_m = c_y \cdot \left(m - \frac{M}{2}\right) \\ u_n = c_u \cdot \left(n - \frac{N}{2}\right), & v_m = c_v \cdot \left(m - \frac{M}{2}\right) \end{cases}$  and  $\begin{cases} c_u = \frac{2}{N} \arcsin\left(\frac{\lambda}{2c_x}\right) \\ c_v = \frac{2}{M} \arcsin\left(\frac{\lambda}{2c_y}\right) \end{cases}$  (11.12)

In the above equation,  $U'(u, v)$  is the angular spectrum of the incoming complex amplitude wavefront  $U(x, y)$ , and has the same number of pixels as the original sampled window (note that when using standard FFT algorithms, the number of samples in the  $x$  and  $y$  directions has to be a power of 2, but the numbers do not need to be equal).

When computing the adequate sampling rate  $N$  for the  $x$  direction and  $M$  for the  $y$  direction (keeping in mind that these numbers have to be powers of 2 for FFT-based propagators), one can derive the basic cell sizes  $c_x$  and  $c_y$ . For a simple CGH simulation process (see also Chapter 6), these cell sizes can actually be the physical cell sizes of the CGH itself (rectangular or square basic cells). The resulting pixel interspacings in the spectrum plane in both directions are  $c_u$  and  $c_v$  in Equation (11.12). Note that the factor 2 accounts for the fact that the maximum diffraction angle can be either positive or negative.

#### *Fresnel FFT-based Propagators: Fresnel Windows Scaling*

As described in Equation (11.7), a Fresnel propagator can be built through FFT algorithms in two different ways: the direct method and the convolution-based method [10]. As seen in Equation (11.7), the convolution method requires three FFTs to be computed and the direct method only one FFT. The difference lies in the way in which the near-field plane (Fresnel reconstruction window) is scaled when using either of these FFT-based propagators [11]. Once an adequate sampling rate is chosen (the Nyquist criterion), one can derive (similarly to the Fraunhofer propagator) the cell sizes  $c_x$  and  $c_y$  in the initial complex amplitude window ( $N$  and  $M$  also have to be powers of 2 here). The two different near-field reconstruction window scaling at a distance  $d$  from the initial window are derived as follows:

$$\begin{cases} U(x_n, y_m) = A(x_n, y_m) \cdot e^{\varphi(x_n, y_m)} \\ U_1(x'_{1n}, y'_{1m}) = A'_1(x'_{1n}, y'_{1m}) \cdot e^{\varphi'_1(x'_{1n}, y'_{1m})} = \text{Quad}_2(x'_{1n}, y'_{1m}) \cdot \text{FFT}[\text{Quad}_1(x_n, y_m) \cdot \text{FFT}[U(x_n, y_m)]] \end{cases}$$

where  $\begin{cases} x_n = c_x \cdot \left(n - \frac{N}{2}\right), & y_m = c_y \cdot \left(m - \frac{M}{2}\right) \\ x'_{1n} = c'_{1x} \cdot \left(n - \frac{N}{2}\right), & y'_{1m} = c'_{1y} \cdot \left(m - \frac{M}{2}\right) \end{cases}$  and  $\begin{cases} c'_{1x} = \frac{\lambda d}{c_x} \\ c'_{1y} = \frac{\lambda d}{c_y} \end{cases}$  (11.13a)

$$\begin{cases} U(x_n, y_m) = A(x_n, y_m) \cdot e^{\varphi(x_n, y_m)} \\ U_2(x'_{2n}, y'_{2m}) = A'_2(x'_{2n}, y'_{2m}) \cdot e^{\varphi'_2(x'_{2n}, y'_{2m})} = \text{FFT}^{-1}[\text{FFT}[\text{Quad}_1(x_n, y_m)] \cdot \text{FFT}[U(x_n, y_m)]] \end{cases}$$

where  $\begin{cases} x_n = c_x \cdot \left(n - \frac{N}{2}\right), & y_m = c_y \cdot \left(m - \frac{M}{2}\right) \\ x'_{2n} = c'_{2x} \cdot \left(n - \frac{N}{2}\right), & y'_{2m} = c'_{2y} \cdot \left(m - \frac{M}{2}\right) \end{cases}$  and  $\begin{cases} c'_{2x} = c_x \\ c'_{2y} = c_y \end{cases}$  (11.13b)

While the direct Fresnel propagator window size is proportional to the reconstruction distance as well as to the wavelength, the convolution-based propagator window remains unchanged with regard to the original window dimension. It is worth noting that the direct Fresnel propagator window is inversely proportional to the initial cell size (inter-pixel spacing).

Figure 11.6 summarizes the scaling process for sampled reconstruction windows from FFT-based numeric propagators in the far and near fields.

However, for many practical cases, these fixed window scalings can be a strong limitation when it comes to reconstructing a particular area of interest in the near- or far-field window. In many cases, DFT-based propagators (in the near and far field) give much more flexibility in defining reconstruction windows that are adapted to the target application.

#### *Near-field Window Size and Resolution Modulation*

We have seen in the previous section that Fresnel propagators based on FFT (the most used propagators, since they are the fastest) suffer from severe drawbacks in terms of scaling and resolution of the reconstruction window size in the near field (see, e.g., Figure 11.6).

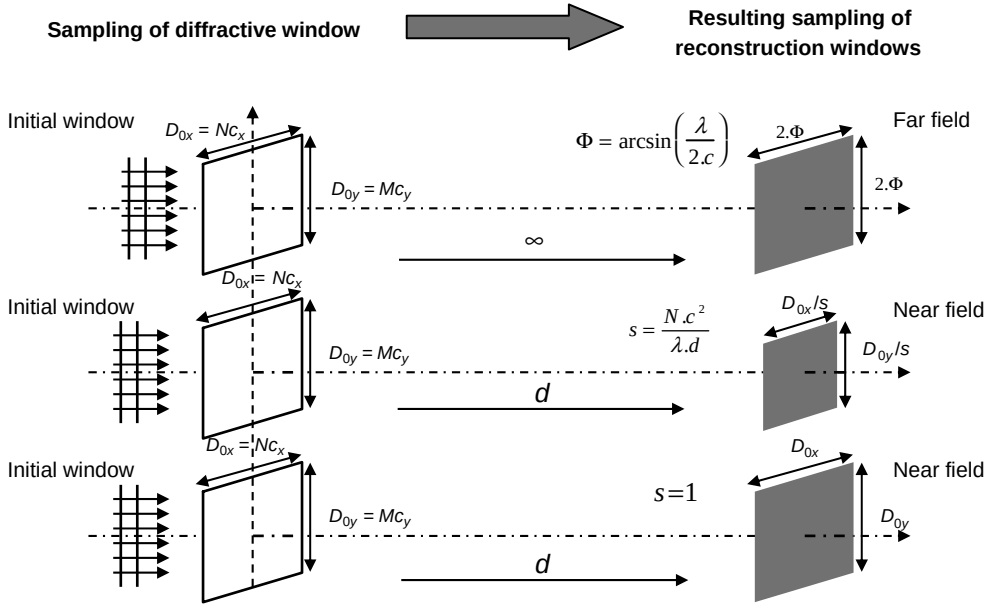


**Figure 11.6** The scaling of FFT-based reconstruction windows in the near and far fields

In order to get around these drawbacks and be able to define a desired reconstruction window size and a desired reconstruction resolution, best suited for a specific application, one can choose between three methods:

- the use of intermediate reconstruction windows;
- modulating the number of cells in the initial plane; or
- modulating the size of the cells in the initial plane.

The second and third methods are called the ‘embedding method’ and the ‘oversampling method’, respectively, and are described in the next section.

The first technique consists of propagating the field into an intermediate window location, and then propagating the window onto the final destination. The size of the final window is therefore a function of the location of the intermediate window in regard to the initial propagation plane (i.e. the diffractive optical element plane). Figure 11.7 depicts this technique.

In the first example in Figure 11.7, the reconstruction pixel interspacing is set by

$$C_f = \frac{\lambda d}{NC_0} \quad (11.14)$$

If we use an intermediate plane (the second example in Figure 11.7), located after or before the final plane, the pixel interspacing in this plane is dictated by the distance from this plane to the initial plane, which is

$$C_i = \frac{\lambda D_{01}}{NC_0} \quad (11.15)$$

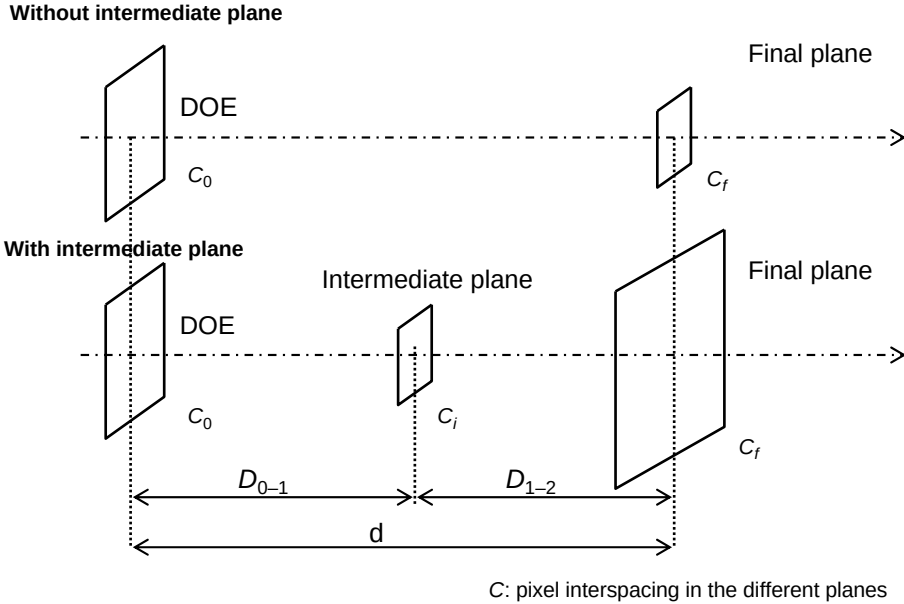


**Figure 11.7** The intermediate plane method for near-field propagation

When propagating to the final plane from this plane, the pixel interspacing in the final plane will then become

$$C_f = \frac{\lambda D_{12}}{NC_i} = \frac{\lambda D_{12} NC_0}{N \lambda D_{01}} = \frac{D_{12}}{D_{01}} C_0 \quad (11.16)$$

where

$$D_{01} + D_{12} = d \quad (11.17)$$

Thus, by varying carefully the position of the intermediate plane, one can increase or decrease the size of the final reconstruction window, without changing any parameters in the initial plane or the initial plane sampling.

Now, this seems to be a very attractive technique, but as usual, it has its limitations and constraints, namely that one cannot place the intermediate plane anywhere, since it has to incorporate all (or most) of the complex information about the diffracted field. If this plane gets too small, there is not enough room to incorporate all this information. Also, if this plane gets too large, the information is diluted over the small amount of pixels in this large plane, and will therefore also produce loss of diffracted field information.

If  $D_{01} = D_{12}$ , the reconstruction size has the same size as the CGH, and is therefore similar to the convolution-based single Fresnel propagator (it retains the initial window size). Note that such an intermediate window can be located before or after the final reconstruction window, since a Fresnel transform can proceed as a forward transform or a backward propagator.

### **Field Sampling with DFT-based Propagators**

In Equations (11.8) and (11.9) of Section 11.2.2.3, we have seen that the reconstruction window of a DFT-based propagator (near- or far-field propagators), can be sampled in a quasi-arbitrary way. Here, we will derive here the limitations of such sampling.

For DFT-based Fraunhofer reconstructions (see Equation (11.8)), the spectral extent of the reconstruction window is  $D_u$  by  $D_v$ , and it can be set off-axis by  $u_0$  and  $v_0$  spectral shifts. The reconstruction cell sizes are thus defined as  $c_u = D_u/(K - 1)$  and  $c_v = D_v/(L - 1)$ . In order to satisfy the Nyquist criterion, the largest extent of the Fraunhofer plane should not cross over the limit set by the initial window sampling  $c_x$  and  $c_y$ :

$$\begin{cases} c_u \cdot K \leq 2(u_{\max} - u_0) \\ c_v \cdot L \leq 2(v_{\max} - v_0) \end{cases} \quad \text{where} \quad \begin{cases} u_{\max} = \pm \arcsin\left(\frac{\lambda}{2c_x}\right) \\ v_{\max} = \pm \arcsin\left(\frac{\lambda}{2c_y}\right) \end{cases} \quad (11.18)$$

As an optical example, take a Fourier CGH that is sampled according to its physical cell size. The maximum reconstruction window (the maximum angle in which this CGH can diffract) is  $u_{\max}$  in the  $x$  direction and  $v_{\max}$  in the  $y$  direction (as described in Equation (11.18)). This spectral window defines the fundamental order reconstruction window, in which both fundamental orders can appear as well as the zero order. One can enlarge this basic reconstruction window and see the higher orders by oversampling the CGH plane and therefore relaxing the conditions described in Equation (11.18), and therefore see the higher orders appear (see also Section 11.2.3.2).

Similarly in the near field, the following equation defines the near-field reconstruction window sampling conditions for DFT-based Fresnel propagation (a simple FT-based propagator):

$$\begin{cases} c'_x \cdot K \leq 2(x'_{\max} - x_0) \\ c'_y \cdot L \leq 2(y'_{\max} - y_0) \end{cases} \quad \text{where} \quad \begin{cases} x'_{\max} = \pm d \cdot \tan\left(\arcsin\left(\frac{\lambda}{2c_x}\right)\right) \\ y'_{\max} = \pm d \cdot \tan\left(\arcsin\left(\frac{\lambda}{2c_y}\right)\right) \end{cases} \quad (11.19)$$

Note that there is no real limitation on the number of cells ( $K, L$ ) or the cell sizes ( $c_u, c_v$  or  $c_x, c_y$ ) individually: the condition on DFT-based reconstruction windows (in the near or far field) is only set in terms of the global size and position of these target windows.

Figure 11.8 summarizes the sampling process for planar 2D windows from near- or far-field DFT-based propagators.

Note that in the case of near-field DFT-based propagators, the reconstruction window does not need to be parallel to the initial window, but can be orthogonal to this window, or in any other orientation (see Figure 11.9).

#### 11.2.2.4 A Numeric Propagator for Real-world Optics

In previous sections, we have described a series of numeric propagators that the optical design engineer can use to compute reconstruction windows both in the near and the far field, and the limitations thereof. These are mainly windows that are two-dimensional reconstruction planes parallel or not parallel to the initial window (usually the position of the optical element).

We will now present a general extrapolation of these propagators (especially the DFT-based ones for near field reconstructions) in order to model real-world optics. By real-world optics, we mean optics that are not necessarily simple planar diffractive elements or lenses, but that can be more complex hybrid refractive/diffractive elements on arbitrary (curved) 2D surfaces or in arbitrary 3D volumes, for applications where we are interested in intensity and/or phase distributions over arbitrary 2D surfaces or even in arbitrary 3D volumes.

In order to implement such a generic numeric propagator, we will refer again to the Rayleigh–Sommerfeld integral (as it has been derived in Appendix B). The Rayleigh–Sommerfeld diffraction

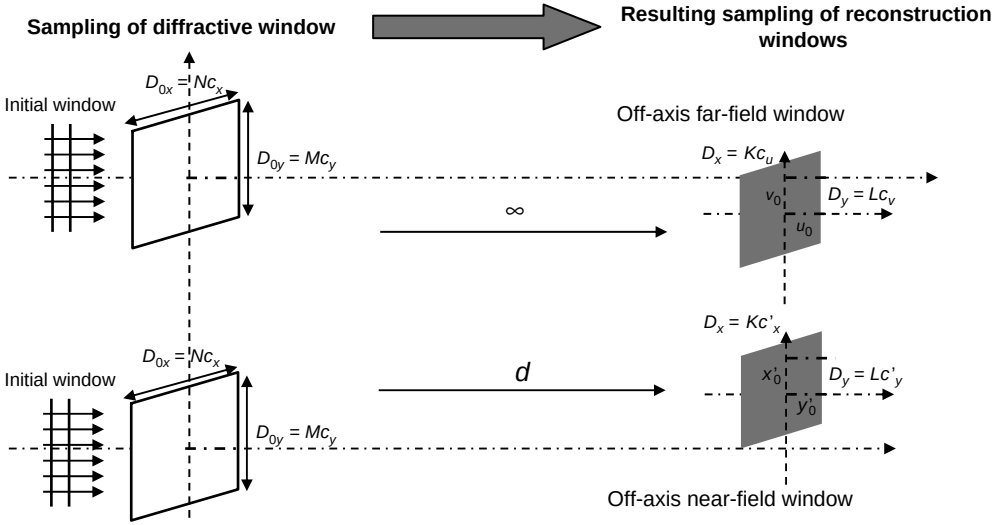


**Figure 11.8** The sampling and orientation of DFT-based reconstruction windows in the near and far fields

integral can be rewritten as follows, using the notation depicted in Figure 11.10:

$$U'(x'_1, y'_1, z'_1) = \frac{-j}{2} \iint_{S_0} \frac{e^{jk r_{01}}}{|\vec{r}_{01}|} \cos(\vec{n}, \vec{r}_{01}) ds \quad (11.20)$$

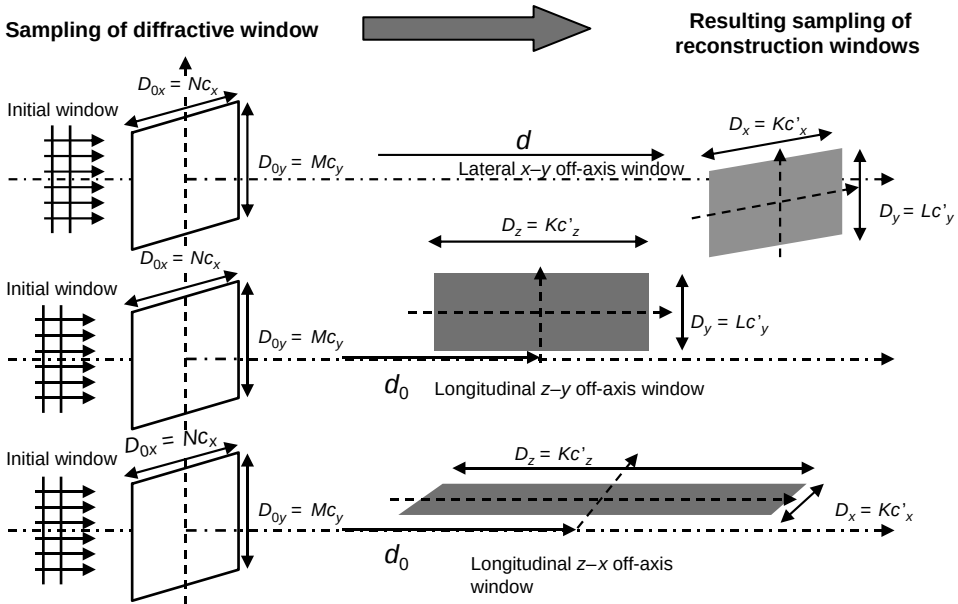


**Figure 11.9** Various potential orientations of near-field DFT-based windows

The vector  $|\vec{r}_{01}|$  is pointing from a location  $P_0(x_0, y_0, z_0)$  on the initial surface to a point  $P_1(x_1, y_1, z_1)$  on the reconstruction surface where the complex amplitude has to be computed. If the reconstruction plane lies far away from the initial window, the vector  $|\vec{r}_{01}|$  can be rewritten as

$$|\vec{r}_{01}| = \sqrt{(z'_1 - z_0)^2 + (x'_1 - x_0)^2 + (y'_1 - y_0)^2} \cong z_1 \cdot \left( 1 + \frac{1}{2} \left( \frac{x'_1 - x_0}{z_1} \right)^2 + \frac{1}{2} \left( \frac{y'_1 - y_0}{z_1} \right)^2 \right) \quad (11.21)$$

Note that we have in Equation (11.20) the obliquity factor (cosine expression) that was omitted in the previous propagators, which we have implemented in the previous sections. However, although Equation (11.16) takes the obliquity factor into account (and thus allows reconstructions closer to the initial aperture, or higher off-axis angles), we still remain in the realm of validity of scalar theory.

The obliquity factor can be rewritten as

$$\cos(\vec{n}, \vec{r}_{01}) = \frac{\vec{r}_{01} \cdot \vec{n}}{|\vec{n}| \cdot |\vec{r}_{01}|} \approx \frac{1}{\sqrt{1 + \frac{x_r^2 + y_r^2}{z_r^2}}} \quad \text{where} \quad \begin{cases} x_r = x'_k - x_n \\ y_r = y'_l - y_m \\ z_r = z'_{k,l} - z_{n,m} \end{cases} \quad (11.22)$$

By substituting Equation (11.22) into Equation (11.20), we can derive the general numeric propagator for near-field scalar reconstruction windows as follows:

$$U'(x'_k, y'_l, z'_{k,l}) = \sum_{n=0}^N \sum_{m=0}^M \left( \frac{z_r}{x_r^2 + y_r^2 + z_r^2} \right) U(x_n, y_m, z_{n,m}) e^{j \frac{2\pi}{\lambda} \cdot (\sqrt{x_r^2 + y_r^2 + z_r^2})} \quad (11.23)$$

Equation (11.23) is a general propagator that can compute the complex amplitude on any surface or within any volume from any surface, provided that the reconstruction geometry holds true in the paraxial domain (see the next section).

### 11.2.3 How Near Is Near and How Far Is Far?

In Sections 11.2.1 and 11.2.2, we have listed five numeric scalar propagators, three of which are used to compute near-field reconstruction windows and two which are used to compute far-field reconstruction windows. While the application of all five propagators to real-world simulations seems quite straightforward, the ‘near field’ and ‘far-field’ regions have to be defined with caution, since all five propagators are based on the scalar theory of diffraction and must therefore remain within the limits of the paraxial domain. Figure 11.11 shows how ‘near-field’ and ‘far-field’ regions can be defined within the boundaries of the scalar diffraction domain.

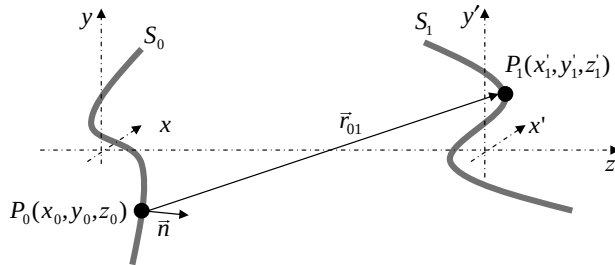


**Figure 11.10** Rayleigh–Sommerfeld diffraction from and to arbitrary surfaces

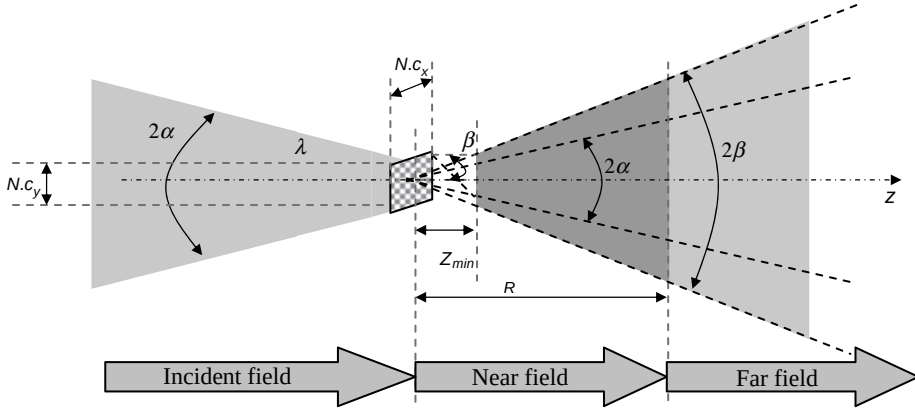


**Figure 11.11** ‘Near-field’ and ‘far-field’ definitions

The scalar diffraction regime is defined by the ratio  $\Lambda/\lambda$  between the smallest grating period  $\Lambda$  in the element and the reconstruction wavelength  $\lambda$  (see also Figure 11.1). In the paraxial domain, this ratio should be larger than 3 (extended-scalar regime) or 5 (scalar regime). As the sine of the diffraction angle is actually this inverse of this ratio, the maximum diffraction angles for extended scalar and scalar regimes are, respectively,  $\alpha = 15^\circ$  and  $\beta = 30^\circ$  (see Figure 11.11).

Figure 11.11 also shows the incoming  $2\alpha$  cone that validates the scalar approximation. However, for the incoming beam, the limitations are more severe in real life, since this angle also modulates the diffraction efficiency of surface-relief elements, as described in Chapter 6.

We have seen previously (see also Appendix B) that all the sampled points in the initial window contribute to the complex amplitude in the reconstruction window. Therefore, the minimum distance  $Z_{\min}$  at which a scalar near field could be computed can be defined as the intersection point between the scalar cone  $2\alpha$  emerging from the edges of the initial window and the extended scalar cone  $2\beta$  from the center window:

$$Z_{\min} \approx Nc \cdot \frac{1}{2(\beta - \alpha)} \propto 5 \cdot Nc \quad (11.24)$$

This is a qualitative expression, but it gives an approximate insight into that smallest distance that can be called the near-field location. Note that this minimum distance is a function of the absolute size of the element. Therefore, if the element is enlarged, this distance increases.

Now, in order to derive the smallest distance that can be called ‘far-field’ for a specific initial window, let us consider an initial circular opening of diameter  $d$  in this window and light of wavelength  $\lambda$  passing through. Much of the energy passing through the small aperture – which is the smallest period  $p$  in the window, or twice the critical dimension ( $p = 2 \times CD$ ) for a diffractive element – is diffracted through an angle of order  $\alpha = \lambda/d$  from its original propagation direction. When we have traveled a distance  $R$  from the aperture, about half of the energy passing through the opening will have left the cylinder made by the geometric shadow if  $d/R = \alpha$ . Putting these together, we see that the majority of the propagating energy in the ‘far-field region’ at a distance greater than the Rayleigh distance  $R$ ,

$$R = \frac{2p^2}{\lambda} \Rightarrow \begin{cases} R_x = 8 \frac{c_x^2}{\lambda} \\ R_y = 8 \frac{c_y^2}{\lambda} \end{cases} \quad (11.25)$$



will be diffracted energy. In this region, then, the polar radiation pattern consists of diffracted energy only, and the angular distribution of the propagating energy will then no longer depend on the distance from the aperture. We are thus in the far field.

Equation (11.24) shows that if the smallest period is different in the  $x$  and  $y$  directions, the Rayleigh distance is also different for these two directions. This is why a Fourier CGH can produce a far-field pattern in  $x$  prior to the far-field pattern in  $y$ , if the CGH is fabricated as rectangular cells rather than square cells (see Chapter 6).

Note also that the Rayleigh distance  $R$  is not a function of the size of the initial window (as opposed to  $Z_{\min}$ , which is a function of the size of the window), but only a function of the smallest period in the initial window and the reconstruction wavelength. We will see in the following sections that replicating an element in  $x$  and  $y$  does not change its Rayleigh distance, but only its  $Z_{\min}$  distance.

### 11.2.4 The Dynamic Range of Reconstruction Windows

In the previous sections, we have described various propagators that can reconstruct numerically complex amplitude windows in both the near and the far field.

Usually, the initial window is some sort of optical element, or optical aperture, composed of refractive or diffractive structures (or a combination thereof). In many cases (see Chapter 6), this element can be a Fresnel lens, a grating, a computer-generated hologram or a more complex hybrid element. The sampling most often used (which takes care quasi-automatically of the Nyquist criterion) is to take the smallest feature in the diffractive element – also known as the Critical Dimension (CD) – as the sampling pixel interspacing. This is especially easy for binary elements – for example, binary gratings, where the smallest structure is actually half the smallest period, which is exactly the Nyquist sampling limit. When using such a sampling, one will reconstruct the fundamental diffraction window in which the two fundamental diffraction orders (the fundamental negative ( $-1$ st) or fundamental positive ( $+1$ st) and potentially the zero (0th) order) can appear. This is a very crude way of modeling a diffractive element, which does not take into consideration fabrication issues such as basic cell shape and aperture shape.

We list below three numeric techniques that allow the optical design engineer to get a more comprehensive view (and one closer to what he or she will actually observe once the element is fabricated) of the intensity and phase maps in both the near and the far fields.

#### 11.2.4.1 The Oversampling Process

Oversampling is a straightforward process that reduces the pixel interspacing in the initial plane (CGH or DOE plane). As the reconstruction window size is a function of the initial cell size (and not the plane size), the reconstruction window will be multiplied by the same factor. Figure 11.12 depicts the oversampling process.

If the diffractive element has been sampled according to the Nyquist criterion (see the previous section), oversampling this same window will still satisfy this criterion.

When oversampling the initial window, one will create higher-order windows that are stitched to the fundamental window, therefore displaying the potential higher orders. This is a useful simulation technique since in many applications, the position of higher orders in regard to the fundamental orders is of great importance (especially when considering the high SNR ratio around the main fundamental order).

Figure 11.13 shows an oversampling process performed on the binary off-axis Fourier CGH that was optimized in Figure 6.9.

Without any oversampling, the only visible reconstruction in the numeric window is the two fundamental orders. When a  $2\times$  oversampling is performed on the CGH, the secondary higher orders become visible, as well as the global sinc envelope that modulates all of the orders. The sinc envelope

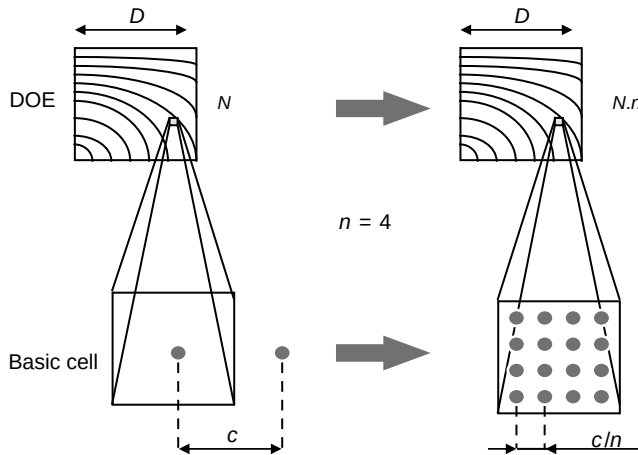


**Figure 11.12** The oversampling process in numeric propagators

seems to be symmetric, which means that the basic cell used to fabricate the CGH is certainly square (if that cell had been rectangular, the sinc would have been asymmetric).

When a more severe  $4\times$  oversampling process is performed, almost all the higher orders that are visible show up on the reconstruction window. The sinc envelope is now clearly visible. Figure 6.15 describes how the optical designer can compensate for this sinc envelope effect by pre-compensating (distorting) the desired pattern before optimizing the CGH (using the sinc compensation technique), if the uniformity of the reconstruction is a critical parameter for the application.

**11.2.4.2 The Embedding Process**

In the embedding process, we literally embed the diffractive element in an array of ‘zeros’, which physically means inserting an aperture stop around the element. The aperture stop is sampled at the same rate as the diffractive itself and creates a single initial sampled plane (see Figure 11.14).

In this case, the basic cell size does not change, and the overall reconstruction window size will not change either. However, since there are more cells in the initial window, and therefore more pixels in the

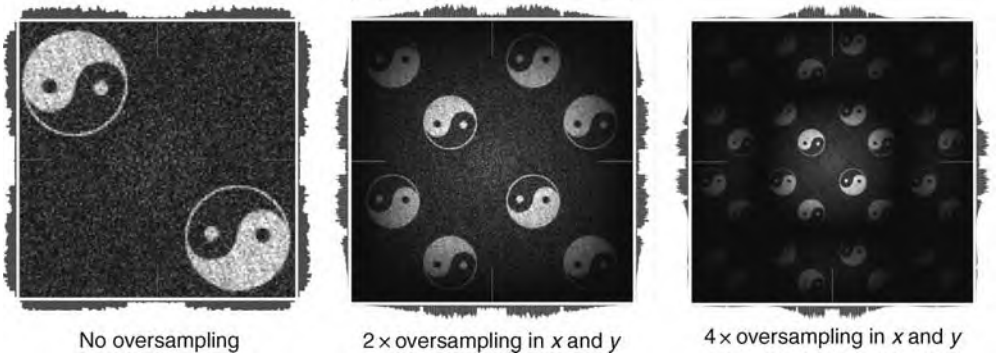


**Figure 11.13** An example of an oversampling process on a binary Fourier CGH

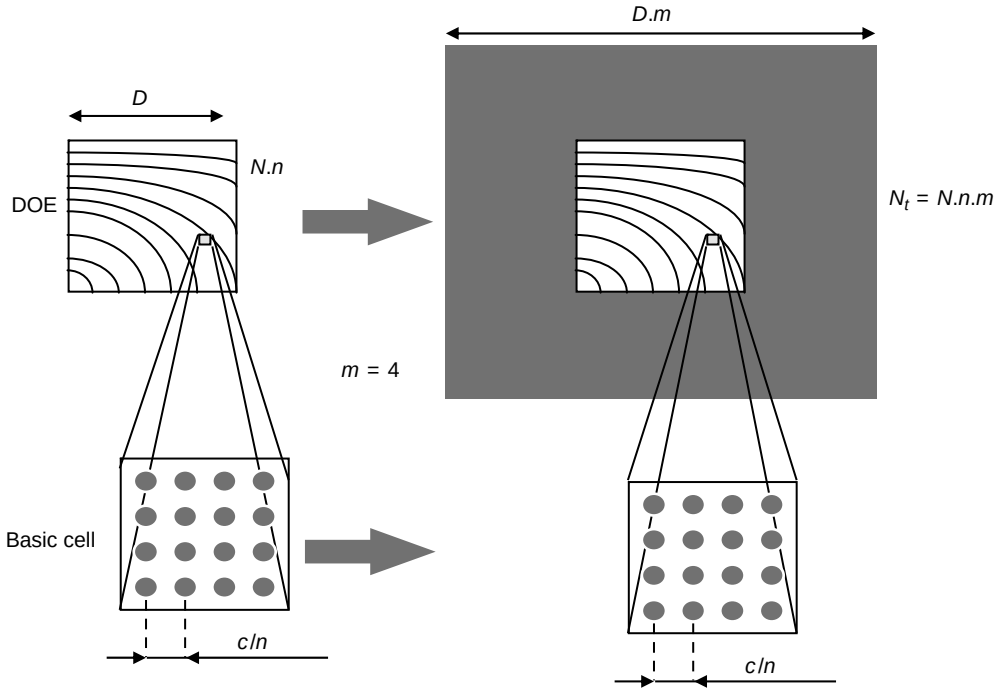


**Figure 11.14** The embedding process

reconstruction window, the resolution of the reconstruction window will be multiplied in accordance with the embedding (or the embedding factor).

By using the FFT algorithm in the design process (see, e.g., the IFTA algorithm in Chapter 6), one can only access certain spot locations within a rigid two-dimensional grid, in both the Fourier and the Fresnel regimes. With the embedding technique, it is possible to reconstruct the field in between the grid used for the design, and therefore access the local SNR in the reconstruction plane.

Figure 11.15 shows an embedding process performed on the binary off-axis Fourier CGH that was optimized in Figure 6.9.

The embedding process does not change the size of the window, but shows more resolution in this same window, especially by reconstructing the speckle-like patterns that are a characteristic of Fourier pattern projections, and one of the main problems to be solved when using laser illumination to form and project images (see, e.g., Chapter 16 on pico-projectors).

The zoom on the reconstruction from a  $2\times$  embedding process in  $x$  and  $y$  clearly shows the speckle grains and the associated hot spots, which do not come out clearly with a standard reconstruction (on the left-hand side of Figure 11.15).

The oversampling and embedding processes can be used together in order to reconstruct an intensity pattern that looks as close as possible to the actual pattern after the CGH has been fabricated and lit by a laser.

Figure 11.16 shows the numeric reconstruction window with a logarithmic intensity profile (the previous reconstruction and a linear intensity profile). This allows us to see clearly the sinc envelope function, and where the first zeros of this function are located.

Such processes can be performed on any type of element – Fourier CGH, DOE lenses, or even micro- or macro-refractive elements and micro- or macro-hybrid lenses – as we will see later on, in Section 11.7.

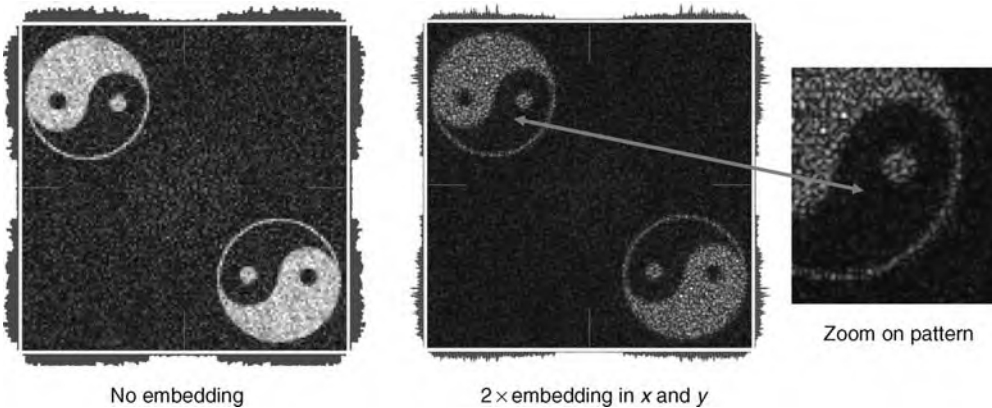


**Figure 11.15** An example of an embedding process on a binary Fourier CGH

**11.2.4.3 The  $x$ - $y$  Replication Process**

The replication process is straightforward process and is used for Fourier diffractive elements as well as for Fresnel diffractive elements.

In the case of Fourier elements, the replication of the basic element, calculated by one of the techniques described in Chapter 6, does not change the overall diffraction pattern, but increases the SNR. This is why when using slow optimization techniques such as GA, DBS or SA, the basic Fourier CGH element can be small, and then replicated in  $x$  and  $y$  to the final size, without changing the target functionality.

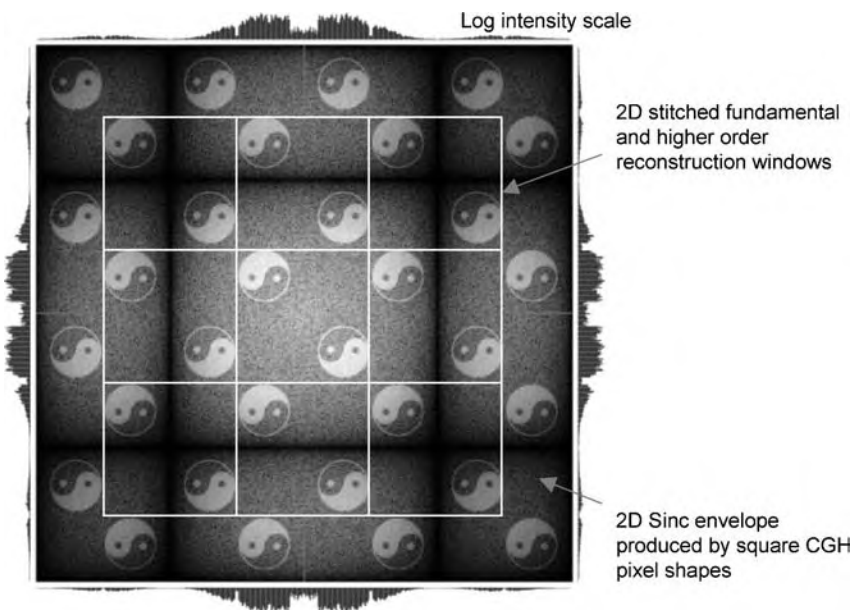


**Figure 11.16** A combined  $4\times$  oversampling and  $4\times$  embedding process (logarithmic intensity profile)

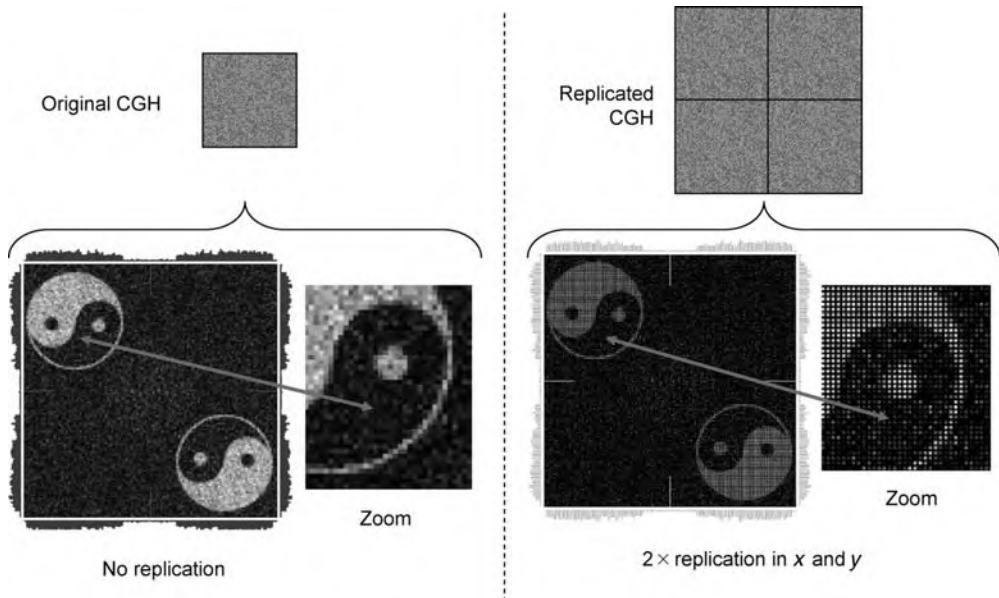


**Figure 11.17** A 2D replication process on a binary Fourier CGH

By replicating the Fourier element in  $x$  and  $y$ , one generates a similar effect in the far-field reconstruction window as in the embedding process: the size of the reconstruction window remains unchanged but the resolution in that window is increased (by the replication factor in  $x$  and  $y$ ). Figure 11.17 shows such a replication process on a binary Fourier CGH.

Figure 11.17 shows an interesting effect of replication. When replicating a Fourier CGH, the overall image does not change – nor does the reconstruction window – but the image gets pixelated (see the zoom on the  $2\times$  replicated CGH reconstruction). A Fourier pattern generated by a digital CGH is always pixelated; however, in a straightforward reconstruction this pixelation does not appear. This is simply the Nyquist criterion kicking in. There cannot be light in between two individual pixels in the sampled Fourier plane. When oversampling more, the reconstruction gets more crisp, but the individual pixels forming the angular spectrum become more and more visible (as they are in the optical reconstruction).

Now, this looks like a bad feature for Fourier pattern generators (like laser point pattern generators), but it is a very nice feature for fan-out gratings or beam splitters, where the key is to achieve the formation of single pixels (single angular beam diffraction). Therefore, fan-out gratings and beam splitters are usually calculated as small CGHs, and then massively replicated in  $x$  and  $y$  in order to smooth out the phase and produce sharp pixels in the far field (i.e. clean sets of diffracted plane wavefronts). If required by the application, one can replicate the CGH in one direction only in order to produce the effect in that direction only (see holographic bar code projection in Chapter 16).

In the case of Fresnel elements, the replication effect is different. Replication of a Fresnel element replicates the optical functionality, but only in the near field. In the far field, the optical functionality is not replicated, but averaged. This is how one can implement beam shapers and beam homogenizers by the use of microlens arrays of fly's eyes.

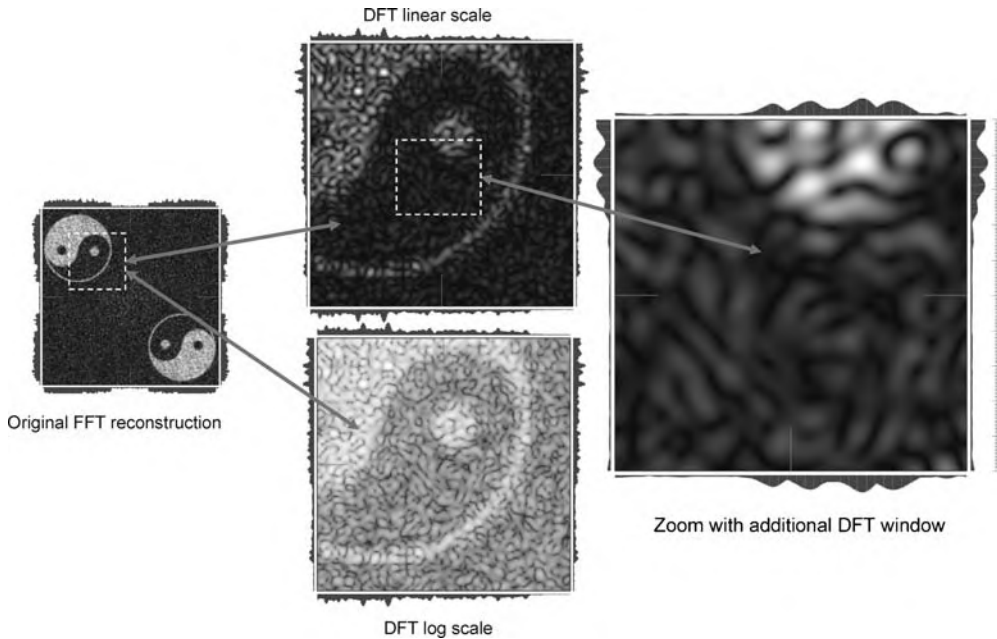


**Figure 11.18** DFT-based numeric reconstructions of a binary Fourier CGH

### 11.2.5 DFT-based Propagators

The previous reconstruction examples have been performed by FFT-based numeric propagators. We will now show the potential of DFT-based numeric propagators. For this, we will take the same binary Fourier CGH examples as for the FFT-based reconstructions (Figure 11.18).

Figure 11.18 shows that one can arbitrarily choose the size and position of a DFT reconstruction window, and arbitrarily set the resolution in this window. As seen in Figure 11.18, one can define smaller and smaller reconstruction windows with increased resolutions. It is interesting to note in the logarithmic reconstruction in Figure 11.18 that the speckle grains appear even if there is not much light (speckle is a phase phenomenon and is therefore not affected by light intensity but, rather, by interferences). In the smallest reconstruction window, however, the speckle seems to be discontinuous between regions with large amounts of light and regions with less light.

The potential of zooming numerically in the reconstruction window indefinitely seems to be very desirable. However, as we have pointed out earlier, such a process requires a much longer CPU time and thus cannot be implemented in a design procedure involving iterative optimization algorithms, but only in the final simulation step (of course, if one has access to a super-computer, one might use a DFT-based algorithm in the design process). The various processes described previously (oversampling, embedding and replication) can also be used in conjunction with DFT-based propagators.

### 11.2.6 Simulating Effects of Fabrication Errors

As we will see in Chapter 15, simulation of the effects of systematic fabrication errors is a very interesting feature that prevents unpleasant surprises when testing a diffractive element fabricated by a foundry,

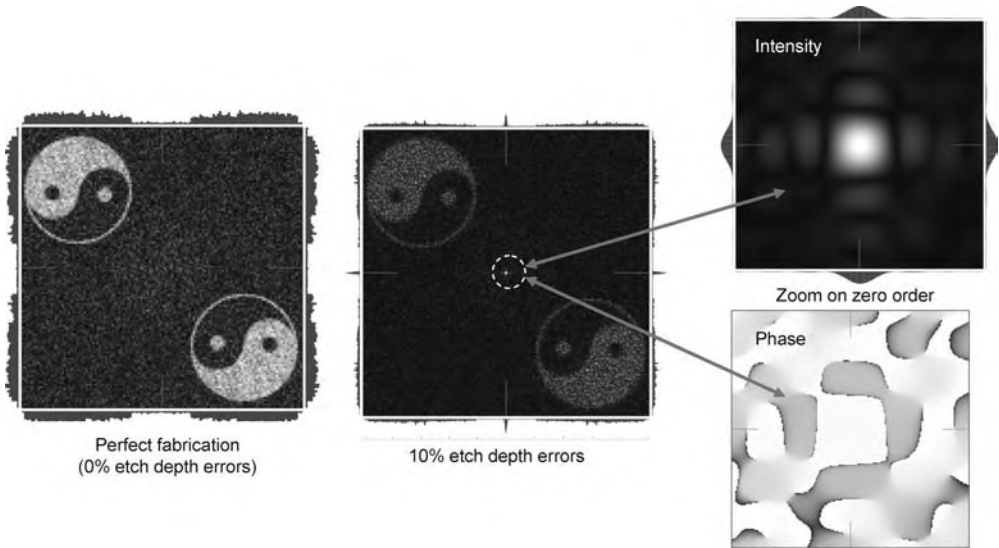


**Figure 11.19** The effect of a  $\pm 10\%$  etch depth error on the reconstruction

although the foundry may have specified  $\pm 5\%$  random etch depth errors over the entire wafer (see also Chapters 12 and 13).

By using FFT- or DFT-based tools, it is very easy to simulate such typical errors, and plug in the tolerancing values that one can get from a foundry prior to fabricating the element. Figure 11.19 shows the effect of an etch depth error of  $-10\%$  on a binary Fourier CGH.

The effect is, of course, an increase in the zero order. This reconstruction has been performed by using an embedding factor rather than an oversampling factor, since what we are interested in here is seeing and evaluating the zero order. One could also zoom over the zero order by using a DFT-based propagator for more information on the amount of light lost.

One can also use the analytic expressions for the diffraction efficiency with the detuning factor (which is related to the phase shift in the diffractive element, and thus related to the etch depth error) as derived in Chapter 5 for gratings, with extrapolation to CGHs and diffractive lenses. Similarly, one can simulate the effects of side-wall angles, surface modulations, nonlinear etch depth variations and so forth.

Figure 11.20 shows an optical reconstruction of a structure illumination CGH (a grid projection in the far field). As the etch depth was not perfect, there is a zero order appearing. The fundamental window is depicted, and one can see the utility of having a zero padding around the main reconstruction, so that the fundamental order lies away from the higher orders.

In some applications where large angles are required, it might actually be desirable to have the fundamental and higher orders stitching to form a single large reconstruction in the far field (though still modulated by the sinc envelope).

### 11.2.7 A Summary of the Various Scalar Numeric Propagators

Table 11.1 shows a summary of the various scalar numeric propagators used here, from the analytic expression to the FFT-based and DFT-based propagators, both for Fourier (far-field) and Fresnel (near-field) operation.

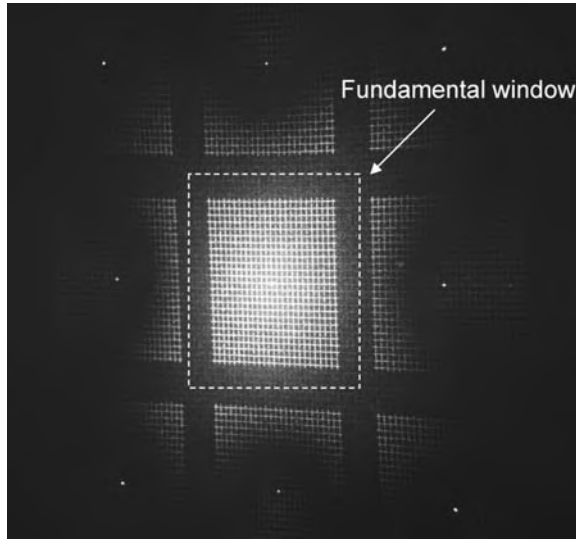


**Figure 11.20** The optical reconstruction of a Fourier grid generator CGH

### 11.2.8 Coherent, Partially Coherent and Incoherent Sources

Partially coherent sources such as LEDs can be a factor of 20 times more efficient than incandescent sources and a factor of five times more efficient than fluorescent lighting technologies. They also have a much longer lifetime than regular incandescent bulbs. This basically proves the case for LED-based lighting technologies for a greener tomorrow. Lighting is becoming a strong market pool for LEDs and LED arrays, especially in the automobile sector and, increasingly, public and domestic lighting. Homogenizing and shaping the LED beam is one of the main tasks in either automotive or standard lighting (hot spots are not desirable).

**Table 11.1** A summary of scalar numeric propagators in the near and far fields

#### Fourier elements

Propagator is based on Fraunhofer approximation of angular spectrum

Analytic propagator:  $U_D = \int_{-\infty}^{\infty} U_I \cdot e^{-2i\pi(x \cdot u + y \cdot v)} \partial x \partial y$

Numeric propagator (FFT-based):  $U_D = \text{FFT}(U_I)$

Numeric propagator (DFT-based):  $U_D = \sum_{k=-N/2}^{N/2} \sum_{l=-M/2}^{M/2} \sum_{i=-A/2}^{A/2} \sum_{j=-B/2}^{B/2} U_I(i, j) \cdot e^{-2i\pi(\frac{ki}{A} + \frac{lj}{B})}$

#### Fresnel elements

Propagator is based on Huygens–Fresnel approximations of near-field diffraction

Analytic propagator:  $U_D(x_1, y_1) = \frac{e^{ikz}}{ikz} \cdot \iint_{\infty} U_I(x_0, y_0) \cdot e^{\frac{i\pi}{\lambda z}((x_0 - x_1)^2 + (y_0 - y_1)^2)} \partial x_0 \partial y_0$

Numeric propagator (FFT-based):  $U_D = e^{-i\pi(x_1^2 + y_1^2)} \cdot \text{FFT}(U_I \cdot e^{-i\pi(x_0^2 + y_0^2)})$

Numeric propagator (DFT-based):  $U_D = \sum_{k=-N/2}^{N/2} \sum_{l=-M/2}^{M/2} \sum_{i=-A/2}^{A/2} \sum_{j=-B/2}^{B/2} U_I(i, j) \cdot e^{\frac{2i\pi}{\lambda} \left| \frac{\vec{r}_{klj}}{r_{klj}} \right|} \cdot \cos(\vec{r}_{klj}, \vec{n}_{ij})$



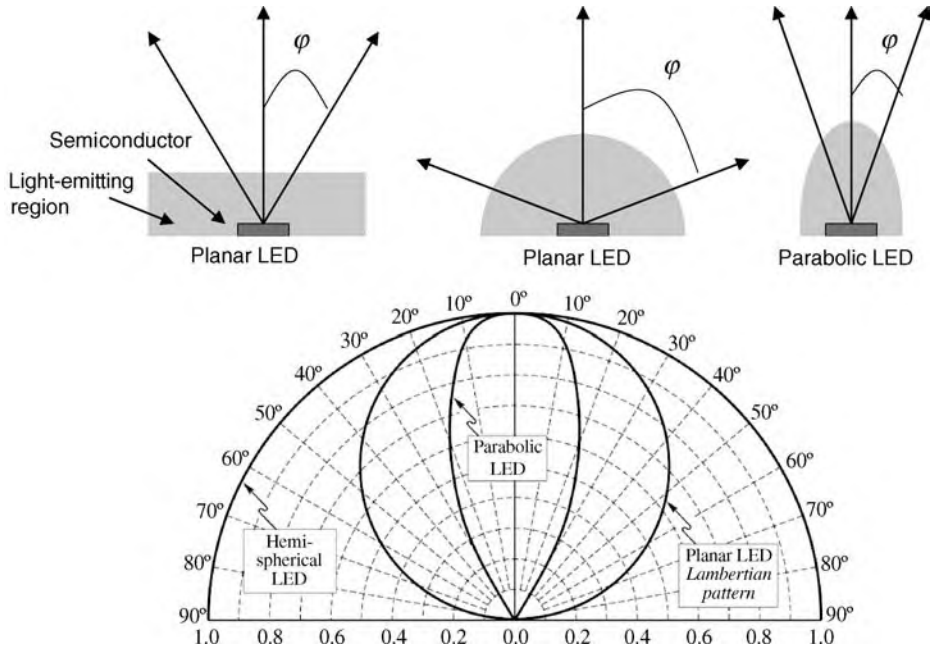


**Figure 11.21** The far-field intensity distribution of various LED geometries

All the reconstructions that we have performed previously have made use of a highly coherent source (spatially and temporally); for example, a monochromatic point source (or a laser). We will now review how one can implement a partially coherent beam configuration (an LED or RC-LED source) in the numeric reconstruction processes that we have described previously.

### 11.2.8.1 LEDs and RC-LEDs

LEDs are partially coherent sources, and therefore one cannot model a digital optic illuminated by an LED with a point spread function (PSF) response, as we have done for laser-illuminated digital optics. Figure 11.21 shows the far-field pattern (angular spectrum) of three typical LEDs: the planar, spherical and parabolic surface LEDs. The planar, and most used, one yields a Lambertian pattern of angular distribution of energy.

Resonant Cavity LEDs (RC-LEDs) are designed to overlap the natural emission band with an optical mode. They yield a much narrower spectral bandwidth and also much narrower angular emissions, as depicted in Figure 11.22. Their behavior is thus much closer to that of lasers and thus more adapted to digital optics (and easier to model).

We will see in the next section that LEDs might suffer from some drawbacks in some configurations compared to laser diodes and VCSELs when used together with digital optics. Thus, when laser-like quality is required in the reconstruction but LED-like specifications are required by the product, an RC-LED might be the solution.

### 11.2.8.2 Modeling Digital Optics with LED Sources

The effects of finite LED die size (usually rectangular), die geometry and nonuniformity of the collimation can be modeled by a convolution process between the perfect reconstruction from a point spread source

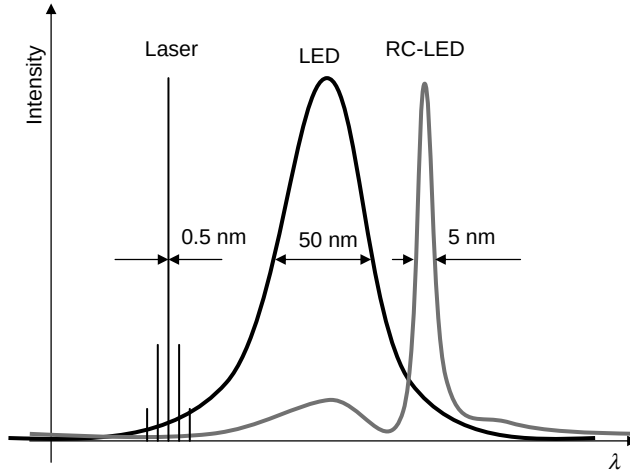


**Figure 11.22** A comparison of emission spectra between lasers, LEDs and RC-LEDs

(PSF, or laser reconstruction) and a 2D comb function that covers the LED die area modulated by a Gaussian distribution (to a first approximation). The area represents the physical LED die size, and the Gaussian distribution represents the intensity over the LED die (higher in the center and decreasing on the edges). The convolution process between this distribution and the PSF function of the digital element is described as follows:

$$U_{LED}(u, v) = U_{PSF}(u, v) \otimes \left( \text{FT} \left( \text{comb} \left( \frac{\Delta_x}{X_{\max}}, \frac{\Delta_y}{Y_{\max}} \right) \cdot e^{-\left( \frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right)} \cdot e^{-\frac{2jR}{R}(x^2 + y^2)} \right) \right) \quad (11.26)$$

where  $\Delta_x$  and  $\Delta_y$  are the sampling points of the Dirac comb function over the LED die,  $\sigma_x$  and  $\sigma_y$  are the Gaussian beam waists of the beam on that die, and  $R$  is the effective radius of curvature of the divergence angle. The PSF is the response of the system (the optical reconstruction) when illuminated by a monochromatic point source located at infinity or by a collimated laser beam.

As an example, Figure 11.23 shows a digital Fresnel CGH beam-shaper lens example optimized by the GS algorithm described in Chapter 6.

The convolution process of the perfect top-hat reconstruction by a two-dimensional comb function of increasing size that has a Gaussian distribution produces a reconstruction that gets more and more fuzzy when the LED die size increases. When the LED die is quite large (e.g. 2 mm square), the reconstruction has actually lost its geometry and is no longer square – nor is it uniform. (For the sake of clarity, we have assumed that the LED beam is more or less collimated by additional optics.)

Another example is shown in Figure 11.24, where a  $4 \times 4$  fan-out grating is illuminated with a laser beam and then with beams from larger and larger LED dies.

Figure 11.24 shows that as the size of the LED die increases (thus reducing the spatial coherence), the reconstruction deteriorates more and more (from the quasi-perfect laser PSF reconstruction on the left to the large LED die on the right). Note that the phase map does not deteriorate as much as the intensity map. These reconstructions have been performed with an embedding factor (no oversampling factor).

Small die LEDs actually produce good results, as shown in the second row in Figure 11.24, where the far field basically bears the LED die geometry, maintaining divergence and limited spatial coherence over the  $4 \times 4$  spot array PSF.

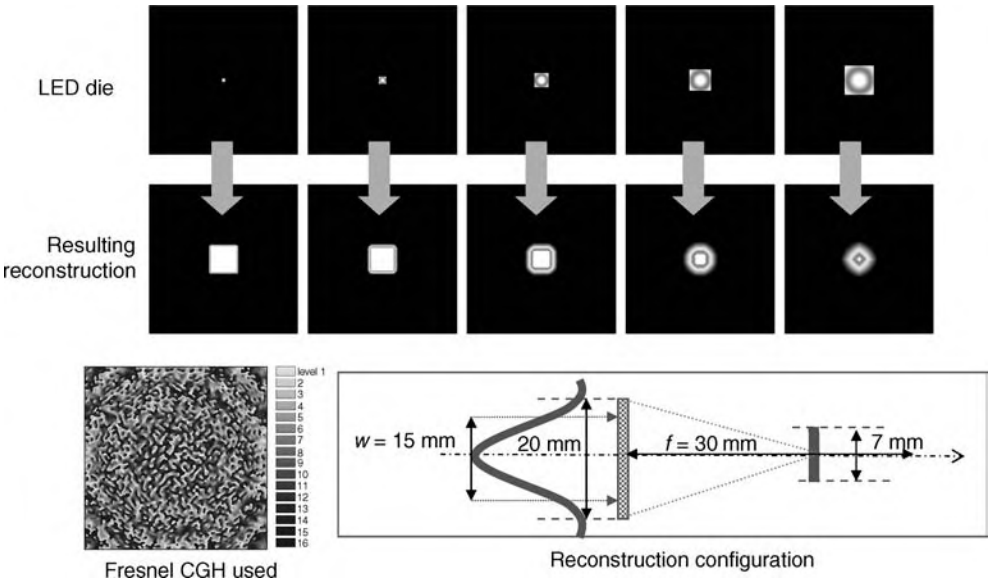


Figure 11.23 An LED reconstruction of a Fresnel CGH top-hat beam shaper

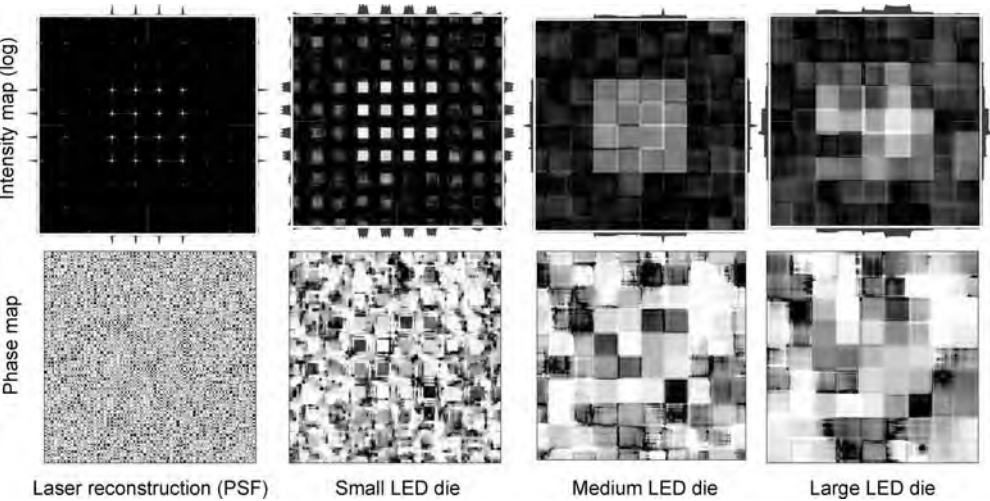


Figure 11.24 The numeric reconstruction of a fan-out grating illuminated by (a) a laser and (b) an LED

11.3 Beam Propagation Modeling (BPM) Methods

In Chapter 3 we have discussed the fundamentals of digital waveguide optics, and their implementation in Photonic Lightwave Circuits (PLCs). Such PLCs can be based on guided waves only or on hybrid free-space/guided-wave configurations, such as the AWG or WGR architectures described in Sections 4.6

and 4.7. The scalar theory of diffraction derived in Appendix B can be applied to model such complex waveguide structures, and in a similar way we have applied it to the modeling of free-space structures in the previous sections. In this section, we will focus on a specific technique called the Beam Propagation Method (BPM).

Beam propagation methods, especially scalar BPMs (in 2D or 3D configurations), are best suited for beam propagation including diffraction effects combined with strong refractive effects. Thus, BPMs are especially useful for the modeling of graded-index micro-optics (Chapter 4) and integrated waveguides (Chapter 3). The BPM method is a split-step method in which the field is propagated and diffracted, and then refracted, in alternating steps [12–14].

This method is increasingly popular because of the ease with which it can be implemented for complex waveguide geometries such as AWGs. Vector BPM methods can also be derived, based on rigorous electromagnetic modeling as described in Appendix A.

The split-step iterative propagation method is based on a core algorithm that includes a Fourier transform of the complex field, phase correction and an inverse Fourier transform. The refraction step is implemented by multiplying the field by a phase correction factor at each point in the transverse grid. Below, the BPM formulations for both the TE and TM polarization modes are presented.

### 11.3.1 BMP for the TE Polarization Mode

The standard Helmholtz wave equation for a transverse field can be written as

$$\nabla^2 E_y(x, z) + k^2 \varepsilon(x, z) \cdot E_y(x, z) = 0 \quad \text{where} \quad \begin{cases} \varepsilon(x, z) = \varepsilon_0(1 + \varepsilon_1(x, z)) \\ \varepsilon_1(x, z) = M \cdot \cos(K_x x + K_z z) \end{cases} \quad (11.27)$$

where  $M$  is the maximum modulation of the relative permittivity and  $K_x$  and  $K_z$  are the projections of the grating vector  $K$  in the  $x$  and  $z$  directions, respectively.

Equation (11.15) can be solved directly to yield an expression for the optical mode with field distribution  $E_y(x, \Delta z)$  for a field at an infinitesimal distance  $\Delta z$  from the origin, in terms of the field at that origin. The BPM is then just the recursion algorithm. The propagation step  $\Delta z$  is typically of the size of the wavelength.

The underlying assumptions of the classical BPM method are as follows:

- the paraxial condition must hold true; and
- the index modulation must be small (i.e.  $M \ll 1$ ).

The TE mode propagator  $\rho_{TE}$  can then be described as follows:

$$E_y(x, z + \Delta z) = \rho_{TE} \cdot E_y(x, z) \quad (11.28)$$

where  $\rho_{TE}$  is the combination of the three previous split steps that we have described, namely

$$\rho_{TE} = \rho_1 \cdot \rho_0 \cdot \rho_1 \quad \text{where} \quad \begin{cases} \rho_0 = e^{-j \frac{k \Delta z}{2 \varepsilon_0} \left( \varepsilon \left( x, z + \frac{\Delta z}{2} \right) - \varepsilon_0 \right)} \\ \rho_1 = e^{-j \frac{\Delta z}{4k} \cdot \frac{\partial^2}{\partial x^2}} \end{cases} \quad (11.29)$$

### 11.3.2 BMP for the TM Polarization Mode

The wave equation for the TM polarization case is slightly different than for the TE polarization mode and is described by

$$\nabla^2 H_y(x, z) + k^2 \varepsilon(x, z) \cdot H_y(x, z) - \left( \frac{\delta \varepsilon}{\varepsilon} \nabla \right) H_y(x, z) = 0 \quad (11.30)$$

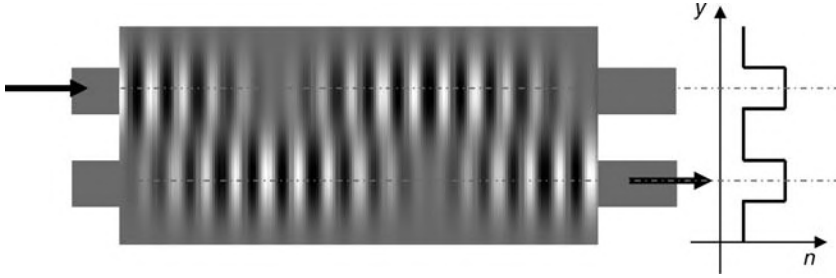


**Figure 11.25** Mode coupling modeling between adjacent ridge waveguides using BPM

where  $H_y(x, z)$  is the transverse magnetic field. The method of solution of this equation (Equation (11.29)) is similar to that for TE, except for the term  $(\delta\epsilon/\epsilon)\nabla$ , which needs to be taken into account. Thus, further assumptions than for the TE mode must be made owing to the additional term. For  $M \ll 1$ , the additional term  $(\delta\epsilon/\epsilon)\nabla$  can be approximated by a constant

$$\frac{\delta\epsilon}{\epsilon} \nabla H_y(x, z) \cong \frac{\epsilon_0(-K_x M \sin(K_x x))}{\epsilon_0(1 + M \cos(K_x x))} \frac{\partial}{\partial x} H_y(x, z) \approx cst. \frac{\partial}{\partial x} H_y(x, z) \quad (11.31)$$

which is important in deriving the TM propagator  $\rho_{TM}$ . Hence, the TM BPM propagator can be defined similarly to the TE BPM propagator as

$$H_y(x, z + \Delta z) = \rho_{TM} \cdot H_y(x, z) \quad (11.32)$$

where  $\rho_{TM}$  is the combination of the three previous split steps that we described, namely

$$\rho_{TM} = \rho_2 \cdot \rho_0 \cdot \rho_2 \quad \text{where} \quad \rho_2 = e^{-j \frac{\Delta z}{4k} \frac{\partial^2}{\partial x^2} - cst. \frac{\partial}{\partial x}} \quad (11.33)$$

Even in its most rudimentary form (as expressed here), the BPM method is a powerful and accurate numeric calculation method to model the functionality of waveguides and more complex PLCs, as well as GRIN optics, micro-refractive optics and even thick holograms constituted by slow index modulations.

As an example of BPM modeling, we show in Figure 11.25 the mode coupling between two identical ridge waveguides.

The beat frequency between the two waveguides in Figure 11.25 is a function of the indices, the wavelength and the waveguide geometry and separation.

The classical BPM method described here can easily be modified if wide-angle propagation is required (that is, beyond the paraxial domain). It can be modified further to handle diffraction problems where the medium may be anisotropic.

## 11.4 Nonparaxial Diffraction Regime Issues

Rigorous vector electromagnetic theories (see Appendix A and Chapter 10) and scalar theory both in the Fresnel and Fraunhofer regimes cover a very wide range of diffractive structures and applications, as they are reviewed in this book. Scalar diffraction theory as reviewed in this chapter is suitable for the modeling of diffraction through structures with periods larger than the reconstruction wavelength and structures that have small aspect ratios. Rigorous diffraction theories based on Maxwell's equations are suitable for the modeling of diffraction through structures with periods similar to the wavelength or shorter, and structures that have high aspect ratios. Such rigorous techniques are also suitable for oblique illumination and numeric reconstructions very close to the microstructures (high angles) [13, 14].

However, there is a region that is not covered accurately by either of these methods: the intermediate region, where the lateral dimensions of the diffracting structures are in the range of several times the wavelength (typically 2 to 10 times), and where the angles are large but still acceptable for scalar theory. An important feature of these intermediate structures is their finite thickness, because of their extended aspect ratio. In scalar theory, Kirchhoff's complex plane screen (see Appendix B) does not take the surface-relief structure into consideration. In rigorous EM theory, we saw that the exact surface-relief profiles can be considered in the modeling process. Intermediate models can incorporate these considerations into scalar theory to some extent.

According to scalar theory, the maximum diffraction efficiency in the first order occurs when the diffracting structures introduce a total phase shift of  $2\pi$  in the analog phase-relief case; that is, an etch depth of  $d = \lambda/(n - 1)$  in a substrate having an index of refraction  $n$ . As the scalar theory assumes that the thickness of the diffractive structures is zero, this theory is only valid for substrates with quite high indices of refraction (see Section 11.1.2). Several propositions have been presented in the literature to model diffraction phenomena through multilevel phase DOEs within this intermediate region. Nonuniform grating etch depths and effects of geometrical shadow duty cycles are some of the keys.

### 11.4.1 The Optimum Local Grating Depth

Let us consider a blazed surface-relief grating with local varying period  $\Lambda$  used to produce a Fresnel lens (Figure 11.26).

We have seen in Chapter 1 that in order to get optimum efficiency, the refraction angle and the diffraction angle should coincide. The two angles coincide for an optimum depth given by

$$d_{opt} = \frac{\lambda}{n - \sqrt{1 - \left(\frac{\lambda}{\Lambda}\right)^2}} \quad (11.34)$$

Note that this 'optimized' etch depth is different from the one that we derived from scalar theory, and is a function of the period of the grating  $\Lambda$ . Besides, it converges to the depth derived from scalar theory when the wavelength-to-period ratio  $\lambda/\Lambda$  approaches zero. Thus, for a blazed Fresnel lens, where the period decreases radially from the optical axis, the depth of the optimized grooves also decreases.

Equation (11.34) was derived for normal incidence. A more general equation takes into account the incidence angle  $\theta_i$ :

$$d_{opt} = \frac{\lambda}{n\sqrt{1 - (\sin(\theta_i)^2)} - \sqrt{1 - \left(\frac{\lambda}{\Lambda} + n \cdot \sin(\theta_i)\right)^2}} \quad (11.35)$$

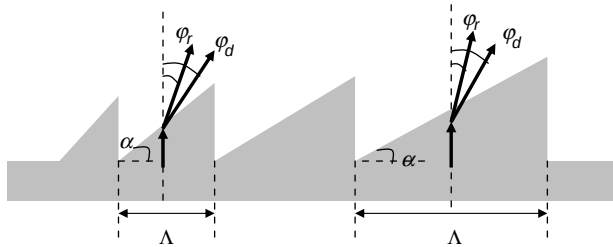


**Figure 11.26** A surface-relief blazed Fresnel lens

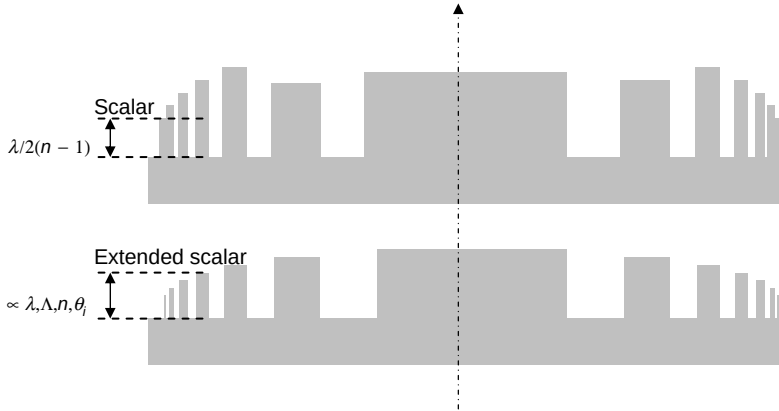


**Figure 11.27** A binary Fresnel lens optimized with scalar theory and extended scalar theory

When the incoming beam is not collimated, but diverging or converging, the local grooves of the Fresnel lens can be optimized in order to get optimal efficiency for that particular prescription. If the lens needs to be optimal for a wide range of configurations, this approach is no longer possible.

Figure 11.27 shows a binary Fresnel lens that has been designed using scalar theory and a second one with extended scalar theory, taking the local groove depths into consideration.

Note that modulation of the depth of the grooves in an analog way is quite possible when using a diamond-turning fabrication technique, but very difficult if using standard binary microlithography techniques.

#### 11.4.2 The Geometrical Shadow Duty Cycle

If we use geometrical ray tracing through a blazed grating [15], we note that a geometrical shadow occurs in the space right after the local prism (see Figure 11.28).

The geometrical shadow duty cycle  $c$  of Figure 11.28 can be expressed as

$$c = 1 - \frac{d\lambda}{\Lambda^2 \cdot \sqrt{1 - \left(\frac{\lambda}{\Lambda}\right)^2}} \quad (11.36)$$

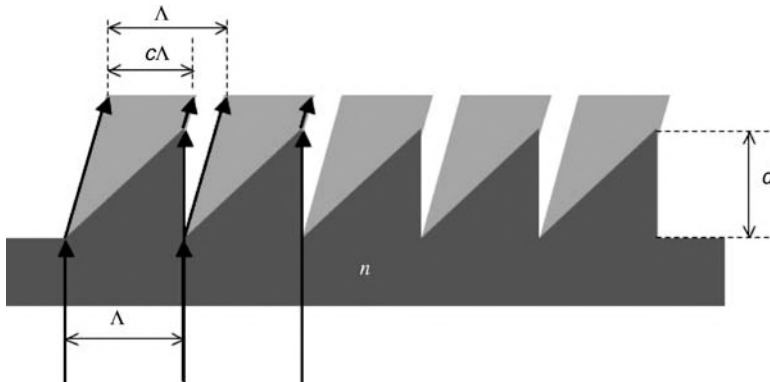


**Figure 11.28** A local geometrical shadow

Thus, for more realistic values of the diffraction efficiency in the first order within the intermediate region (and hence for diffractive structures that are very large compared to  $\Lambda/\lambda \gg 1$ ), the value of the duty cycle  $c$  should be taken into account as a modulating factor of the efficiency  $\eta$  derived from scalar theory (see Chapter 5), and the extended scalar diffraction efficiency can be expressed as

$$\eta_{\text{extended}} = c^2 \cdot \eta \quad (11.37)$$

When the wavelength-to-period ratio approaches zero, the DC reduces to unity, and the diffraction efficiency is again well described by scalar theory.

## 11.5 Rigorous Electromagnetic Modeling Techniques

The scalar theory of diffraction as expressed in the early sections of this chapter as a thin phase plane approximation is an accurate modeling tool for elements that have smallest feature dimensions on the order of the reconstruction wavelength. Figure 10.1 shows the realm of validity of scalar, extended scalar and rigorous vector modeling techniques.

When contemplating the possibility of using a rigorous vector technique to design and model a diffractive element (see Appendix A and Chapter 10), it is important for the optical engineer to remember that due to the high CPU time and complex implementation of the numeric tools, anything more complex than a linear grating is almost impossible to model as a whole element. Therefore, rigorous theories are not often used to design DOEs or CGHs. However, CGHs and DOEs can be approximated locally to linear gratings, as long as they produce sets of smooth fringes (see also the local grating approximation earlier in this chapter).

The diffraction angles (and thus the position and geometry of the optical reconstruction) predicted by scalar theory still hold true even if one is way outside of the realm of scalar theory. Only the diffraction efficiency predictions are different from those predicted by more accurate vector theories.

The conditions that are required for effective use of scalar theory are the following:

- the lateral periods are large with regard to the reconstruction wavelength ( $\Lambda \gg \lambda$ );
- the heights of the structures are comparable the wavelengths used ( $h \approx \lambda$ );
- the polarization effects are neglected (no coupling between the electric and magnetic fields); and
- the region of interest is located far away from the diffracting aperture, and near the optical axis (within the limits of the paraxial domain).

Although these conditions are not too restrictive (most of the practical problems can be solved by scalar theory), there are some cases where it is very desirable to have DOEs that may have the following characteristics:

- sub-wavelength lateral structures ( $\Lambda < \lambda$  or  $\Lambda \approx \lambda$ );
- high aspect ratio structures (deep structures;  $d > \lambda$ );
- polarization-sensitive structures (form birefringence); and
- the region of interest located very close to the DOE ( $z \approx \lambda$ ).

Thus, the vector EM theory model is best suited for applications that include the following:

- very fast lenses ( $f\# < 1$ );
- anti-reflection structures (ARS);
- polarization-dependent gratings;
- very high efficiency binary gratings;
- zero-order gratings; and
- resonant waveguide gratings.



Microlithography fabrication technologies have enabled the fabrication of such sub-wavelength structures (especially with the help of recent DFM and RET techniques as described in Chapter 13). The larger the reconstruction wavelength is, the easier the fabrication of such structures. Thus, diffractive elements that can be modeled by scalar theory in the visible spectrum may need vector EM theory to be modeled with CO<sub>2</sub> laser wavelengths.

One of the most desirable features is the high diffraction efficiency that can be achieved with simple binary sub-wavelength gratings. In scalar theory, a binary element will only give 40.5% efficiency in the best case. A sub-wavelength grating can yield an efficiency greater than 90%.

In principle, by solving Maxwell's time-harmonic equations (see Chapter 10 and Appendix A) at the boundaries of the sub-wavelength structures, one would get an exact solution of the diffraction problem. Unfortunately, in most cases, there is no analytic solution to that problem, and these partial derivative coupled equations have to be solved numerically:

- Appendix A gives the basis of Maxwell's time-harmonic coupled equations, and shows how scalar theory has been derived from this point on.
- Chapter 8 (Holographic digital optics) shows an implementation of the coupled wave theory, which is applied to sinusoidal slanted Bragg gratings, and can be solved analytically with the two-wave coupled wave Kogelnik approximation.
- Chapter 10 (dedicated to sub-wavelength digital optics) summarizes the various models used to solve Maxwell's time-harmonic equations in various different configurations.

## 11.6 Digital Optics Design and Modeling Tools Available Today

There are various optical CAD tools available on the market today that claim to be able to design and simulate digital optics. However, there are few people who can actually design and model digital optics with the right tools, namely scalar, semi-scalar and vector diffraction theory as we have discussed them throughout this book.

The most widely used tools (CodeV™, Zemax™, Asap™, Fred™, Optis™ etc.) have been developed for classical optical design and are thus based on ray-tracing algorithms [16–19]. Their incursion into physical optics territory remains very rare. Most of these software programs can design hybrid diffractive optics that are fabricated as perfect blazed structures very well.

For scalar propagation, Diffract, Virtual labs and DS-CAD provide good tools. As for CGHs and complex DOEs, which are heavily based on scalar theory and iterative optimization algorithms, there are a few software programs that can perform these tasks, namely Virtual labs and DS-CAD. As for sub-wavelength gratings and photonic crystals, there are very few software programs available today, but their number will increase significantly in the coming years, especially by integrating RCWA or FDTD techniques (R-Soft, Grating Solver, GD-Calc etc.).

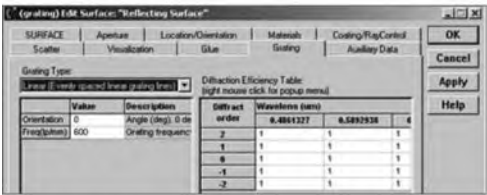
Figure 11.29 shows how a Fresnel lens can be specified in a classical CAD tool (FRED™) [19]. In the first window, the various grooves forming the Fresnel lens are defined, and the second window is used to specify how many diffraction orders are present and how their respective efficiencies are influenced by the wavelength.

Figure 11.30 is a screenshot of how ray tracing is performed over a multi-order diffraction grating.

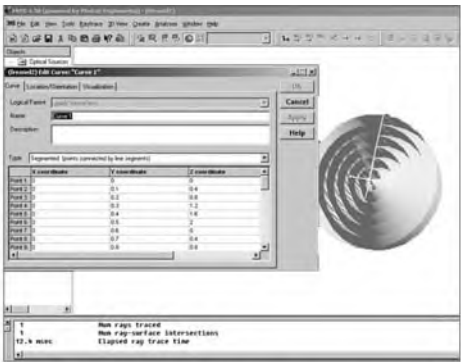
The diffraction orders that appear in Figure 11.30 have been set by hand, as have their respective diffraction efficiencies. Also, such a modeling task does not take multi-order interferences into account.

Table 11.2 summarizes the various tools available on the market today.

In parallel to the digital optics endeavor that we emphasize in this book, it is interesting to note that there is a tremendous software effort in the Electronic Design Automation (EDA) industry to provide vector diffraction tools to model the behavior of sub-wavelength reticles and photomasks in lithographic projection tools. These EDA software packages generally optimize the patterned features on the reticle using optical proximity correction (OPC), phase-shift masking (PSM), oblique illumination and so forth (see Chapter 13), in order to push the resolution envelope without changing the lithographic hardware



This window is used to define each single groove forming the sawtooth-type Fresnel lens



This window is used to specify how many diffraction orders are present and how their respective efficiencies are influenced by wavelength

**Figure 11.29** Specifying a Fresnel lens in a classical optical CAD tool

tools. Such software packages are marketed as Design For Manufacturing (DFM) EDA tools for the IC industry. The core of these software packages is what we are describing in this chapter, namely scalar, semi-scalar and rigorous diffraction models to model diffraction through the reticle in order to compute the aerial image formed on the wafer. The price level of these software packages is such that they are out of

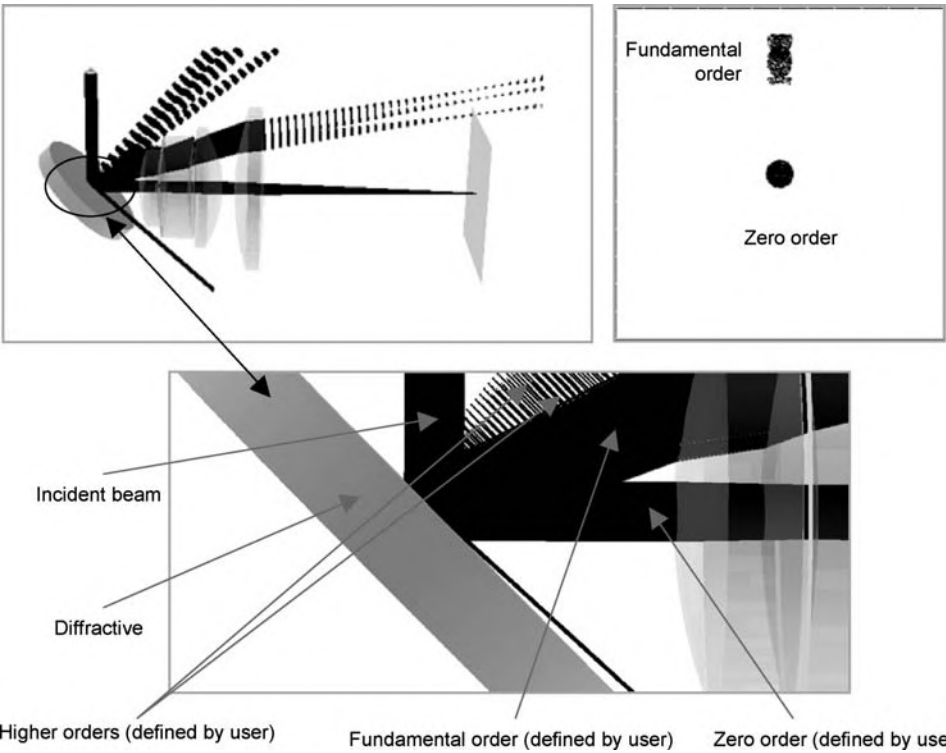


**Figure 11.30** Ray tracing through a diffraction grating in classical optical CAD tool

**Table 11.2** The digital optics design and simulation tools available on the market today

| Company               | Design tool           | CGH design | DOE design | Holographics | Ray trace | Scalar propagation | Vector EM models | Optomechanical tolerancing | Mask layout generation |
|-----------------------|-----------------------|------------|------------|--------------|-----------|--------------------|------------------|----------------------------|------------------------|
| ORA                   | <i>Code V</i>         | No         | Yes        | Yes          | Yes       | Some               | No               | Yes                        | No                     |
| Optima                | <i>Zemax</i>          | No         | Yes        | Some         | Yes       | Some               | No               | Yes                        | No                     |
| Breault               | <i>ASAP</i>           | No         | No         | Some         | Yes       | Yes                | No               | Yes                        | No                     |
| Photon design         | <i>Fred</i>           | No         | Yes        | Yes          | Yes       | No                 | No               | Yes                        | No                     |
| Optis                 | <i>Solstis</i>        | Some       | Yes        | Some         | Yes       | Some               | No               | Yes                        | No                     |
| MMS                   | <i>Diffract</i>       | No         | No         | Some         | Yes       | Yes                | Some             | Yes                        | No                     |
| LightTrans            | <i>Virtual Lab</i>    | Yes        | Yes        | Some         | Yes       | Yes                | No               | Some                       | Some                   |
| EM Photonics          | —                     | No         | NO         | No           | Yes       | Yes                | Yes              | Yes                        | No                     |
| Rsoft                 | <i>Diffract Mod</i>   | Some       | Some       | Some         | Yes       | Yes                | Yes              | Yes                        | No                     |
| Diffraction Solutions | <i>DS-CAD</i>         | Yes        | Yes        | Yes          | Yes       | Yes                | Some             | Yes                        | Yes                    |
| gsolver               | <i>Grating solver</i> | No         | No         | No           | No        | No                 | Yes              | No                         | No                     |
| <i>GD-Calc</i>        | <i>GD-Calc</i>        | No         | No         | No           | No        | No                 | Yes              | No                         | No                     |

reach for an industry that relies on diffractives as only a sub-component of its products. Such DFM/EDA software packages can only be purchased (or licensed) if the industry has to rely heavily on them to stretch out the reach of their multi-million dollar lithography machines to access the multi-billion dollar IC fab market. Some such companies are Synopsis, Cadence, Mentor Graphics and KLA-Tencor.

### 11.7 Practical Paraxial Numeric Modeling Examples

In this section, we will present the following examples of the previous numeric propagators (both in the near and the far field), applied to real-life applications:

- a hybrid imaging objective lens for a digital camera;
- the modeling of a hybrid lens for CD/DVD Optical Pick-up Units (OPUs); and
- the modeling of a laser diode to lensed fiber coupling.

The first two applications are taken from consumer electronic products and the third one from optical telecoms. All three examples include refractive elements. The first two also include diffractive elements.

#### 11.7.1 A Hybrid Objective Lens for a Compact Digital Camera

Here, we consider a hybrid refractive/diffractive lens used in an IR digital camera. The lens has a spherical/aspheric convex/convex surface with an aspheric diffractive surface on the aspheric refractive surface (second surface). The lens is optimized and modeled using a standard optical design software tool as far as ray tracing is concerned.

We will perform a classical lens field-of-view analysis, which means that we will launch a collimated wavefront on this lens at increasing angles and compute the resulting focus plane, both with standard ray-tracing tools and with a numeric Fresnel propagator based on the previously described analytic formulation of diffraction in the near field. The lens and the illumination geometry are shown in Figure 11.31. The resulting lateral spot diagrams for incident angles from 0° to 25° are shown in Figure 11.32.

Both modeling results (ray-traced and scalar numeric Fresnel field propagators) agree well on all spot diagrams, even with angles up to 25° that the limit of paraxial approximation). However, the resolution in the numeric analysis tool when using the scalar model is much greater than when using only geometrical ray-tracing methods. Also, effects of multiple diffraction order interferences (see the fringes created in the reconstructions) can be observed, which cannot be demonstrated with conventional ray-tracing algorithms.

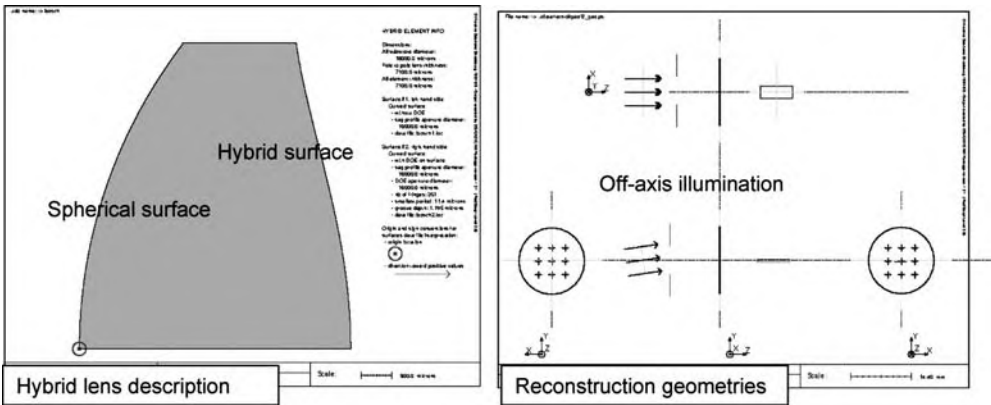


Figure 11.31 The hybrid lens used in the following numeric reconstructions

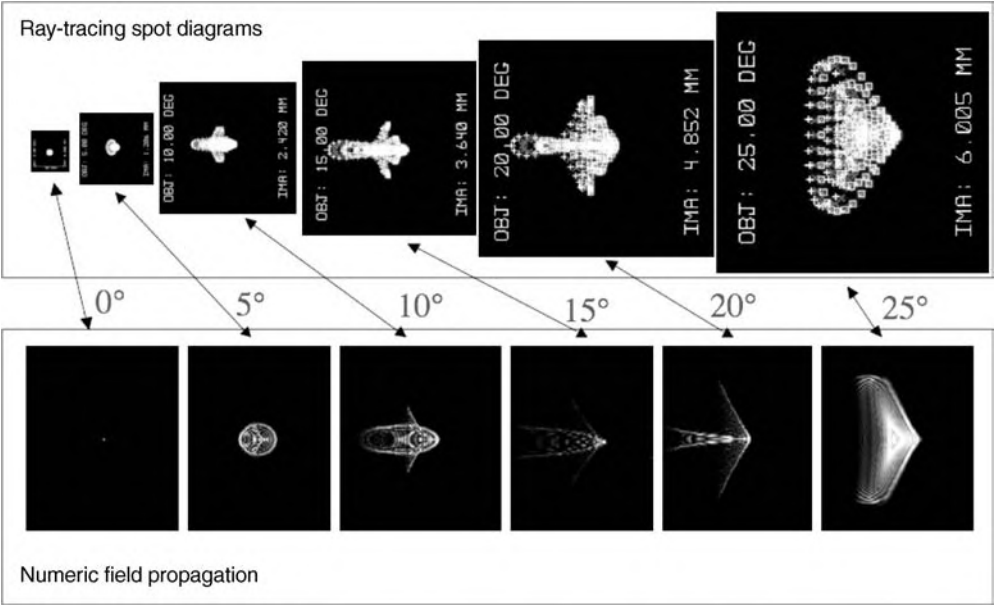


**Figure 11.32** Ray-traced spot diagrams and scalar field propagation in lateral planes for various field angles

In order to show the level of detail one can obtain by using such scalar field propagators instead of ray-traced algorithms, we show the 3D intensity maps of the on-axis and off-axis illumination (Figure 11.33).

In Figure 11.33, on the right-hand side, a small aliasing is showing in the off-axis reconstruction, owing to the Nyquist criterion (which was just about to be violated in that off-axis case).

Scalar propagation of diffracted/refracted/reflected wavefronts actually gives much more flexibility in modeling diffractives (and any other micro-optical elements), in the sense that not only can lateral spot diagrams be computed, but also longitudinal intensity profiles, as shown in Figure 11.34.

The numeric reconstruction can actually be computed on any type of surface, planar or curved, even in a 3D volume, as long as it can be described in a computer. This gives a much deeper insight into the modeling aspects, which was not possible with only ray tracing.

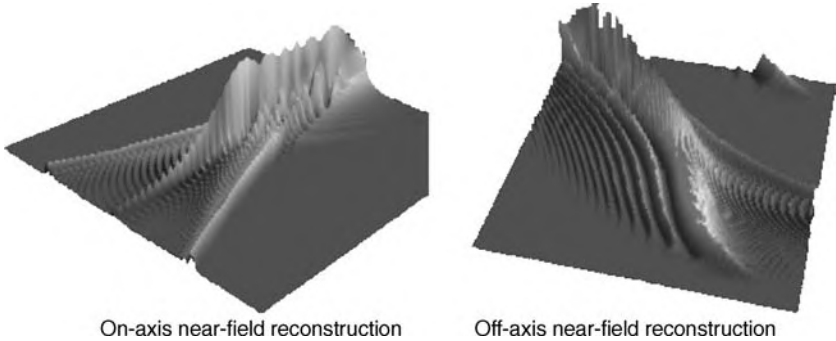


**Figure 11.33** Three-dimensional intensity maps of the previous reconstructions

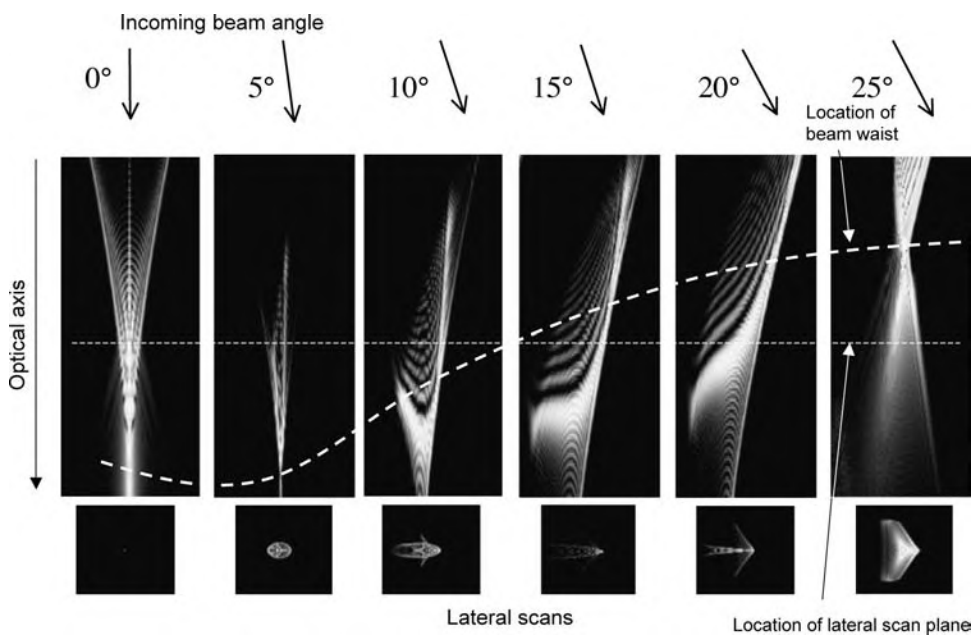


**Figure 11.34** Longitudinal reconstructions via numeric DFT-based scalar propagators

11.7.2 The Modeling of a Hybrid CD/DVD OPU Objective Lens

Here, we model the hybrid dual-focus objective lens used in various CD/DVD drives on the market today (see also Chapters 7 and 16).

Figure 11.35 shows numeric reconstructions along the optical axis for various implementations of a dual-focus lens – as a simple dual lens, as a super-resolution lens (center of the lens obscured) or as an extended-focus lens (for details on DOF modulation techniques, see Chapter 6).

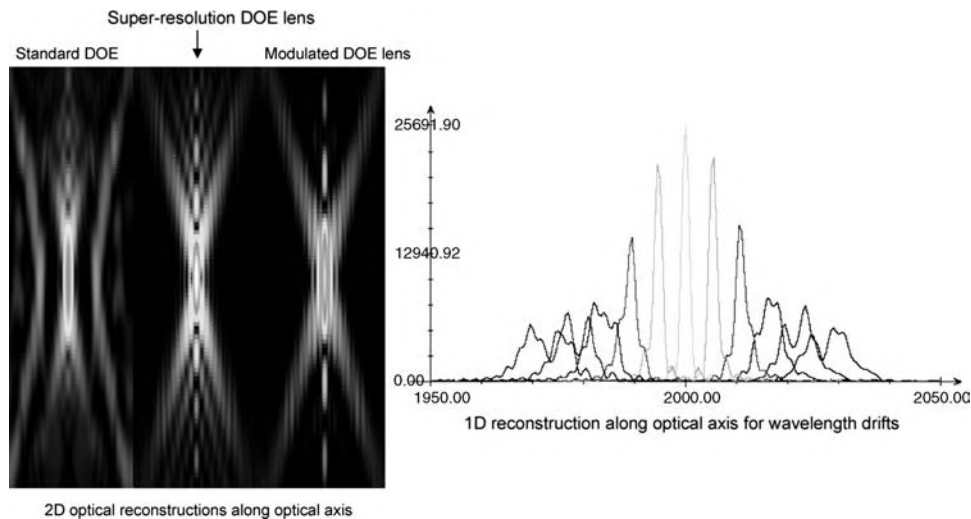


**Figure 11.35** The modeling of a dual-focus CD/DVD OPU objective lens

One can see that the focusing properties for each configuration are different, and that the Strehl ratio is linked to the DOF (the number of side lobes at focus). Figure 11.35 also shows a tolerancing analysis along the optical axis when the wavelength varies. One can see not only that the focus location varies (longitudinal aberrations), but also that the quality of the beam deteriorates (Strehl ratio).

### 11.7.3 The Modeling of a Laser Diode to Lensed Fiber Coupling

Finally, we will show some simulations using Fresnel propagators on an example including only refractive micro-optics and waveguide optics (no diffractive optics here), and see that such tools are still very valuable, for a wide variety of simulation tasks.

The task here is to couple a laser diode into a single mode fiber, without any external optics (see also Chapter 3). Here, we use a lensed fiber tip that is actually a chiseled fiber tip in one direction (see Figure 11.36). Such a fiber tip can be shaped using a CO<sub>2</sub> laser ablation system.

First, we will propagate the diverging field escaping from the laser diode waveguide into free space. Figure 11.37 shows the free-space propagation of the beam along the fast and slow axes of the laser, both in intensity and phase maps.

Second, we will use the field expression hitting the fiber end tip and the fiber end tip geometry to calculate the coupling efficiency between the free-space wave and the fiber end tip region. The coupling is expressed as follows:

$$\eta = 100 \frac{\left| \iint \Psi_{\text{sys}}(x, y) \cdot \Psi_{\text{fiber}}^*(x, y) \partial x \partial y \right|^2}{\iint \Psi_{\text{sys}}(x, y) \cdot \Psi_{\text{sys}}^*(x, y) \partial x \partial y \cdot \iint \Psi_{\text{fiber}}(x, y) \cdot \Psi_{\text{fiber}}^*(x, y) \partial x \partial y} \quad (11.38)$$

$$\text{where } \psi_{\text{fiber}}(x, y) = e^{-\frac{r^2}{w_0^2}}$$

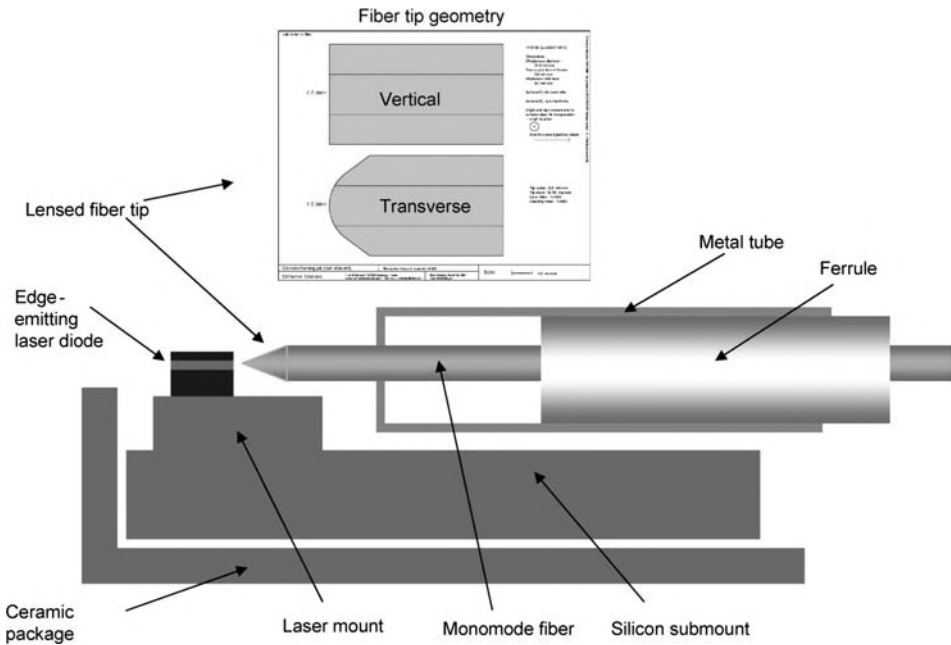


**Figure 11.36** Laser diode to fiber coupling via a lensed chiseled fiber tip

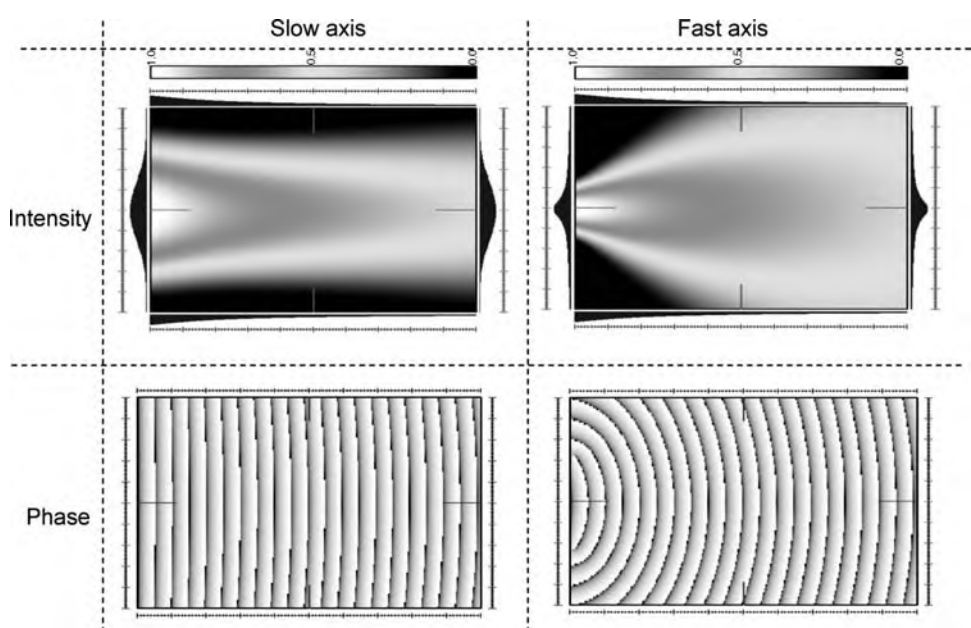


**Figure 11.37** The divergence of the laser diode beam at the exit of the waveguide

Figure 11.38 shows the coupling efficiency as a function of the position of the fiber with reference to the laser diode.

As can be seen, there is a coupling maximum at around 20  $\mu\text{m}$ , and one can position the fiber at the optimal position to achieve such a coupling. One can also optimize the fiber end tip geometry in order to increase the coupling efficiency. Such optimizations are performed in Figure 11.39, where we change first the fiber tip radius and then the fiber tip angle.

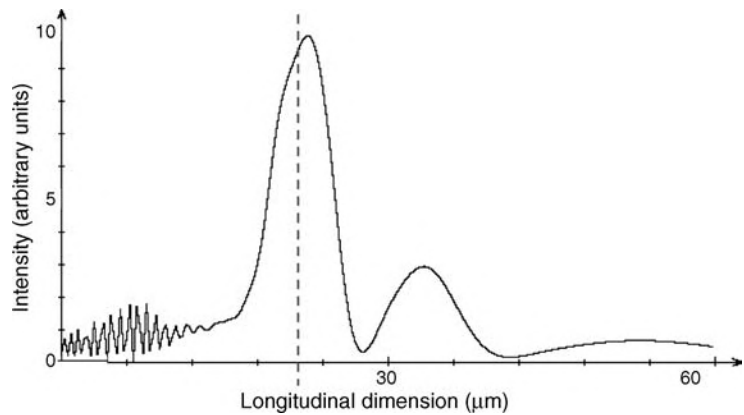


**Figure 11.38** The coupling efficiency between the laser diode and the fiber



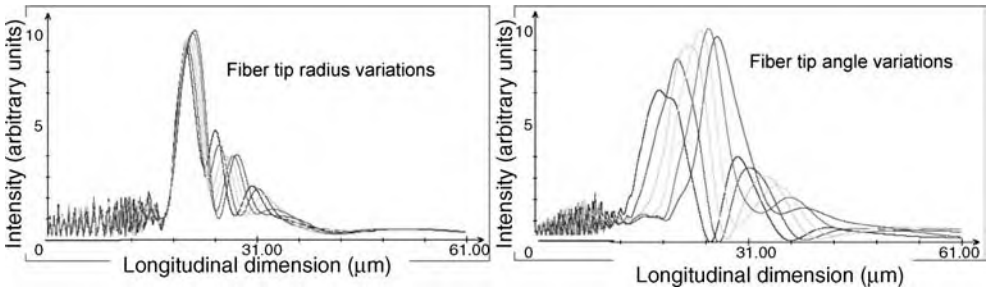


**Figure 11.39** The coupling efficiency as a function of the fiber tip radius and fiber end tip angle

One can see a trend in the variation of the fiber tip geometries, and optimize such a fiber to achieve the best coupling. This is only possible via numeric simulation.

Finally, we model the propagation of the coupled field inside the lens fiber, before the field gets completely guided by the regular fiber core (see Figure 11.40), for the lensed axis and the chiseled axis.

Furthermore, Figure 11.40 shows the modeling of the backscattered complex field from the end tip fiber into free space for both dimensions, so that one can get a feeling of how much light might enter the laser diode guide again and create potential laser perturbations.

Back-propagation is easily implemented in the scalar propagators that we have listed in this chapter, by reversing the sign of the exponential either in the Fraunhofer expression or in the Fresnel expression (type 1 or type 2, even in the DFT-based propagators).

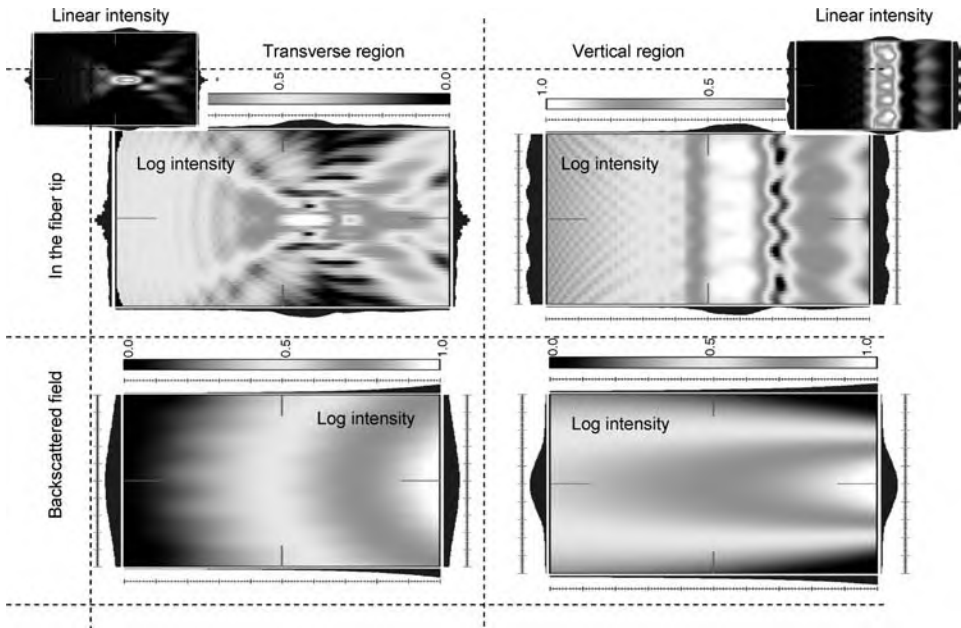


**Figure 11.40** Propagation of the coupled field inside the fiber tip and the backscattered field from the fiber tip end

### 11.7.4 Vortex Lens Modeling Example

In this last example, we show how one can model nonconventional beams, such as optical vortex beams. Here, we investigate optical vortex beam modes 1 and 2, produced by the diffractive vortex lenses described in Section 5.4.6.6, which can be used in various applications ranging from plastic graded-index fiber coupling to high-resolution interferometric metrology of high aspect ratio microstructures.

In order to model the behavior of such vortex beams in the vicinity of the focal plane, we reconstruct numerically the intensity and phase maps in transverse windows located at successive distances from the focal plane (see Figure 11.41), as well as longitudinal reconstructions along the optical axis (see Figure 11.42).

These numerical reconstructions have been performed through the use of a near-field DFT-based algorithm described earlier, in Section 11.2.5. The sizes of the windows are  $200\mu\text{m} \times 200\mu\text{m}$  and the sampling step is  $1\mu\text{m}$ . Numerically propagating the beam into various lateral windows in the vicinity of the focal plane informs us about how the beam focuses down and the beam waist. For example, one can see clearly that both beams share the same focus, and produce a doughnut beam profile at the center with a Gaussian profile, as well as having an optical axis that is absolutely free of any light. However, the mode 2 vortex beam has a larger waist than the mode 1 vortex beam, although the lenses have the same focal

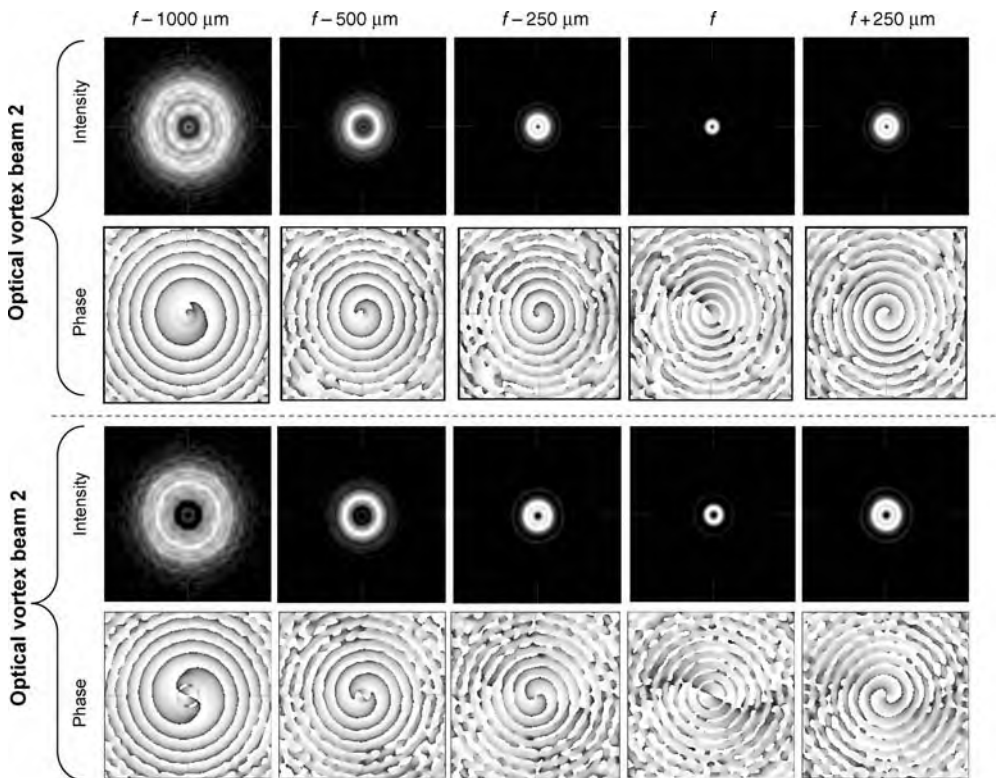


**Figure 11.41** Lateral reconstructions of vortex beams of modes 1 and 2, in intensity and phase, at various distances from the focal plane ( $f = 30\text{ mm}$ )

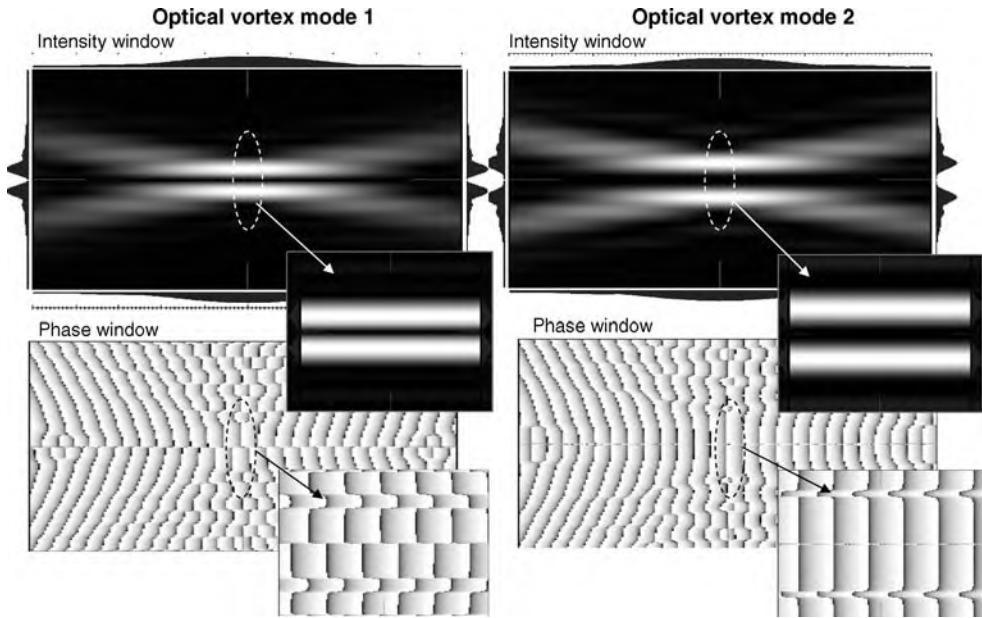


**Figure 11.42** Longitudinal reconstruction windows of vortex beams of modes 1 and 2, in intensity and phase

length, the same aperture and are used with the same wavelength. The phase maps show also that the phase vortex changes direction after the focal plane. At the focal plane, the phase map shows an interesting pattern, which is quite similar to the phase map of a spherical lens (no phase vortex present at that particular location).

The same vortex beams have also been reconstructed along the optical axis with the same near-field DFT algorithm as previously (see Figure 11.42). In this case, the windows are not square as in the previous lateral windows, but rectangular – 500  $\mu\text{m}$  long by 200  $\mu\text{m}$  wide, with  $200 \times 200$  pixels. It can be seen that the two beams have similar intensity patterns and depths of focus, but different phase patterns. The mode 1 vortex beam produces a clear-cut phase shift of  $\pi$  on the optical axis, whereas the mode 2 vortex beam has no phase discontinuities at the optical axis. In order to see what happens more closely, we have reconstructed another set of windows, this time rectangular in the opposite direction (10  $\mu\text{m}$  long by 50  $\mu\text{m}$  wide), again with the same number of pixels ( $200 \times 200$ ). This shows the versatility of the near-field DFT numeric propagator, which can set the resolution arbitrarily in both axes, depending on what has to be analyzed. It is also interesting to notice the sharp  $\pi$  phase jumps occurring in the vicinity of the highest intensity over the doughnut rings, in both vortex beam modes.

In this chapter, we have summarized the various tools available to the optical engineer to implement scalar, semi-scalar and vector diffraction theory for the modeling, simulation and optomechanical tolerancing of binary optics in various configurations. We have also given several examples of the numeric modeling of digital micro-refractive optics, hybrid optics and diffractive optics for various applications used in industry today.

Chapters 8 and 10 show various additional examples of the implementation of vector EM theories for both holographic and sub-wavelength digital structures.

## References

- [1] G.H. Spencer and M.V.R.K. Murty, 'General ray-tracing procedure', *Journal of the Optical Society of America A*, **41**, 1951, 650.
- [2] W.H. Southwell, 'Ray tracing kinoform lens surfaces', *Applied Optics*, **31**, 1992, 2244.
- [3] A.J.M. Clarke, E. Hesse, Z. Ulanowski and P.H. Kaye, 'Ray-tracing models for diffractive optical elements', *Optical Society of America Technical Digest*, **8**, 1993, 2–3.
- [4] J.L. Rayces and L. Lebach, 'Modeling of diffractive optical elements for lens design', Personal communication, 1993.
- [5] W.C. Sweatt, 'Mathematical equivalence between a holographic optical element and an ultra-high index lens', *Journal of the Optical Society of America A*, **69**, 1979, 486–487.
- [6] W.C. Sweatt, 'Achromatic triplet using holographic optical elements', *Applied Optics*, **16**, 1977, 1390.
- [7] R.D. Rallison and S.R. Schicker, 'Wavelength compensation by time reverse ray tracing', in *'Diffractive and Holographic Optics Technology II'*, I. Cindrich and S.H. Lee (eds), SPIE Press, Bellingham, WA, 1995, 217–226.
- [8] F. Wyrowski, 'Design theory of diffractive elements in the paraxial domain', *Journal of the Optical Society of America A*, **10**(7), 1993, 1553–1561.
- [9] F. Wyrowski and O. Bryngdahl, 'Digital holography as part of diffractive optics', *Reports on Progress in Physics*, **54**, 1991, 1481–1571.
- [10] G.W. Forbes, 'Scaling properties in the diffraction of focused waves and an application to scanning beams', Personal communication, 1993.
- [11] W. Singer and M. Testorf, 'Gradient index microlenses: numerical investigations of different spherical index profiles with the wave propagation method', *Applied Optics*, **34**(13), 1995, 2165–2171.
- [12] G.C. Righini and M.A. Forastiere, 'Waveguide Fresnel lenses with curved diopters: a BPM analysis', in *'Lens and Optical Systems Design'*, H. Zuegge (ed.), SPIE Vol. 1780, 1992, 353–362.
- [13] S. Ahmed and E.N. Glytsis, 'Comparison of beam propagation method and rigorous coupled wave analysis for single and multiplexed volume gratings', *Applied Optics*, **35**(22), 1996, 4426–4435.
- [14] C.A. Brebbia, J.C.F. Telles and L.C. Wrobel, *'Boundary Element Techniques, Theory and Application in Engineering'*, Springer-Verlag, Berlin, 1984.
- [15] G.J. Swanson, *'Binary Optics Technology: Theoretical Limits on the Diffraction Efficiency of Multi-level Diffractive Optical Elements'*, MIT Lincoln Laboratory Technical Report 914, 1991.
- [16] BeamProp User Guide, Version 4.0, Rsoft Inc., 2000; [www.rsoftinc.com](http://www.rsoftinc.com).
- [17] Breault Research Organization Inc.; [www.breault.com](http://www.breault.com).
- [18] Optical Research Associates, *Code V*; [www.opticalres.com/](http://www.opticalres.com/).
- [19] Photon Engineering, *FRED*; [www.photonengr.com/](http://www.photonengr.com/).

# 12

## Digital Optics Fabrication Techniques

The first synthetic grating assembled by man was probably made out of animal hair (binary amplitude transmission gratings). These would diffract direct sunlight faintly into its constitutive spectra (rainbow colors). Later synthetic gratings were probably made from scratches in shiny surfaces, metal or glassy rocks (as binary phase gratings), in order to generate interesting 3D visual effects (for more details on such ‘scratch-o-grams’, see Chapter 2).

Today, industry is becoming a fast-growing user of diffractive and micro-optics, and therefore needs a stable, reliable and cheap fabrication and replication technological basis, in order to integrate such elements in industrial applications as well as in consumer electronics products (Chapter 16), especially when the products and applications using such elements are to become commercial commodities.

Depending on the target diffraction efficiency and the smallest feature size in the digital optical element, the optical engineer has a choice between a vast variety of fabrication technologies, ranging from holographic exposure and diamond machining to multilevel binary lithography and complex gray-scale masking technology.

The flowchart in Figure 12.1 summarizes the various fabrication technologies as they have appeared chronologically.

For a single optical functionality – for example, a spherical lens – the optical engineer can decide to use various different fabrication technologies [1], which have their specific advantages and limitations, mainly in terms of diffraction efficiency. Figure 12.2 shows different physical implementations of the same diffractive lens (and the refractive lens counterpart on top), with their respective fabrication technologies and their respective diffraction efficiencies (both theoretical and expected practical).

It is important to note that the notion of diffraction efficiency is closely related to the final application (see also Chapters 5 and 6). One can therefore increase the efficiency without moving to a more complex fabrication technology, by just implementing a specific optical reconstruction geometry. One might also think that the maximum diffraction efficiency is the most desirable in all cases. Chapter 16 shows that there are many applications that require a specific diffraction efficiency, which is much lower than the maximal achievable by the technology chosen to fabricate the element.

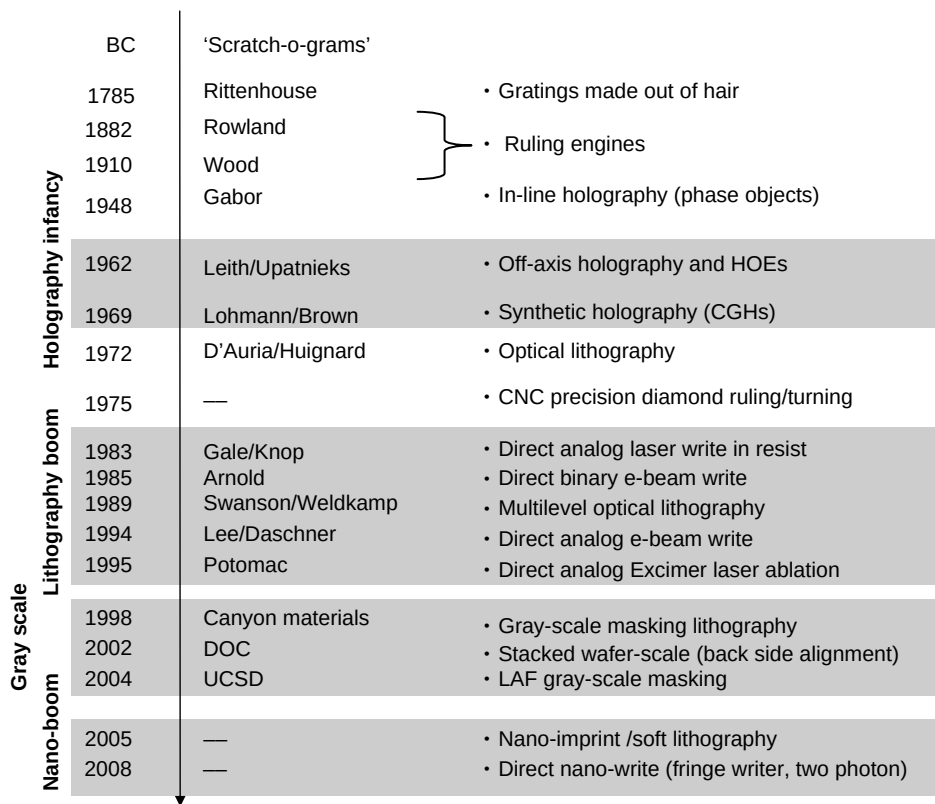


**Figure 12.1** A chronology of fabrication techniques and technologies for digital optics

## 12.1 Holographic Origination

Holographic origination, or holographic exposure of gratings, display holograms or HOEs, can yield a variety of optical functionalities, either as thin surface-relief HOEs as shown in Figure 12.3 (by the use of photoresist spun over a substrate) or as more complex Bragg grating structures in a volume hologram emulsion.

Chapter 8 is dedicated to holographic optical elements, and describes in detail the various holographic origination techniques used today in industry.

Holographic recording, especially as index modulation, is not in essence a micro-fabrication process, as compared to diamond ruling or optical lithography. However, a holographic origination process can be followed by a lithographic fabrication process; for example, by transferring a sinusoidal pattern in resist into a binary pattern in the underlying substrate (see Figure 12.4).

For many applications, it is desirable to produce a nonsinusoidal profile in the resist; for example, to produce asymmetric profiles such as sawtooth structures, blazed or triangular grooves for high efficiency. A multiple recording process can thus be used, which is similar to the Fourier decomposition of an arbitrary periodic signal into various sinusoidal basis functions (see Figure 12.5).

The first example in Figure 12.5 shows the origination of an *échelette* grating (or sawtooth grating), via a series of standard holographic exposure processes (a Fourier decomposition process). The second

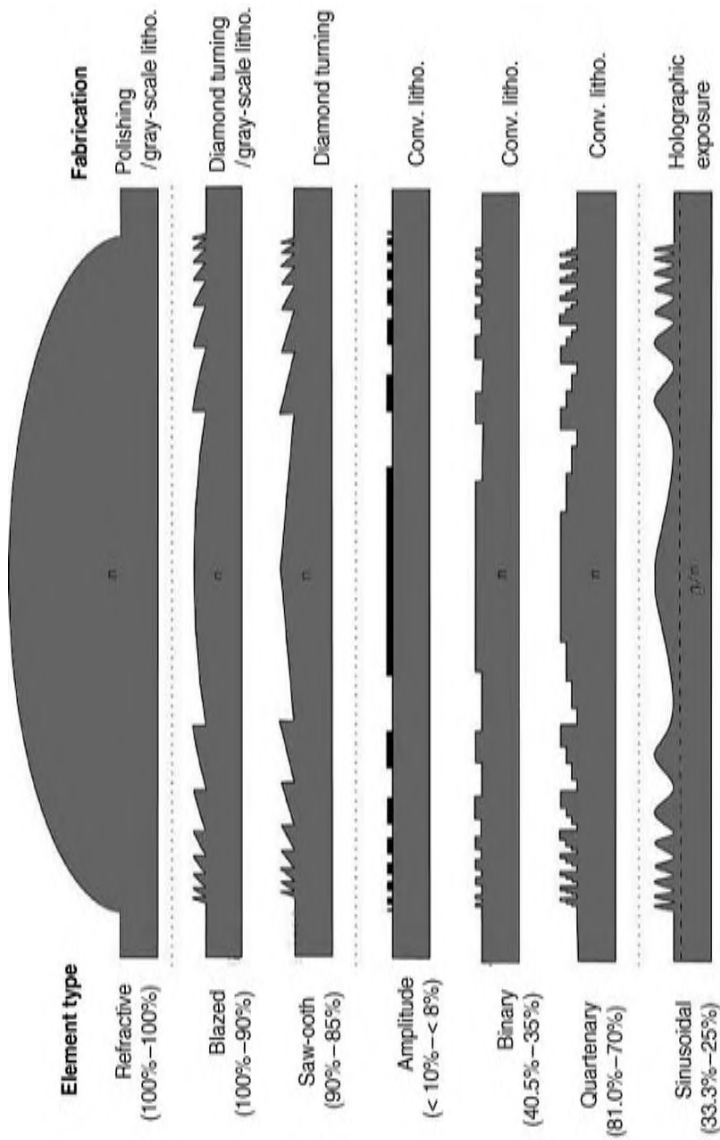


Figure 12.2 Different physical encoding schemes for a same Fresnel lens

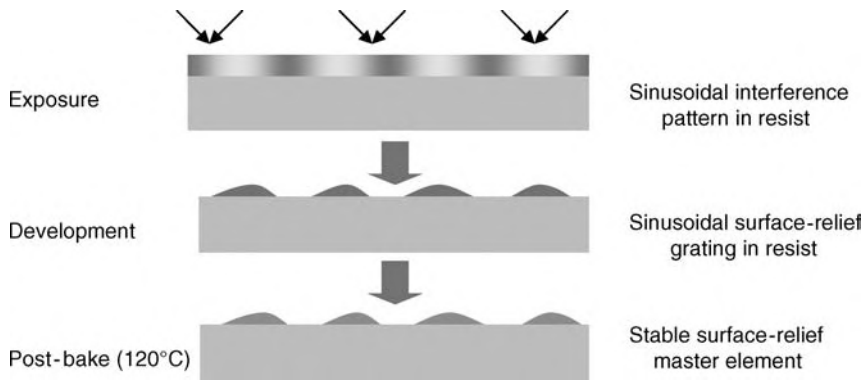


**Figure 12.3** Holographic exposure of photoresist

example in Figure 12.5 shows the origination of a blazed grating that can produce high diffraction efficiency in the scalar regime, although such an element is considered as a thin hologram.

Moreover, 3D holographic exposure can produce 3D periodic structures in thick photoresist (SU-8 resist, for example). Such 3D structures can implement photonic crystals or metamaterials (for more details on these issues, see Chapter 11).

## 12.2 Diamond Tool Machining

### 12.2.1 Diamond Turning

Diamond turning via a computer-controlled high-resolution CNC lathe can be a very effective way to produce high-quality efficient diffractives in a wide range of materials (plastics, glass, metals, ZnSe, Ge etc.; see Figure 12.6).

However, such elements are limited to circularly symmetric elements such as on-axis symmetric lenses, with smallest fringes several times greater than the diamond tool. They can be rather large (Fresnel reflectors) or relatively small (Figure 12.7 shows an example of a hybrid refractive/diffractive lens turned in styrene plastic, with a diameter of 2.5 mm and a sag of 300  $\mu\text{m}$ , to be used as a hybrid dual-focus OPU lens in a CD/DVD drive – see also Chapter 16).

More complex multi-axis CNC lathes can generate complex anamorphic fringes and profiles, as is also performed for their refractive counterparts. Minimum feature sizes (fringe widths) for diamond-turned Fresnel lenses can actually be quite small, below 2  $\mu\text{m}$ . The diamond tool tip combined with the trajectory of that tip controls the local blaze angle. The smallest fringe should have a blaze equal to or larger than the diamond tip wedge. For a blazed grating, it is convenient to have the diamond end tip geometry matching the grating blaze angle (see also Figure 12.8 below). However, due to wear of that tip (especially when using hard materials such as glass), it is necessary to change the tip every now and then in order to prevent blaze angle changes that would affect the diffraction efficiency.

Diamond turning is also a versatile tool to fabricate deep groove grating and broadband or multi-order harmonic diffractive lenses (see Chapter 5), since the depth of the element is not limited to a preset of quantized depths as in standard microlithography, but can reach almost any depth (provided that the tip can reach down and the material can sustain the aspect ratio).

### 12.2.2 Diamond Ruling

Diamond ruling consists of ruling grooves into a hard material, for the implementation of either prism arrays (refractive) or linear gratings. Depending on the diamond tool used, one can fabricate an accurate



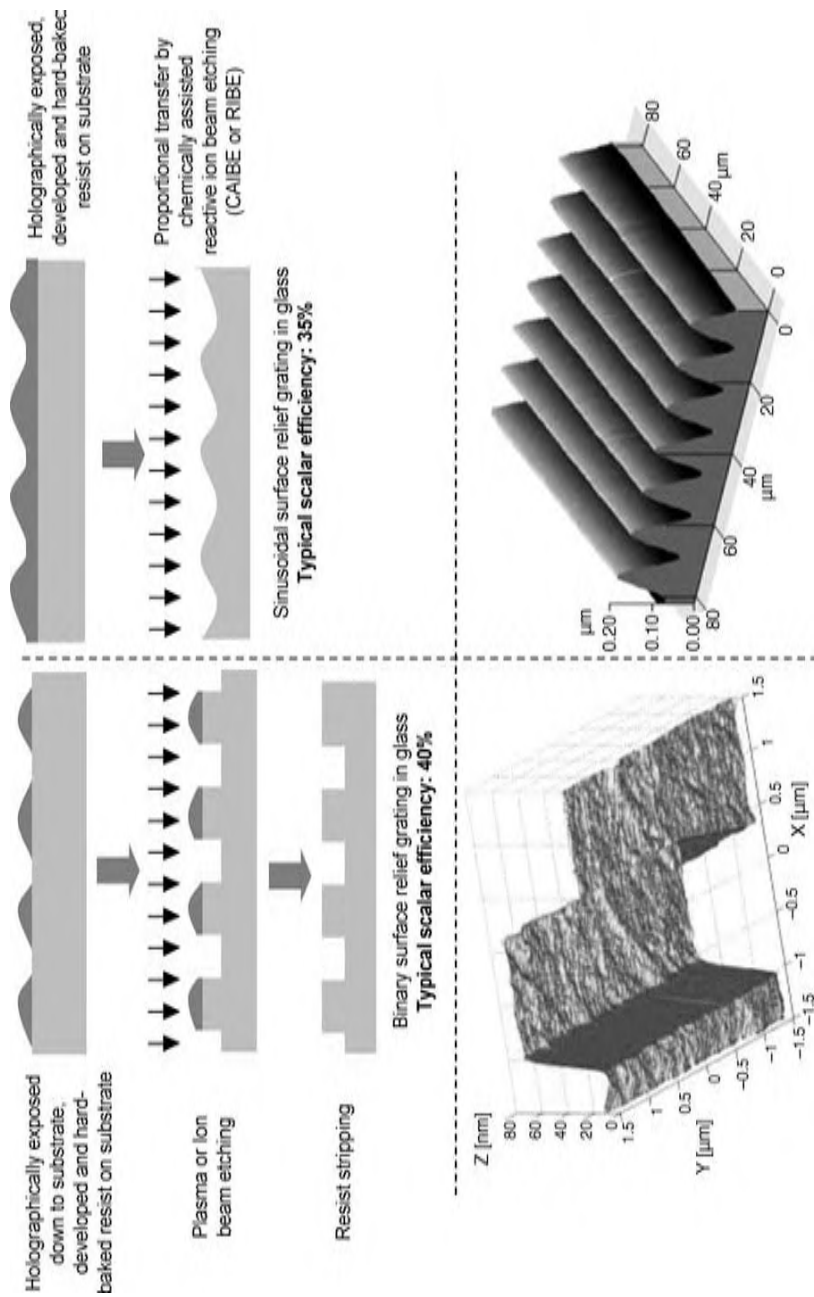


**Figure 12.4** The analog and binary RIE transfer of a hologram in resist

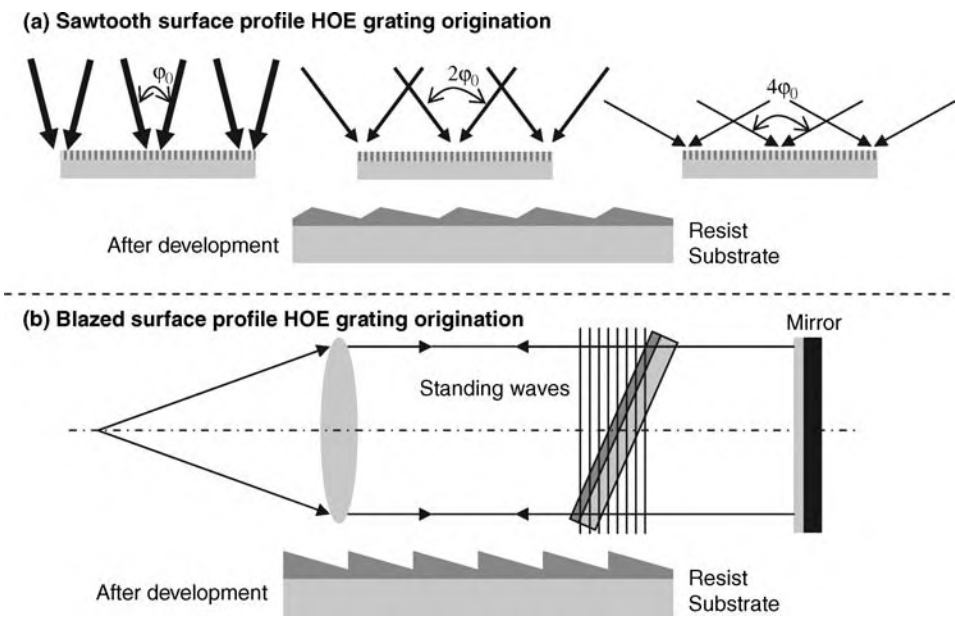


**Figure 12.5** Multiple exposures to produce arbitrary-period structures in photoresist

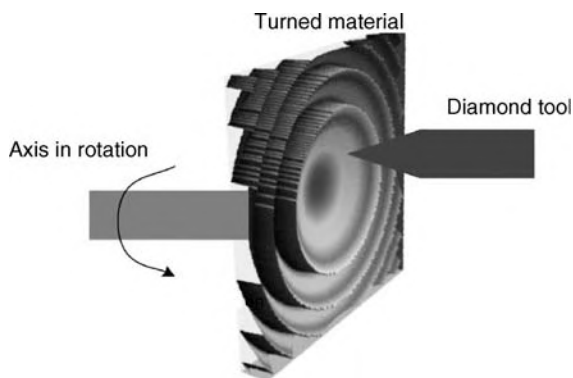


**Figure 12.6** Diamond turning of a spherical blazed Fresnel lens

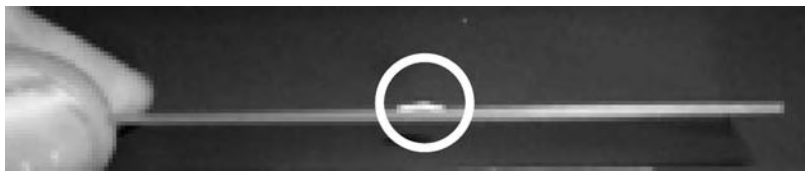


**Figure 12.7** An example of a small hybrid refractive/diffractive dual-focus OPU lens turned out in plastic

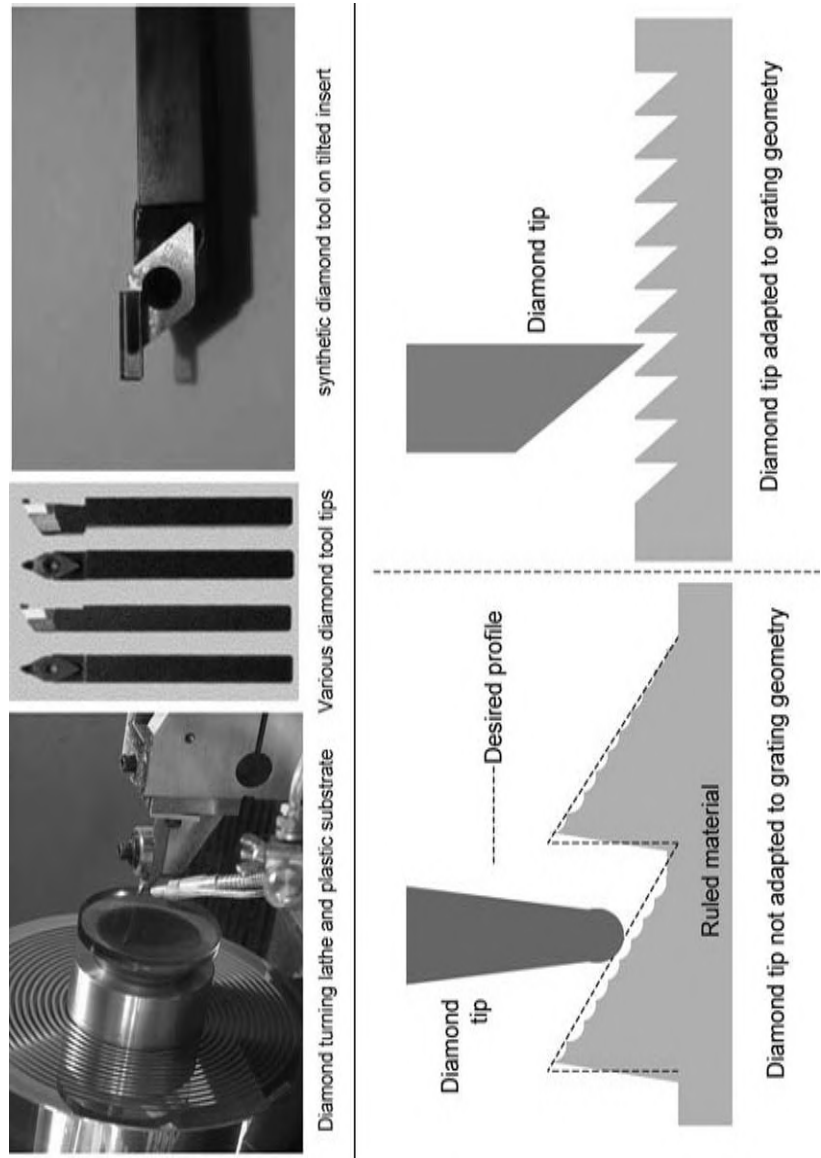


**Figure 12.8** A diamond end tip tool and the resulting fringe geometry

replica of the tool end tip geometry or can use that tool to carve out the desired fringe geometry (see Figure 12.8).

Large-area micro-optics, such as Brightness Enhancing Films (BEF), prism gratings or refractive micro-Fresnel lenses, can also be manufactured efficiently by the use of diamond ruling and diamond turning, respectively. Such BEF films are depicted in Chapter 16.

Diamond-turned masters remain expensive, but are perfect positive molds for subsequent replication by injection molding (see also Chapter 14).

## 12.3 Photo-reduction

Photo-reduction was one of the first techniques used to produce synthetic holograms or CGHs [2, 3]. In the early days of CGHs, the independently derived Lohmann and Burch detour phase encoding methods (see Chapter 6) on an amplitude substrate were perfect candidates to be tried out on  $24 \times 36$  mm slides.

### 12.3.1 $24 \times 36$ mm Slides

In 1967, Lohmann and Paris tried out their first CGH encoding technique on a large piece of paper, blackened with ink (by hand at first and then by a computer-controlled ink plotter). A photograph is taken of this black and white pattern, and transferred onto a  $24 \times 36$  mm transparency slide. The white part is then transparent and the black part is opaque. The photo-reduction can be quite large when transferring a photograph of a  $1 \text{ m} \times 1 \text{ m}$  piece of paper onto a  $24 \times 36$  mm slide ( $40 \times$  reduction). If the photographic slide is of high resolution (a low ASA number), the smallest features can be as small as  $25 \mu\text{m}$ , provided that the original artwork had a resolution of 1 mm.

Several problems arise with such transparencies:

- very low efficiency, due to amplitude encoding and physical amplitude media;
- low contrast in the slide (silver grains), leading to diffusion and random noise; and
- low diffraction angles, due to limited resolution.

Such amplitude-encoded early CGHs are no longer used, since lithographic techniques have been widely distributed for the fabrication of diffractive optics.

### 12.3.2 PostScript Printing

As an alternative to photo-reduction, direct PostScript laser printing on transparencies [3] has been proposed as a cheap desktop prototyping and testing tool for diffractives. Again, here the functionality is binary amplitude. However, as laser printers can print gray scales, gray-scale amplitude is possible with such tools, by either using pulse width modulation, pulse density modulation or other fancy diffusion algorithms to produce gray tones with binary pulses. High-quality direct PostScript printers can print features down to 2400 dpi or lower ( $10 \mu\text{m}$  pulses and lower). Such desktop production methods are very popular among graduate students in Fourier optics, but are not well suited for industry.

One can think about direct PostScript printing as the prehistory of lithography. Actually, half a century ago, the first photomask patterning machine to be used to produce a printed circuit photomask by Intel at that time was graciously produced by the Anheuser-Busch beer label printing contractor in St. Louis, Missouri.

However, the PostScript shape description language is a very powerful tool with which to describe curved and smooth fringes for analytic-type diffractives (lenses, curved gratings, Bézier curves, splines etc.), much more powerful and adapted to our needs than the standard IC industry formats (GDSII etc.).

Such a high-level vector description language has not been integrated with laser or e-beam mask patterning machines up to now. The market for such elements is too small for the microlithography fabrication industry to decide to change their formats, which are adapted to IC features.

This leads us to microlithography fabrication techniques, which are today the major technology used for the production of diffractive optics for industry.

12.4 Microlithographic Fabrication of Digital Optics

The recent and rapid development of digital optics in industry has been made possible thanks to fabrication technologies that have been developed for other fields, such as the Integrated Circuits (IC) manufacturing industry for master element production, the Compact Disk (CD) industry for plastic replication of these masters, or holographic embossing for embossing replication of these same masters.

The following sections review the various microlithographic techniques used today to produce digital optics, namely:

- binary and multilevel lithography;
- direct write lithography; and
- gray-scale lithography.

The various surface-relief digital optical elements that can be produced by these lithographic techniques [4–6] are shown in Figure 12.9, with the example of a simple on-axis Fresnel diffractive lens.

Figure 12.9 also shows the various maximal diffraction efficiencies that one can access when using either binary, multilevel or gray-scale lithographic techniques. Theoretical and practical (verified) efficiency figures are given.

12.4.1 Wafer Materials

Choosing the material on which to perform the lithography is usually the first task on hand when deciding to produce digital optical elements. Figure 12.10 shows some of the most popular wafers used for digital

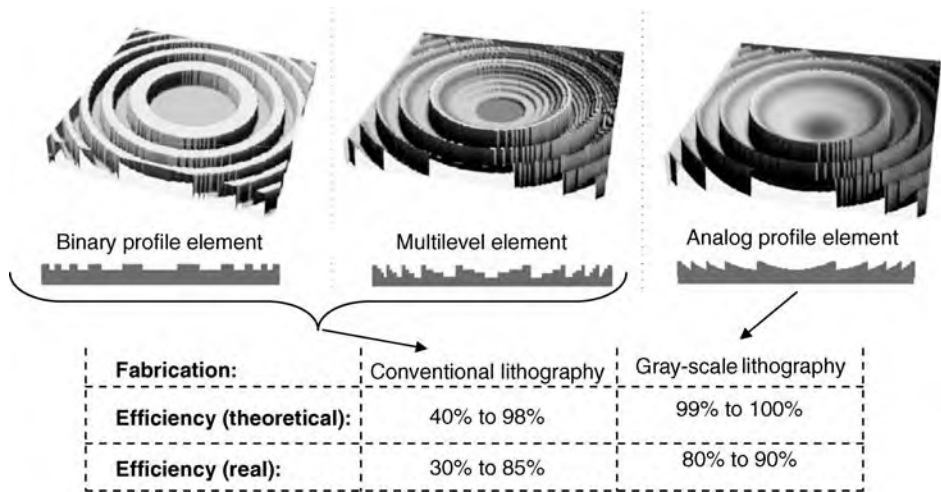


Figure 12.9 Binary, multilevel and analog surface-relief lithographic fabrication (Fresnel lens)

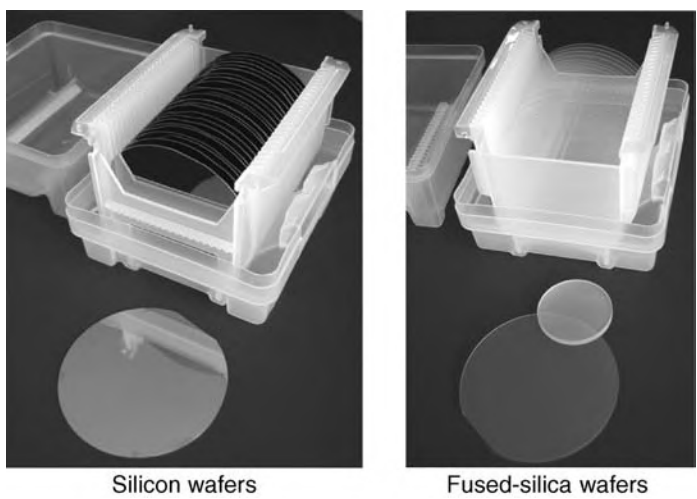


**Figure 12.10** Four-inch diameter silicon and quartz wafer cassettes

optics fabrication: silicon wafers and quartz (or fused silica) wafers. When the final application requires the use of visible or near IR wavelengths, fused silica wafers (quartz) are the best choice.

Silicon wafers (which are cheaper, and the process for which is well understood by the IC industry) can also be used for many purposes in photonics: to produce digital optics to be used in reflection mode (with adequate reflective coating), for digital optics to be used in the IR region, to produce alignment structures such as fiber V-grooves, or to produce master elements to be replicated in plastic. The silicon wafer is then not used as the final element, but as an intermediate element (see Chapter 14). Figure 12.11 shows the transmission graphs of typical UV-grade fused silica wafers used in industry.

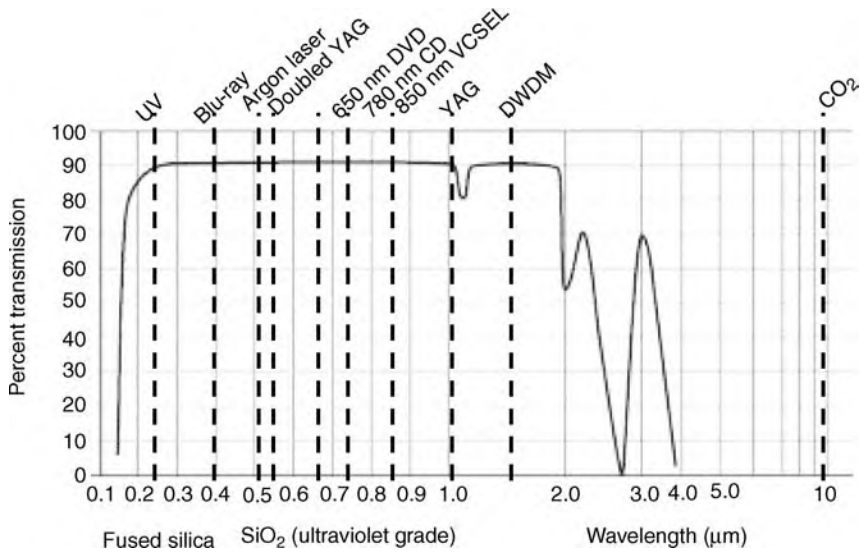


**Figure 12.11** Transmission graphs for UV-grade fused silica

Figure 12.11 shows that such wafers can be used to produce digital optics for a very wide range of wavelengths (from 180 nm UV to 1.8 μm), which covers the standard lithographic applications (excluding DUV or EUV lithographic applications), up to the C and S telecom bands (excluding thermal IR imaging and CO<sub>2</sub> laser light). The standard fused silica wafers have a slightly shorter transmission curve than the UV-grade fused silica, especially in the UV region.

These wafers are usually shipped in cassettes of 20 wafers (and are also processed in batches of 20). There are usually some broad misconceptions in the optics community when attempting to use lithographic fabrication techniques to produce digital optics, one of which is:

*If the processing of a batch of 25 wafers costs X amount of dollars, one wafer should then cost X/25 amount of dollars.*

This is only true if one pays for the batch fabrication with a full number of wafers in the cassette (20 or 25) and then considers each wafer individually (the typical industrial fabrication case). It is not the case if one needs only one wafer all in all (the typical academic prototyping case). In this case, a whole batch process has to be used to produce a single wafer, and then the single wafer will cost as much as 25 wafers.

Many other wafer materials can be used, including the following:

- soda lime wafers;
- Borofloat® and Pyrex® wafers;
- sapphire wafers or very resistant optics in the visible region;
- ZnSe, ZnS or Ge wafers for IR optics; and
- GaAs wafers for PLCs.

However, before choosing an initial wafer material for a target applications, one has to make sure about the following critical points (in addition to the final wavelength transmission plots):

- Is the wafer polished to standard wafer specifications?
- Can the lithography tool see the wafer even though it is transparent (sometimes it needs an additional reflective coating for the litho handling robot to actually see it)?
- Does the wafer have a flat for field alignment?
- Does the wafer have the standard thickness and size for the lithography tool to be used?
- Can the wafer be dry etched (e.g. amorphous glass does not etch as well as quartz, ZnSe produces bad contamination in a standard etching chamber etc.)?

When ordering a batch of wafers, one has to specify various parameters, as described in Table 12.1. The Total Thickness Variation (TTV) is an important parameter, and is usually linked to the target resolution to

**Table 12.1** A standard wafer specifications list

| Specification list                      | units                                                  | 100 mm<br>(4 inch) | 200 mm<br>(6 inch) | 300 mm<br>(8 inch) |
|-----------------------------------------|--------------------------------------------------------|--------------------|--------------------|--------------------|
| Wafer thickness                         | μm                                                     | 300/375/525        | 725                | 775                |
| Thickness tolerance                     | μm                                                     | ±15                | ±30                | ±50                |
| LLS (localized light scattering defect) | μm                                                     | >0.3               | >0.2               | >0.15              |
| TTV (Total thickness variations)        | μm                                                     | <5                 | <4.5               | <4                 |
| Local flatness                          | μm                                                     | <0.4               | <0.3               | <0.2               |
| Surfaces (front/back side)              | Double side polished for transparent single for opaque |                    |                    |                    |
| Scratch/dig (surface quality)           | Standard 60/40 or 80/50 or custom polished 10/5        |                    |                    |                    |
| Flats                                   | Required for multiple field alignment                  |                    |                    |                    |

**Table 12.2** Scratch/dig numbers reflecting wafer surface polish quality

| Scratch/dig<br>reference number | Scratch width<br>(mm) | Maximum dig (or<br>bubble) diam. (mm) | Dig or bubble<br>separation distance (mm) |
|---------------------------------|-----------------------|---------------------------------------|-------------------------------------------|
| 120                             | 0.12                  | 1.20                                  | 20                                        |
| 80                              | 0.08                  | 0.80                                  | 20                                        |
| 60                              | 0.06                  | 0.60                                  | 20                                        |
| 50                              | 0.05                  | 0.50                                  | 20                                        |
| 40                              | 0.04                  | 0.40                                  | 20                                        |
| 30                              | 0.03                  | 0.30                                  | 20                                        |
| 20                              | 0.02                  | 0.20                                  | 20                                        |
| 15                              | 0.015                 | 0.15                                  | 20                                        |
| 10                              | 0.010                 | 0.10                                  | 1                                         |
| 05                              | 0.005                 | 0.05                                  | 1                                         |
| 03                              | 0.003                 | 0.03                                  | 1                                         |

be achieved by the lithographic process. The resolution will be deteriorate if the TTV increases. Localized Scattering Defects (LLS) are usually a critical parameter for IC, but not critical for photonics, especially for free-space digital optics.

When using a transparent wafer for a target transmission operation (no reflective coating), it is always desirable to have both surfaces polished. If only one surface is to be used (a Si wafer or a reflective coating on a quartz wafer), a single polished surface is acceptable. The quality of the surface is usually expressed by means of an associated number called the scratch/dig specification. Table 12.2 explains what the scratch/dig numbers actually mean. The scratch/dig specification is important since any defect on the wafer will affect the uniformity of the resist coating, produce typical comet tail features in the resist uniformity, and thus affect the quality of the lithography over these areas.

*12.4.2 The Multilevel Optical Lithography Process*

Optical microlithography fabrication technologies and techniques are derived from the standard IC industry, and have been applied to the fabrication of diffractives since the mid-1980s [7–10].

These techniques are used to fabricate diffractive optics when either the following conditions cannot be met by the previously described fabrication technologies (diamond turning and holographic exposure):

- the fabrication of arbitrary features (CGHs);
- complex surface-relief structures that cannot result from holographic interference;
- precise spatial multiplexing of various diffractives (planar optical bench, arrays);
- the integration of precise digital microstructures with additional alignment or integration features (fiducials, marks, resolution targets etc.);
- the use of hard materials such as quartz, glass, silicon, sapphire, ZnSe, Ge and so on; and
- the need for mass replication of the master element in these same materials.

Microlithographic fabrication techniques can result in a variety of microstructured phase-relief elements, as depicted in Figure 12.8 (from binary to multilevel to quasi-analog surface relief). In many cases, the diffraction efficiency is the most important criteria to be considered, and therefore a



multilevel or quasi-analog surface relief is often desired rather than a simple binary element, but this would also require a high fabrication budget.

When the optical designer does not have the adequate CAD/CAM tools to generate the required mask layouts, the microlithography fabrication process can become quite cumbersome. The lithographic fabrication process can be split into five major tasks:

1. Generation of the phase profile (via analytic or numeric techniques).
2. Fracture of the resulting fringes into GDSII polygons (which is the IC industry's standard format for a mask layout).
3. Photomask patterning via a laser or e-beam patterning system.
4. Optical lithography, etching and dicing.
5. Potential replication of master elements by embossing or injection molding.

Binary and multilevel lithographic fabrications are now the most widely used diffractive fabrication techniques for digital optics. Although they are not ideal, they are the most widely available techniques. While the binary process is quite simple (one lithography step for two phase levels), the multilevel process is a little more complex in the way it uses  $N$  consecutive field alignments in order to produce more than  $N$  surface-relief steps.

#### 12.4.2.1 Successive Mask Alignments

Figure 12.12 shows this process, where  $2^N$  phase-relief levels can be fabricated using a set of  $N$  binary amplitude photomasks.

For each masking process, there is a resist coating process (including an adhesion promoter), a field alignment process (except for the first one), an exposure process (the taking of the picture) and the development process (including post-bake), an RIE etching process, and a resist strip-off process followed by a cleaning process. The successive etching depths depend on the mask index used. Usually, the lowest mask index has the deepest etching depth.

This is one of the drawbacks of this technique. The multilevel technique is good for low number of phase levels, from 2 to 8 or 16 [11]. A greater number does not make sense, either in terms of performance or in terms of fabrication costs. As will be seen later, there are systematic fabrication errors that arise from each field alignment and subsequent etching. The addition of several of these errors can lead to a severe drop in optical performance (noise, higher orders, ghost orders etc.). If one requires more than 16 phase levels, one should use either direct write techniques or gray-scale lithography (see the following sections).

As an example of a multilevel lithography process, Figure 12.13 depicts the generation of the successive mask layouts for an eight phase level diffractive lens. For such a lens, three successive binary masks ( $8 = 2^3$ ) are generated, the fringes of which are fractured into polygons. These three masks are then used in multilevel lithography to generate the eight phase levels.

It is not always necessary to produce  $N$  sets of physical masks. If one uses only a quadrant of a mask, and then turns the mask  $90^\circ$  for each new lithography process, one can achieve the desired 16 phase levels (see Figure 12.14), in the right order, in a quarter of the wafer. The other three quadrants are unusable, since they do not have the right etch depth orders.

In order to produce the sets of photomasks in Figure 12.13, the optical designer has to produce a set of layout files (also called fabrication files or mask layouts).

#### 12.4.2.2 Optimal Phase Levels for a Given Fabrication Constraint

As the index of the photomask increases, the minimum feature size also decreases, as does the required alignment accuracy (see Table 12.3). In Table 12.3,  $h_r$  and  $h_t$  refer, respectively, to the etching depths for

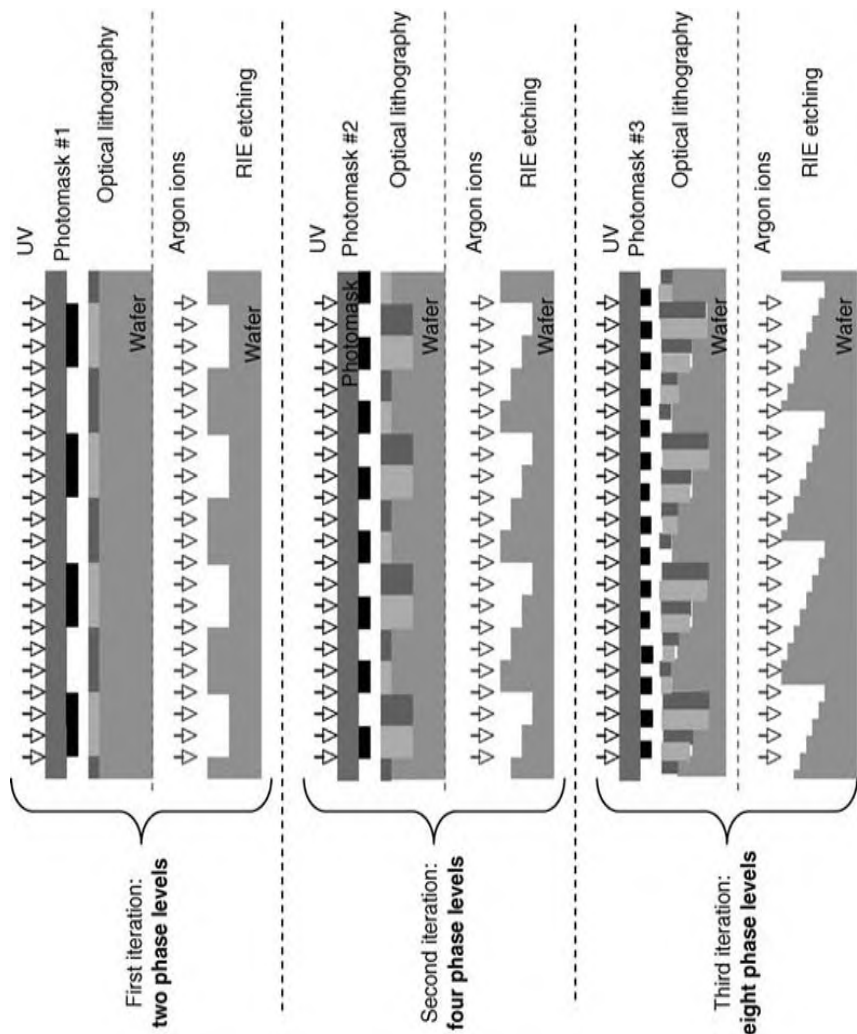


Figure 12.12 Conventional multilevel lithographic fabrication techniques

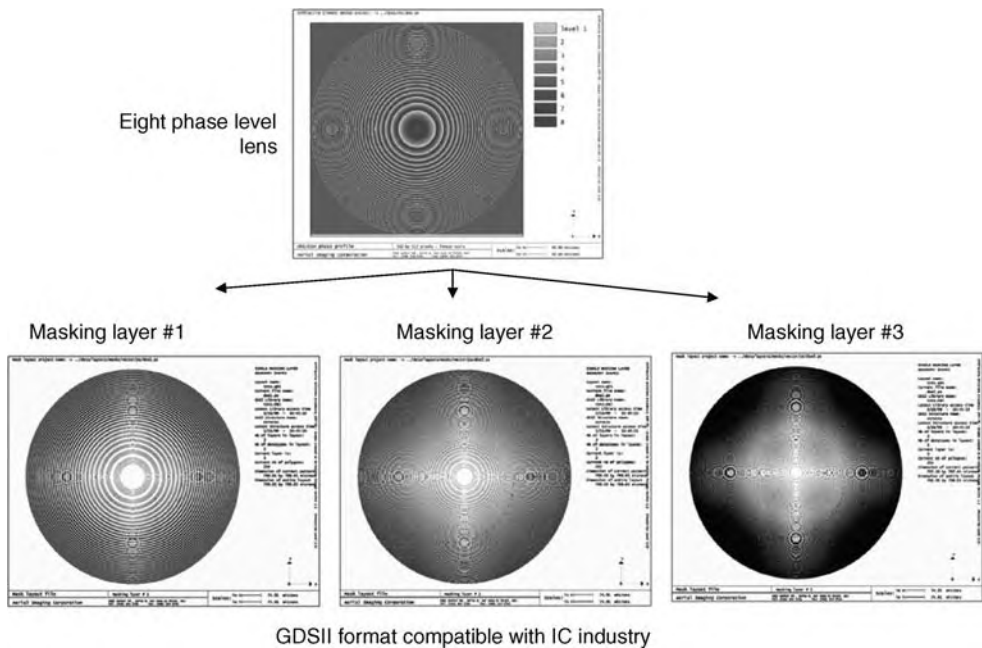


**Figure 12.13** The generation of GDSII layouts (successive masks) from the initial lens description

reflective and transmission operation and  $N$  is the photo-reduction factor in the lithographic projection step ( $N = 1\times, 4\times, 5\times$  or  $10\times$ , depending on the lithography tool to be used).

Table 12.3 shows a standard alignment accuracy of  $\pm 25\%$  for the smallest grating period or fringe width in the current mask layer) and a standard etching depth accuracy of  $\pm 10\%$  (these are the usual specifications in industry). These values are a good rule of thumb when it comes to specifying field alignment and etching tolerances.

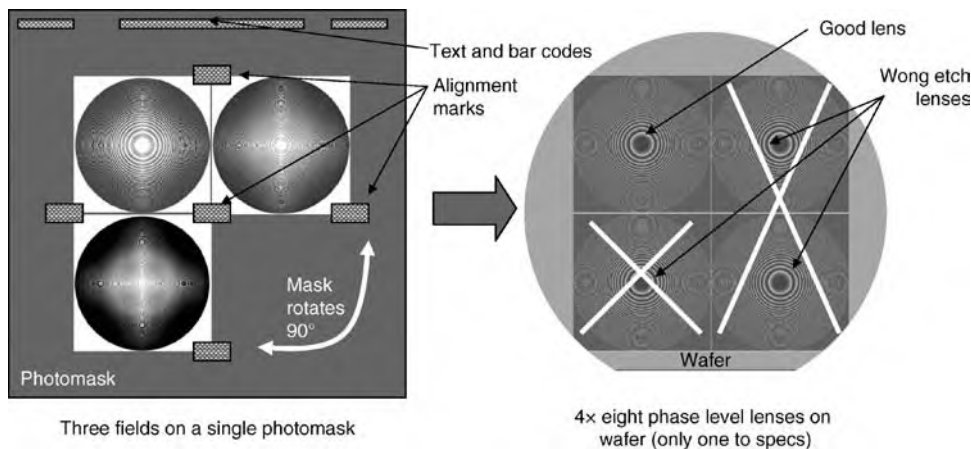


**Figure 12.14** Multiple field alignment in a single physical mask

**Table 12.3** The CD, field alignment accuracy and etching depth accuracy for multilevel fabrication

| Mask index | Lithographic reduction factor ( $N\times$ ) | Number of phase levels | Smallest mask feature (CD) | Smallest feature on wafer | Wafer field alignment accuracy ( $\pm$ ) | RIE etching depth $h_t = \lambda/(n - 1)$ ,<br>$h_r = \lambda/4$ | Etching depth accuracy ( $\pm$ ) |
|------------|---------------------------------------------|------------------------|----------------------------|---------------------------|------------------------------------------|------------------------------------------------------------------|----------------------------------|
| 1          | $N$                                         | 2                      | $N\Delta/2$                | $\Delta/2$                | None                                     | $h/2$                                                            | $h/20$                           |
| 2          |                                             | 4                      | $N\Delta/8$                | $\Delta/8$                | $\Delta/32$                              | $h/4$                                                            | $h/40$                           |
| 3          |                                             | 8                      | $N\Delta/16$               | $\Delta/16$               | $\Delta/64$                              | $h/8$                                                            | $h/80$                           |
| 4          |                                             | 16                     | $N\Delta/32$               | $\Delta/32$               | $\Delta/128$                             | $h/16$                                                           | $h/160$                          |

If one contemplates the possibility of fabricating a transmission diffractive Fresnel lens as a binary, four phase level or even 16 phase level lens, for 650 nm laser light in fused silica ( $n = 1.45$ ), with a minimum fringe width of 12  $\mu\text{m}$  (which lies well within the realm of scalar diffraction – see Chapter 5 for details of how lens parameters are linked to fringe widths), one would need to produce a set of one, two or four masks, respectively. Such a lens would require the following specifications:

- For a binary lens version (40% efficiency):
  - CD is 6.0  $\mu\text{m}$
  - Field alignment: no alignment necessary
  - Etch depth: 722 nm  $\pm$  36 nm
- For a four phase level lens version (80% efficiency):
  - CD is 3.0  $\mu\text{m}$
  - Field alignment:  $\pm 370$  nm
  - Etch depth: 361 nm  $\pm$  18 nm
- For a 16 phase levels lens version (99% efficiency):
  - CD is 0.75  $\mu\text{m}$
  - Field alignment:  $\pm 10$  nm
  - Etch depth: 90 nm  $\pm$  9 nm

One can see that a binary lens version is quite simple to fabricate, whereas a 16 phase level lens requires a submicron mask (at  $1\times$ ), a prohibitive 10 nm alignment accuracy and an etch depth accuracy of 9 nm. Although the etch depth accuracy could be achieved, an alignment accuracy of 10 nm is very difficult. A quadratic version yielding 80% efficiency is thus a good compromise.

A more efficient way to fabricate a lens given a specific fabrication constraint (such as the CD at  $1\times$ ), is to fracture the consecutive fringes (or rings) into the maximum allowed number of steps, to stay within the fabrication constraints (see Figure 12.15). Such a lens can produce 64 phase levels (quasi-analog) in the center, which are then reduced to eight, then to four, and finally to a binary version for the smallest fringes at the edges of the lens. The efficiency would be highest in the center (almost 100%, where most of the light would fall anyway for most applications), and would be lower and lower, down to 40%, at the edges.

Figure 12.16 shows an optical microscope photograph of an array of microlenses that have been fabricated with this technique, in the center over four levels and at the edges over two levels. The individual lenses are coded over a rectangular aperture in order to be packed closely together for a laser array collimation task.

This technique takes care of the minimum feature size, but does not take care of the field accuracy requirements for multilevel fabrication, which still have to be addressed for the central fringes that are fractured over many levels.

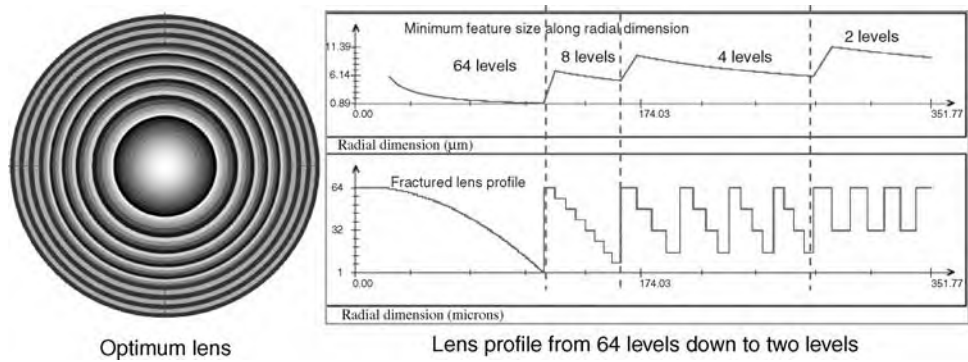


**Figure 12.15** The optimal fracture of a Fresnel lens according to the critical dimensions for fabrication

*An Important Note for Numeric-type Digital Diffractives*

For a numeric-type digital optical element such as a CGH, in which fringes are not defined (see Chapter 6), the CD does not decrease with the number of phase levels as with an analytic element (like the lens described before), intriguing as this might be. A binary CGH can have the same CD as a multilevel CGH, for the exact same optical functionality (provided that the Nyquist criterion is satisfied). Here, the optimal phase level fracture is automatically taken into consideration when one designs the CGH with one of the iterative optimization algorithms listed in Chapter 6. These algorithms take the number of phase levels into consideration and produce the optimal phase map for that specific number of phase levels.

This is one important difference between analytic-type diffractives and numeric-type diffractives when it comes to fabrication considerations.

Once the number of phase levels has been set, in order for the individual masks to be fabricated by conventional mask-generation tools (e-beam or laser beam patterning machines), the data have to be formatted in the standard format used by the IC industry

**12.4.3 Mask Layout Generation**

There are many mask pattern formats developed by the IC industry, depending on the mask-patterning machines to be used. Most of them are not optimum for digital optics, since they have been developed for the IC industry, with IC features in mind. As the digital optics sector is using IC fabs for its own fabrication

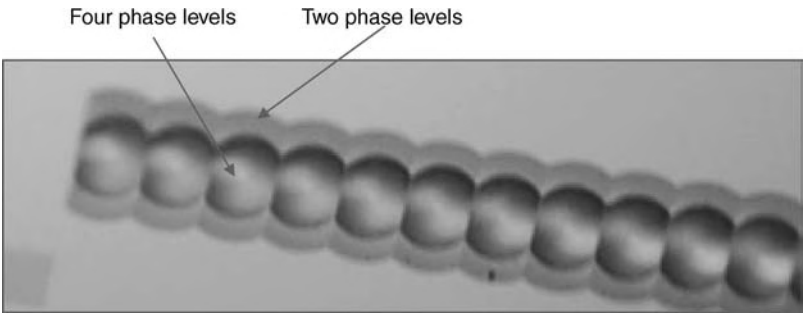


**Figure 12.16** Fresnel lens array fractures with a variable optimal number of phase levels

needs, it has to conform to IC pattern standards. Up to now, there has been no high-throughput mask-patterning machine that has been solely optimized for digital optics other than direct laser write machines in research laboratories and universities.

There are two format levels, the high-level formats – used by the majority of the Electronic Design Automation (EDA) software –and the low-level formats, which are directly readable by the mask-patterning machine.

### 12.4.3.1 Low-level Formats

Usually, the optical engineer would fracture the DOE or CGH into a high-level format such as Gerber Data Stream Information Interchange (GDSII), Caltech Intermediate Format (CIF) or Autocad (DXF), or even PostScript level 2.0 (PS 2.0). However, sooner or later, the CAD system must produce a chip description that can be read by a mask-patterning machine. With today's smaller IC features, it is becoming clear that electron-beam lithography methods are preferable to optical techniques. Within the domain of electron-beam devices, a very popular format is the Electron Beam Exposure System (EBES) developed at Bell Laboratories.

The low-level formats are the ones that best describe the actual and exact shape that will be patterned by the e-beam over the resist layer. There are two types of low-level format: the vector type and the raster type. The vector type is the most desirable, since it reduces the amount of data and speeds up the writing process. This type is used for vector scan mask-pattern generators, such as Cambridge Instrument's EMBF, EBES machines, Lion and ZBA machines, and so on, which are mostly e-beam patterning machines.

EBES formats are readable directly by a number of different mask-patterning systems and this explains why no hierarchy is allowed (see a high-level format such as GDSII, which accepts a hierarchy). Although these machines have computers in them, they spend all their time driving the electron beam and cannot accept a complex input format. In fact, there are only three different figures that can be drawn: rectangles, parallelograms and trapezoids (see Figure 12.17). Everything else must be composed from these figures. Although some machines have allowed extensions to EBES, such as arrays and extended addressing, the original format is the only guaranteed standard.

The major drawback of the EBES format is that it is binary and that the binary word size is restrictively small for modern tasks. With only 16-bit words, a complete chip must be described in a square of only 65 536 units on a side. This used to be a large field but it is easily exceeded today, forcing chips to be described as multiple abutting dies. Another problem with EBES is that the geometry must be sorted so that it is presented to the electron-beam machine in spatial order. This geometry must be clipped into stripes and then all the pieces must be sorted by their  $y$  coordinate (see Figure 12.13). Rounded figures

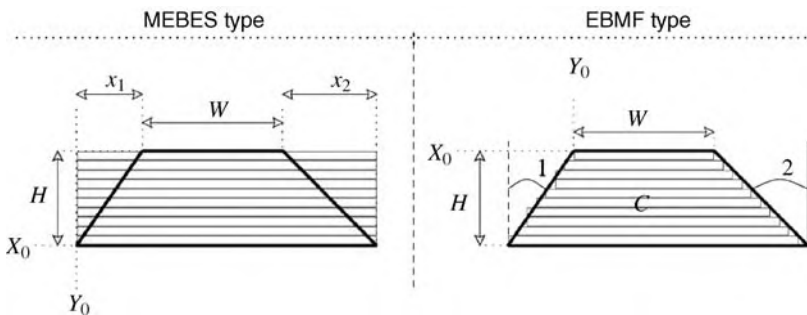


**Figure 12.17** Low-level vector format basic shape specifications

must be approximated by polygons and complex polygons must be broken into four-sided figures. Combined with a lack of hierarchy, these problems cause EBES files to be large and slow to produce, especially for rounded structures, which are widely used in digital optics.

The raster scan type low-level formats are used for JEOL e-beam patterning systems, Micronic laser patterning systems such as the Omega system (.MIC format), the Heidelberg Laser Patterning systems (.LIC formats) and so on. These are mostly laser beam patterning machines.

However, the vector formats can be restrictive in their elementary shape geometries. Besides, they are binary formats and they do not include any language description. Typically, they only allow trapezoidal shapes, with additional restriction to the lateral angles. The basic shapes of the Cambridge Instruments' EBMF and Perkin Elmer's MEBES I and II formats are described in Figure 12.17.

The MEBES type is actually a raster/vector scan patterning machine (it patterns stripe by stripe and blanks its beam when no shape is to be exposed), although the patterns are described in a vector form (trapezoids instead of raster pixels as for JEOL machines); whereas the EBMF is a true vector scan type patterning machine.

The latter goes to the upper left edge of the polygon to pattern, and patterns it in the bosphedron way. After having patterned it, the machine moves to the closest next polygon to be patterned and so on. The number of addressing excels (or pixels) on the e-beam grid is defined as follows for the MEBES raster scan along stripes:

$$\begin{cases} 0 < X_0 < 2^{15} & \text{pixels} \\ 0 < Y_0 < 2^8 & \text{pixels} \end{cases} \quad (12.1)$$

and for the EBMF vector scan within a single main EBPB field as

$$0 < X_0, Y_0 < 2^{15} \quad \text{pixels} \quad (12.2)$$

The MEBES machines can pattern any lateral angle in the trapezoidal description (limited, however, by the e-beam addressing grid, i.e. the excel size), whereas EBPB machines have a limited choice of seven angles:

$$\alpha_1, \alpha_2 = -63^\circ, -45^\circ, -27^\circ, 0^\circ, 27^\circ, 45^\circ \quad \text{and} \quad 63^\circ \quad (12.3)$$

Next, the MEBES file format cannot store any information about the e-beam clock rate (i.e. the e-beam dosage), unlike the EBMF format (parameter  $C$  in Figure 12.13):

$$C = 0, 1, \dots, 15 \quad (12.4)$$

Each of the 16 different values of  $C$  refers to a look-up table defined previously by experimental characterization of the resist depth versus the e-beam dosage, which is especially interesting for either direct write or gray-scale analog write. Therefore, it is very difficult to perform a direct analog e-beam write with MEBES type machines: one would have to use multiple patterning techniques instead of the direct analog write.

Typical values of the data length for the description of a trapezoid (in EBMF format) are 64 bits (32 bits for the specification of the upper left-hand corner, and 32 bits for the height, width, angles, writing clock and inversion parameters of the actual trapezoid).

These low-level mask formats introduce a tremendous amount of information, most of it not adapted to curved patterns (i.e. micro-optics), and thus dramatically increase the file size. On the other hand, for the description of a very high density SDRAM module, for instance, the file size would remain low, since the data format has been optimized for IC patterns only. This builds up a strong case for optimum intermediate and low-level fracturing algorithms that decrease the final file size by keeping a moderate fracturing CPU time.

### 12.4.3.2 High-level Formats

High-level formats can be quite versatile: for example, in PS format, a circular shape can be described by its center position and its radius. Low-level formats, on the contrary, can be very rigid: for example, an EBMF format is composed of trapezoids, with a limited choice of only four different angles on each side. However, each of them are optimized for the IC industry; in other words, for patterning integrated circuits. The GDSII format is the most widely used in industry, either for IC fabrication or fabrication of micro-optical elements.

High-level mask description formats include the best-selling file formats in the microlithographic industry: the technology-independent formats such as GDSII and CIF. Note that DXF are sometimes also considered as intermediate file formats (as are the Gerber, MANN and associated formats).

The Caltech Intermediate Format (CIF) is a simple method for describing integrated circuits. The goals of CIF are to be concise, to be readable by humans and machines and, most of all, to remain constant so that the method can be relied on as a standard. Its authors have allowed 'user extensions', but do not permit any change in the basic commands.

A CIF file is a set of textual commands to set layers and draw on them. CIF describes only the geometry that is found on IC masks: rectangles, polygons, wires and circular pads. There are no higher-level graphics operations such as text or curves. There is hierarchy, however, which makes CIF easy to generate, and it can be used to build complex graphics. In addition, the language is structured to allow multiple CIF files to be easily aggregated into a single larger file, so that multi-project chips can be specified. The only drawback of CIF is that it is not readable by any mask-making machines and thus must be further processed before it can be used in fabrication.

The Calma GDSII Stream Format is older than CIF and is more widely used. Its early entry on the market caused subsequent vendors to accept it for compatibility reasons; thus it became a standard. In addition to representing mask geometry, it also has facilities for topological information and other arbitrary attributes. The reason for this is that GDSII is the complete database representation for Calma CAD systems, and not just the output specification for mask-making. Thus it includes text, arrays, structures and hierarchy. GDSII has changed over the years as new constructs have been introduced. Nevertheless, backward compatibility has been maintained, which has allowed the format to remain the industry standard. Like CIF, this is an intermediate format that must be converted before it can be made into a mask. The other drawback of GDSII is that it is binary, which makes it unreadable by humans. Some would argue that these formats are not meant for humans and that the binary format provides more compaction of data. However, the CAD programmer does appreciate a readable format and, as far as compaction is concerned, a concise text language such as CIF does not use much more space.

GDSII and CIF are moderately restrictive in their data description since, for example, GDSII can only describe line segments, squares, rectangles and polygonal structures that have up to 200 edges. In any case, these file formats need further processing by technology-dependent compilers that are not optimized for DOE fabrication, but rather for IC fabrication.

This said, there is hope on the horizon, with new formats that are going to become mainstream in the IC industry, and that will overcome the GDSII format constraints, including:

- 32-bit shape definition;
- 32-bit integers;
- and 256 layers.

Such new formats are also eagerly awaited, since the data sizes of masks are increasing in volume, mainly due to the aggressive Optical Proximity Correction (OPC) techniques that are used by the IC industry today (see Chapter 13) to reduce the printable feature on the wafer below the Rayleigh resolution limit without investing in new lithographic hardware.



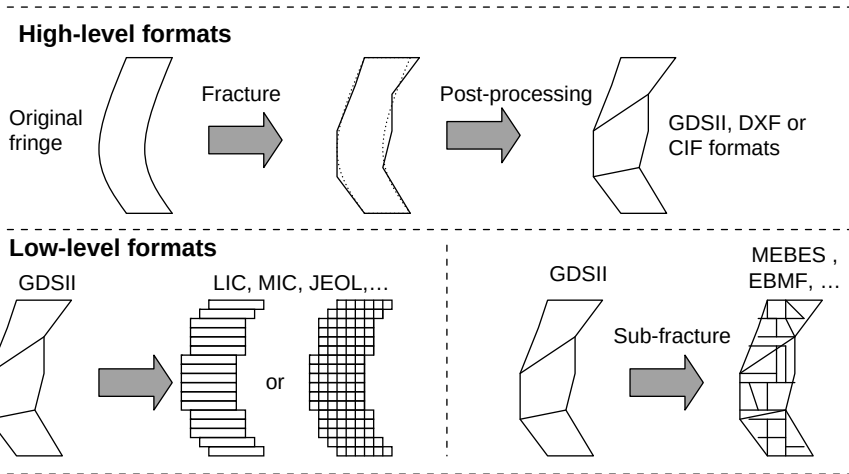


**Figure 12.18** Fracturing smooth fringes into basic lithographic shapes for mask generation

For more information on preparing the job-deck file to be sent to the mask shop, including the Process Definition – Manufacturing Instructions (PDMI) document (see Chapter 15).

The new OASIS format [12] is not constrained by the older GDSII format limitations, but the current EDA tools available in industry still are. It will take some time for current EDA tools to be adapted to OASIS. But this is not so critical for digital-optics EDAs, since there are no current EDA tools for micro-optics. Therefore, the first digital optics EDA tools will be OASIS compatible right from the beginning.

### 12.4.3.3 Analytic-type Element Data Fracture

Integrated circuits are made of simple geometrical features, such as squares and rectangles. Diffractive structures, especially diffractive lenses, are made of smooth circular fringes. The optical designer therefore has to fracture smooth fringe shapes into basic geometrical features (see Figure 12.18).

It is important to remain as close as possible to the original fringe shape in order to get the desired optical functionality, but fracturing the fringe into very small polygons dramatically increases the size of the final file. Therefore, there is a compromise between the size of the fabrication file (and thus the mask patterning time and fabrication budget), and the final resolution attained (the quality of the diffracted wavefront).

When fracturing a diffractive element, one also has to consider the size of the writing beam, not just the resolution of the constitutive polygons. Sometimes, the writing spot is not a spot but, rather, an imaged shape (in variable-shaped e-beam writers).

When considering the fracture of an analytic element such as a lens, what one should never do is to fracture an analytic lens in a raster way (rasterize the phase profile as a pixelated image). This error is done over and over in industry, and the resulting effect is disastrous, with the lens not performing as planned. When the lens is modeled with a Fresnel propagator as described in Chapter 11, without using the oversampling factor, one can fool oneself very easily into thinking that a raster lens would perform well (see Figure 12.19).

Figure 12.19 shows that a rasterized lens gives rise to many artifacts in the lens (moiré fringes) and its reconstruction plane, although the pixel used for rasterizing the lens is smaller than the smallest fringe in the lens. The effects on the reconstruction are considerable deterioration of the Strehl ratio, to which can be added the quantization noise, high-frequency noise and so on. On the other hand, a vector fractured lens

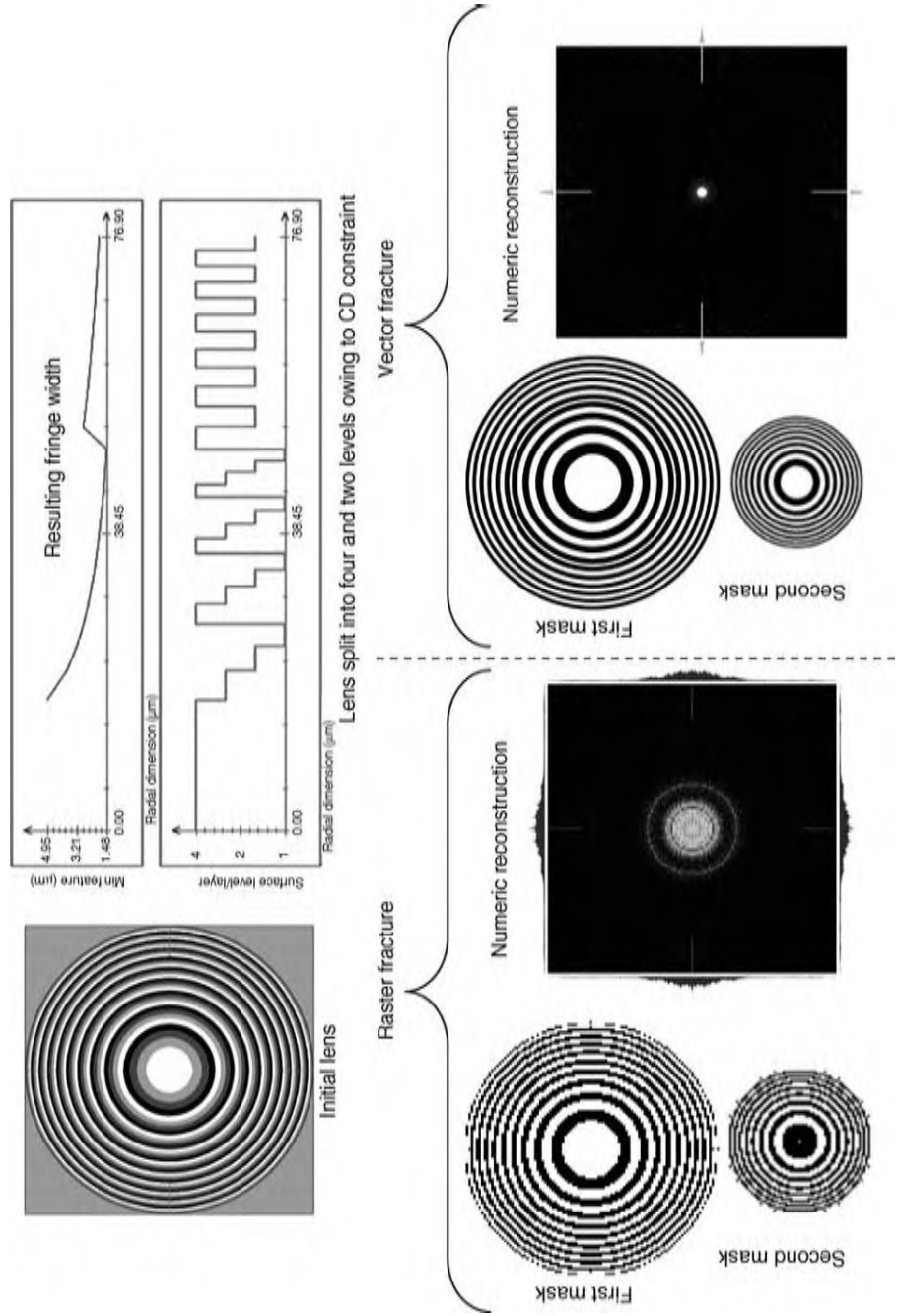


**Figure 12.19** The effects of fracturing a lens into raster or vector formats

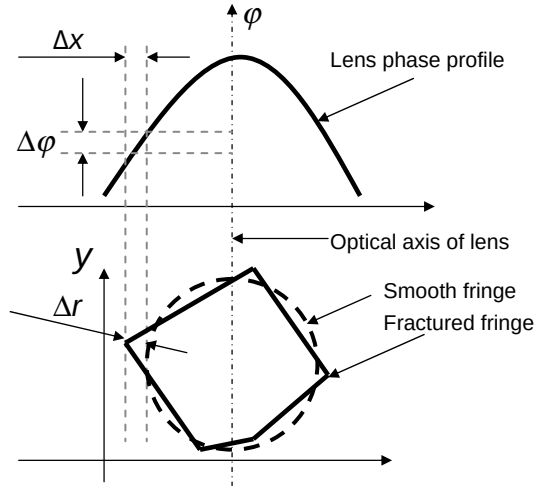


**Figure 12.20** The relation between fracture and local phase error

gives good results. Although the raster method is the way to go for numeric-type elements (see CGH fracture in the next section), it is definitely not the way to go for analytic-type elements such as lenses, curved gratings and so on.

It is important to understand that both the vector and raster fractures have been performed with the exact same CD in mind, the CD constraint being limited by the target mask patterning system to be used. The following paragraphs show how a lens can be fractured in a vector way.

When fracturing a diffractive lens fringe, a curved grating fringe or a micro-optical curved shape, one has to do this according to the largest allowed phase error for such a lens. When fracturing such shapes, a lateral shift between the smooth fringe and the segment describing that fringe locally is equivalent to a phase error, as depicted in Figure 12.20. In the figure,  $\Delta\varphi$  is the local phase error, which is linked to the local fringe fracture error  $\Delta r$ .

When fracturing a lens (or any other optical element), the fracture should be performed according to the following condition:

$$|\Delta\varphi| < \Delta\varphi_{\max} \quad \text{where} \quad \Delta\varphi_{\max} = \frac{2\pi}{N} \quad \text{or} \quad \left( \Delta\varphi_{\max} \Leftrightarrow \frac{\lambda}{N} \right) \quad (12.5)$$

The maximum phase error in the waves is set to  $\lambda/N$  by Equation (12.5).

The lens fracture task on hand is thus to proceed to fracture all the fringes while producing a minimum number of segments, and while not yielding any phase error greater than  $\varphi_{\max}$ , this being in order to reduce the file size, which can quickly become prohibitive for large lenses (several Gb of GDSII data). Thus, the segment length has to be modulated for each fringe. Figure 12.21 shows a lens that has been fractured with phase errors of  $\lambda/2$ ,  $\lambda/5$ ,  $\lambda/10$  and  $\lambda/20$ .

One can see in Figure 12.21 that a lens with a maximum fracture of  $\lambda/3$  or even  $\lambda/5$  does not give good (visual) results, especially in the central rings. The central rings will always be coarser than the outer rings, since the local phase slope in the central region is nearly flat (the center of the lens), and therefore a segment can be located far away from the original fringe and still give a good phase error.

When fracturing a lens, the maximum phase error  $\varphi_0$  has to be carefully chosen, in regard to the optical functionality (phase error) but also in regard to the final file size. Indeed, the resulting file size grows exponentially, proportional to the inverse of the phase error (see Figure 12.22).

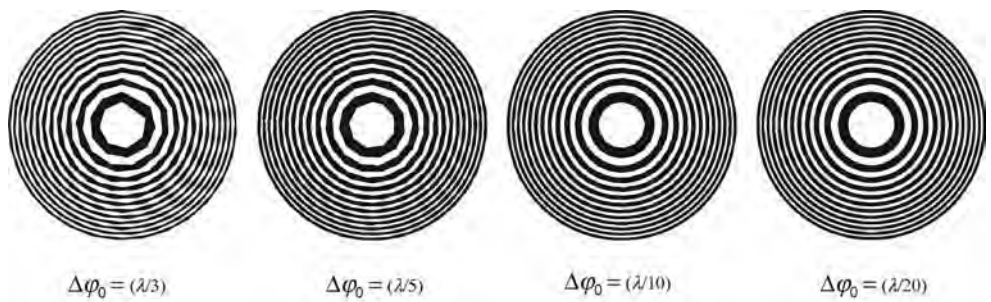


**Figure 12.21** Lens fracture with increasing phase errors

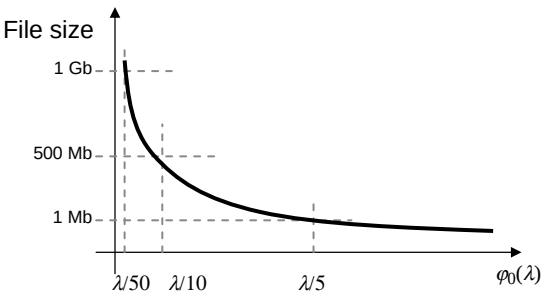


**Figure 12.22** The GDSII file size as a function of the maximum phase error during fracture for a diffractive lens

For a given phase error, the fringes are fractured into their own sub-constituents, which are the various masks for a given fringe. Figure 12.23 shows the resulting fracture of the lens described in Figure 12.19. Thus, the fracture process for an analytic element such as a lens is a critical process that dictates the efficiency of such a lens when fabricated.

These previous fracture considerations were targeted for the high-level fracture process into the GDSII, CIF (or even OASIS) formats, which is performed outside the mask shop, on a CAD EDA tool for digital optics. There is another fracture process that is performed in the mask shop and is dependent on the machine to be used. This fracture process is not part of the lens design process but, rather, part of the PDMI: therefore, it is presented in Chapter 15 (see Section 15.2.1).

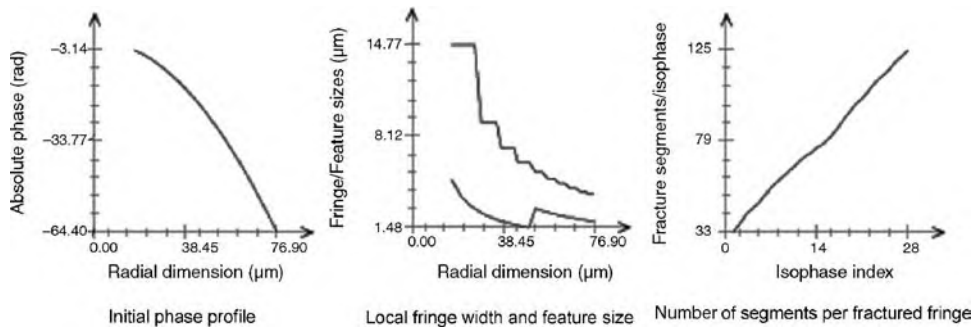


**Figure 12.23** The resulting fracture data for a multilevel lens with a CD constraint

12.4.3.4 Numeric-type Element Data Fracture

The fracture process for numeric-type elements is much easier, since numeric-type elements such as CGHs are already optimized at the very beginning as rasterized elements for rectangular or square cells. The cells, as well as the quantized number of phases, are already taken into account in the iterative optimization process. Thus, a numeric-type element (a CGH) can be fractured simply as a raster element, as opposed to analytic-type elements, which have to be fractured in a vector format (see the previous section). However, there is one major drawback for such a fracture: the file size!

Let us consider a typical CGH size of  $1024 \times 1024$  pixels. The file size for a  $1024 \times 1024$  JPEG image can be as low as 0.05 Mb in gray-scale format, and 1 Mb in uncompressed BMP format. However, a  $1024 \times 1024$  array of cells in GDSII format is 32 Mb! These numbers are for a binary fabrication (only one mask). If a 16 phase level element is desired, the JPG file would still be around 0.05 Mb, but the equivalent GDSII mask file description file can be as high as 100 Mb (about 500 times the size of the original JPG file). If several tens of such elements are to be inserted in the mask layout, the size of the layout can thus jump to several tens of Gb! For most mask shops, a conventional mask layout file size is around 150 Mb for an average writing time of about 45 minutes (which also depends on the desired fracture grid and resolution needed). If the mask layout data size is much higher, therefore yielding longer writing times, the additional effort is usually invoiced to the customer as extra charges. Unlike an IC mask layout, an optical mask layout is usually flat (no hierarchy), thus increasing the file size and yielding prohibitive writing charges (data size = writing time = \$\$\$). Now, we will see that there are several ways to decrease such a file size:

- The first technique would be to use the hierarchy in the GDSII file, as one would do to place transistors in an IC file (by defining a library of shapes and then placing them). Unfortunately, a CGH is always a custom element, and nothing is replicated, other than the basic square or rectangular pixel.
- The second technique is to reduce the number of CGHs on the layout and if these are to be patterned twice on the mask, use a stepper to step them twice during lithography.
- The third technique might be the most effective one, and is based on merging contiguous cells together in x, in y or in both directions.

Figure 12.24 shows a simple CGH example in which the cells are merged into various geometries in order to reduce the overall GDSII file size.

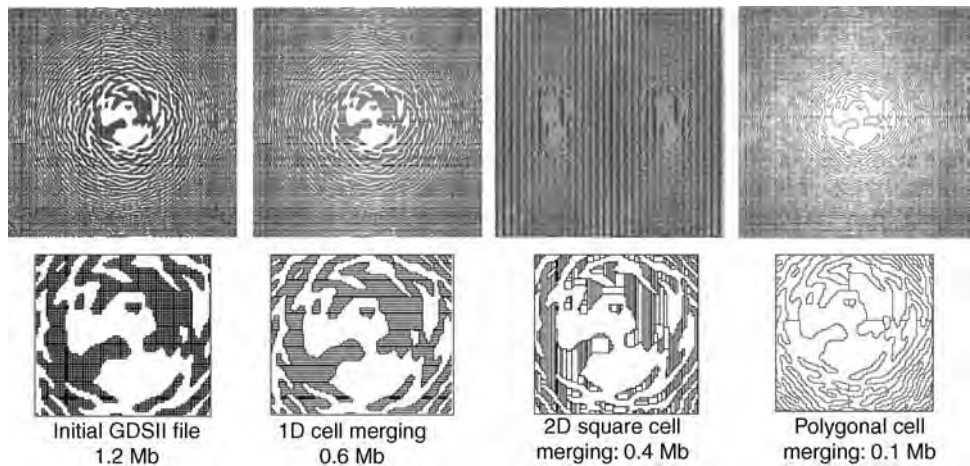


Figure 12.24 The merging of CGH cells in GDSII format to reduce the final file size



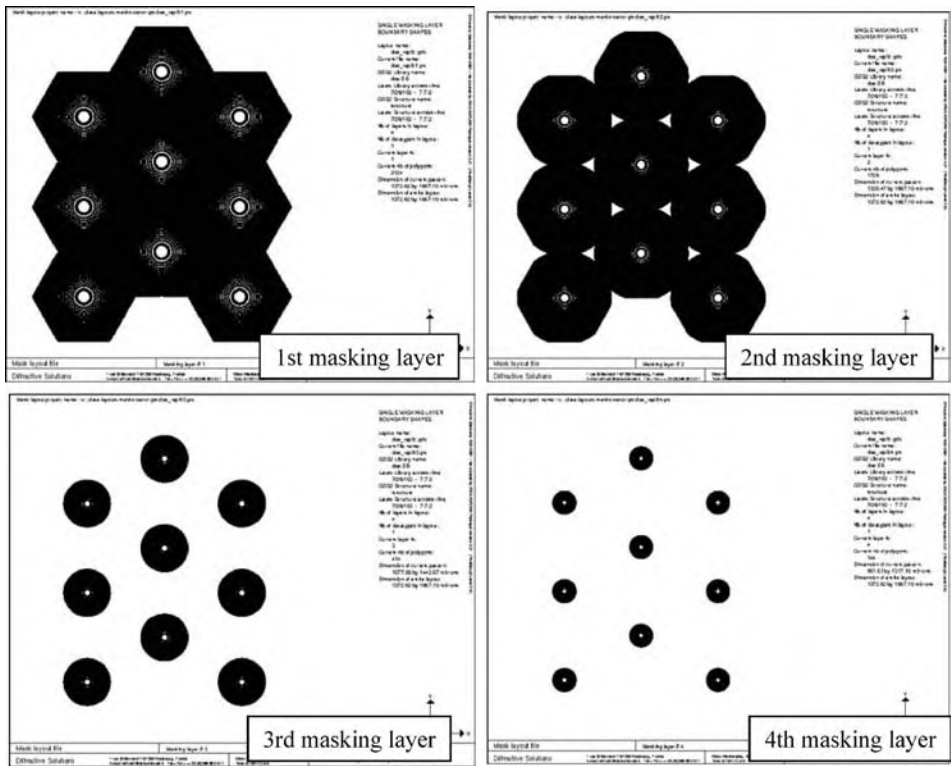


**Figure 12.26** The first four layers of an array of hexagonal close-packed lenses for beam homogenization

which is why the consecutive masks show a smaller and smaller lens aperture. The center lens is thus encoded over 16 levels but the edges only over two levels.

Such a lens is cut into a hexagonal shape in order to replicate it in a hexagonal close-packed array with a 100% fill factor, in order to produce a fly’s eye array for beam homogenization (Figure 12.26).

A photograph of the final fabricated lens array is shown in Figure 12.27.

Such arrays are used in UV laser beam homogenization for steppers, in light engines for digital projectors, in Shack–Hartmann wavefront sensors and so on (for more applications of lens arrays, see Chapter 16).

### 12.4.4 Mask Patterning (Tooling)

Once the mask layouts are generated, in one or more GDSII layers, for binary or multilevel fabrication, respectively, one has to specify the additional features in order to get the mask shop to fabricate the right mask for the right tools and the right process down the road (see also Section 15.2.1).

#### 12.4.4.1 Photomask Terminology

A photomask is basically an amplitude stop that incorporates the fractured data to be imaged onto the wafer. The photomask is not the final element; rather, it is a tool to produce the final elements on the

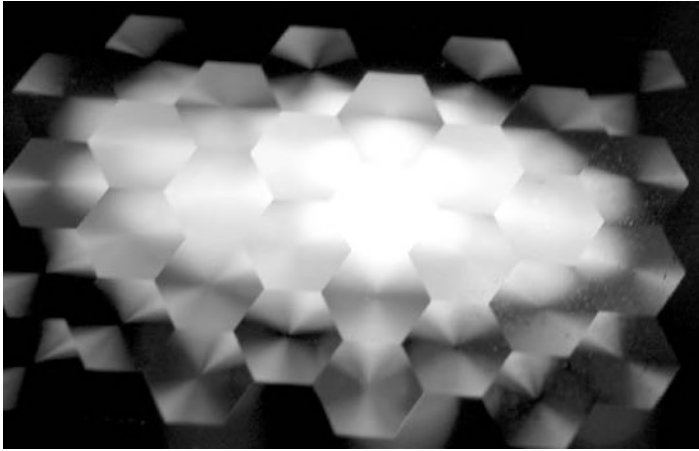


**Figure 12.27** A hexagonal close-packed array fabricated in fused silica wafers

wafers. This is why photomasks are also referred in IC fabs as the ‘tooling’. Usually, for a given job, there is a set of ‘toolings’ for the various fields to be aligned.

Let us first define below some of the lingo used in mask shops, since such IC fab language is usually Chinese or Greek to optical engineers who are accustomed to conventional optical fabrication specifications. Refer also to Chapter 15 for the various terms to define how a mask job should be processed.

#### ***Substrate and Coating***

The physical implementation of a photomask is usually a chrome pattern on a transparent substrate (the substrate can be soda lime, which is cheaper, or quartz, which is more expensive), depending on the quality required. When multiple field alignment is required with tight alignment accuracy (as is usually required in digital optics), it is best to use quartz, since the temperature expansion coefficient of quartz is very low, and thus would produce very small variations in size if the temperature increased between lithographic exposures.

High-reflectivity Cr (about 2000 Å thick) is used for the vast majority of masks. However, other materials can be used depending on the projection wavelength (an Au layer for X-ray lithography, for example). Section 12.7 describes how exotic materials can be used to implement gray-scale photomasks (HEBS etc.).

#### ***Reticle or Photomask?***

There are many different photomask types, and many names. People generally speak about ‘tooling’ or masks or photomasks or reticles, depending on the lithography job. ‘Tooling’ is a general term. ‘Photomask’ is a term that usually refers to a  $1\times$  mask, usually for mask aligners (contact, proximity or projection). One could also speak about a  $1\times$  master.

A ‘reticle’ is usually a mask used in projection lithography tools, with a specific reduction factor, such as stepper and scanners ( $2\times$ ,  $4\times$ ,  $5\times$  or even  $10\times$  reduction factors). That means that the pattern written on the reticle is 2, 4, 5 or 10 times larger than the final pattern on the wafer. The  $5\times$  reduction projection is the most used in industry. There are also other masks especially for the Ultratech stepper, which are called UT1x reticles and are  $1\times$  masks, usually organized in rectangular fields rather than arrays.

There are pros and cons of choosing a reticle rather than a  $1\times$  master photomask: the good news is that the CD to be patterned on the reticle is much larger than the final CD, therefore reducing the costs of patterning the mask. The bad news is that the real estate on the final wafer is much reduced, generally



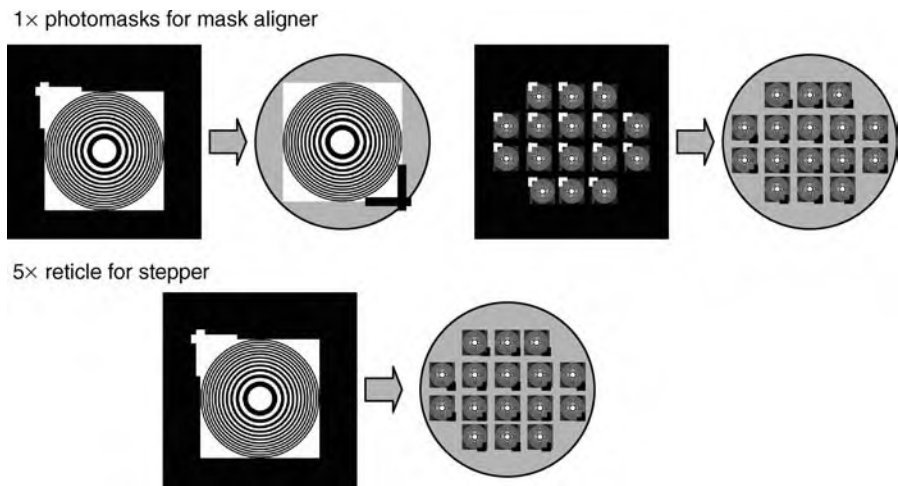


Figure 12.28 The various toolings used in the IC industry

to about a 1 square inch maximum. So if one has a large Fresnel lens to cover more than one square inch on a wafer, one would have to use a 1 × lithography tool (mask aligner) rather than a stepper. Figure 12.28 shows the various toolings used in industry today and the resulting pattern after transfer onto the wafer.

One can see in Figure 12.28 that a 1 × mask can either hold a large element or an array of elements. The drawback of having an array on the mask is that the writing time might be longer, and thus the price of the mask higher. However, if one needs a final element that is larger than 1 square inch, one has to go with a 1 × projection tool (a 1 × mask and mask aligner).

For example, when patterning a large AWG (see Chapter 3), which is larger than the stepper field size, one has the choice of stitching several stepper fields with successive masks or using a single mask in a mask aligner. The second choice is usually the cheapest, but does not provide the best resolution. One also has to be careful with systematic stepper field stitching errors.

Figure 12.28 also shows that the patterns are mirror imaged on the wafer. Therefore, it is necessary to mirror image the pattern prior to patterning the photomask. The resolution and the field-to-field alignment accuracy are much better for steppers than for mask aligners (see Section 12.4.5).

In general, the wafer is about 1 inch smaller than the mask. Table 12.4 shows the various mask sizes and the respective wafers used in industry today.

Table 12.4 Standard mask and wafer sizes

| Tooling (mask) | Wafer         | Lithography                 |
|----------------|---------------|-----------------------------|
| 3 inch         | 2 inch        | Mask aligner                |
| 4 inch         | 3 inch        | Mask aligner                |
| <b>5 inch</b>  | <b>4 inch</b> | <b>Mask aligner/stepper</b> |
| 6 inch         | 5 inch        | Mask aligner                |
| <b>7 inch</b>  | <b>6 inch</b> | <b>Mask aligner/stepper</b> |
| 9 inch         | 8 inch        | Stepper/scanner             |
| 13 inch        | 12 inch       | Scanner                     |

Therefore, when using a 5 inch mask, it is very unlikely that one can project the pattern onto anything other than a 4 inch wafer, and so on. Also, it is very unlikely that a mask with alignment marks specific to one stepper can be used in another stepper. Therefore, it is crucial for the optical designer to know the whole underlying fabrication flow before designing the photomask or the reticle. It is, however, possible to include hybrid alignment marks that can be used by more than one stepper (e.g. hybrid ASML/Nikon alignment marks) or more than one mask aligner (e.g. a hybrid Karl Suess/Canon set of mask alignments).

Although 8 and 12 inch lithography tools are available in industry, their use is usually prohibitive for digital optics (cost-wise). For most of the digital optics fabrication performed today, industry uses either 4 inch or 6 inch wafers (see the bold type in Table 12.4). Chapter 15 defines the various ways in which one can specify a mask: the CD, tone, polarity and parity.

**Alignment Fiducials**

Alignment fiducials are critical features of any photomask or reticle, especially if they are to be used in a series (or set) of masks or toolings. Usually, the field alignment is as good as the alignment patterns. If the alignment marks are not drawn correctly, although the lithography machine might have the potential to align tightly, the alignment will not be good. Figure 12.29 shows typical alignment marks that are used in a stepper system (in a Canon FPA 2500i stepper and a Perkin Elmer Censor Micra III stepper).

A reticle usually has two sets of alignment marks: the reticle pre-alignment marks, which are used to align the reticle in the reticle holder, and the field-to-field alignment marks – the field-to-field alignment marks are the most critical (see Figure 12.29).

Usually, each lithography tool has its own sets of specific alignment marks. Such alignment marks can be completed with sets of alignment verniers and alignment analysis features, which can be part of the Process Control Monitor (PCM) dies (see Chapter 15). When using a mask aligner, one can use the standard set of alignment marks used for that specific machine, but one can also choose to include custom sets of alignment marks, since the alignment is usually performed manually and each technician has his (or her) preferred set of marks.

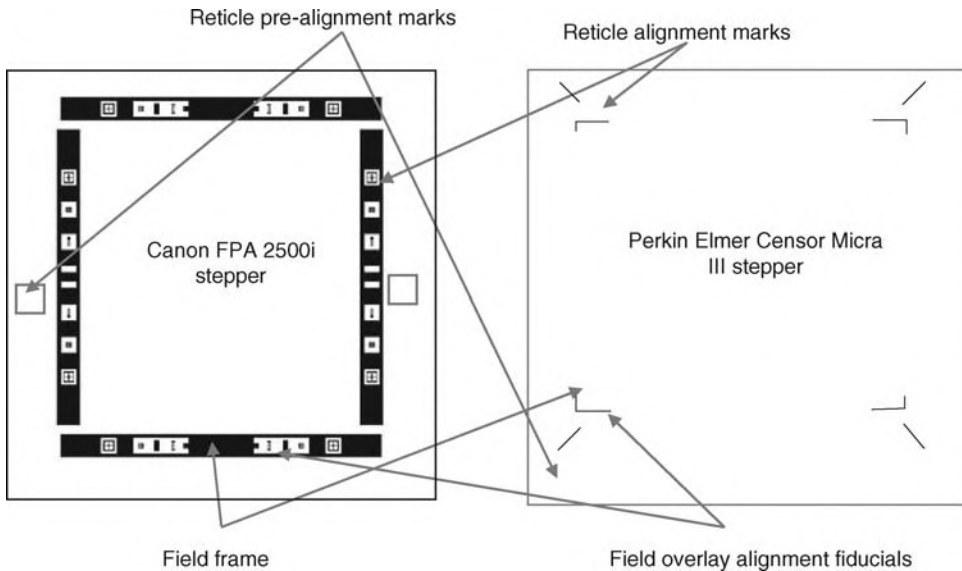


**Figure 12.29** Reticle and field alignment marks

### The Pellicle

A pellicle is a metal frame, over which a thin transparent film (membrane) is stretched, placed on top of the photomask. Pellicles come in many shapes and sizes to accommodate the different array shapes and sizes of photomasks, and with different film types to meet the stepper exposure requirements. The pellicle film keeps contamination out of the focal plane of the wafer lithography system, so that it will not damage the wafer.

#### 12.4.4.2 Laser Beam Pattern Generators

Laser beam pattern generators are used to pattern masks with CDs down to  $0.8\ \mu\text{m}$ . A typical laser beam patterning machine is shown in Figure 12.30.

Such machines include Applied Materials' Alta and Core (formerly Etec) lines of patterning systems, the Sigma and Omega systems from Micronic, the LBWs from Heidelberg Instruments (from the tabletop  $\mu$ -101 system to very wide area writing machines for LCD screens), and a whole list of custom-made laser beam patterning machines in many universities and research institutes. It is interesting to point out that laser beam patterning machines can have their beam intensity modulated easily during the writing process – for example, by the use of an acousto-optical modulator – in order to perform analog direct write (see the following sections). The acousto-optical modulator can also be used to deflect the beam in real time. Also, many laser beam patterning machines use multiple beam writing in parallel (e.g. the Etec Core 2000). Finally, there are also various excimer laser patterning machines and direct write machines used today (see, e.g., Heidelberg's LBA system).

Laser beam writers have three options that are very desirable for micro-optics:

- the potential to write on curved surfaces (for hybrid optics);
- the potential to turn the substrate instead of moving the beam (for diffractive lenses); and
- the potential to modulate the laser beam in real time over 256 levels.

These three features are implemented in many direct write laser beam systems, but are very difficult (or impossible) to implement in e-beam mask patterning systems, which are optimized for IC shapes

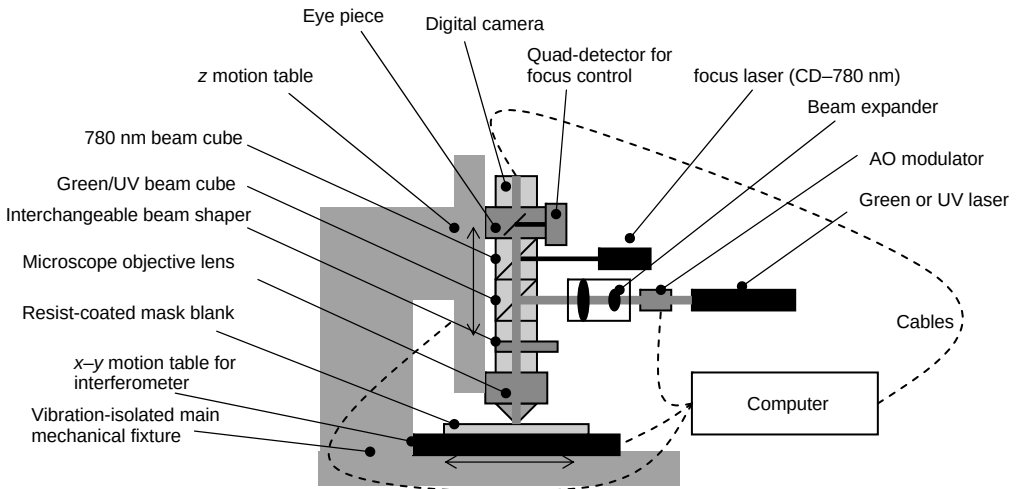


**Figure 12.30** A laser beam mask-writing machine

(binary Cartesian coordinate shapes). The 3D surface writing can be implemented as a CD OPU focus control system.

The second feature (turning the substrate instead of moving the beam) is a very desirable feature for the implementation of perfectly circular shapes (Fresnel lenses, circular gratings, circular waveguides etc.) that do not have to be fractured in vector or raster formats.

Another feature that is usually integrated in laser systems is the conformation of the laser beam through a diffractive beam shaper in the confocal microscope axis, in order to produce in the resist a uniform illumination spot, rather than a sinc, Bessel or Gaussian beam (see Chapter 15).

#### 12.4.4.3 Electron Beam Pattern Generators

When the CD in the mask is below  $0.8\text{ }\mu\text{m}$ , electron-beam patterning machines are used [13]. E-beam patterning machines can focus the spot down to sizes below  $0.1\text{ }\mu\text{m}$  – even down to  $5\text{ nm}$ ! A typical e-beam lithography system is shown in Figure 12.31.

Variable-shaped illumination is especially interesting if these patterns to be printed are circular or curved. In many systems, the variable-shaped exposure is set to be trapezoids, which are suitable for IC features but less so for digital optics. Typical machines are the MEBES, EBMF, JEOL and Hitachi e-beam patterning machines.

However, there are also severe limitations to e-beam patterning, some of which are listed below:

- infrastructure costs (the vacuum chamber and e-beam column);
- the writing time, due to single beam exposure (versus multiple beam exposures with a laser beam);
- the rigidity of the writing architecture and the writing formats (some e-beams are still linked to 20-year-old mainframe computers – VAX machines, IBM clusters etc.); and
- e-beam proximity effects in e-beam resist.

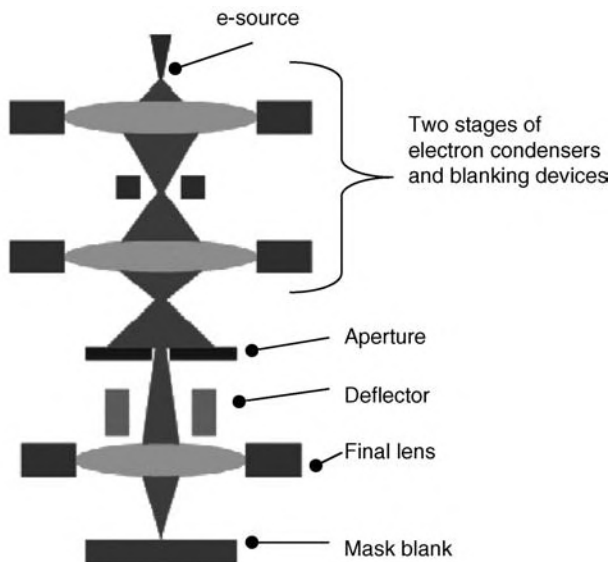


**Figure 12.31** An electron-beam mask patterning machine

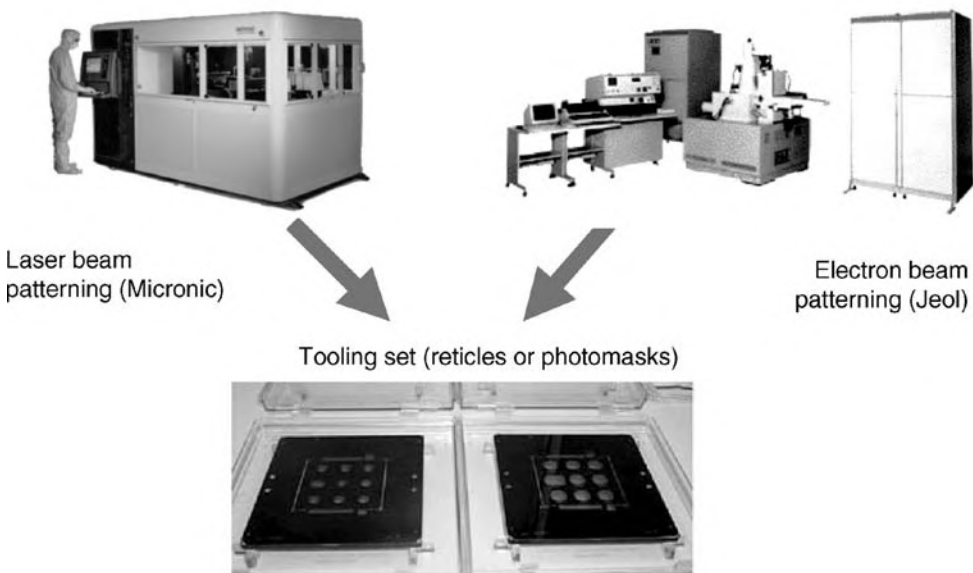


**Figure 12.32** Laser and e-beam mask patterning systems

The first three limitations can be overcome by increasing the writing time and thus also the required budget. The e-beam proximity effects in 2D and 3D are more difficult to address, and need some compensation algorithms (for more details on proximity effects and compensation algorithms such as OPC, see Chapter 15).

Figure 12.32 shows examples of laser beam and e-beam patterning machines and resulting tooling sets. The various optical lithography tools used in industry today, which make use of the previously described reticles or photomasks, will be reviewed below.

*12.4.5 Optical Lithography*

While mask patterning is limited by the direct writing resolution of the laser beam or e-beam (which can be optimized by numerous techniques, including the proximity error corrections listed in the next chapter), the optical lithography is almost always based on imaging.

**12.4.5.1 Projection Imaging Resolution Limits**

There are mainly two ways to perform an optical lithography process:

- contact or proximity alignment; and
- optical projection via a reduction factor ( $1\times$  to  $10\times$ ).

Table 12.5 lists the specifications of each lithographic technique, together with their advantages and limitations. Both are now widely used for diffractive optics fabrication (see also Figure 12.25 for the schematics of the projection architectures).

Figure 12.33 shows the differences between proximity, contact and projection lithography. Contact alignment is often used for low production runs and R&D efforts. It yields better resolution than proximity alignment, but the mask has to be cleaned thoroughly after several exposures due to resist contamination.

**Table 12.5** Contact and projection lithography for multilevel diffractive fabrication

| Technique                                | Contact or proximity lithography       | Projection lithography                                                                                        |
|------------------------------------------|----------------------------------------|---------------------------------------------------------------------------------------------------------------|
| Lithography tool                         | Mask aligner                           | Stepper–scanner                                                                                               |
| Demagnification factor                   | 1×                                     | 1× to 10× (mostly 5×)                                                                                         |
| Minimum feature size on wafer            | 1.0 μm                                 | 1.2 μm (g-line)<br>0.85 μm (h-line)<br>0.35 μm (i-line)<br>0.18 μm (deep UV)<br><0.1 μm (extreme UV or X-ray) |
| Maximum field size on wafer              | 90% of the wafer size                  | <25 × 25 mm                                                                                                   |
| Minimum lateral field misalignment       | Approx 0.5 μm                          | <0.25 μm                                                                                                      |
| Cost of machine                          | Medium (250K)                          | Very high (>5 M)                                                                                              |
| Costs per batch                          | Low to medium                          | High to very high                                                                                             |
| Minimum batch                            | Individual wafers/cassette to cassette | Cassette to cassette                                                                                          |
| Wafer size versus mask (or reticle) size | 2 inch/3 inch                          | 4 inch/5 inch                                                                                                 |
|                                          | 3 inch/4 inch                          | 6 inch/7 inch                                                                                                 |
|                                          | 4 inch/5 inch                          | 8 inch/9 inch                                                                                                 |
|                                          | 5 inch/6 inch                          | 12 inch/13 inch                                                                                               |

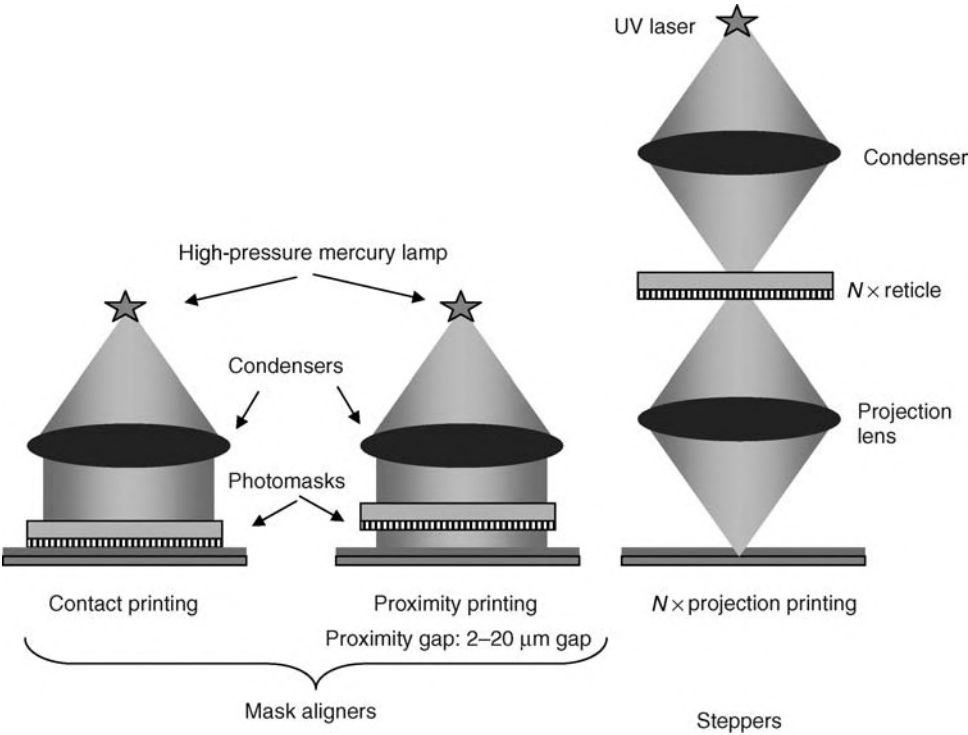


**Figure 12.33** Contact, proximity and projection lithography

Since the mask is in contact with the resist-coated wafer, the mask needs to be cleaned often, and this can reduce its performance over time. In proximity or projection lithography, the mask is not in contact with the wafer, and therefore remains clean and free of defects.

### 12.4.5.2 Mask Aligner Architectures

A mask aligner is what its name implies: it aligns masks to wafers. There are three ways to implement a mask aligner:

- contact alignment;
- proximity alignment; and
- $1\times$  projection alignment.

A mask aligner does not step the field as a stepper would, since the field is nearly as large as the mask itself. The field alignment in a mask aligner can be automated through a set of specific alignment targets, but is also often performed manually when alignment tolerances are critical, as they usually are for digital micro-optics fabrication.

Figure 12.34 shows a contact aligner (Karl Suess MJB55) and a projection aligner (PE 300HT).

Such mask aligners are perfect for digital optics fabrication, since they can be operated in manual mode (most of the complex mask aligners and steppers operate only in automatic mode for high production volumes).

A very important aspect of quartz (or fused silica) wafers is that they are actually transparent! On lithographic tools, the robotic arm grabbing the wafers from the cassette and placing them on the chuck for exposure can rely on two type of sensors: optical or pneumatic. If the sensor is optical, it is obvious that the robotic arm will not see the wafers, and therefore cannot proceed. If the sensor is pneumatic (measuring the air pressure reflected from the wafer) the robotic arm can thus sense transparent wafers. One solution to prevent disastrous crashes in an optical sensor based aligner is to coat the wafers with a reflective coating prior to resist coating (e.g. a 200 nm Cr or Al coating). The coating can be applied on the back side or on the front side of the wafer. If the coating is applied on the front side, wet Cr etching has to be performed prior to RIE etching of the substrate. This has pros and cons, as seen in Chapter 15.

### 12.4.5.3 Stepper Architecture

Steppers are one of the most expensive microlithography machines that one can find in a chip production facility. A stepper transfers the binary reticle pattern by image projection onto a wafer, creating an aerial

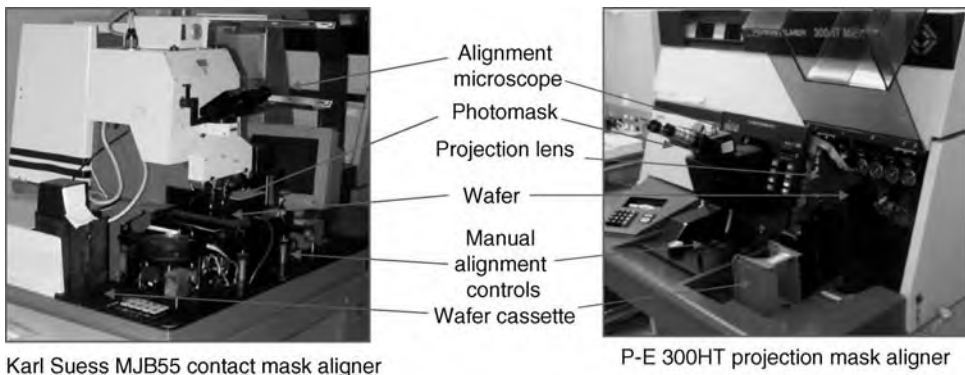


**Figure 12.34** Contact and projection mask aligners

image in the vicinity of the photoresist layer (the photoresist coatings are around  $1\text{ }\mu\text{m}$  thick), with a reduction factor between  $1\times$  or  $10\times$  – typically  $5\times$  for most steppers today.

The word ‘stepper’ comes from ‘step and repeat’, since these lithography tools are able to step and repeat the projected pattern or field from the reticle in a two-dimensional array onto the wafer. Typical sizes of the reticles are 5 or 6 inches square, and typical sizes of the wafer are 4, 6 and now even 8 inches in diameter. Reticles are usually made out of thick fused quartz (90 mil, 125 mil or 250 mil thick – or thousands of an inch). The size and thickness of both reticle and wafer are very critical and cannot be changed, unlike with contact printing mask alignment tools. Steppers use complex alignment marks or fiducials to align successive photomasks onto the same wafer, which is then processed in between the successive exposures – to produce, for example, a multilevel phase-relief DOE. The lateral alignment accuracy of typical IC foundry steppers is within a tenth of a micron, and the smallest lateral dimensions are about  $150\text{ nm}$  at wafer scale (below  $150\text{ nm}$ , one has to use the RET techniques described in Chapter 13).

New type of steppers called scanners, which combine both stepping and scanning processes, have been introduced to industry. These machines move around both reticle and wafer to project a small part of the reticle onto the wafer at a given time, in order to decrease the smallest transferable lateral feature size  $\delta x$  even more by using continuously the central part of the imaging optics, which yields the best performance during the entire large reticle field transfer. Note that some mask aligners are also scanners (e.g. the PE 300HT).

Conventional steppers use UV light sources from the g-line, h-line or i-line of a mercury lamp (see Table 12.5). However, deep-UV excimer laser sources are used more and more often, with wavelengths shorter than  $200\text{ nm}$ , in order to decrease the diffraction effects occurring in the projection stage and thus decrease  $\delta x$ . The imaging lenses used in these steppers thus also have to be optimized for these very short wavelengths, and made of optical material transparent to DUV, which make them very expensive optics (CaF). Figure 12.35 shows an UV ASML stepper and its internal components.

Table 12.6 summarizes the specifications of steppers used in industry today.

For more details on how to increase the resolution of steppers through the use of hardware optimizations, process optimizations and especially Design For Manufacturing (DFM) solutions such as Reticle Enhancement Techniques (RET), see Chapter 15.

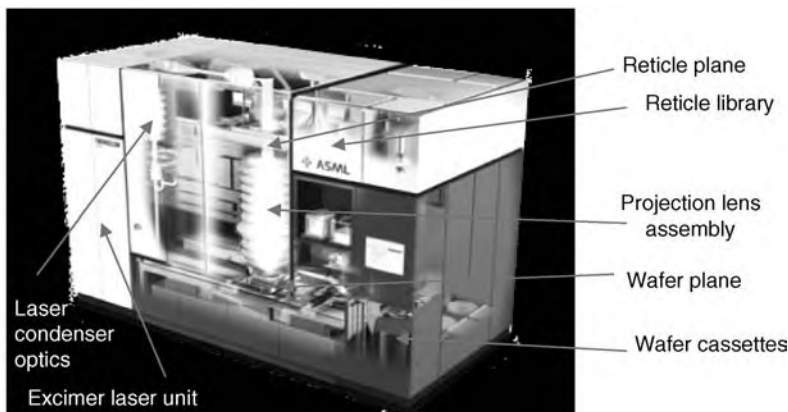


**Figure 12.35** An example of a stepper



**Table 12.6** A summary of steppers used in industry today

| Stepper brand   | GCA                   | Canon                  | SVGL                   | Nikon                  | ASML                   |
|-----------------|-----------------------|------------------------|------------------------|------------------------|------------------------|
| Model           | <i>DSW4800</i>        | <i>ES3</i>             | <i>Micrascan III +</i> | <i>NSR-S205C</i>       | <i>5500/750E</i>       |
| M:1             | 10×                   | 4×                     | 4×                     | 4×                     | 4×                     |
| Wavelength      | 436 nm                | 248 nm                 | 248 nm                 | 248 nm                 | 248 nm                 |
| Max. NA         | 0.28                  | 0.73                   | 0.60                   | 0.75                   | 0.70                   |
| Field size      | 10 × 10 mm            | 26 × 33 mm             | 26 × 34 mm             | 25 × 33 mm             | 26 × 33 mm             |
| Overlay         | 500 nm                | 25 nm                  | 45 nm                  | 30 nm                  | 30 nm                  |
| Lens distortion | 250 nm                | 20 nm                  | 25 nm                  | 20 nm                  | 20 nm                  |
| Wafer size      | 3, 4, 5 inches        | 200, 300 mm            | 150, 200 mm            | 200, 300 mm            | 150, 200 mm            |
| Throughput      | 20 wafers<br>per hour | 100 wafers<br>per hour | 90 wafers<br>per hour  | 140 wafers<br>per hour | 120 wafers<br>per hour |

12.4.6 Wet Bench Processing

Wet bench processing is an important step in any lithographic process. Prior to lithography, the wafers are cleaned in a sulfuric acid bath with hydrogen peroxide (agitated and heated), followed by a DI ultrasonic water bath and a rinse/spin process, all in wafer batch holders (see also Chapter 15). The development of the wafers is performed on a wet bench, as well as the resist stripping and cleaning after the etching step (see the next section). Figure 12.36 shows a typical wet bench, including an acid bath, DI water cleaning, positive resist development and rinse/dry spin.

12.4.7 Etching

Once the lithography process is done, the resist bears the binary profile of the photomask, and is ready to be etched down the substrate. Very often, optical engineers tend to use the photoresist surface profile as the final element, for reasons mainly related to a limited fabrication budget, a relatively high speed of fabrication turnaround and so on. However, even though the photoresist layer might be more or less

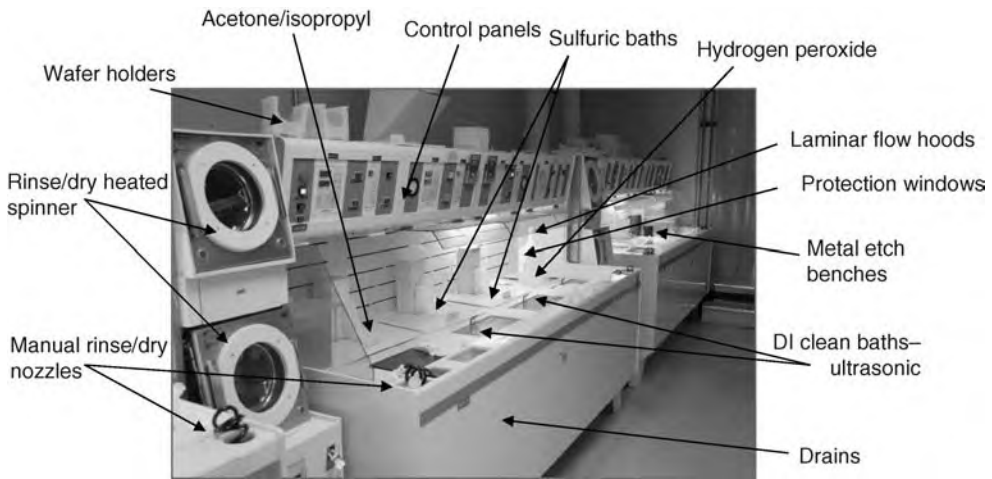


**Figure 12.36** A wet bench line example. Reproduced by permission of USI Inc.

transparent to the wavelength to be used in the final product, this is strongly not recommended for the many following reasons:

- photoresist is a polymer and even after a hard bake, it deteriorates over time;
- photoresist side walls show standing-wave profiles;
- photoresist does not 'resist' high laser powers;
- photoresist swells with ambient humidity;
- photoresist swells with a temperature increase;
- photoresist is not totally transparent to visible light;
- photoresist surfaces can be rough, which creates diffusion and other parasitic effects;
- photoresist layer can peel off the substrate over time;
- one cannot fabricate multilevel profiles in photoresist without using direct gray-scale write or gray-scale masking lithography; and
- photoresist can be very harmful to health (cancerous).

Figure 12.37 shows these various effects in photoresist layers.

For some practical examples of resist lift-off problems, see Section 15.4.7.

For all these reasons, it is very unlikely that a resist profile can be used for anything other than prototyping tests, and it should be etched in the underlying the substrate. There are two ways to etch a photoresist profile down the substrate: by wet or by dry etching.

#### 12.4.7.1 Wet Etching

A wet etching process is performed in an acid bath (hydrofluoric acid, KOH etc.). The wet etching process can be either anisotropic (etching is performed in specific spatial directions) or isotropic (etching is performed in all spatial directions).

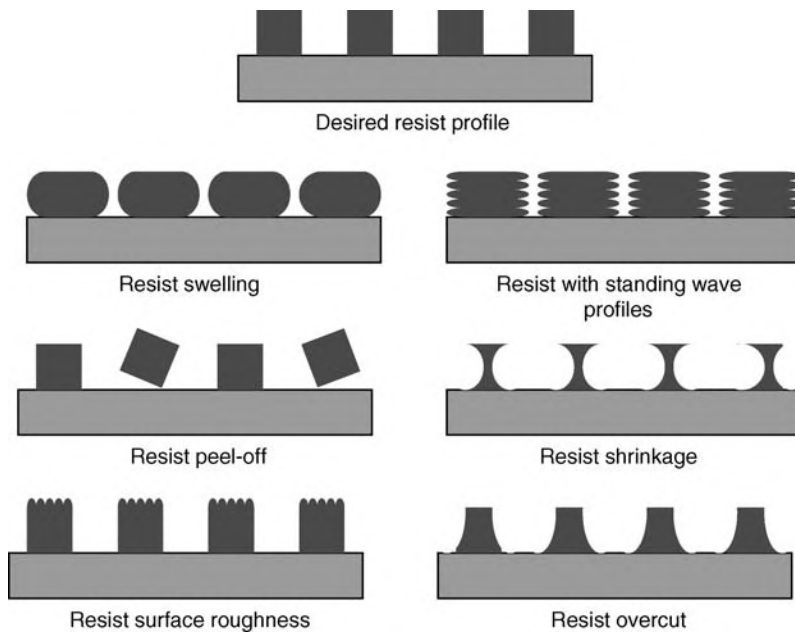


**Figure 12.37** Problems related to photoresist layers

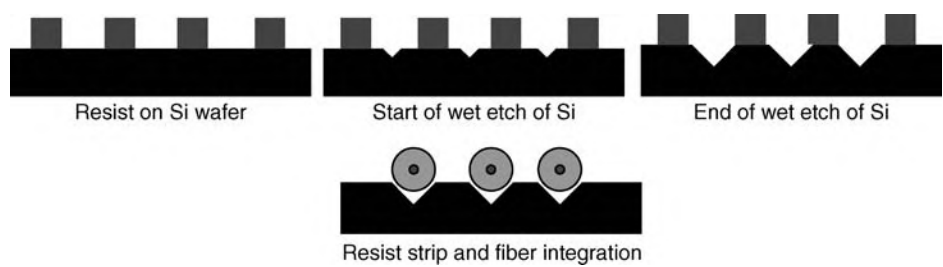


Figure 12.38 Wet etching of silicon substrates

**Anisotropic Wet Etching**

Wet anisotropic etching occurs when the wafer material has particular crystalline structures, such as Si wafers with specific crystalline planes (111–101 etc.). Wet etching of Si wafers can be a valuable process to produce particular structures and grooves in the wafer cheaply and quickly, for specific crystal orientations (see Figure 12.38).

Such wet isotropic etching of silicon substrates can be very useful; for example, to produce V-grooves for the accurate positioning of fiber arrays. However, due to the fact that the geometry of the groove cannot be engineered, this technique is very seldom used for the production of digital optics.

Etch stop layers can also be used in conjunction with anisotropic etching to produce different structure sections with a more or less flat bottom but with angled side walls. Laser-assisted wet etching can also be used to control the etching process.

**Isotropic Wet Etching**

Wet isotropic etching occurs when there is no particular crystal geometry (as in glass, fused silica etc.). However, wet etching of digital optics structures in glass substrates provides quite weak results in terms of optical efficiency, due to the isotropic nature of the etch yielding undercuts and other parasitic effects (see Figure 12.39).

Figure 12.39 shows the undercut, or even overcut, etch effects that occur during isotropic wet etching of quartz, fused silica or other glasses. This is why wet etching of glass (e.g. through HF acid) is very seldom used to produce high-quality digital optics in industry. Chapter 11 shows that a flat-bottomed structure produces the exact phase shift for local destructive interference that produces high diffraction efficiency. Rounded grooves produce all kind of phase shifts, and thus are not very efficient.

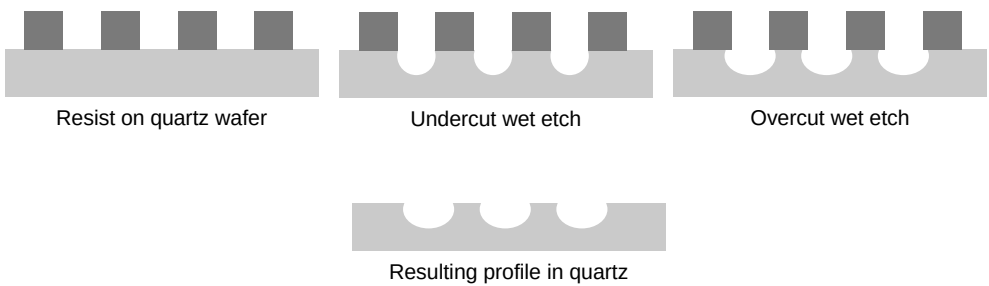


Figure 12.39 Wet etching of quartz, fused silica or glass substrates

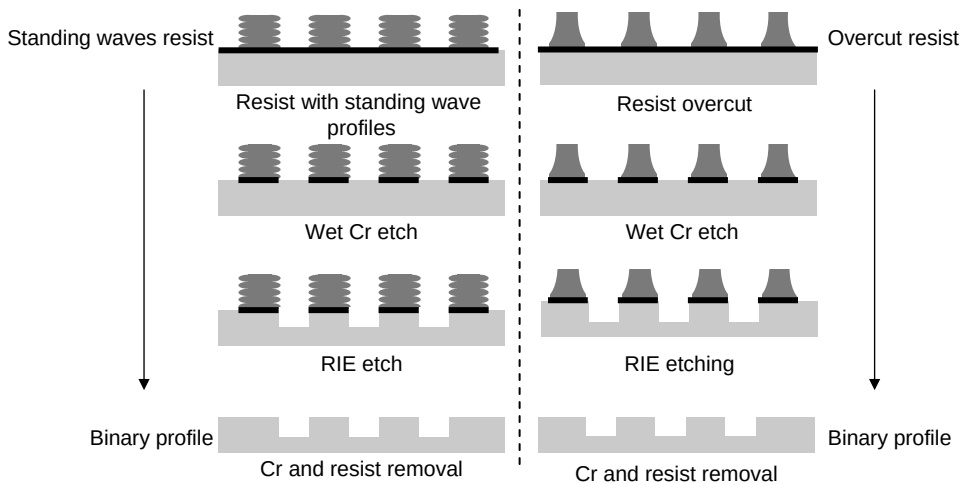


**Figure 12.40** A binary RIE etching process with a Cr layer

### 12.4.7.2 Dry Etching

Dry etching is an anisotropic etching process (etching performed in only one spatial direction), due to ion assistance, as opposed to wet etching, which is an isotropic process. Dry etching is the *de facto* solution to produce high-quality profiles for digital optics. Dry etching of fused silica can be performed in a Reactive Ion Etching (RIE) plasma chamber, with added gas, this usually being referred to as Chemically Assisted Ion Beam Etching (CAIBE). Figure 12.40 shows how the RIE process can transfer a photoresist pattern with a nonoptimum surface profile (with overcut and standing waves) into a perfect binary profile into the underlying substrate. The process utilizes an additional Cr layer under the resist in order to produce the sharp edges. The Cr layer is wet etched before the RIE can begin (since the RIE process will not etch the Cr).

An RIE etching process is more a mechanical grinding process than a chemical etching process. The argon ions are accelerated between a cathode and an anode, and are bombarding the substrate. Typically, the photoresist ‘resists’ the etching process (this is not really true, since one can also etch down the resist profile into the wafer in an proportional etching step).

Figure 12.41 shows a typical RIE etching chamber (from an Applied Materials AM 8110 RIE machine). In this plasma chamber, the wafers are set vertically. In other RIE chambers, the wafers can be set horizontally. The plasma chamber can be accepting one or more wafers (in the AM 8110, up to 32 4 inch wafers can be placed simultaneously in the chamber).

Fused silica or quartz etching is usually referred as an oxide etching process. A proper fused silica RIE process is usually performed in three steps:

- the first step is an asher priming process, with only  $O_2$  gas – this removes unwanted resist residues, usually in 30 seconds to one minute;
- the second step is the actual oxide RIE etching step with  $CHF_3$  gas (and  $O_2$ ) – the etching time is proportional to the desired etch depth; and
- the third step is again a finalizing asher step (with only  $O_2$ ) – this usually takes from 30 seconds to one minute.

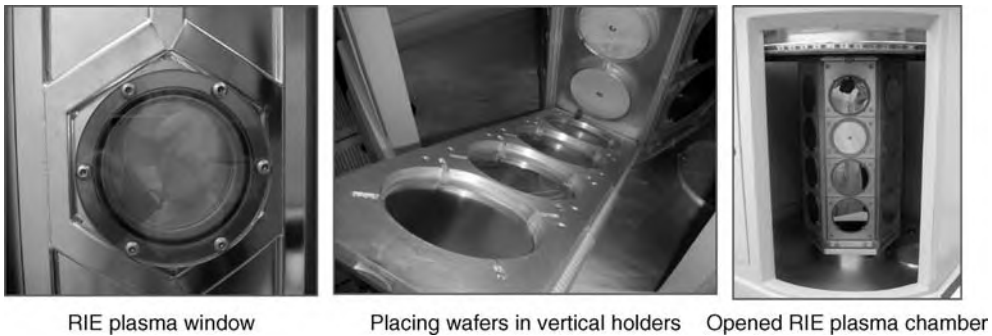


**Figure 12.41** A typical RIE etching chamber for an oxide etch ( $\text{CHF}_3$ )

Figure 12.42 shows some results of the dry etching process in fused silica substrate (on binary and multilevel diffractive Fresnel lenses). Vertical side walls are very desirable features for reaching maximum efficiency (optimal local destructive interference) for digital optics. Such features are achieved by RIE etching in the scans in Figure 12.42. Dry etching does not work well on simple glass or borosilicate substrates. This is why most digital optics are etched into quartz and fused silica rather than glass wafers.

**12.4.7.3 Etch Depth Accuracy**

Etch depth errors are systematic fabrication errors (they always happen). Chapter 11 shows that etch depth errors produce a drop in efficiency, and an increase of the light coupled in the zero order, which in most cases is a negative effect. Many techniques to control etch depth have been reported in the literature (see

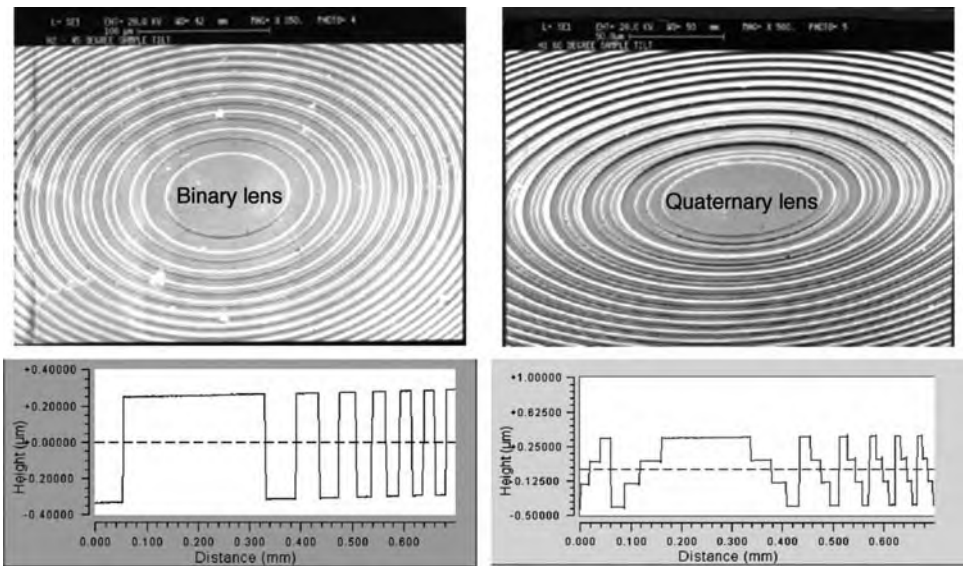


**Figure 12.42** An example of dry etched binary and multilevel digital optics in fused silica

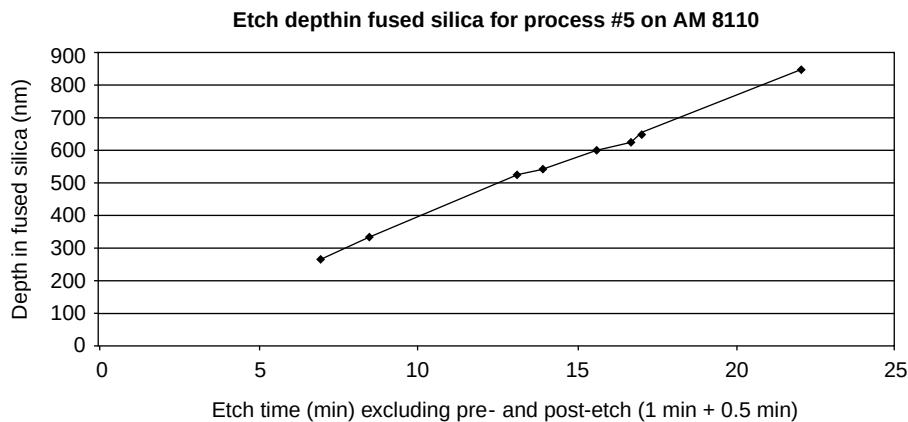


**Figure 12.43** The etch rate of fused silica wafers in a  $\text{CHF}_3$  RIE chamber. Reproduced by permission of USI Inc.

also Chapter 15). The most straightforward is to characterize the etch rate accurately (see Figure 12.43). One can see that (fortunately) such an etching rate is actually a linear process. The etching rate in fused silica in Figure 12.43 results in about 38 nm per minute.

Another method to control the etch depth accurately is to deposit a layer of material on top of the wafer and selectively etch that layer rather than the substrate itself. Such a layer can be  $\text{SiO}_2$  or another oxide layer, on top of a glass substrate (see Figure 12.44). Such a layer is also called an etch stop layer.

Such a layer is also very useful if one wants to fabricate a ridge waveguide with a higher index of refraction. BPSG layers can also be easily coated on the substrate. The interesting feature with BPSG is that the index of refraction can be accurately controlled by boron or phosphorous doping.

**12.4.7.4 An Example of Digital Optics Fabrication**

Figure 12.45 shows diffractive microlens arrays fabricated via stepper lithography on a 6 inch silicon (left) and quartz (right) substrate. Such microlens arrays are then diced out as individual elements (see the next section). The resolution achieved with an i-line stepper (here a Canon FPA 2500 i-line stepper) can be as small as 350 nm and the field-to-field accuracy as good as  $\pm 100$  nm.

There are many ways to produce diffractive structures, in reflection and in transmission, some having low efficiency, some high, some working in dual transmission and reflection modes. Figure 12.46 shows five wafers with similar elements, but fabricated in slightly different ways:

- Wafer A is a resist pattern on a silicon wafer, which is optimal for nickel electroplating for subsequent plastic replication of the wafer pattern (see Chapter 14).

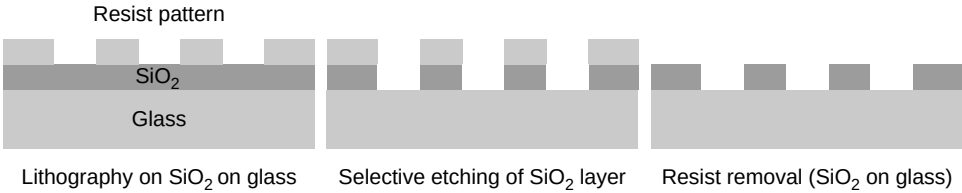


**Figure 12.44** An  $\text{SiO}_2$  etch stop layer on top of a glass wafer

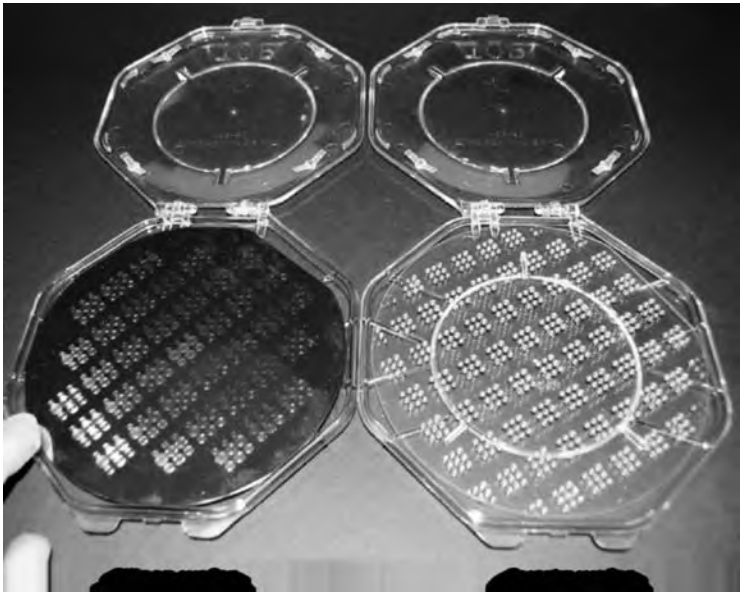


**Figure 12.45** Silicon and quartz wafers with multilevel diffractive lenses

- Wafer B is a wet etching process for Cr on a glass wafer. Such a process can produce a diffractive that has low efficiency, but works in both reflective and transmission modes with similar efficiencies (about 8%).
- Wafer C is an etched quartz wafer with optimal efficiency.
- Wafer D is an etched quartz wafer with a gold coating for optimal efficiency in reflection mode (the etch depth is different from that of wafer C).

In order to show the surface quality that one can achieve with the previously described multilevel fabrication techniques, a confocal white light interferometric microscope was used in Figure 12.47 to produce 3D profilometry scans of various diffractive lenses, after the first etching pass and after the second etching pass. The lens diameter is around 4 mm, and the smallest feature size is around 0.8  $\mu\text{m}$  (for the second masking layer).

In some cases, it is desirable to have multiple phase levels on a single wafer. A binary element might be required to produce symmetric patterns, and a four-level element for the suppression of the third

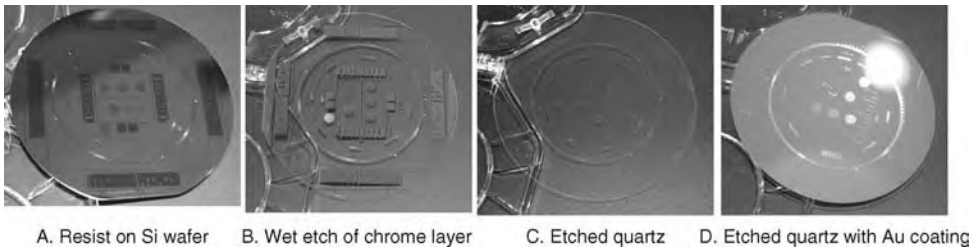


**Figure 12.46** Various process variations around multilevel lithography

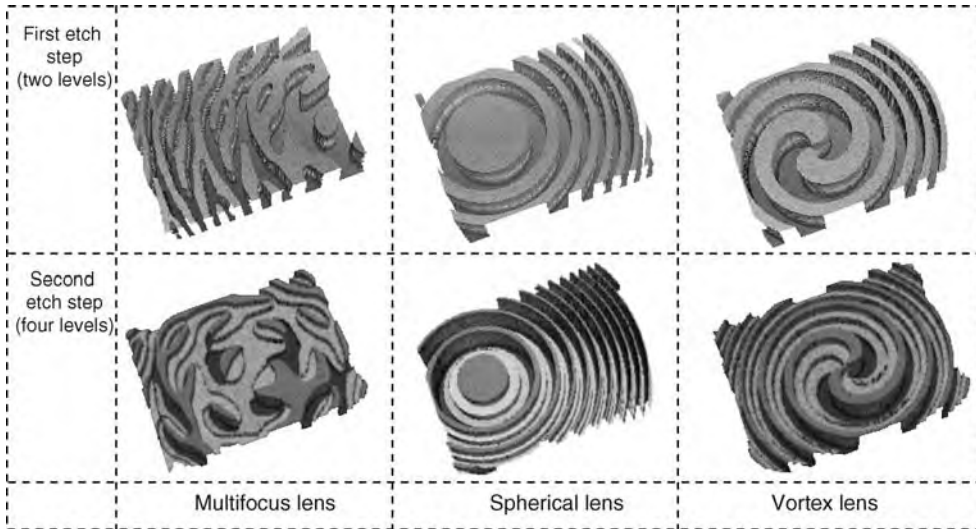


**Figure 12.47** Profilometry scans for quarternary diffractive lenses

harmonics, and then an eight-level element for maximum efficiency. For example, one can fabricate binary elements, four-level elements and eight-level elements on a single wafer, all depending on how the photomasks have been organized. Figure 12.48 shows such a hybrid fabrication, where three masks (clear field) have been used to fabricate a set of several dozen elements with two, four and eight levels as final elements.

The optical microscope photographs show the three different surface-relief configurations on the same wafer (appearing in different gray shades on the wafer). The elements in this wafer were beam-shaping CGHs, and have been fabricated with basic rectangular pixels of  $2 \times 20 \mu\text{m}$  (lateral aspect ratio of 10), since the burden on the diffraction angles in the  $x$  and  $y$  directions were very different. Once all the lithography steps have been processed, the final step in the fabrication would then be the dicing and packaging of the individual dies.

#### 12.4.8 Wafer Dicing/Wafer Scribing

Prior to dicing, the wafer can go through an optional AR coating process (on the nonetched surface), in order to minimize Fresnel reflections through this surface. For silicon wafers, one usually refers to this process as wafer scribing, since the wafer is scribed before it is broken into individual dies. Scribing does not work well with glass and quartz substrates. For such substrates, one has to use diamond saw cutting (dicing). The wafer is put onto adhesive tape (usually referred to as blue tape), and set underneath a diamond saw, which cuts through the wafer without cutting into the underlying adhesive tape (see Figure 12.49).

Now, a question comes to mind: ‘How can one see the dicing lines of these are etched transparent lines in transparent substrate?’ The problem is that they are almost invisible. The solution is to etch small gratings in the dicing lines. For example, if the dicing line is  $250 \mu\text{m}$  wide, one can use a grating with a  $10 \mu\text{m}$  pitch and a width of  $250 \mu\text{m}$ , with the total length required (e.g. through the whole substrate).

Such grating dicing lines diffract the white light and create sharp colorful structures on the otherwise transparent wafer. Thus, the dicing lines can be seen very easily either by a technician or by the robot



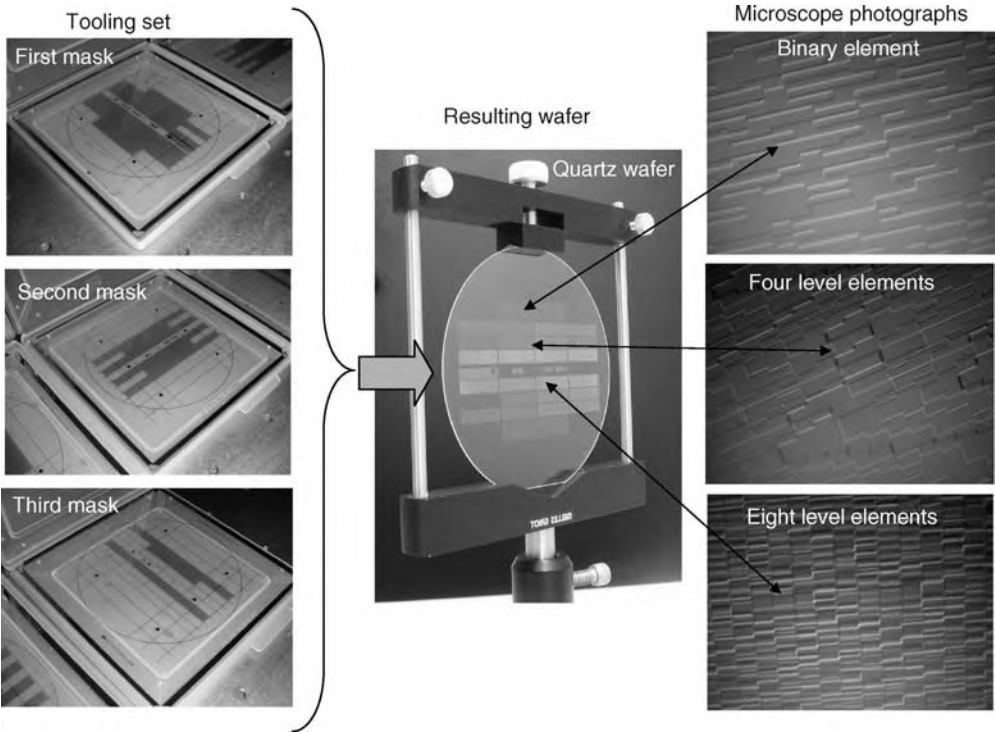


**Figure 12.48** A single-wafer multiple surface-relief configuration fabrication

vision controlling the dicing process. Figure 12.50 shows such grating dicing lines (or holographic dicing lines).

If one inserts the wafer with the structures on the adhesive tape, due to the relative index matching of the tape and the wafer itself, the structures become invisible, even though gratings dicing lines are used. The wafer has to be put on top of the adhesive tape structures. After dicing, the tape is introduced into a UV beam, which reduces the stickiness of the tape, and thus the individual dies can be remove by a vacuum tweezers or even by hand dies (see Figure 12.51). Following this process, the individual dies have be cleaned to remove any remaining tape or glue.

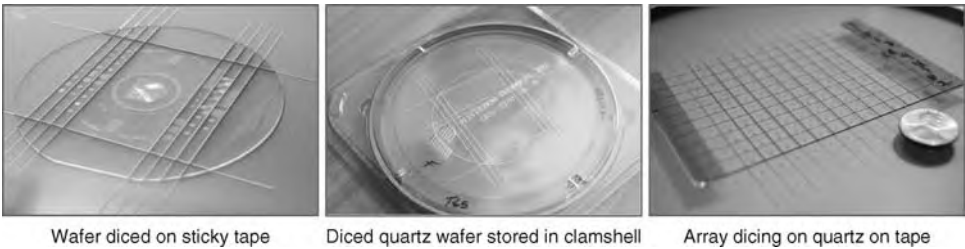


**Figure 12.49** Quartz wafers diced out on adhesive tape and stored in clamshells

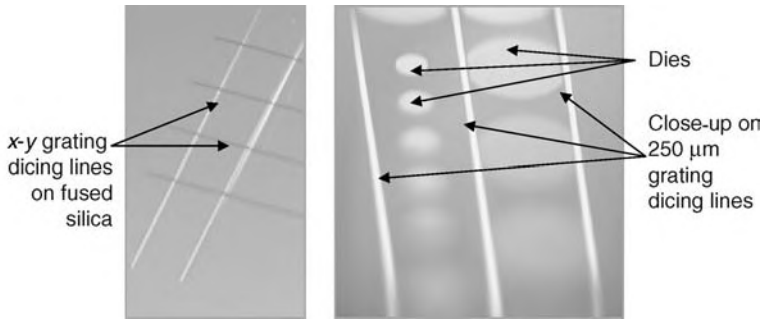


**Figure 12.50** Holographic grating dicing lines in transparent substrates

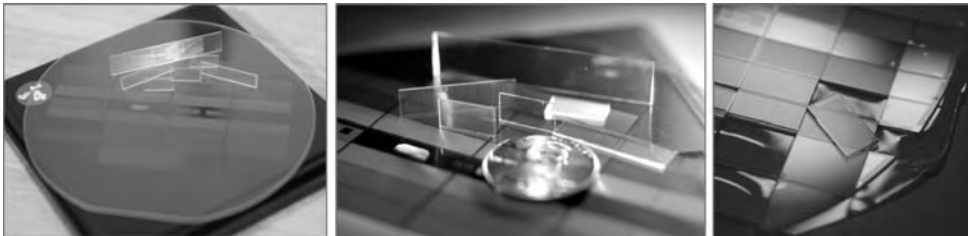


**Figure 12.51** Individual dies and uncut wafer

### 12.4.9 Labeling/Packaging

After the dicing and cleaning process, the dies have to be sorted out, because the dies simply resemble each other, even though the elements might be different (see Figure 12.52). Usually, one adds text labeling or a reference on the mask layout beneath each element, so that one can sort out the dies easily after dicing.

Individual die packaging can be done efficiently in GelPacks<sup>TM</sup>. Such packages hold the die securely by surface tension on a specific gel, with a retention coefficient that can be specified (from 0, the weakest, to 4, the strongest). If the wafers are the final devices, and do not need to be diced out, they are shipped in individual spring-loaded wafer carriers. GelPacks<sup>TM</sup> and spring-loaded wafer carriers are depicted in Figure 12.53.

There is another step involved, which usually comes before the wafer dicing process, and that is the fabrication and optical characterization of the wafers, in order to see if they meet the target specifications. The characterization processes are addressed in Chapter 15.



**Figure 12.52** Individual dies to be sorted out



**Figure 12.53** Individual dies on GelPacks™ and spring-loaded wafer carriers

The next three sections are dedicated to analog or quasi-analog processes using conventional binary lithographic techniques, as described in the previous section. The first two techniques are dedicated to the fabrication of refractive profiles, either in surface-relief or index-gradient form. Next will be described the various direct write techniques that can produce quasi-analog surface-relief diffractive and hybrid diffractive/refractive profiles in resist. A review of the various gray-scale lithographic techniques using gray-scale masks will conclude this section.

## 12.5 Micro-refractive Element Fabrication Techniques

The fabrication of micro-refractive elements with large sags is a key element of numerous applications today. However, such elements cannot be fabricated by multilevel lithography as presented previously. The multilevel lithographic techniques are well suited for surface-relief diffractives.

### 12.5.1 The Ion Exchange Process

The first technique presented here, the ion exchange process, uses binary lithographic techniques and ion diffusion to produce, directly in the substrate, an index modulation that can implement a variety of micro-optical elements, from microlenses to beam shapers to buried waveguide structures.

In a typical ion exchange process, one would use a binary chrome mask transferred over the desired substrate. Molten salt is then poured over the binary pattern. The diffusion operates through the openings in the resist pattern down to the substrate (see Figure 12.54).

Today, industry mainly uses two different ion exchange technologies:

- the thermal ion exchange process; and
- electric-field assisted ion exchange.

In the thermal ion exchange process, the index distribution is generated by thermal exchange of  $\text{Ti}^+$  ions in a glass substrate for  $\text{K}^+$  or  $\text{Na}^+$  ions in a molten salt. Other ions can be exchanged in the same way. Ion stuffing is produced at  $\sim 400^\circ\text{C}$  for several days, and ion unstuffing at  $\sim 430^\circ\text{C}$  for several hours. This process produces a smoothly varying index profile within a single mask opening location.

In the electric-field assisted ion exchange technique, an additional electric field is used to diffuse the ions further down in the glass substrate. This technique is well adapted for the fabrication of arrays of GRIN microlenses (see Chapter 4) and integrated (buried) waveguide structures (see Chapter 3). Such integrated buried waveguide structures are very interesting, since the position of the core inside the

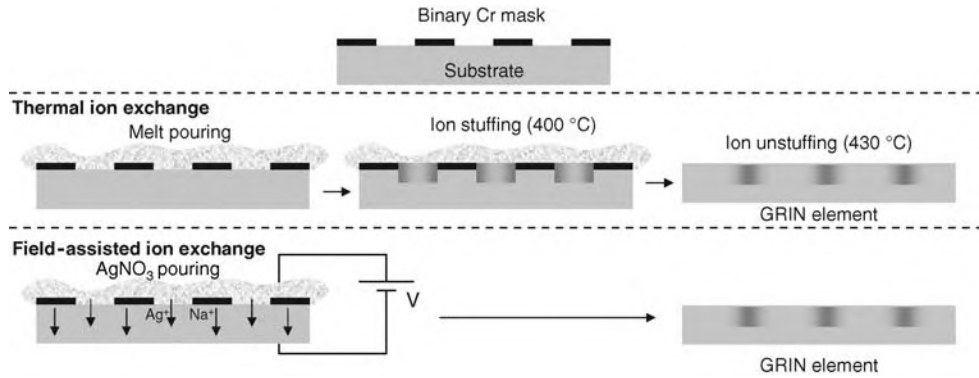


**Figure 12.54** The ion exchange process for the fabrication of index modulation

substrate can be positioned accurately, and can even be very close to the surface, and thus efficient PLC devices can be produced based on, for example, Bragg grating coupling (see also Chapters 3 and 16). Despite the long processing time (several days for the thermal ion exchange process), this fabrication technology gives rise to high-quality elements with analog index distribution profiles.

There are other ion diffusion fabrication methods used in industry today; for example, ion-beam injection by means of high-voltage accelerators.

### 12.5.2 The Resist Reflow Process

The resist reflow process can produce smooth surface-relief profiles in photoresist or BPSG. Such structures cannot be very small, unlike the index modulations produced with the previous ion exchange technique. They can produce refractive micro-optics with smoothly varying profiles [14–16]. However, it is impossible to produce blazed or *échelette* gratings with these techniques, since the profile of such elements needs accurate angles for optimum efficiency.

In a resist reflow process, one uses the surface tension effects on the substrate by melting and reflowing a material that melts at much lower temperatures than the substrate. Therefore, a quartz substrate is very desirable due to its very high melting point. Resist has very low melting point: the melting point for BPSG is higher, but still much lower than that for quartz.

The straightforward way to produce a resist reflow is to pattern a simple binary resist structure on the substrate, etch down a pedestal and begin to melt the resist so that the surface tension forms a dome, which can be approximated to a microlens (see Figure 12.55).

The resist profile can be further etched down the substrate by proportional RIE etching (see also the next section). The pedestal is etched down the substrate, and forms the base of the profile, since the surface tension will force the resist to remain on that pedestal. Therefore, it is possible to produce arrays of close-packed lens arrays – but with fill factors less than 100%, since these lenses cannot touch each other. The angle formed by the smooth resist profile at the interface of the substrate is dictated by the wetting angle.

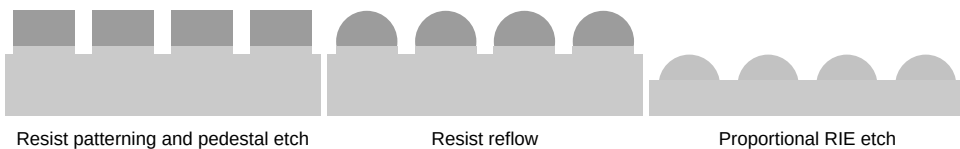


**Figure 12.55** The resist melt and reflow process and the subsequent proportional RIE etch

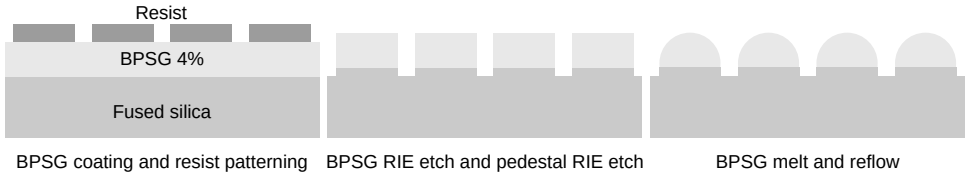


**Figure 12.56** The BPSG reflow process

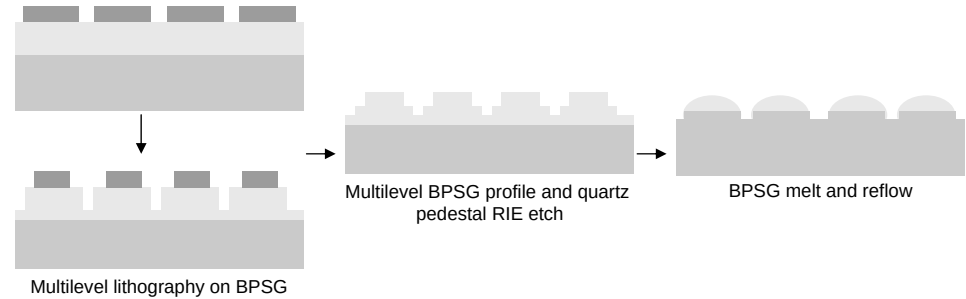


**Figure 12.57** The pre-shaping of the BPSG profile prior to melt and reflow

A better process is to use a BPSG layer on the substrate, where the amount of boron and phosphorous not only dictates the refractive index, but also the melting point (see Figure 12.56). A typical BPSG layer can be 4% phosphorous and 2% boron.

The deposition of BPSG can be done at 800 °C. The melting point of BPSG with 4% concentration is around 450 °C and the reflow can be done at around 100 °C. The BPSG etching can be done either with BOE etching or with standard oxide RIE, with  $\text{CHF}_3$  or  $\text{SF}_6$ . BPSG can be used as the final material, so there is no need to proportionally etch the profile into the substrate.

However, it is quite difficult to produce an exact profile with binary masks. An optimization to this process is to begin to shape the BPSG profile with the previously described multilevel lithographic techniques, and then proceed to melt and reflow (see Figure 12.57). Thus, one can produce profiles that are not only domes but can be more complex and, to a certain degree, include prescribed aspheric profiles.

Figure 12.58 shows an hexagonal array of refractive microlenses fabricated by BPSG 4% melt and reflow over a quartz substrate. The figure also shows the focal spots when the wafer is placed in direct sunlight over a piece of paper.

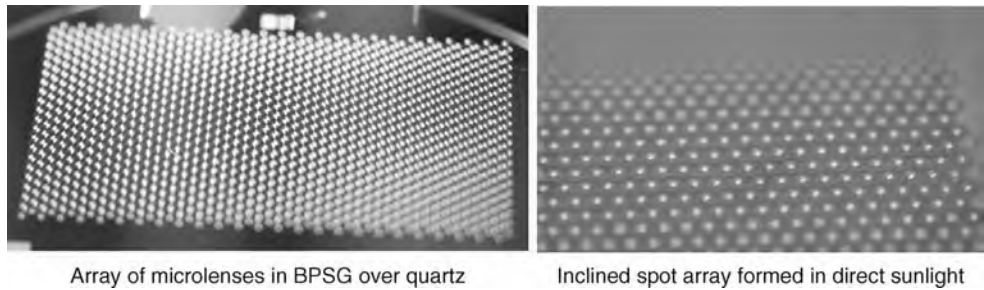


**Figure 12.58** An example of an array of refractive microlenses fabricated by the BPSG 4% melt/reflow process

## 12.6 Direct Writing Techniques

In order to avoid the systematic lateral field misregistration errors that occur when using conventional multilevel lithography (see previous sections), direct write methods have been reported [17–20]. These methods consist of directly writing into a resist layer a binary, multilevel or quasi-continuous profile, either by using either an e-beam or a laser beam mask-patterning system.

### 12.6.1 Direct Binary Write

In direct binary write, an e-beam or laser beam is used to expose the resist to form a pattern directly in the final wafer, without any optical lithography step (see Figure 12.59).

Direct write techniques are useful when very fine patterning is required, or when a stepper cannot be used. It has been shown previously that when the desired element is larger than the limited field of the stepper, such an element cannot be produced on the wafer, other than by using multiple masks and field-stitching techniques. In this case, direct binary write is a good alternative.

This said, there is no tooling. So, if one device is required, it is fine. If many devices are required, many similar writing processes have to be performed. This process is thus very unlikely to be used in production mode, but only in research and development mode.

A Cr layer was used in the direct write process shown in Figure 12.59, since this produces much better side-wall profiles after RIE etching (see also Section 12.4.7.2). If the writing tool used is an e-beam patterning machine, there needs to be a conductive layer underneath the e-beam resist in order to discharge the electron beam to the ground and prevent resist charging. Thus Cr is a good choice. If the final element remains in resist, without etch, that layer can be an Indium Tin Oxide (ITO) layer, which is conductive and transparent to visible light. Also, due to proximity effects in the resist, if the resolution is approaching the limits of the resist, e-beam proximity compensation should be used to optimize the binary pattern.

### 12.6.2 Direct Analog Write

When multilevel or quasi-analog surface profiles are desired, along with very small features (thus requiring very tight field-to-field alignment tolerances), direct analog write techniques can be used. Such techniques are also used when the device is larger than the stepper field. In this technique, the standard lithographic process is used without proceeding to any photo-lithographic step, which means that this fabrication technique can only produce a single element. Therefore, direct analog write cannot be used in production mode.

In order to produce the multilevel profile, two different techniques are investigated:

- direct dosage modulation in real time; and
- the multi-pass exposure technique.

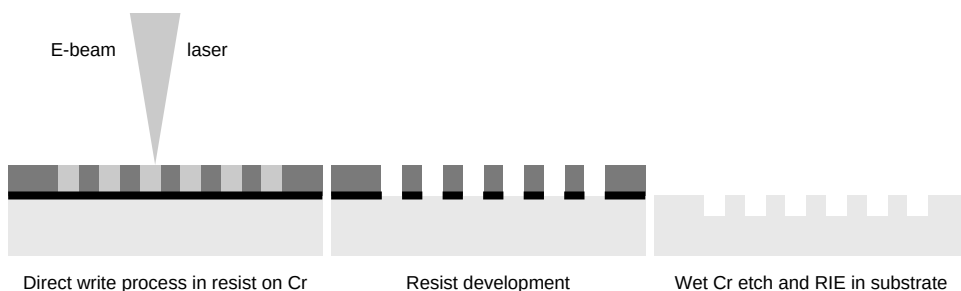


**Figure 12.59** The direct binary write technique

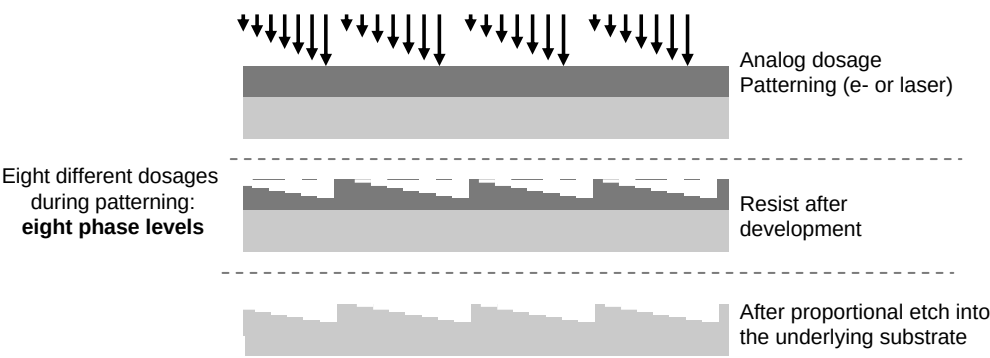


Figure 12.60 The direct dosage modulation write technique

12.6.2.1 Dosage Modulation

In the dosage modulation technique, the writing beam (laser or e-beam) is directly modulated to the appropriate dosage when writing the desired shape. It is therefore a single-pass writing scheme (see Figure 12.60).

To do this, the writing system has to be able to modulate the beam in real time. Section 12.1 showed that the Cambridge Instruments EMBF patterning system can produce 16 different clock rates, or e-beam dosages, for a given shape (trapezoid). A laser beam can be modulated in real time by an acousto-optical modulator. The calibration of the resulting depth in the developed photoresist versus the e-beam dosage has to be carried out carefully, as shown in Figure 12.61.

The depth versus e-beam dosage is nearly linear, but for accurate results, a look-up table has to be used to associate a specific dosage to each clock rate.

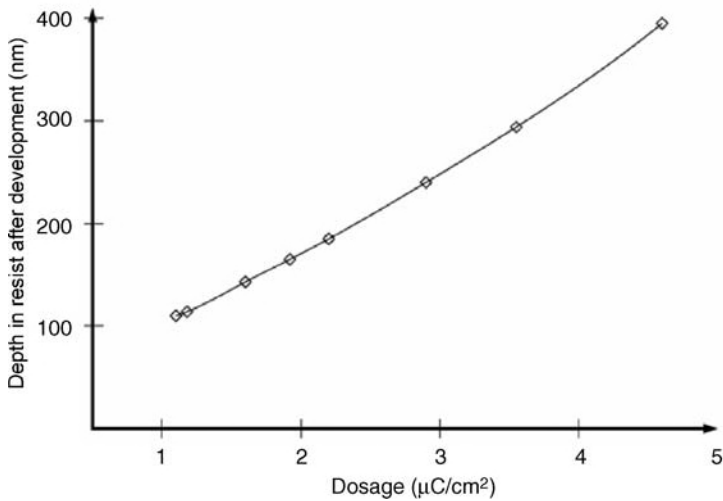


Figure 12.61 Depth in resist versus the e-beam dosage

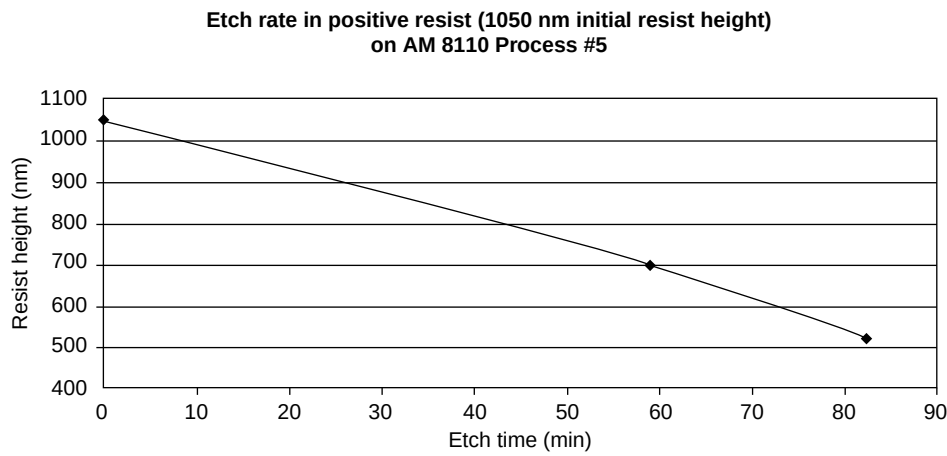


**Figure 12.62** A proportional etch from an analog resist profile into a fused silica substrate

**12.6.2.2 Proportional RIE Transfer**

The proportional RIE transfer [21] into the underlying substrate can be performed by Chemically Assisted Ion Beam Etching (CAIBE). A typical etching rate of quartz versus resist in an oxide etcher ( $\text{CHF}_3$ ) is shown in Figure 12.62.

When comparing the 6.3 nm/min etching rate to the previously described 38 nm/min rate in fused silica, one can calculate the proportional etching rate ratio  $38/6.3 = 6$ . Thus, in order to produce a surface depth of 6  $\mu\text{m}$ , one has to produce a surface profile of 1  $\mu\text{m}$  in the resist. This large ratio is particularly appreciated when it comes to etching down a refractive profile in which the required sag is much larger than the resist coating on the substrate. Figure 12.63 shows the real scale of the transfer from resist into the underlying substrate for micro-refractive profiles and diffractive profiles. An Inductive Coupled Plasma (ICP) – RIE chamber can also be used for proportional etching transfer.

In the case of diffractive profiles, it is important to try to reduce this proportional etching rate, since any error in the resist will be amplified by a factor of six in the underlying substrate. However, this might also be a positive effect for producing broadband (multi-order) diffractives, with deep grooves, provided that one can control the resist depth very accurately (which is a difficult task).

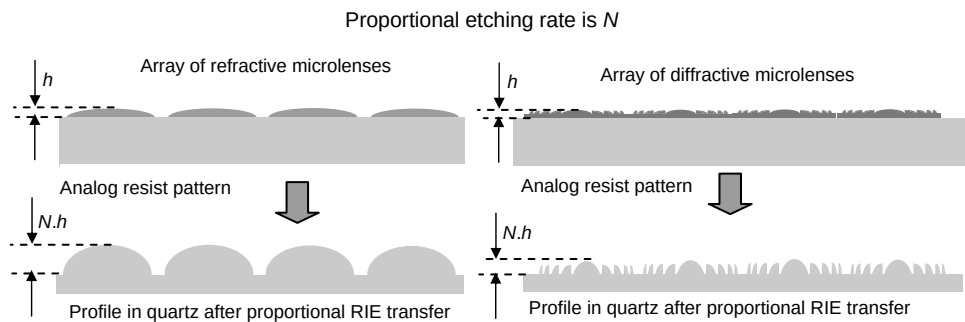


**Figure 12.63** A proportional RIE transfer from a resist profile into substrate



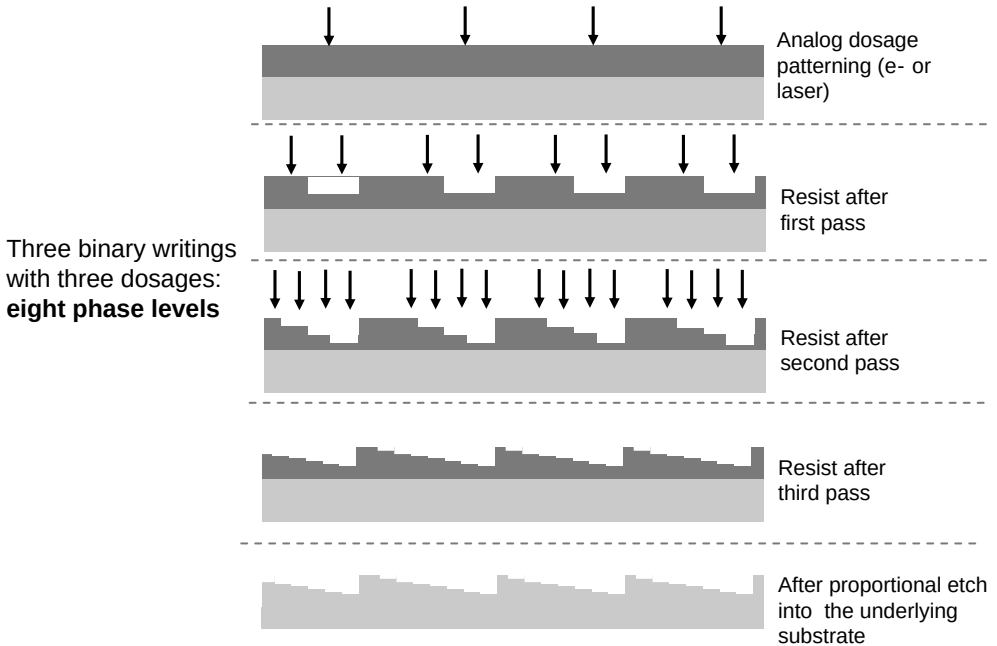


**Figure 12.64** The multi-pass direct write technique

### 12.6.2.3 Multi-pass Exposure

In the direct write multi-pass exposure technique, the same writing strategy is used as for traditional multilevel exposure (see the previous sections), but this time using direct binary write. The patterning system writes a first binary pattern (which is the binary pattern on the first mask that would be used in standard multilevel lithography), with a strong dosage. Then the writing system realigns to alignment fiducials on the substrate, and begins to pattern the second binary pattern (which, incidentally, is exactly the second masking layer for conventional multilevel lithography), this time with a lower dosage, and so on, to produce, over  $N$  binary write passes,  $2^N$  surface profiles in the resist (see Figure 12.64).

This technique can be used by more patterning systems than the previously described dosage modulation technique, since it does not require the dosage to be changed during the writing, but only between consecutive writings. Also, it requires many fewer different dosages for the same number of final surface-relief levels in the resist layer.

However, the disadvantage of this technique is that it needs to register  $N$  times to a pre-written pattern (or sets of etched alignment marks), and therefore it produces systematic misregistrations due to the limited accuracy of such alignment techniques (even though they might be interferometric alignment techniques).

As an example, Figure 12.65 shows an array of 16 phase level Fresnel diffractive lenses fabricated via a dosage modulation direct write technique.

### 12.6.3 The CD Authoring Technique

CD authoring techniques are pseudo-lithographic techniques developed by the CD industry, which have been used for many years to produce master disks. A glass disk coated with photoresist is made to spin in front of a laser beam that is vibrating at a precise frequency and is blanked in order to produce the various CD pits and tracks used in audio and data CDs and DVDs.

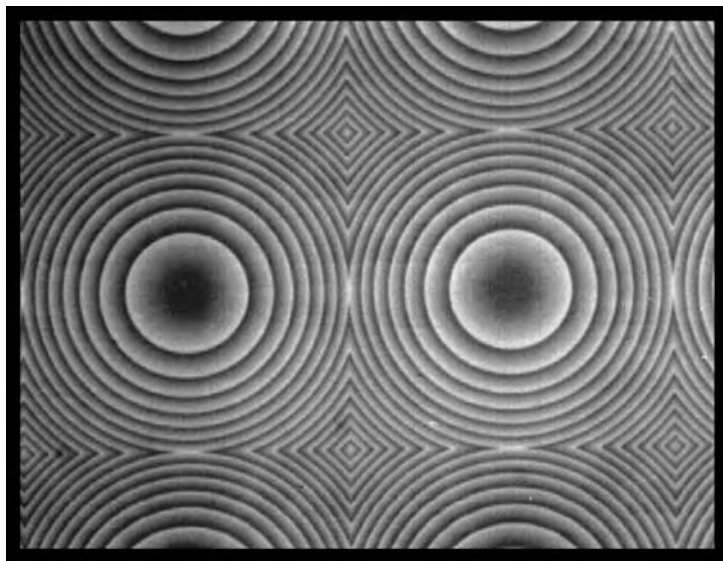


**Figure 12.65** An array of Fresnel lenses fabricated by direct write with dosage modulation over 16 levels. Reproduced by permission of Walter Daeschner

Although such techniques are not well suited for conventional diffractives and micro-refractives, recent developments of such CD authoring machines can take in Cartesian coordinate patterns, and produce a multi-pass write technique as described in the previous section. It is worth noting that Blu-ray pits can be as small as 300 nm.

#### 12.6.4 Direct Fringe Writers

A fringe writer is basically a step-and-scan repeat process, which exposes in a interferometric (holographic) way a very small photosensitive region in 3D space [22]. The step-and-scan process is well known in lithography, and gave rise to steppers and scanners. However, unlike lithographic steppers controlled by standard interferometers, the phase error from holographic pixel to pixel has to be orders or magnitude lower than the most precise stepper in industry today.

Steppers and other sub-micron lithographic fabrication equipment that has been optimized for the IC industry is not well suited to controlling the local phase. However, it is well suited to controlling the overall intensity (the pattern). Therefore, if one is more interested in local phase accuracy than global intensity accuracy, one has to use a phase-locking device to write such gratings.

In a Bragg grating and other high-quality sub-wavelength gratings for metrology applications (see Section 16.4.1.1 on diffractive linear encoders), it does not really matter if the absolute position of a fringe 10 cm away from another fringe is located within 10 nm, 100 nm or even 1000 nm. However, it is very important that there is no phase dislocation and that there are no phase jumps anywhere in between these two fringes.

Therefore, direct fringe writers have been developed that use a phase-locking technique. This reads the phase error on the pixel, and compensates the error in the  $x$ - $y$  table in real time by using a local nanometric positioning device on top of the  $x$ - $y$  interferometric stage, such as a local piezo-electric transducer. There are many ways to implement a phase-locked fringe writer, and one such example, from the University of Arizona, Optical Science Center, is shown in Figure 12.66.

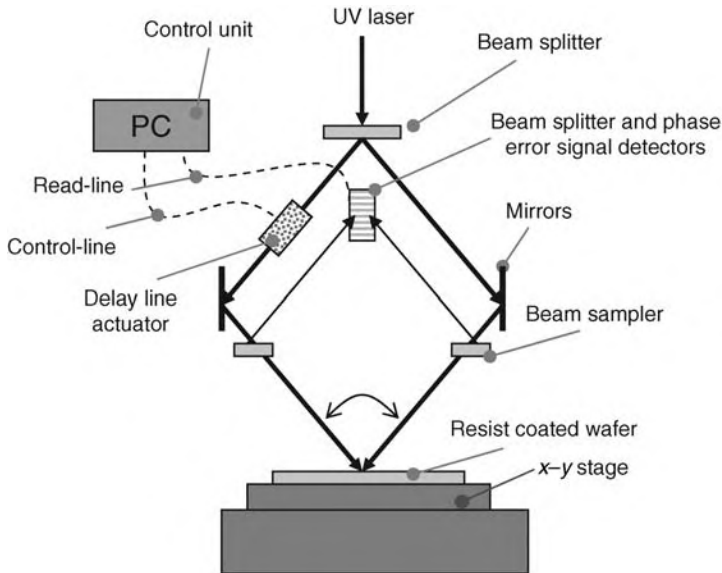


**Figure 12.66** A simple phase-locked fringe writer architecture

For example, fringe writers are used today to write Bragg gratings directly into optical fiber cores, where the phase accuracy has to be very high in order for them to be used in ultra-dense DWDM applications, as well as high-quality sub-wavelength gratings for fine metrology applications. Other fringe writers are used to produce 3D pixelated volume holograms of objects that are described in three dimensions in a computer (see, e.g., Zebra Imaging). Finally, fringe writers are also used to write 2D or 3D photonic crystals or sub-wavelength slanted surface-relief gratings directly into thick photoresists (see, e.g., Holox Technologies) for various applications in display, solar applications and sensors.

### 12.6.5 Two-photon Lithography

Three-dimensional photonic crystals can be produced based on materials, such as polymers, which are sensitive to two-photon excitation. Two-photon polymerization is initiated when femtosecond laser pulses are tightly focused in a photosensitive material. Two-photon excitation is a process that depends on the square of the intensity. In this process, photochemical or photophysical changes are induced in the material structure that are spatially localized near the high-intensity region of the focal point due to simultaneous absorption of two identical photons. Thus the energy transferred to the material by two photons of wavelength  $\lambda$  corresponds to that transferred by a single photon of wavelength  $\lambda/2$ . At the same time, one takes advantage of long-wavelength exciting light with a larger penetration depth, which enables 3D patterning of bulk material. Scanning the focus within the material allows one to produce 3D structures at the micron scale and sub-micron scale (down to about 200 nm when using an 800 nm Ti: sapphire laser). Figure 12.67 shows a commercially available two-photon nanolithography device from Nanoscribe<sup>TM</sup>, and some of the nanometric fabrication examples.

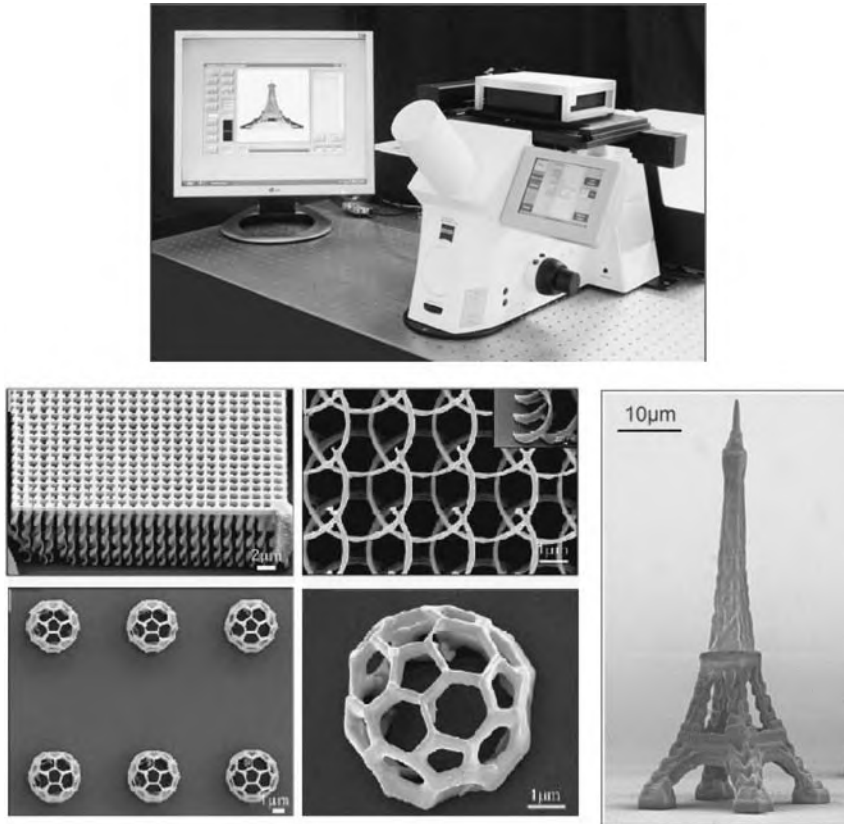


**Figure 12.67** The Nanoscribe™ two-photon 3D lithographic tool and realization examples. Reproduced by permission of Nanoscribe

Sub-wavelength structures can be produced because the absorption profile for the two-photon process is narrower than the beam profile. In addition, this method is well suited to introducing localized defect structures in a PC in order to generate waveguiding structures. The most common structures that have been fabricated using two-photon lithography are logpile-type photonic crystals.

## 12.7 Gray-scale Optical Lithography

The previous sections described two main techniques to produce quasi-analog surface-relief profiles in resist (and subsequent transfer into the substrate by proportional etching):

1. Multilevel binary lithography with  $N$  masks for  $2^N$  levels.
2. Direct multilevel write in resist over  $2^N$  (multi-pass) or  $N$  levels (direct modulation).

The first technique (multilevel lithography) is easily done using traditional IC fabrication techniques and conventional masking tooling sets. However, when the number of phase levels increases, the successive systematic fabrication errors in both lateral misregistrations and etch depths dramatically reduce the performance of the resulting elements. Besides, it is also a quite expensive and lengthy process when the number of masks involved reaches three or more (see Figure 12.68).

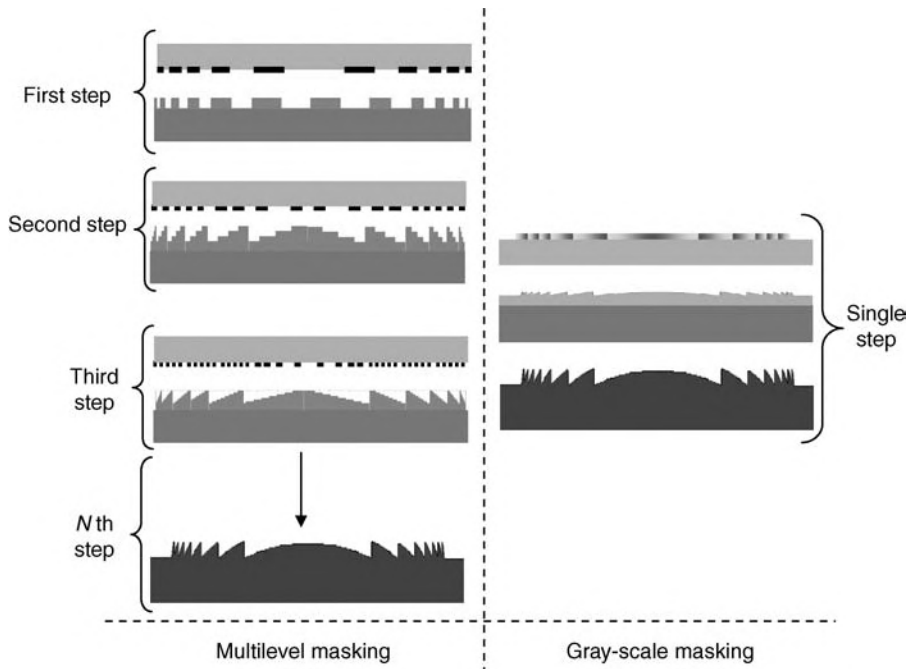


**Figure 12.68** The gray-scale masking process compared to the multilevel binary masking process

The second technique (especially direct dosage modulation) can provide much better alignment accuracy and etch depth control than the first technique. However, the characterization of the process is more difficult, and the requirements on the writing machines are more critical: therefore such a technique cannot only be implemented in traditional IC fabrication units. Furthermore, as the final element is the wafer directly, there is no tooling involved, and therefore it is not a technique that can be used in a production mode.

Below is a fabrication scheme that uses the best of both techniques, namely:

- direct write to reduce the successive alignment and etch depth errors; and
- the use of a tooling in order to enable a production mode.

Such a fabrication scheme should use a single mask to produce a quasi-analog surface relief directly. As the mask therefore has to have gray-scale amplitude modulations, such techniques are called gray-scale masking lithography.

When it comes to producing a multilevel or quasi-analog surface-relief profile via optical lithography, gray-scale lithography seems to be the holy grail. Such techniques are especially well suited for refractive micro-optics, hybrid refractive/diffractive optics and continuous-profile diffractives.

There have been numerous gray-scale lithographic techniques presented in the literature, some using standard binary Cr on glass masks and some using exotic gray-scale materials [23–29]. These various techniques will be reviewed in this section.

### 12.7.1 ‘Binary Gray-scale’ Lithography

The use of small binary structures to implement gray shades is not a new technique. In the Middle Ages, engravers used this technique to produce magnificent engravings that can reveal to the human eye a

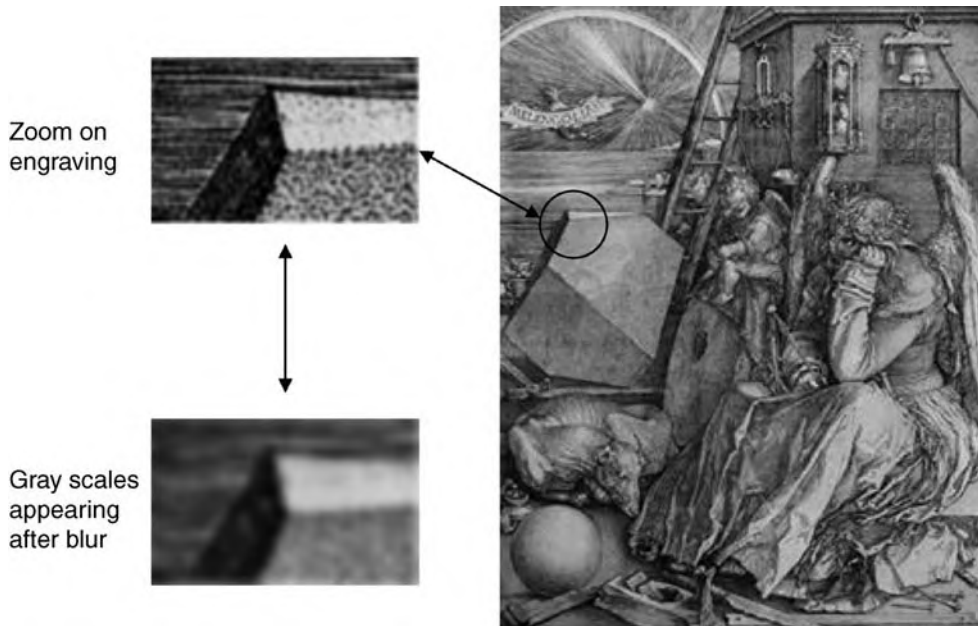


**Figure 12.69** ‘Melancholia I’, a binary gray-scale engraving by Albrecht Dürer (1514)

wide range of gray tones, giving the impression that the image is actually really composed of shades of gray.

Figure 12.69 shows the famous engraving ‘Melancholia I’ by the German Renaissance master Albrecht Dürer (1471–1528). It is an allegorical composition that has been the subject of very many interpretations. The figure shows a portion of a polyhedron, which produces a nice gray-scale range when a blurring filter is applied to the binary engraving (see the bottom inset).

It is interesting to note that the theme of Dürer’s engraving is alchemy (attempting to synthesize substances described as possessing unusual properties) and that the polyhedron coincides with the 3D arrangement of elements in a photonic crystal or a metamaterial (see Chapter 11), which, in essence, is optical alchemy (attempting to synthesize materials described as possessing unusual properties – i.e. a negative index).

Lithographic binary gray-scale masks have been investigated for many years. They use standard Cr on glass binary masks to integrate gray scales exactly as a laser printer would use binary black pixels to implement gray shades on an image printed on a paper. Among the various gray-scale encoding methods, two of the main ones are presented in Figure 12.70:

- pulse density modulation encoding; and
- pulse width modulation encoding.

Such pulse modulation encodings are used in various physical output devices, such as laser printers, newspaper printing machines and so on.

Figure 12.71 shows some examples of elements produced by gray-tone masking through a rasterized Cr mask.

It is very interesting to point out the analogy between binary gray-scale encoding and the effective medium theory (EMT), which is described in Chapter 11. In the EMT method, the phase is encoded by

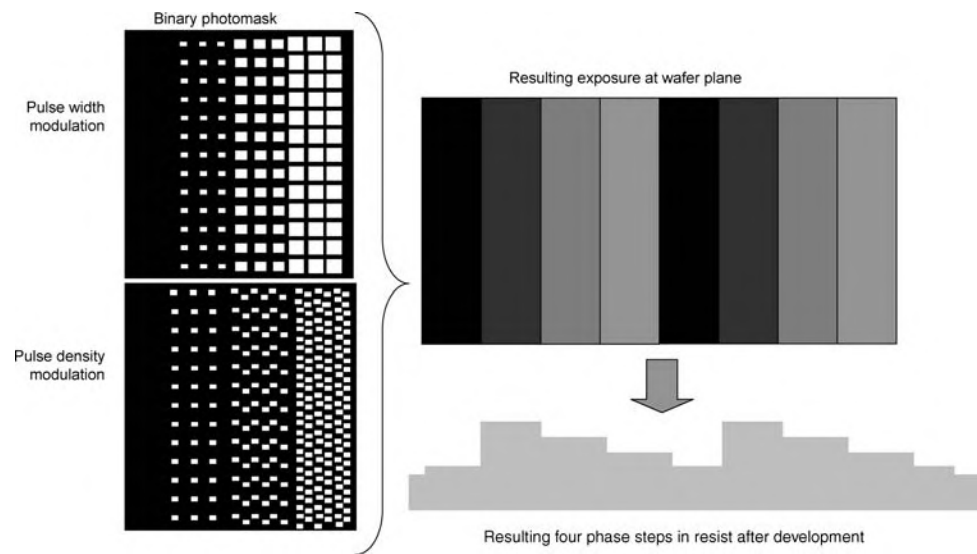


**Figure 12.70** Binary gray-scale pulse modulation encoding in a binary Cr on glass mask

pulse density or pulse width modulations of phase structures rather than by using amplitude structures as is done here.

Although this is a relatively easy technique for implementing gray scales, there are many drawbacks to such an approach to gray-scale masking:

- the smallest feature size in the mask (the CD) has to be much smaller than the smallest feature in the final element after resist development;
- the size of the fabrication data file can become very large; and
- the resolution lithographic system has to be intentionally reduced to produce averaging in the aerial image (blur).

Therefore, several real gray-scale masking techniques have been developed. They are reviewed in the following sections.

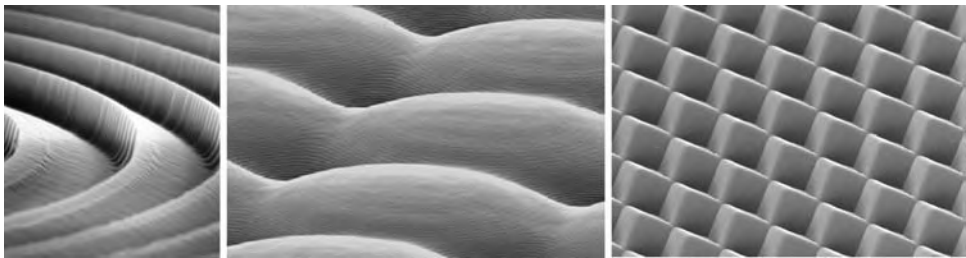


**Figure 12.71** Examples of analog surface-relief elements produced by binary gray-tone masking (courtesy of Dr K. Reimer, of the ISIT Fraunhofer Institute für Silizium Technologie)

### 12.7.2 Analog Gray-scale Masking Lithography

Several gray-scale lithography techniques have been reported since the mid-1990s. They have in common the fabrication of a gray-scale mask or reticle used in projection lithography in order to expose a photoresist directly in an analog way and thus yield an analog surface-relief modulation.

#### 12.7.2.1 The Inorganic GeSe Resist Masking Technique

One of the critical features in gray-scale masking, besides the number of gray scales fabricable, is the resolution that one can achieve with this technology. In the previous halftone binary masking technique, the resolution remains very poor, and suitable only for micro-refractive elements such as micro-prisms and microlenses or refractive Fresnel lenses. In many applications, it is important to combine analog surface relief and small feature sizes down to the micron.

Inorganic gray-scale GeSe resists have been developed, and are mostly referred to as chalcogenide glasses. In the 1970s, inorganic resists were very commonly used in IC fabrication, before industry moved to organic resists, which are now massively used in the semiconductor industry. Inorganic resists are coated in the wafer by either sputtering or CVD methods. Selenium is one of the key ingredients of such resists. These resists are then exposed, and developed in a dry process (RIE) rather than in a wet process like organic resists (and photographs). Exposure does not therefore involve any polymerization process.

The resist used here is a compound belong to the IV main group of the Periodic Table of Elements, which includes germanium selenide and arsenic sulfide (not the healthiest resist). The structure of a GeSe inorganic resist is columnar vertical, therefore providing a very good resolution (no diffusion takes place, and the flow is concentrated in a vertical way). In the GeSe gray-scale technique, such a layer is sputtered on top of the wafer (about 100 nm thick). An additional silver selenide layer (about 10 nm thick) is deposited on top of the inorganic resist.

The exposure of the resist can be done either via e-beam or laser beam writing, using the same dosage modulation technique as described earlier for direct analog write methods. During exposure, the silver selenide is diffused into the underlying GeSe layer, and the diffusion strength is function of the e-beam or laser beam dosage. The resist is then dry developed (that is, RIE etched). The etching rate is dependent on the quantity of silver in the GeSe layer, and thus provides a smooth surface profile in the inorganic resist. This surface profile is very thin (10–100 nm); however, due to the relative opacity of such inorganic resist, it is sufficient to provide a real amplitude gray scale through the mask for UV light. Figure 12.72 shows the GeSe gray-scale resist fabrication and exposure process.

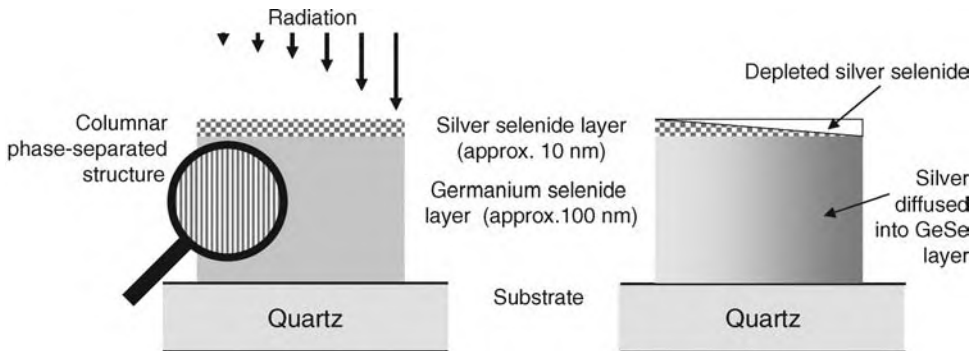


**Figure 12.72** An inorganic GeSe resist coating and exposure for producing an amplitude gray-scale mask



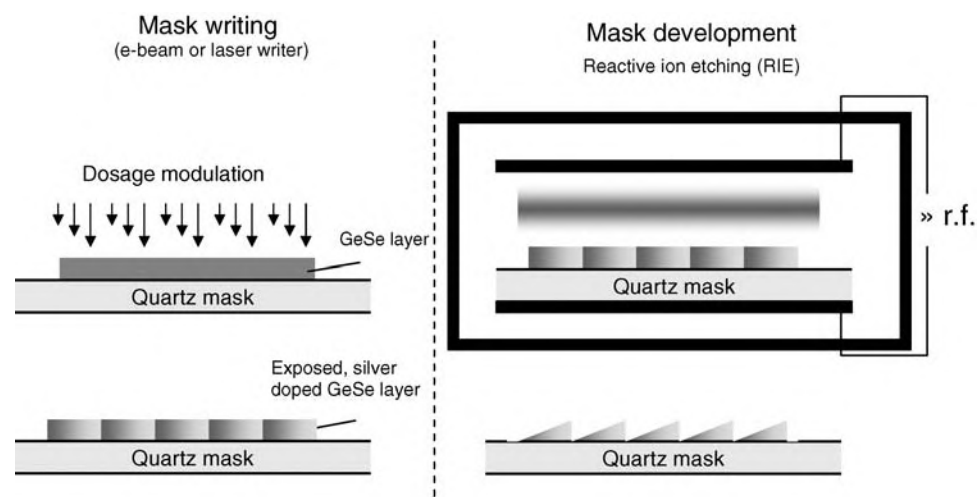


**Figure 12.73** The dry development of a GeSe inorganic resist

The mask is then developed in an RIE chamber to reveal the amplitude modulation in the resist layer (see Figure 12.73).

Figure 12.74 shows such an amplitude gray-scale inorganic resist mask (a detail of a Fresnel diffractive lens).

The mask (or rather, reticle) is patterned at  $5\times$  and then used in an i-line stepper lithographic system to transfer the analog gray-scale modulation into an analog surface-relief modulation in conventional organic resist spun on the wafer (see Figure 12.75).

Following this lithographic step, a proportional RIE etch transfers the surface profile into the underlying wafer, as has been described for previous direct write methods (see previous sections). Figure 12.76 shows a blazed grating fabricated by the GeSe gray-scale masking technique and transferred into the underlying substrate by proportional RIE.

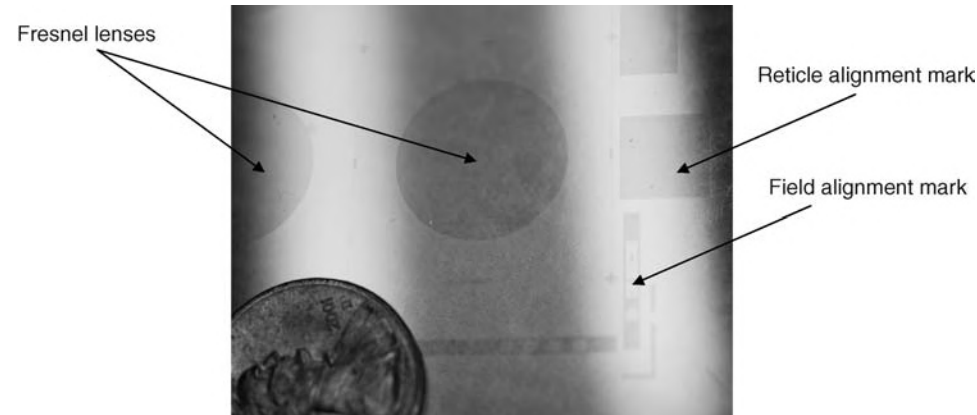


**Figure 12.74** An example of a GeSe inorganic gray-scale mask

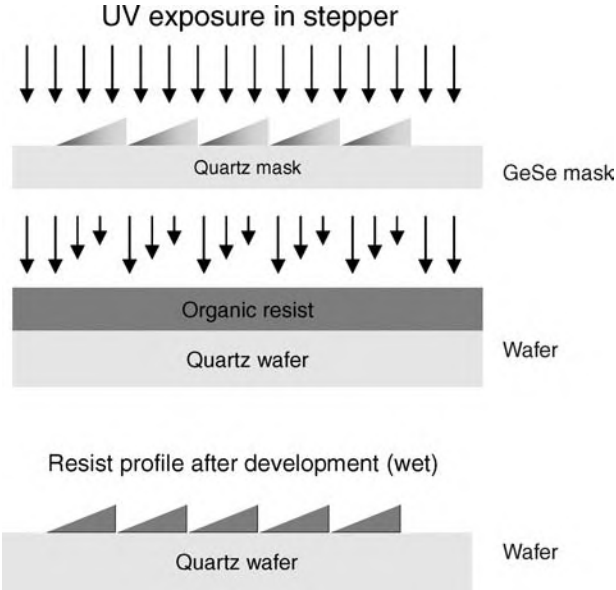


**Figure 12.75** The lithographic transfer of gray-scale amplitude into analog surface relief

This technique produces high-resolution features along with 256 or more gray scales. However, due to the complexity of this technique (using inorganic resist and dry development), an alternative technique is described below – the HEBS glass gray-scale masking technique, which is now widely used in industry.

**12.7.2.2 The HEBS Glass Masking Technique**

High Energy Beam Sensitive Glass (HEBS) glass is a gray-scale masking technique used in many applications today, ranging from MEMS and microfluidics to digital optics and micro-optics.

HEBS glass is made of low-expansion zinc borosilicate. This glass contains alkali to facilitate ion exchange reactions that make the glass sensitive to high-energy beams, electron or laser beams. The ion exchange process is carried out for long enough to cause silver ions to diffuse into the glass to a depth of up to 3  $\mu\text{m}$ . This process is conducted at a temperature above 320  $^{\circ}\text{C}$ , and complex crystals containing silver-alkali halide are formed. The chemical reaction to produce opaque regions of silver atoms is achieved by exposing the glass to high-energy beams; that is, electron beams of more than 100 kV. So the principle behind this mask is that HEBS glass changes its opacity when exposed to a beam of high-energy electrons or a laser (see Figure 12.77).

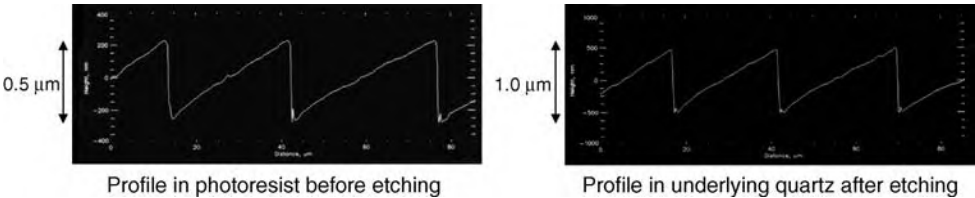


**Figure 12.76** An example of a blazed grating fabricated by a GeSe gray-scale masking technique. Reproduced by permission of Walter Daeschner

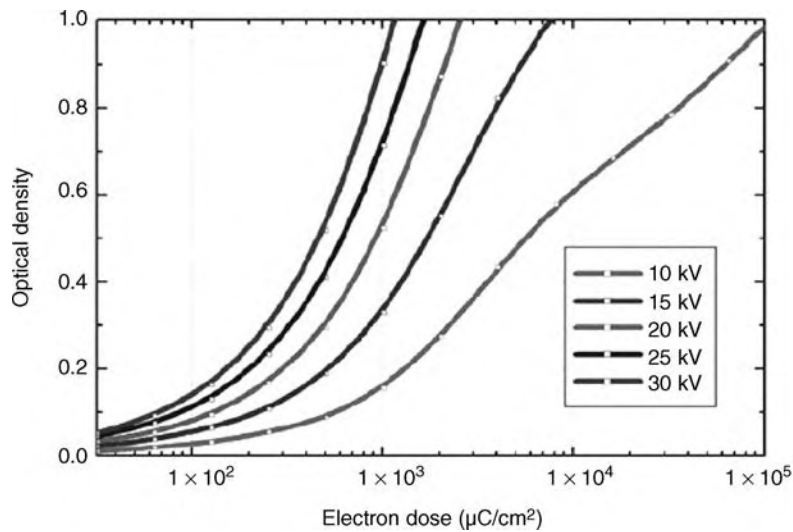


**Figure 12.77** Optical density versus electronic dosage in HEBS glass

For example, gray-scale masks may be generated in an electron-beam writer by varying the dosage of electrons striking different areas of a HEBS glass plate, in a similar way to the direct analog e-beam write of the previous section. Figure 12.78 shows such a HEBS glass including two CGHs and a diffractive Fresnel lens exposed through e-beam writing with 16 different dosages.

Figure 12.79 shows examples of final micro-optical structures in a wafer fabricated with HEBS glass gray-scale masking process and subsequent proportional RIE transfer into quartz.

However, a HEBS gray-scale mask cannot provide enough optical density at wavelengths shorter than 350 nm for high-resolution component fabrication, and it is even not applicable when the exposure wavelength is below 300 nm (which is the case for most steppers) because of the intrinsic properties of HEBS glass. Also, due to its internal structure, HEBS glass does not reproduce the resolution of the

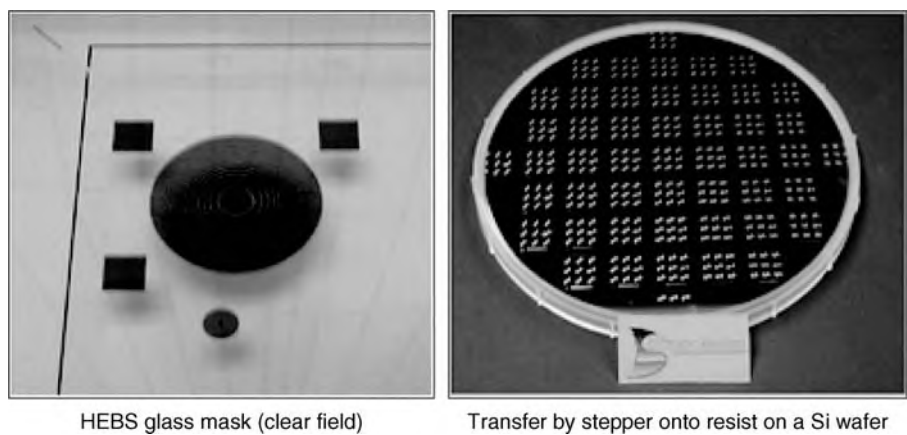


**Figure 12.78** An example of a HEBS grayscale glass mask. Reproduced by permission of Walter Daeschner

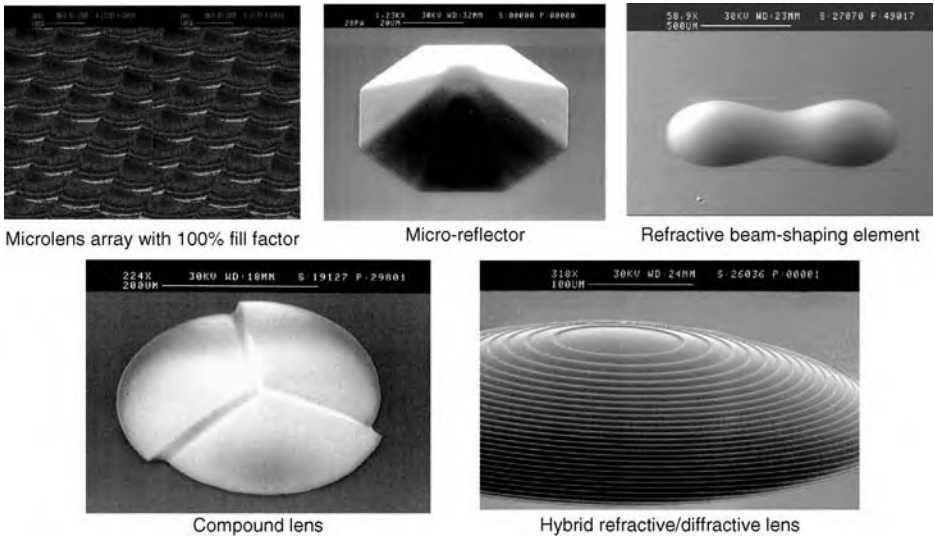


**Figure 12.79** Etched micro-optical elements fabricated via HEBS gray-scale masking (courtesy of Ernst-Bernhart Kley, FSU Jena)

previous GeSe inorganic resist. The beam is more or less diffused in all directions during exposure. Another gray-scale masking technology has been developed recently, which has the advantages of HEBS and yields high resolution.

**12.7.2.3 The Carbon-based Glass Masking Technique**

The carbon-based Light Attenuating Film (LAF) technique has been proposed recently [30, 31] to overcome the limitations of HEBS glass when high-resolution features are required in the final element. Figure 12.80 illustrates the schematic fabrication process of a gray-scale technology utilizing a new carbon-based mask material and improved lithographic/etching processes.

An e-beam writer is used to write on a ZEP 7000A e-beam resist. After exposing the ZEP/LAF/quartz plate in the e-beam writer, the resist is developed in ZED 750 developer. The developed resist pattern is then transferred into the light-attenuating film by anisotropic etching (RIE). The final mask is produced after the etching step. The transmittance of the mask is related to the LAF thickness:

$$T = e^{-\kappa d} \tag{12.6}$$

where  $\kappa$  is the absorption coefficient of the LAF. Figure 12.81 shows the transmittance of a carbon-based LAF gray-scale mask and the transmittance of HEBS glass as a function of wavelength. The figure shows a higher optical density than can be achieved by a HEBS mask when using a wavelength below 300 nm (the wavelengths used in steppers). The shorter the wavelength, the higher the optical density. Because of the higher optical density of the LAF mask, it can accommodate more phase levels than a HEBS mask.

When the gray-scale mask is used in lithographic and dry etching processes, with special attention paid to preserving the analog features in the mask, many gray-scale device structures can be fabricated without going through the cumbersome lithographic and etching processes involving multiple masks. Figure 12.82 shows an AFM picture of an off-axis diffractive lens with a height of about 1  $\mu\text{m}$  and periodicities ranging from 10 to 12  $\mu\text{m}$ , fabricated in quartz by employing the new gray-scale mask and 248 nm stepper

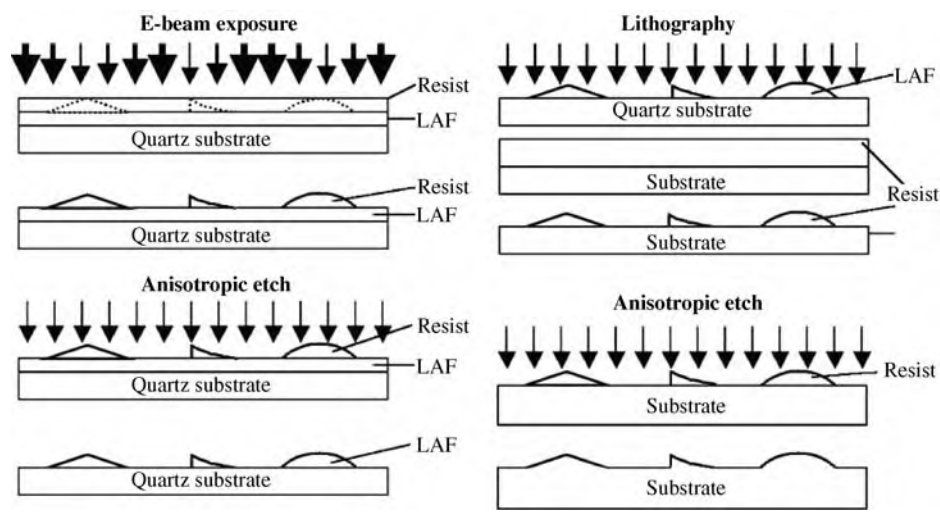


**Figure 12.80** The fabrication process of a carbon-based LAF gray-scale mask. Reproduced by permission of UCSD, courtesy of Professor Sing Lee and Dr Zhou Zhou

lithography. Figure 12.82 also shows a microscopic picture of a corner cube array (with a depth of about  $10\text{ }\mu\text{m}$  at its apex), fabricated in silicon through the use of the gray-scale mask and an i-line aligner.

The gray-scale structures in Figure 12.82 have surface smoothnesses corresponding to 64 thickness or phase levels on the left and 256 levels on the right.

More complex structures can be fabricated via gray-scale lithography. As an example of a 256 phase relief profile, the topology of the Alsace region in France has been fabricated in a carbon-based LAF gray-scale mask. Figure 12.83 shows the desired gray-scale pattern on the left and the fractured GDSII pattern

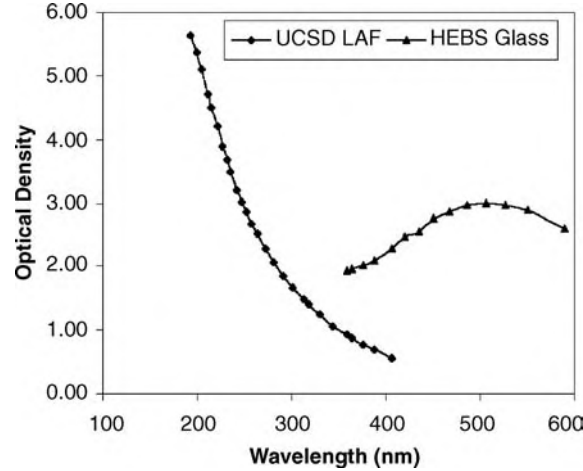


**Figure 12.81** The transmittance of a carbon-based LAF and HEBS glass versus wavelength. Reproduced by permission of UCSD, courtesy of Professor Sing Lee and Dr Zhou Zhou

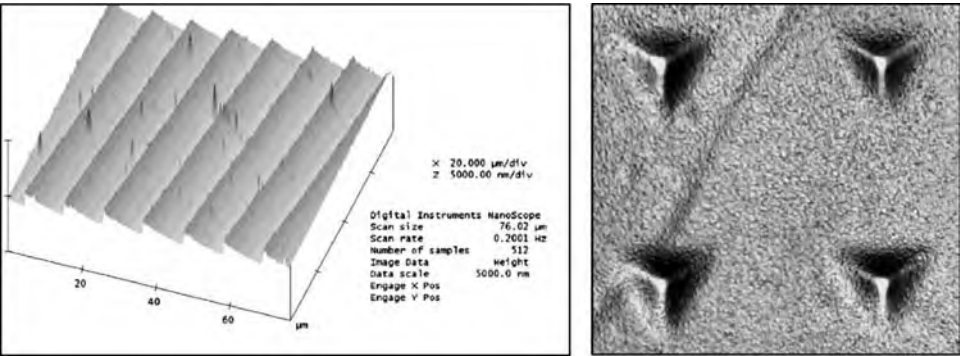


**Figure 12.82** The gray-scale lithography process – an off-axis diffractive lens array of micro-corner cubes. Reproduced by permission of UCSD, courtesy of Professor Sing Lee and Dr Zhou Zhou

on the right (vector format). Such a GDSII file uses the full available 256 GDSII layer. Note that the fracture can be rather complex due to the complexity of the pattern.

Next, the gray-scale mask is fabricated and transferred into an analog resist via a DUV stepper. Figure 12.84 shows the resulting photomask (left) and the final pattern transferred in quartz (see the AFM plot on the right-hand side). The gray-scale photomask here is five times larger than the final pattern, since a 5 $\times$  reduction stepper has been used.

Figure 12.85 shows another 256-level surface-relief profile, which includes similar surface topology and the surface profiles of the three Nobel prize laureates from Alsace (Dr Albert Schweitzer, Professor Alfred Kastler and Professor Jean-Marie Lehn). A zoom over the 3D face of Dr Schweitzer is also shown in Figure 12.85.

Table 12.7 summarizes the four gray-scale masking techniques presented in this section (binary halftone masks, the GeSe inorganic resist, HEBS glass and the carbon-based LAF mask).

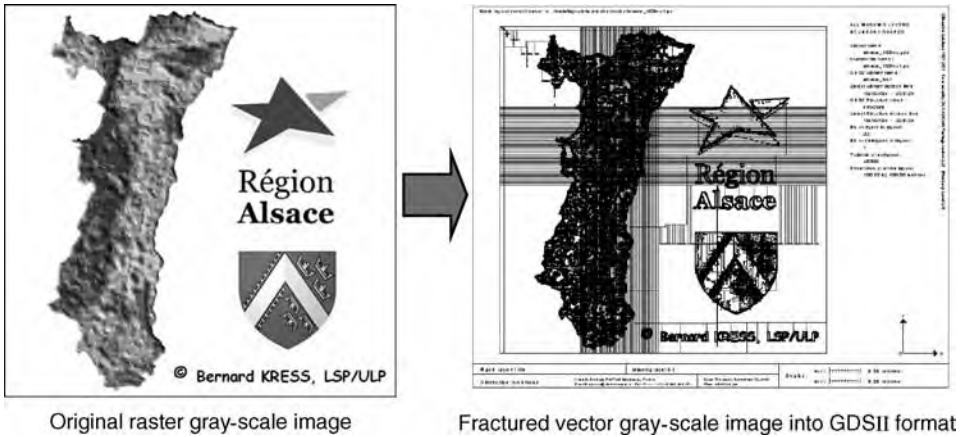


**Figure 12.83** Left: the 256 gray-scale levels of the 3D topology of the Alsace region in France. Right: the resulting vector fractured pattern split into 256 different GDSII levels

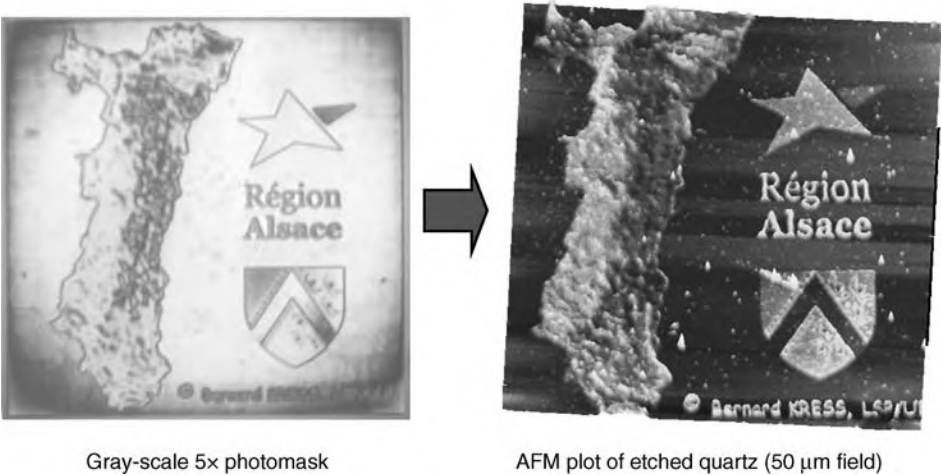


**Figure 12.84** Left: the resulting gray-scale photomask of the Alsace topographic map. Right: the final 256-relief profile in quartz of the Alsace topographic map

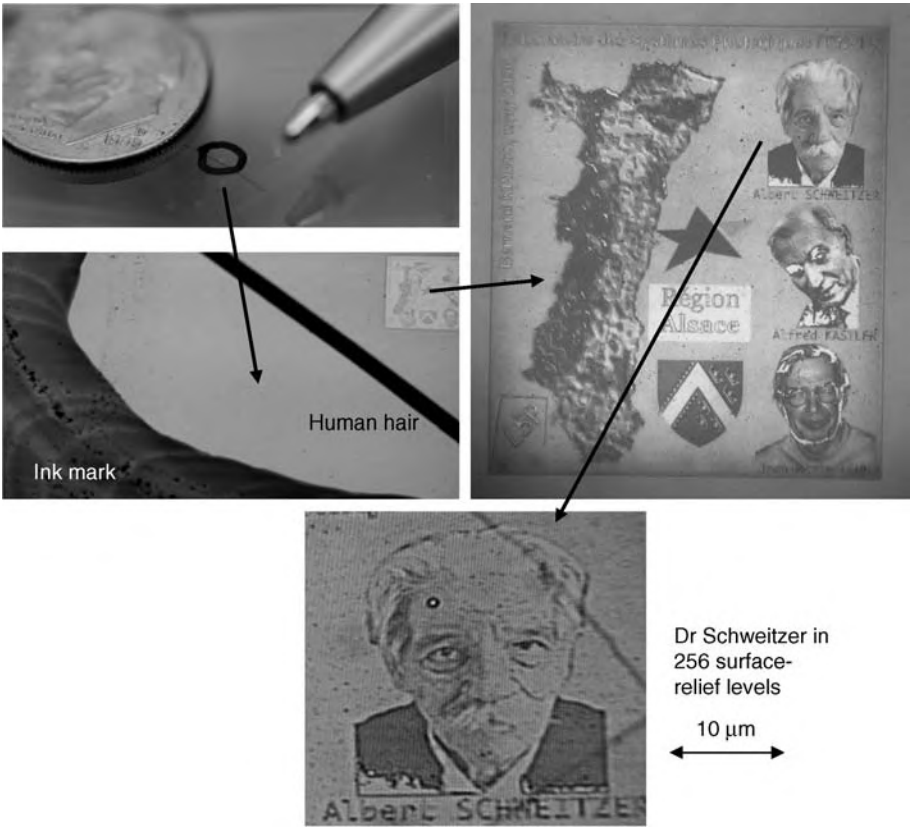


**Figure 12.85** Elements transferred in quartz over 256 levels by gray-scale lithography and proportional etching

**Table 12.7** A summary of the four gray-scale lithographic techniques presented in this section

| Gray-scale masking technique | Resolution | Optical density | Wavelength           | Fabrication |
|------------------------------|------------|-----------------|----------------------|-------------|
| Binary halftone masking      | Poor       | Low             | Nonsensitive         | Easy        |
| Ge–Se inorganic resist       | Very high  | Low             | Acceptable in UV     | Difficult   |
| HEBS mask                    | Medium     | Medium          | Not acceptable in UV | Medium      |
| Carbon-based LAF mask        | High       | High            | Acceptable in UV     | Medium      |

## 12.8 Front/Back Side Wafer Alignments and Wafer Stacks

In some applications, it is desirable to stack processed wafers one on top of the other, or to align patterns front and back on a single wafer.

### 12.8.1 Front and Back Side Wafer Pattern Alignment

Aligning front and back side on a single wafer is useful, for example, to align digital optics lenses. When using Fourier elements, the only alignment to be performed is angular, since these elements are periodic in the  $x$  and  $y$  directions (see Chapter 6). Such front/back side alignment can be performed in a back side mask aligner. However, such aligners are very scarce in industry (they are used especially for MEMS applications), so an alternative solution to this problem is proposed.

Below are described two techniques to align digital elements on the back side of the wafer in a regular mask aligner. One involves using sets of alignment microlenses and the other uses the Talbot self-imaging effect presented in Chapter 5.

#### 12.8.1.1 The Use of Alignment Microlenses and CGHs

Figure 12.86 shows a wafer with alignment microlenses etched on the front side, which have exactly the right focal length in the wafer material to focus incoming collimated light on the back side.

The back side is coated with photoresist, and the wafer is flood-illuminated in a mask aligner. The focus from the microlenses creates a feature exactly on the optical axes of these lenses. After etching of these features, one can therefore precisely align another pattern on the back side of that wafer, by using these marks. The microlenses can be replaced by Fresnel CGHs, which focus into sets of alignment fiducials, such as crosses and so on.

#### 12.8.1.2 The Use of Talbot Self-imaging

Another technique consists of using the Talbot self-imaging effect (see Chapter 5). A set of gratings is patterned within a Process Control Monitor (PCM) region of the photomask. Such gratings can be either circular gratings or orthogonal linear gratings. The periods of the gratings are carefully calculated so that they produce a self-image on the back side of the wafer (see Figure 12.87).

The alignment here is performed in real time, by producing moiré effects between the two gratings (see Figure 12.88). There is no physical grating implemented on the back side, but only an image projected from the front side (by adequate laser illumination). This technique has the advantage of not requiring additional lithographic processes (such as the etching of the alignment marks in the previous microlens alignment technique).

Figure 12.89 shows an example of a diced-out element including an array of 12 collimation and focusing diffractive lenses aligned on the front and back side of a fused silica wafer, for an application in a 10 Gb/s optical ethernet transceiver (see Chapter 16).



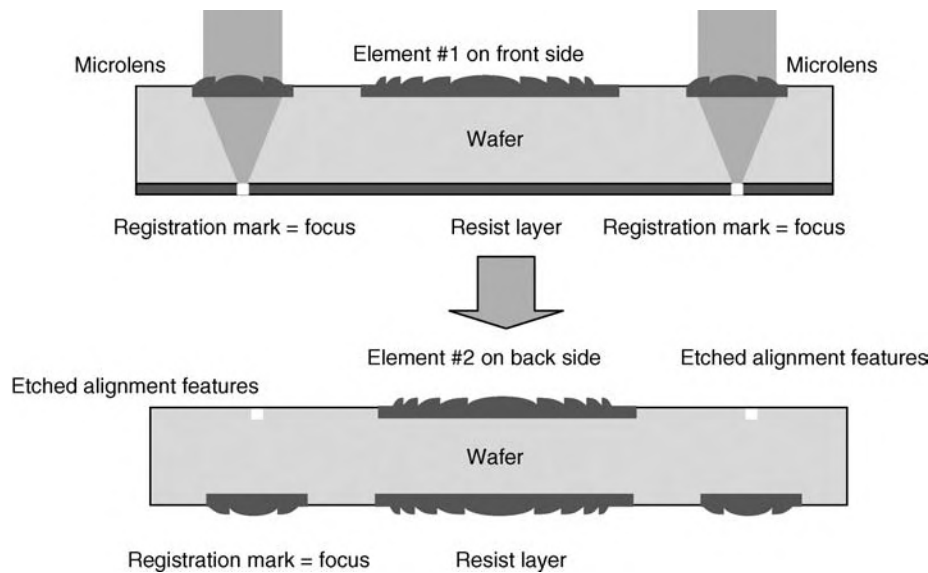


**Figure 12.86** Front/back side alignment using microlenses on glass wafers

*12.8.2 Wafer-to-wafer Alignment*

Aligning a wafer to another wafer, and potentially to a third wafer, is very desirable since many elements can be patterned on a wafer and tightly aligned during lithography. If the elements on the various wafers are to be used in an optical train application, one can dice the various elements after aligning the multiple wafers. Therefore, there is no need to align the individual dies after dicing, since everything is aligned prior to the dicing step.

One can use mechanical alignment techniques such as etched posts in the wafer, which coincide with recesses in the other wafer. Another technique consists of using solder bumps, as used in flop chip bonding, also in etched recesses in touching wafers, and let the forces of UV curing shrinkage align the wafers naturally to the etched recesses in the various wafers.

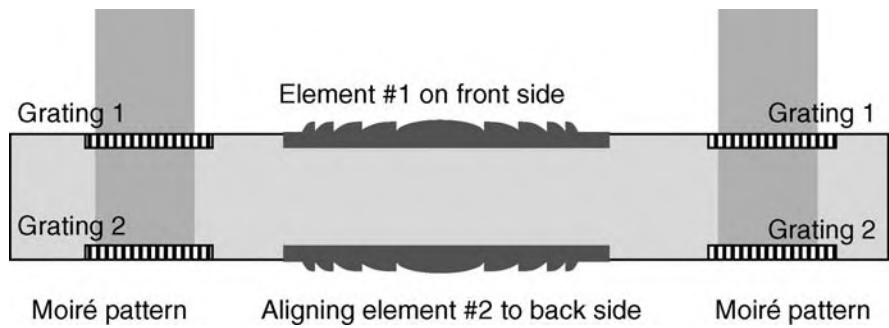


**Figure 12.87** Front/back side alignment using Talbot self-imaging

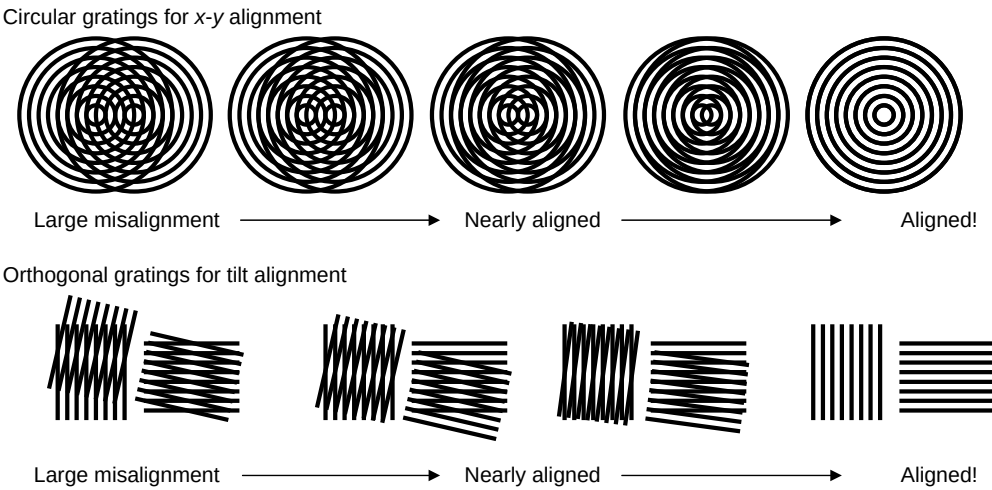


**Figure 12.88** An example of Talbot imaging alignment using moiré effects

Dual wafer-to-wafer alignment can also be performed by using the previously described front/back side single-wafer alignment techniques (through the use of microlens arrays or Talbot self-imaging). Such a technique has already found its way into a commercial product: see the planar stacked camera objective example presented in Chapters 4 and 16.

**12.9 A Summary of Fabrication Techniques**

Table 12.8 shows the various lithographic fabrication technologies investigated in this chapter and the resulting typical diffraction efficiencies, as well as their respective advantages and limitations.

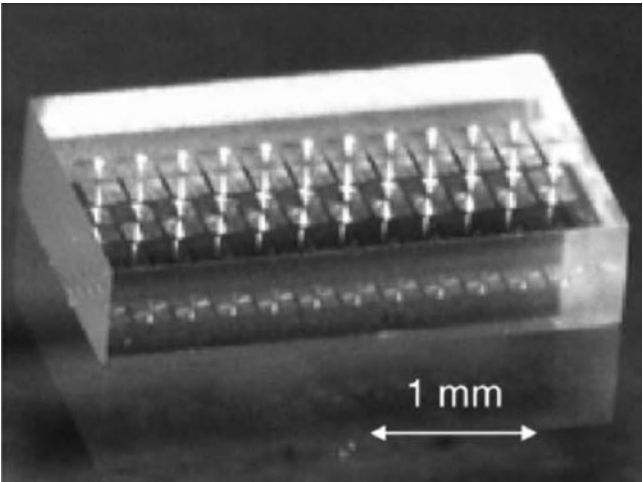


**Figure 12.89** An example of front/back side diffractive lens alignment on a fused silica wafer

**Table 12.8** A summary of the various lithographic fabrication technologies investigated

| Lithography technique                    | Number of surface levels | $\eta$ (%) theoretical | $\eta$ (%) real | Number of mask | Number of litho/etching processes | Resolution                | Fabrication costs                  | Advantages                                                     | Limitations                                                     |
|------------------------------------------|--------------------------|------------------------|-----------------|----------------|-----------------------------------|---------------------------|------------------------------------|----------------------------------------------------------------|-----------------------------------------------------------------|
| Binary amplitude (mask or reticle)       | Binary                   | <8%                    | <8%             | 1              | 0/0                               | high                      | Minimal (only mask)                | Cost and resolution                                            | Only one element, very low efficiency                           |
| Binary phase in resist                   | Binary                   | 40.5%                  | $\approx$ 30%   | 1              | 0/0                               | high                      | Minimal (only mask exposure)       | Cost and resolution                                            | Profile in resist (Short life, very sensitive)                  |
| Direct binary write in resist            | Binary                   | 40.5%                  | $\approx$ 30%   | 1              | 0/0                               | Extremely high            | Minimal (only mask exposure)       | Cost and resolution                                            | Profile in resist (Short life, very sensitive)                  |
| Direct analog write (laser or e-beam)    | Quasi-analog             | >80%                   | $\approx$ 80%   | 1              | 0/0                               | Very high                 | Medium, requires dosage modulation | Direct write with laser or e-beam, no lateral alignment issues | Only one element, profile in resist, requires proportional etch |
| Binary phase etched in wafer             | Binary                   | 40.5%                  | $\approx$ 35%   | 1              | 0/1                               | High                      | Small (only mask and RIE)          | Cost and resolution, stable material                           | Only one element                                                |
| Multilevel ( $N$ masks for $2^N$ levels) | 4                        | 81%                    | $\approx$ 75%   | 2              | 2/2                               | Medium                    | Low                                | High efficiency' many elements                                 | Mask misalignments and etch depth errors                        |
|                                          | 8                        | 95%                    | $\approx$ 85%   | 3              | 3/3                               | High (difficult)          | Medium                             | High efficiency' many elements                                 | Costs of multiple lithography steps and related misalignments   |
|                                          | 16                       | 99%                    | $\approx$ 90%   | 4              | 4/4                               | High required (difficult) | High                               | High efficiency' many elements                                 | Costs and cascaded misalignments                                |
|                                          | >16                      | >99%                   | $\approx$ 90%   | >4             | >4/4                              | Very high required        | Very high                          | High efficiency' many elements                                 | Very high costs                                                 |
| Gray-scale lithography                   | Analog                   | >90%                   | $\approx$ 90%   | 1              | 1/1                               | Medium                    | High                               | High efficiency' many elements' no misalignment                | Costs                                                           |

In addition to the lithographic techniques reviewed in this chapter (mask pattern generators, mask aligners, stepper, RIE etching etc.), there are a variety of other techniques available today in industry, which are, however, not available as standard fabrication methods in industry. The following chapters will discuss some of these exotic fabrication techniques, such as EUV lithography, maskless lithography, soft lithography, immersion lithography and X-ray lithography.

## References

- [1] T.J. Suleski and R.D. Kolste, 'A roadmap for micro-optics fabrication', *Proceedings of SPIE*, **4440**, 2001, 1–15.
- [2] T.J. Suleski and D.C. O'Shea, 'Gray scale masks for diffractive optics fabrication: I. Commercial slide imagers', *Applied Optics*, **34**(32), 1995, 7507–7517.
- [3] D.C. O'Shea and W.S. Rockward, 'Gray Scale Masks for Diffractive Optics Fabrication: II. Spatially Filtered Halftone Screens', *Applied Optics*, **34**(32), 1995, special issue.
- [4] G.J. Swanson, 'Binary Optics Technology: the Theory and Design of Multi-level Diffractive Optical Elements', MIT Lincoln Lab Technical Report 854, 1989.
- [5] W.B. Veldkamp, 'Binary optics: the optics technology of the decade', Presented at the 37th International Symposium on Electron, Ion and Photon Beams, San Diego CA, June 1993.
- [6] M.W. Farn, 'Design and fabrication of binary diffractive optics', Ph.D. thesis, Stanford University, Stanford, CA, September 1990.
- [7] T. Fujita, H. Nishihara and J. Koyama, 'Blazed gratings and fresnel lenses fabricated by electron-beam lithography', *Optics Letters*, **7**, 1982, 578–580.
- [8] J. Jahns and S.J. Walker, 'Two-dimensional array of diffractive microlenses fabricated by thin film deposition', *Applied Optics*, **29**, 1990, 931–936.
- [9] G.J. Swanson and W.B. Veldkamp, 'Diffractive optical elements for use in infrared systems', *Optical Engineering*, **28**, 1989, 605–608.
- [10] L. D'Auria, J.P. Huignard, A.M. Roy and E. Spitz, 'Photolithographic fabrication of thin film lenses', *Optics Communications*, **5**, 1972, 232–235.
- [11] H. Kostal, J.J. Wang and F. Thomas, 'Manufacture of multi-level encoded subwavelength optical data storage media', Presentation to Topical Meeting of EOS: Advanced Imaging Techniques, July 2005.
- [12] J. Petty and J.T. Weed, 'OASIS implementation: a Synopsis perspective', presentation at Synopsis, November 2005.
- [13] R.W. Hawley and N.C. Gallagher, 'Efficient electron beam pattern data format for the production of binary computer generated holograms', *Applied Optics*, **29**(2), 1990, 216–224.
- [14] D. Daly, R.F. Stevens, M.C. Hutley and N. Davies, 'The manufacture of microlenses by melting photoresist', *Measurement Science and Technology*, **1**, 1990, 759–766.
- [15] D. Daly, 'Microlens Arrays', Taylor & Francis, London, 2001.
- [16] E.-B. Kley, H.-J. Fuchs and A. Kilian, 'Fabrication of glass lenses by melting technology', in 'Lithographic and Micromachining Techniques for Optical Component Fabrication', E.-B. Kley and H.-P. Herzig (eds.), SPIE Vol. 4440, 2001, 85–92.
- [17] M.T. Gale, 'Direct writing of continuous-relief micro-optics', in 'Micro-optics: Elements, Systems and Applications', H.P. Herzig (ed.), Taylor & Francis, London, 1997.
- [18] M.T. Gale, M. Rossi, H. Schütz and P. Ehbets, 'Continuous relief diffractive optical elements for two-dimensional array generations', *Applied Optics*, **32**, 1993, 2526–2533.
- [19] T. Hessler, M. Rossi, R.E. Kunz and M.T. Gale, 'Analysis and optimization of fabrication of continuous-relief diffractive optical elements', *Applied Optics*, **37**(19), 4069–4079.
- [20] M.T. Gale, M. Rossi, J. Pedersen and H. Schütz, 'Fabrication of continuous-relief micro-optical elements by direct laser writing in photoresist', *Optical Engineering*, **33**, 1994, 3556–3566.
- [21] T. Shiono and K. Setune, 'Blazed reflection microFresnel lenses fabricated by electron beam writing and dry development', *Optics Letters*, **15**, 1990, 84–86.
- [22] T.J. Suleski, B. Baggett, W.F. Delaney, C. Koehler and E.G. Johnson, 'Fabrication of high spatial frequency gratings through computer generated near-field holography', *Optics Letters*, **24**, 1999, 602–604.
- [23] Y. Oppliger, P. Sixt, J.M. Stauffer *et al.*, 'One step 3D shaping using a gray tone mask for optical and microelectronic applications', *Microelectronic Engineering*, **23**, 1994, 449–454.

- [24] W. Daeschner, P. Long, M. Larsson and S. Lee, 'Fabrication of diffractive optical elements using a single optical exposure with a grey level mask', *Journal of Vacuum Science and Technology B*, **13**(6), 1995, 2729–2731.
- [25] W. Daeschner, P. Long, R. Stein, C. Wu and S. Lee, 'One step lithography for mass production of multilevel diffractive optical elements using High Energy Beam Sensitive (HEBS) grey-level masks', *Proceedings of SPIE*, **2689**, 1996, 153–155.
- [26] Canyon Materials Inc., 'HEBS glass photomask blanks', Product information 96-01, Canyon Materials Inc., 6665 Nancy Ridge Drive, San Diego, CA 92121, USA.
- [27] E.-B. Kley, F. Thomas, U.D. Zeitner, L. Wittig and H. Aagedal, 'Fabrication of micro-optical surface profiles by using gray scale masks', *Proceedings of SPIE*, **3276**, 1997, 254–262.
- [28] C. Wu, 'Methods of making high energy beam sensitive glasses', U.S. Patent 5,078,771, January 7, 1992.
- [29] W. Daeschner, P. Long, R. Stein, C. Wu and S. Lee, 'General aspheric refractive micro-optics fabricated by optical lithography using a High Energy Beam Sensitive (HEBS) glass grey level mask', *Journal of Vacuum Science and Technology B*, **14**(H.6), 1996, 3730–3733.
- [30] Z. Zhou and S.H. Lee, 'Two-beam-current method for e-beam writing gray-scale masks and its application to high-resolution microstructures', *Applied Optics*, **47**(17), 2008, 3177–3184.
- [31] Z. Zhou and S.H. Lee, 'Fabrication of an improved gray-scale mask for refractive micro- and meso-optics', *Optics Letters*, **29**(5), 2004, 457–458.



# 13

## Design for Manufacturing

In the previous chapter, the various lithographic technologies used today to produce digital optical elements on quartz, silicon or other wafers (i.e. mask patterning systems, optical lithography systems, ion etching systems, etc.) were reviewed. The various processes used to produce multilevel and quasi-analog surface relief (multilevel lithography and gray-scale lithography) were also reviewed.

Based on this review, the current chapter provides solutions to optimize the fabrication processes and associated resolutions to produce micro- and nanostructures that are better adapted to the digital optics realm, as has been done for the IC industry. However, one has to remember that basic IC shapes used in standard semiconductor fabrication can be very different to those used in digital optics (parallelogram-type shapes rather than smooth curved shapes).

Design For Manufacturing (DFM) consists of including machine specific characteristics in the design process of an element in order to optimize the fabrication of that element with this specific machine. In microlithography, DFM consists of introducing numeric lithographic models in the design process of the microstructures (micro-electronics or micro-optics) to be patterned on the wafer. Such DFM processes include software and hardware solutions.

### 13.1 The Lithographic Challenge

In optical lithography, the industry constantly attempts to reduce the minimum printable feature size  $\delta x$  (see Equation (13.1) below), either for IC fabrication or for digital optics fabrication. The ever-growing fabrication needs for new nano-electronic devices and new nanophotonic devices puts high pressure on the IC industry to provide lithographic systems that can print smaller and smaller features on larger and larger wafers [1, 2].

Figure 13.1 shows Moore's law in IC fab, applied to the digital-optics realm, which has been active since the beginning of the IC fabrication era in the early 1960s.

#### 13.1.1 Limitations of Optical Lithography

A typical optical projection lithography tool, as described in Chapter 12 (a stepper, scanner or projection mask aligner) is imaging a binary pattern from a reticle onto a wafer with a specific reduction factor (usually 5 $\times$ ). Figure 13.2 shows the internal structure of a typical lithographic stepper.

The reticle diffracts the incoming light (from a KrF laser or a mercury arc lamp in which the specific line has been isolated – g, h or i) into several diffraction orders (see Figure 13.2). Since the projection lens has a finite numeric aperture (NA), a finite number of these diffracted fields will be recombined and partial

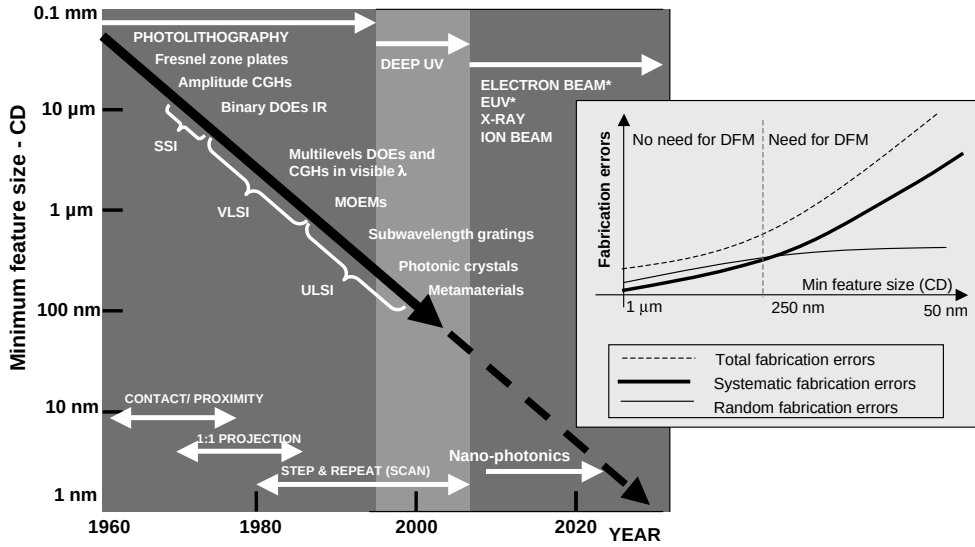


**Figure 13.1** Moore's law in IC fab for digital optics

(reduced) image recovery will occur at the wafer level (focal plane). The finite NA of the projection lens actually acts like a Fourier filter (a low-pass filter).

A reticle pattern contains multiple spatial frequencies (multiple terms in the Fourier expansion of the desired shape). The Modulation Transfer Function (MTF) is a criterion that is used to assess how well the various spatial frequencies are transmitted by the projection lens:

$$MTF(v) = \frac{M_{reticle}(v)}{M_{aerial\ image}(v)} \quad (13.1)$$

Imaging a small pattern on a reticle through a lithographic projection lens onto a wafer thus boils down to a diffraction problem: the condenser lens produces sets of rays that are all diffracted by the reticle in numerous diffraction orders; the lens recombines some of these diffracted rays to produce an aerial image in the vicinity of the resist layer [3, 4]. It is a direct and inverse diffraction problem, solved by analyzing the diffraction phenomenon and calculating the resulting pattern from the rays that are actually used by the projection lens, due to its finite NA. Figure 13.3 summarizes this process.

The physics and the mathematical description of diffraction phenomena are reviewed in detail in the previous chapters in their diverse forms – scalar diffraction (Chapters 1, 2, 5, 6 and 11), extended scalar diffraction (Chapters 7 and 11) and vector diffraction formulations (Chapters 8, 9 and 11).

Thus, according to the set-up in Figure 13.3, one can write

$$\begin{cases} \sin(\theta) - \sin(\theta_i) = \frac{m\lambda}{\Lambda} \\ \Lambda = 2d \end{cases} \quad (13.2)$$



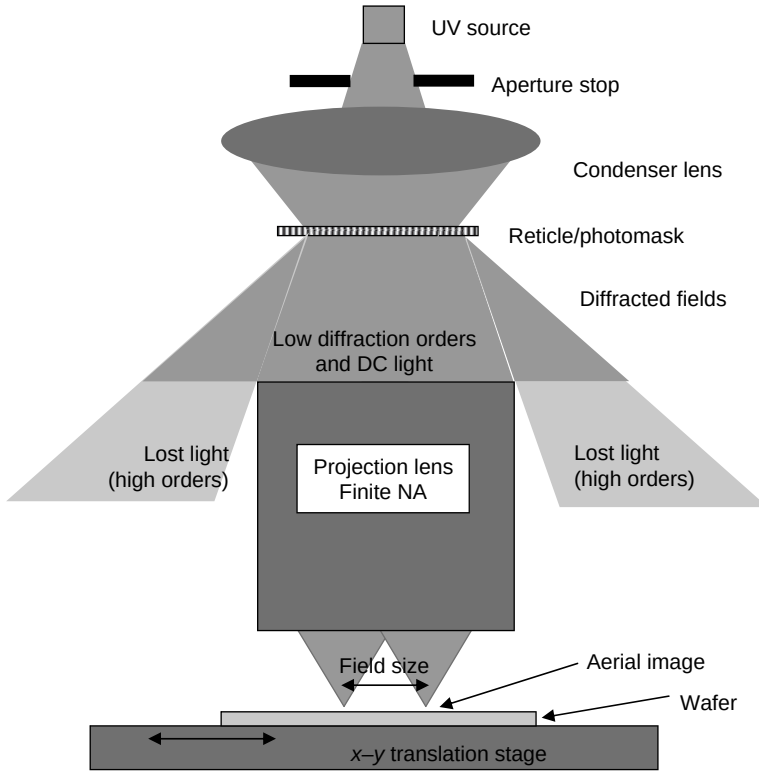


**Figure 13.2** A typical projection lithographic system (stepper)

where  $m$  is the diffraction order considered ( $m = 0, \pm 1, \pm 2, \dots$ ),  $\Delta$  is the period of the pattern considered (locally on the reticle) and  $d$  is the size of the shape (a binary grating with a duty cycle of 50%). The index of refraction around the reticle is 1 (air).

See the binary amplitude Fourier decomposition described in Chapter 5. If all the diffraction orders are intercepted by the projection lens, one can write (for on-axis illumination  $\theta_i = 0$ ):

$$I(x) = I_0 \left( \frac{1}{2} - \frac{2}{\pi} \sum_{m=1}^{\infty} \frac{(-1)^m}{m} \cdot \sin\left(\frac{m\pi}{2}\right) \cdot \cos\left(\frac{m\pi x}{d}\right) \right)^2 \quad (13.3)$$

If only one order is intercepted by the projection lens (e.g. the zero order or the first order), there is no pattern information in the image. In order to produce an image, there have to be at least two beams interfering at the wafer plane. The minimum feature size according to Equation (13.2) is thus given by

$$\delta x = 0.5 \frac{\lambda}{\sin(\theta_0)} \quad (13.4)$$

where  $\theta_0$  is the largest angle captured by the projection lens in on-axis configuration.

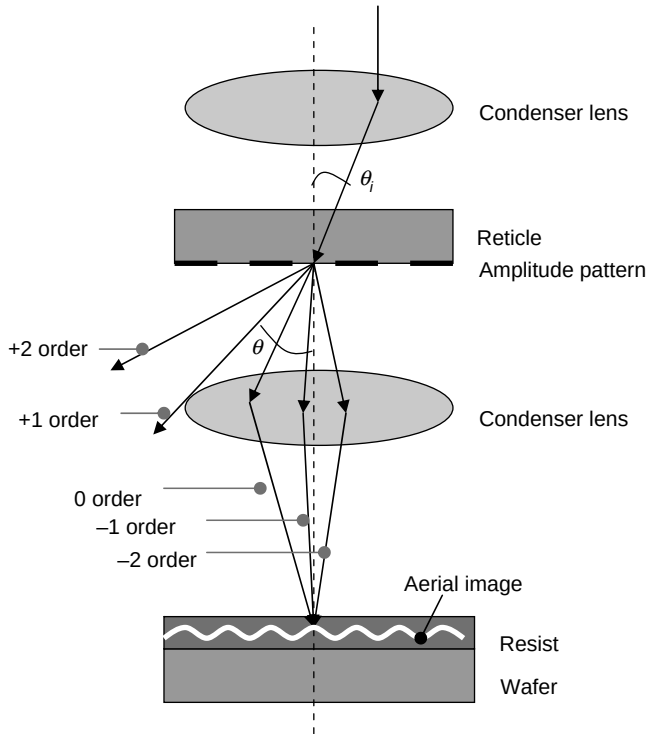


**Figure 13.3** The diffraction problem in lithographic imaging

If all the diffracted angles are used to reconstruct the image (i.e. the lens has an infinite aperture),  $\delta x$  is simply  $d$ , and the shape on the wafer is exactly the shape on the reticle (magnified by the reduction factor of the stepper lens).

When using a finite circular aperture (Appendix C), the Fourier transform of such a shape is a Bessel function. When considering the projection lens as a circular aperture, one can thus write the following expression for the intensity pattern:

$$I(x) = I_0 \cdot \left( 2 \frac{J_1(x)}{x} \right)^2 \quad (13.5)$$

According to the Rayleigh resolution criterion (see also Chapter 5), the maximum resolution is when the Airy disk maximum of one source coincides with the zero of the other Airy disk from another source, which happens at  $x = 2\pi \times 0.61$ . One can then write the following Rayleigh resolution criterion for a lithographic stepper lens:

$$\delta x = 0.61 \frac{\lambda}{NA} = k_1 \frac{\lambda}{NA} \quad (13.6)$$

where the  $NA$  is the sine of the previous angle  $\theta_0$  multiplied by the index of the surrounding medium ( $n = 1$ ; see Equation (13.4)) and  $\delta x$  is the smallest feature achievable by a projection lens with such an  $NA$  for a wavelength  $\lambda$ ; and where  $k_1$  is also called the resolution factor, or 'k-factor', of the projection system, and is equal to 0.61 in the simple case of an on-axis illumination and a circular aperture.

For partial coherence illumination, the previous resolution limit can be rewritten as

$$\delta x = k_1 \frac{\lambda}{(\sigma + 1) \text{NA}} \tag{13.7}$$

where  $\sigma$  is the degree of coherence of the illumination source.

The depth of focus (DOF)  $\delta z$  for an objective lens with specific NA is given by

$$\delta z = \pm 0.5 \frac{\lambda}{\text{NA}^2} = k_2 \frac{\lambda}{\text{NA}^2} \tag{13.8}$$

where  $k_2$  is a factor similar to  $k_1$ , which is dependent on the definition of acceptable imaging and on the type of feature to be imaged.

It is interesting to note that the DOF decreases linearly when the resolution is increased by decreasing the wavelength. However, when the resolution is increased by increasing the NA of the projection system, this same DOF decreases in a quadratic manner.

A typical depth of focus of an i-line stepper can be smaller than the micron, thus demonstrating the need to accurately align the objective lens (400 pounds) to the wafer! However, the depth of focus has to remain on the order of the resist thickness (around 1  $\mu\text{m}$ ), this is why the NA of a stepper should remain below 0.5.

Therefore, when designing a lithographic system, a tradeoff has to be made between the optimal resolution achievable and the optimal DOF of such a system (defining both  $k_1$  and  $k_2$ ).

13.1.2 Increasing the Resolution

In order to follow Moore’s law (Figure 13.1 and Equations (13.6) and (13.7)), one has three ways to reduce the smallest feature size  $\delta x$  that can be printed on the wafer:

- *Reduce the wavelength  $\lambda$ :* reducing the wavelength is a costly process (see Table 13.1). First, one has to produce a source for low  $\lambda$  (such as a UV Excimer laser or an extreme UV source) and make sure all that the projection lenses are transparent to that wavelength (the lenses are usually the critical and most expensive part of a stepper). If they are not transparent, a reflective objective must be used; however, then one needs to work in off-axis mode, and thus the complexity of the objective and the amount of aberrations to control increase dramatically. This is one of the reasons why industry did not use the F<sub>2</sub> DUV laser source (157 nm), which would have required new glass materials, since moving to 248 nm ArF DUV laser illumination produced more problems than it solved.

Table 13.1 Sources used in steppers today

| Year | Source            | Type                | $\lambda$ (nm) | NA   | $k_1$ | $\delta x(\mu\text{m})$ |
|------|-------------------|---------------------|----------------|------|-------|-------------------------|
| 1980 | Mercury arc lamps | G-line              | 436            | 0.28 | 0.96  | 1.50                    |
| 1983 |                   |                     |                | 0.35 | 0.96  | 1.20                    |
| 1986 |                   | H-line              |                | 0.45 | 1.00  | 1.00                    |
| 1989 |                   | I-line              |                | 0.45 | 0.86  | 0.70                    |
| 1992 |                   |                     |                | 0.54 | 0.74  | 0.50                    |
| 1995 | UV laser          | KrF                 | 248            | 0.60 | 0.57  | 0.35                    |
| 1997 |                   |                     |                | 0.93 | 0.50  | 0.25                    |
| 1999 |                   |                     |                | 1.0  | 0.43  | 0.18                    |
| 2001 |                   |                     |                | 0.75 | 0.37  | 0.11                    |
| 2003 | DUV laser         | ARF                 | 193            | 0.85 | 0.45  | 0.09                    |
| 2005 |                   | F <sub>2</sub>      | 157            | 0.90 | 0.45  | 0.06                    |
| 2008 | EUV source        | X-rays (R&D effort) | 13             | 0.20 | 0.50  | 0.03                    |

- *Increase the NA*: increasing the NA (see Table 13.1) is also a costly process, since one will come up against physical limitations in terms of lens sizes and the objective weight. A typical i-line objective lens can weigh up to half a ton of glass (UV-grade fused silica or calcium fluoride) and has a structural metal frame (usually magnesium). However, when migrating to DUV and EUV wavelengths, one has to replace the transmission optics by off-axis reflective optics, which is another challenge to overcome.
- *Reduce the resolution factor  $k_1$* : Reducing the resolution factor  $k_1$  (see Table 13.1) seems to be one of the least complicated solutions to increase the resolution, since there does not seem to be a direct physical implication that would produce a technological dead end, as for the two first propositions. Such  $k_1$  reduction solutions can be linked to stepper hardware optimization or to reticle optimization techniques.
  - Hardware optimization to reduce  $k_1$ : such solutions include the optimization of the illumination scheme and immersion lithography. These solutions are cheaper than changing the source and the projection lenses, so IC fabs like it. They also include process optimizations such as double exposure and other such exposure tricks, as well as photoresist formulation optimization.
  - Reticle pattern optimization: as this solution is not directly linked to the lithographic projection device (the stepper), this is attractive to IC fabs since it does not cost them much. Such techniques are called Reticle Enhancement Techniques (RET) and are mostly software based (OPC etc.). They can also be physical, as in Phase Shifting Masks (PSM). Today, Electronic Design Automation (EDA) software includes most of the RET techniques, and relies heavily on sub-wavelength diffraction modeling software as presented in Chapter 11 (and thus can be very expensive). Such EDA design optimizations are called Design For Manufacturing (DFM), where the design (reticle pattern) is optimized for a specific manufacturing (lithographic) tool.

The various software and hardware optimization tools to implement DFM (RET) and improved lithographic techniques in existing or new types of steppers will be described below.

Figure 13.4 uses the lithographic system description of Figure 13.2, and describes the various software (RET) and hardware solutions discussed in this chapter.

## 13.2 Software Solutions: Reticle Enhancement Techniques

Table 13.1 shows that changing the lithographic physical parameters (wavelength or NA) can be quite challenging, and can become very costly. There are other ways to increase the quality of the lithographic projection without changing the hardware, namely by using Reticle Enhancement Techniques that reduce the resolution factor  $k_1$  (see Figure 13.4).

Reticle Enhancement Techniques (RET) are software techniques, since they attempt to pre-distort the reticle pattern in order to yield the desired pattern in the wafer resist [5–7]. Such software solutions include optical proximity correction in 2D and 3D, biasing, proportional etching compensation and phase-shift mask generation (the latter is actually a hybrid software/hardware method). The next section reviews the hardware methods that can reduce the  $k_1$  factor.

### 13.2.1 Direct Write Proximity Compensation Techniques

Compensation techniques for both e-beam and laser beam proximity effects are derived below. These effects occur, respectively, in e-beam resists or photoresists.

#### 13.2.1.1 E-beam Proximity Compensation

In Chapter 12, laser and e-beam based mask patterning systems were discussed in optical and e-beam resists. The laser writing beam can be focused down to sub-micron sizes (down to 0.4  $\mu\text{m}$  with UV lasers), and the electron writing beam can be focused down to below 10 nm! However, this does not mean that one can directly write a 400 nm feature or a 10 nm feature with such beams. There are many adverse effects that cause deterioration of the resulting feature that such beams can write into resist.

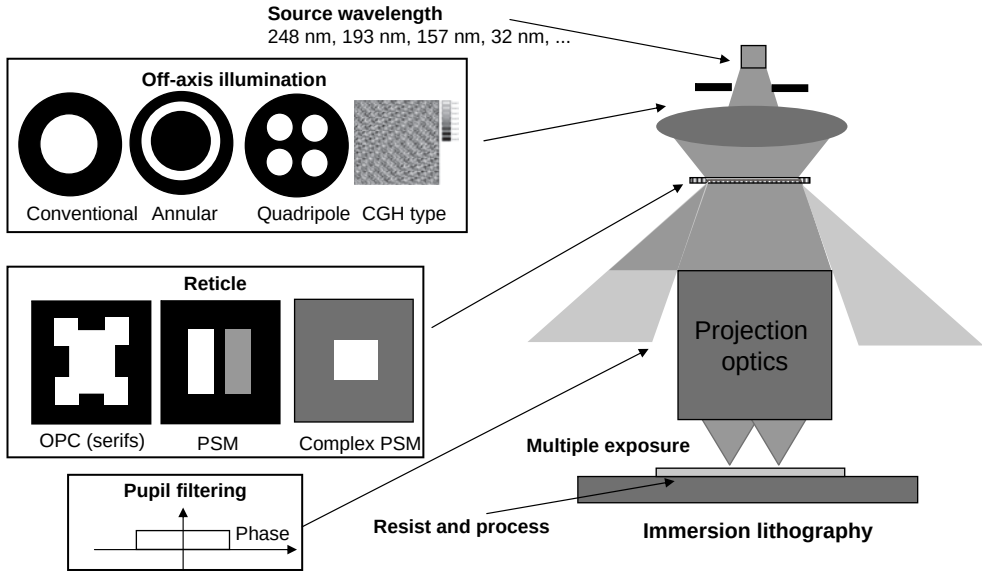


**Figure 13.4** Various software (RET) and hardware solutions to reduce the resolution factor  $k_1$

The writing beams are influenced by the material in the vicinity of the focus (diffusion of the light or electrons in the resist), the artifacts of the beam itself (spot shape, side lobes of the focused beam) or back-reflection from the underlying surface. Such effects are considered as proximity effects.

Figure 13.5 shows the proximity effects occurring in e-beam resist during the mask patterning write, which is especially useful for a direct write, where the beam writes the shape at the final size ( $1\times$ ) directly – see Chapter 12.

The electron beam is usually focused underneath the e-beam resist surface, in order to achieve the best side walls quality (e.g. side wall angles). The electron beam within the resist layer undergoes complex physical phenomena, including absorption, scattering, diffusion, reflections and so on.

Both experimental analysis and theoretical methods (e.g. iterative Monte-Carlo methods) show that the backscattered electron distribution can be expressed, as the initial electron beam distribution, by a Gaussian function. The summation of both e-beam distributions gives us the E-beam Point Spread Function (EPSF) of the e-beam patterning system for a specific type of e-beam resist. The EPSF can be described as follows:

$$\text{EPSF}(r) = \frac{1}{\pi(1+\eta)} \left( \frac{1}{\alpha^2} \cdot e^{-\frac{r^2}{\alpha^2}} + \frac{\eta}{\beta^2} \cdot e^{-\frac{r^2}{\beta^2}} \right) \quad (13.9)$$

where  $\eta$  is the fraction of the incident e-beam energy that is backscattered. The normalization condition is thus

$$\int_0^{\infty} 2\pi r \cdot \text{EPSF}(r) = 1 \quad (13.10)$$

Typically, one order of magnitude differentiates the values for the forward and backscattered radii  $\alpha$  and  $\beta$ , with typical values of approximately  $0.2 \mu\text{m}$  and  $5.0 \mu\text{m}$ . The EPSFs have to be characterized for each new e-beam resist, acceleration voltage, beam current or e-beam dosage.

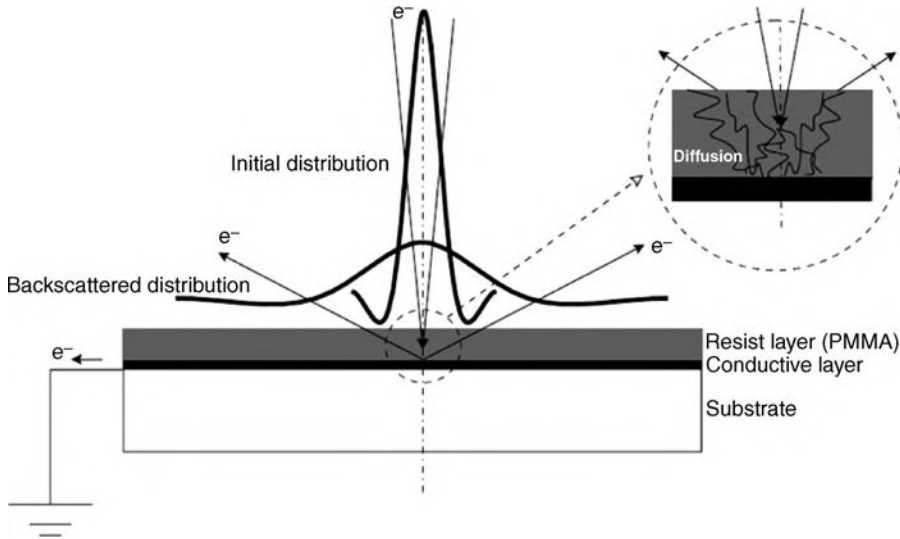


**Figure 13.5** E-beam proximity effects in e-beam resist

The EPSF can be considered as the unit pulse response of the incident electron-beam. The effective absorbed dose  $A(x, y)$  by the resist (resulting from the exposure of the initial calculated (and desired) digitally described multidose-pattern  $D(x, y)$  in the GDSII file) can be expressed by the following convolution product:

$$A(x, x) = D(x, y) * \text{EPSF}(x, y) \quad (13.11)$$

This convolution product will be the basis for the propositions for 2D and 3D e-beam proximity effect compensation methods in this section.

A practical example is used here to quantify the effects of both 2D and 3D compensation on the global digital optical element's performance: the element used here is a Fresnel multifocus lens CGH optimized by an IFTA algorithm presented in Chapter 6. This particular CGH has a cell size of  $1 \times 1 \mu\text{m}$  with four phase levels.

The typical e-beam effects are applied on the raw CGH data. The EPSF parameters to be used in the following simulations are:

- $\alpha = 0.2 \mu\text{m}$ ;
- $\beta = 4.0 \mu\text{m}$ ;
- $\eta = 15\%$ .

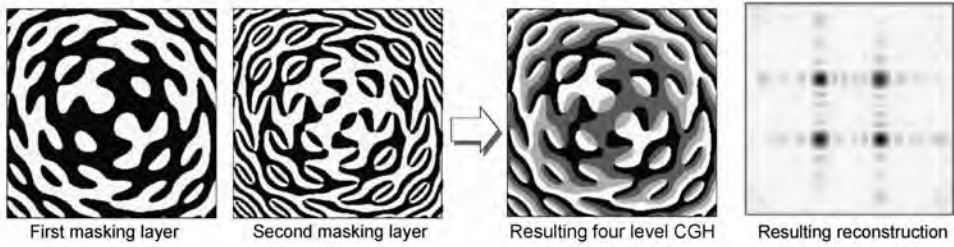
This is a quite 'standard' EPSF that is encountered in most of the e-beam writers used in industry.

### **Two-dimensional E-beam Proximity Effects (2D EPE)**

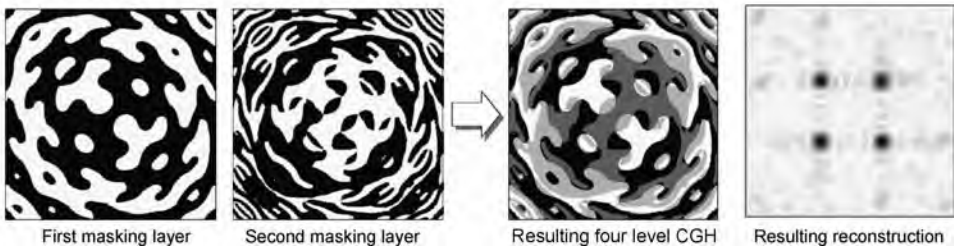
Two- and three-dimensional e-beam proximity effects (2D and 3D EPE) in direct write affect the resulting lateral structure widths on binary masks. Although binary write induced 2D-EPE effects introduce three-dimensional modulations of the resist, as the resist is removed down to the substrate prior to chromium wet etch, the resulting structures are again two-dimensional.

The 2D EPE do affect the lateral extent of the binary structures on standard Cr on glass masks, and 3D EPE affect the surface profile of the final CGH more significantly. Although only 15% of the electrons are

Unaffected masks



2D EPE-affected masks

**Figure 13.6** 2D EPE effects on binary chrome masks

backscattered, Figure 13.6 shows that the first CGH mask pattern can nevertheless be seriously affected by such a 2D EPE.

The first mask is relatively unchanged, whereas the second mask goes through serious transformations. E-beam proximity effects on the lateral structures depend not only on the structure itself, but also on the neighboring structures, as can be seen in the second affected mask pattern. In the corner areas of the square CGH aperture, the affected fringes exhibit high fidelity (not as many neighboring structures as there are on the aperture edge), whereas the fringes at about one-third from the center are much more affected, since they have a heavy surrounding fringe distribution. After the two binary lithographic processes have been performed (see Chapter 12), the resulting four phase levels CGH shows high-frequency structures that are induced by the nonlinear shrinkage and swelling of the affected fringes on the two previous chrome-over-glass masks.

### ***Three-dimensional E-beam Proximity Effects (3D EPE)***

Three-dimensional Electron Proximity Effects (3D EPE) affect both the lateral and longitudinal (etch depths) geometries of the structures when using direct analog e-beam write with various e-beam dosages. This technique is used in direct analog e-beam write (see Chapter 12).

As the resist is proportionally transferred by RIBE after development of the latter, the exact three-dimensional surface-relief profile is transferred from the resist into the substrate (no anisotropic RIE effects are considered yet; they will be addressed later on in this chapter). The affected profile keeps all its characteristics in the final CGH, unlike in binary e-beam write where only the lateral geometry is kept. Also, there is no need to remove the resist down to the substrate. This is a very desirable feature, since it is extremely difficult to monitor whether the resist is entirely removed in the valleys (as in standard lithography). Figure 13.7 shows typical 3D EPE effects in e-beam resist and the subsequent numeric reconstruction.

Once can see that the 3D EPE proximity effects have caused the reconstruction to deteriorate considerably, especially the uniformity over the four spots: two of the spots are simply gone.

When scanning over the central part of the previous four-level CGH, one can more clearly see the effects of 3D EPE. Figure 13.8 shows such scans over a mild (upper) and a severe (lower) 3D EPE effect on the previous CGH.

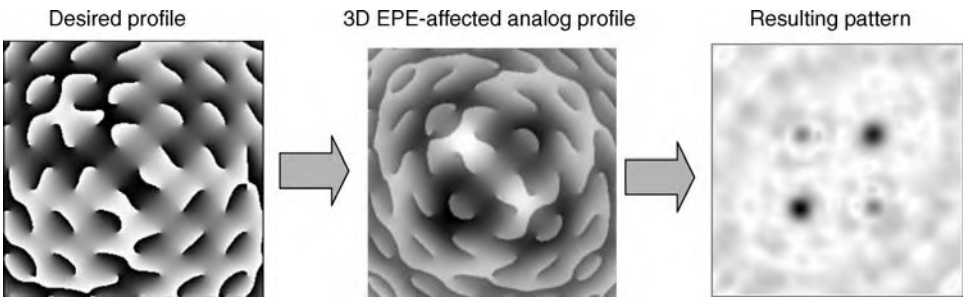


**Figure 13.7** A 3D affected quasi-analog profile in e-beam resist (by direct analog e-beam write)

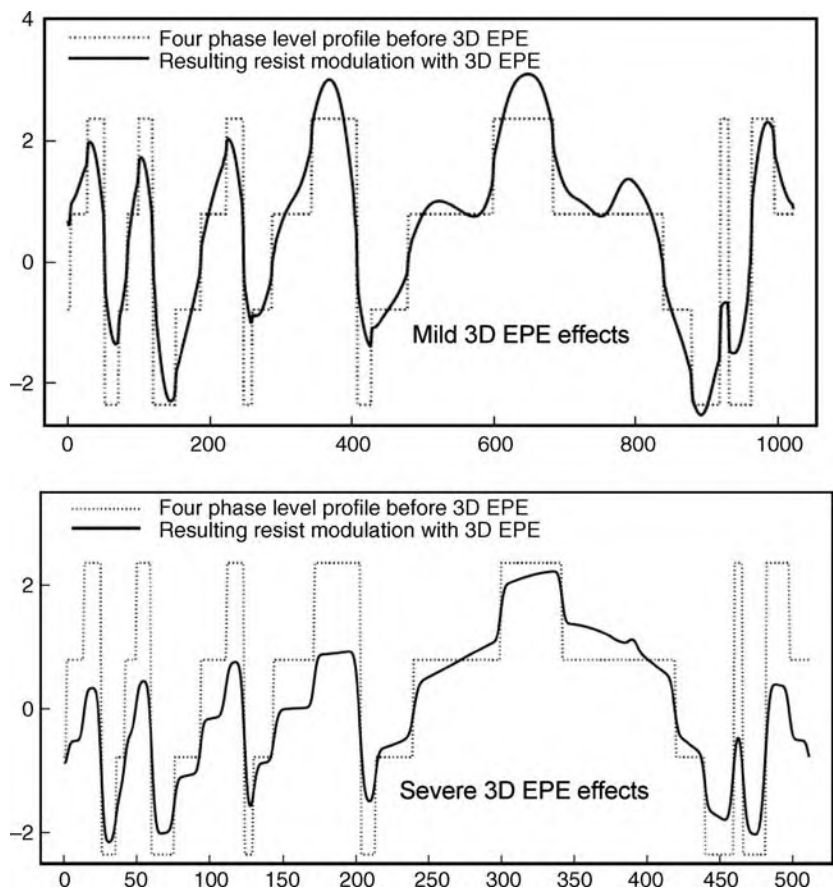


**Figure 13.8** Mild and severe 3D EPE effects on a four phase levels CGH



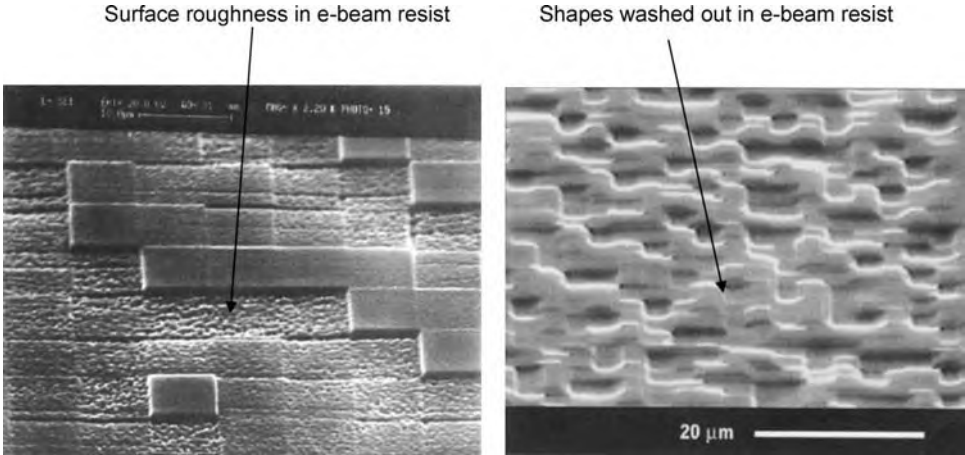


**Figure 13.9** SEM photos of 3D EPE affected e-beam resist profiles

In Figure 13.8, one can see that some of the higher frequencies are lost, and that some parasitic frequencies appear. Also, the depth is getting shallower, therefore affecting the overall efficiency and uniformity of the reconstruction (see Figure 13.7) and also producing a large zero order.

Figure 13.9 shows some SEM photographs of direct analog write into e-beam resist, and corresponding 3D EPE effects.

Figure 13.10 shows the same 3D EPE effects, but this time on a blazed grating, an analytic element (the previous CGH was a numeric-type element).

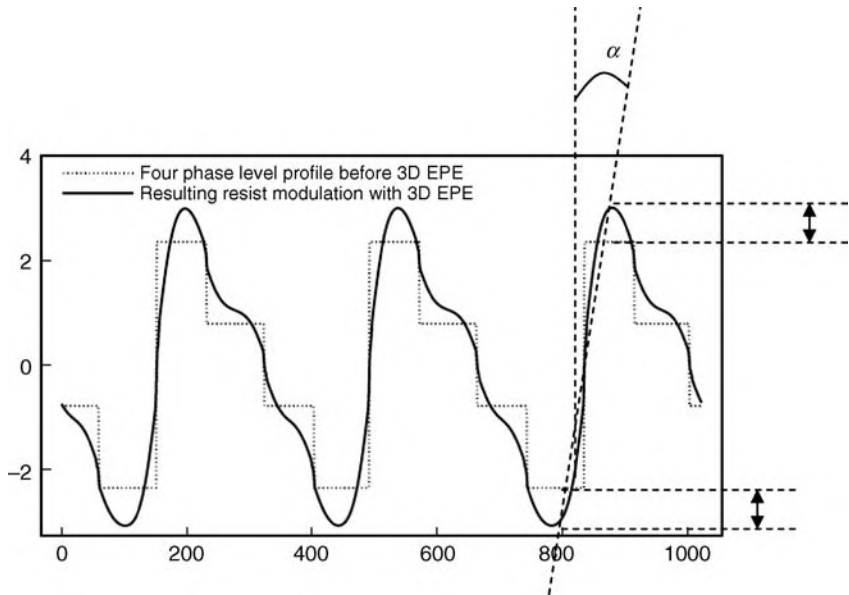


**Figure 13.10** Three-dimensional EPE effects on a blazed grating (analytic element)

The major effect on the blazed grating (or blazed Fresnel lens) is that the side walls get less and less normal to the substrate (thus affecting efficiency), and that the overall depth is actually increasing (thus again affecting the efficiency). The period is kept more or less intact (the diffraction angle remains unchanged).

### *E-beam Proximity Compensation Techniques*

As the affected final pattern  $A(x, y)$  can be expressed as a convolution between the initial pattern  $D(x, y)$  and the E-beam Point Spread Function EPSF (see Equation (13.11)) one can apply the convolution theorem and use the FFT algorithm:

$$A = \text{FFT}^{-1}[\text{FFT}[\text{EPSF}] \cdot \text{FFT}[D]] \quad (13.12)$$

Ideally, one would like to have  $A(x, y) = D(x, y)$ . Thus, by replacing  $A(x, y)$  by  $D(x, y)$  and inverting the matrix expression, one can calculate the optimal pre-compensated  $C(x, y)$  dosage that would provide the desired  $D(x, y)$  resist dosage after 2D EPE affliction:

$$\text{FFT}[D] = \text{FFT}[\text{EPSF}] \cdot \text{FFT}[C] \quad (13.13)$$

One has to keep in mind that these are vector (matrix) operations, not scalar operations.

Based on this analysis, two different methods for proximity effects compensation are proposed, first for analytic-type CGHs and second for numeric-type CGHs (see the classification of CGHs in Chapters 5 and 6). In the first case (analytic-type CGHs), the CGH is optimized regardless of fabrication constraints or errors, and the optimized pattern is compensated afterwards, as a second distinctive process. In the second case, the compensation technique is integrated directly within the CGH optimization process (within an IFTA-type optimization algorithm): the CGH is thus optimized in parallel for its optical function and for fabrication error compensation. Figure 13.11 summarizes these two different software compensation methods.

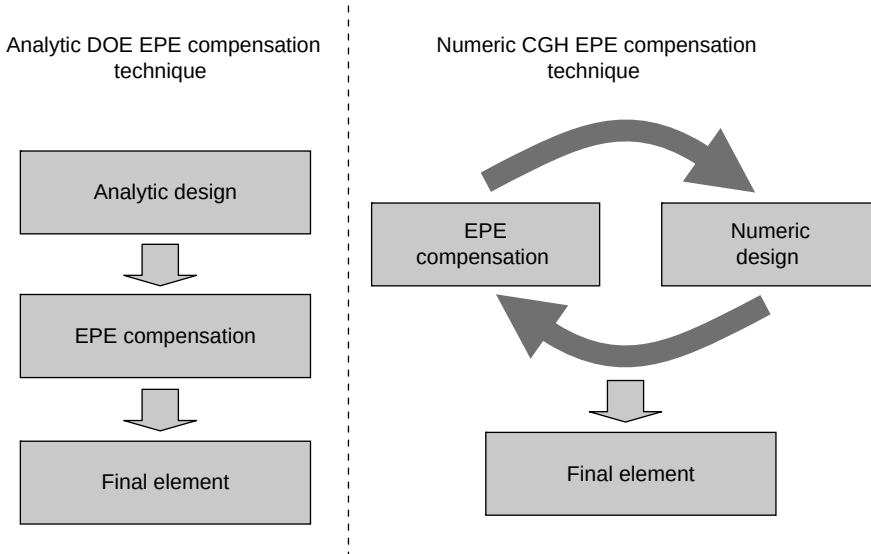


**Figure 13.11** E-beam proximity effects compensation methods for analytic DOEs and numeric-type CGHs

### The EPE Compensation Technique for Analytic-type Elements

Once the analytic-type DOE has been designed, the resulting profile is quantized to  $N$  levels, to produce a binary mask pattern or binary phase element ( $N=2$ ) or a multilevel phase element ( $N>2$ ). An iterative steepest descent method is therefore used (very similar to DBS described in Chapter 6) in order to optimize a pattern (or e-beam dosage distribution), which, if convoluted with the measured or calculated EPSF, would produce, after quantization, a pattern closest to that desired pattern or that of the final Cr mask.

The RMS error in amplitude between the desired pattern  $P$  and the patterns  $C_i$  obtained in the current iteration  $i$  is set to be the current cost function, which has to be decreased by the compensation algorithm:

$$E_i = \sqrt{C_i^2 - P^2} \quad (13.14)$$

The RMS error in amplitude describes the best binary amplitude differences between the two patterns. Figure 13.12 reports the compensation algorithm for analytic-type elements. In the figure,  $\varepsilon$  is a small value below which the cost function is taken to be small enough to stop the compensation process.

The core of the algorithm resides in the way the dose distributions of the binary patterns are modulated (i.e. how the pattern evolves throughout the successive iterations). The value of each pixel (over-sampled CGH cell) is no longer binary, but can take on many values (quasi-analog). Although pixel alterations would be a good way to optimize the pattern (with FFT update algorithms as in DBS: see Chapter 6), the required amount of number crunching would be overwhelming.

It is preferable to modulate the pattern in a parallel way (filters) rather than in a serial way (pixel flipping). In order to find the adequate DOE pattern evolution strategy, several types of rule-based proximity effects compensation have been analyzed. The main characteristics are as follows:

- structures get narrower as the neighborhood dosage increases;
- structures get wider as the neighborhood dosage decreases;

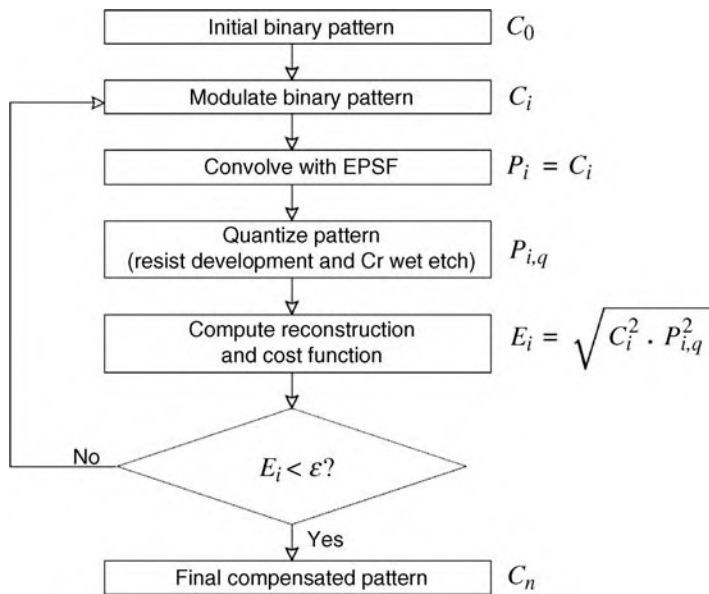


**Figure 13.12** DBS EPE compensation for analytic elements

- sharp edges are smoothed out; and
- smooth serifs are added to structure edges.

A set of filters is applied to the pattern in order to create the previously listed effects artificially:

- smoothing (Fourier filtering),  $\Phi_1$ ;
- edge enhancement (gradient filter),  $\Phi_2$ ;
- edge doubling (Laplacian filter),  $\Phi_3$ ;
- difference mapping between the original pattern and the enlarged smoothed pattern (sharp edge enhancement),  $\Phi_4$ .

By linear combination of these filters, a global transformation  $\Phi$  can be written as follows:

$$\Phi = \sum_{n=0}^N c_n \cdot \Phi_n \quad (13.15)$$

The steepest descent optimization algorithm will converge to a solution (the local cost function minimum). An annealing process of the corresponding factors  $c_n$  is used in order to reduce the evolutions between iterations when convergence has begun to take place.

However, this can still be computationally expensive. Two alternatives to this method have been developed, by splitting the entire CGH aperture into small areas and compensating the small area structures by the use of look-up tables. Hence, to compensate an entire CGH aperture, the dimensions are analyzed locally and the appropriate compensation dose distribution is copied from the look-up tables.

The local compensation is performed over small areas (either in 1D or 2D), by a Direct Binary Search (DBS) algorithm, as described earlier in this section. This local compensation process is viable here since the considered aperture is very small (relatively low CPU time consumption).

The first technique is best suited for diffractive lenses optimized by external optical design software, and the second is best suited for square CGH cell encoding over  $N$  phase levels. In the first method, the problem is reduced to a one-dimensional CGH compensation problem: this is possible since the fringes are usually symmetrical (spherical and aspheric lenses), and the radius of curvature is large enough for the element to be considered as one-dimensional (locally). One can therefore write

$$C(x) = \text{FFT}^{-1} \left[ \frac{\text{FFT}[P(x)]}{\text{FFT}[\text{EPSF}(x)]} \right] \quad (13.16)$$

The inversion process is greatly simplified since vectors are used rather than matrices. In the second method, the CGH is encoded with square cells, and single elements are processed one at a time by considering a cluster of cells formed by the eight neighboring cells. If the fabrication is to be performed over  $N$  levels, the number  $\rho$  of different cluster configurations (i.e. the number of matrices within the look-up table) that can be found within the DOE aperture is  $\rho = N^9$ .

This number can still be reduced, since the cluster can exhibit several symmetries that result in the same configuration (a nonoriented cluster of cells is used, since the cluster is considered separately from the rest of the element's aperture). If  $s$  is the number of symmetries that can occur in the cluster, the number of matrices to fill (compensated clusters) in the look-up table becomes

$$\begin{cases} \rho = \frac{N^9}{2^s} \\ s = 4 \end{cases} \quad (13.17)$$

Therefore, in the case of fabrication over two levels,  $r = 64$ . If the cell size is about  $2.0 \mu\text{m}$  and the e-beam excel grid has a size of  $0.1 \mu\text{m}$ , the single matrix size  $M$  would be of 60 pixels in each direction of space. The CPU time and the memory required to compute and store these different matrices as floating point

numbers can be expressed as follows:

$$\begin{cases} \text{CPU} \propto \rho(I \cdot M^2)(3M^2 \cdot \log_2(M^2)) \\ \text{RAM} = 8\rho M^2 \end{cases} \tag{13.18}$$

where  $I$  denotes the number of global iterations required for the algorithm to converge (typically after 10–20 iterations). In our binary case, these values would be approximately 0.5 CPU hours on a 50 MIPS machine with 0.3 Mb RAM space. When processing elements to be fabricated over more than two levels, the look-up table increases dramatically in size, whereas the CPU time for compensation stays constant. For more than two levels (e.g. four levels), the required RAM space increases to 1 Gb.

Unlike direct compensation methods (where the entire DOE aperture is sampled to the e-beam addressable grid), this method increases the CPU time only linearly with the number of cells constituting the DOE aperture (whereas the CPU time of direct methods increases faster than the square of the total number of cells).

*The Numeric-type CGHs Compensation Technique*

The proximity effects compensation method for numeric-type CGHs is different from the current compensation method because the compensation is directly included in the CGH optimization process (see the IFTA algorithms described in Chapter 6). The compensation process is illustrated in Figure 13.13.

As the proximity effects can be modeled by a convolution process (see the previous section), one can shift the affliction process into the reconstruction plane by simply filtering the CGH's Fourier transform by the Fourier transform of the EPSF. However, this is only possible for CGHs that work in the Fourier regime. In the Fourier regime case, the previous compensation/optimization algorithm becomes the method shown in Figure 13.14.

As an example of the numeric 2D and 3D EPE compensation techniques, Figure 13.15 shows a small portion of a numeric-type CGH to be patterned, first in a binary way (the upper plots) and then in an eight phase levels way (the lower plot).

One can see clearly in Figure 13.15 how the compensated pattern, even though it is affected by either 2D or 3D EPE effects, is much closer to the desired pattern than the uncompensated patterns.

Now, one other important issue that Figure 13.15 clearly shows is that the fracture grid needs to be reduced in order to produce smaller features for the compensation algorithm, therefore increasing the data file and increasing the cost of the reticle patterning. This is one of the drawbacks of any compensation

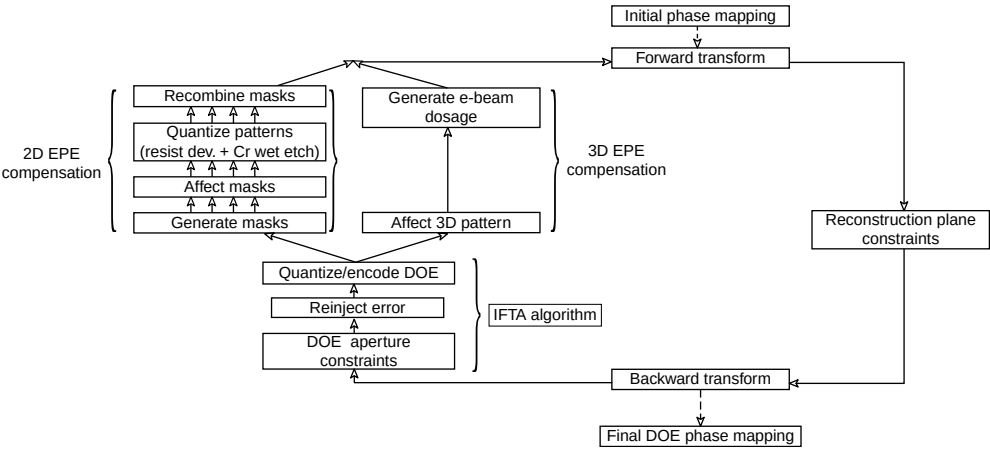


**Figure 13.13** A numeric-type elements compensation scheme

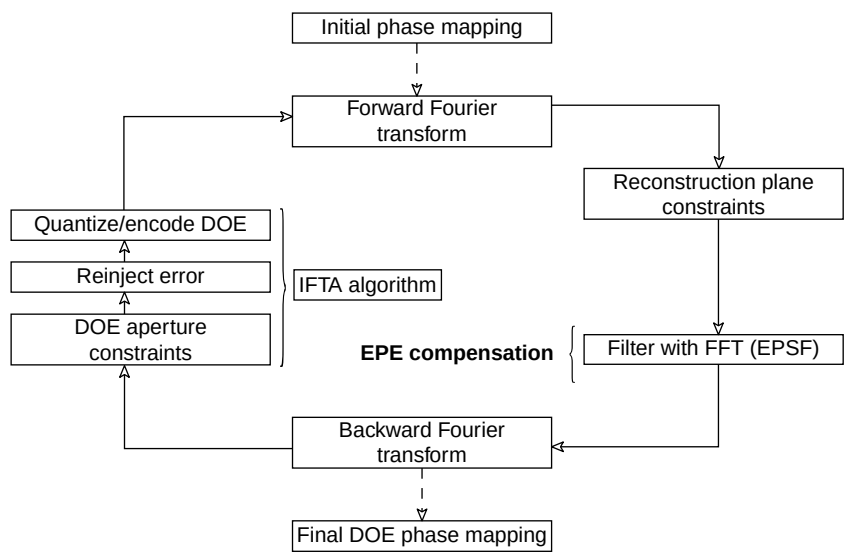


**Figure 13.14** The compensation method for numeric-type Fourier CGHs

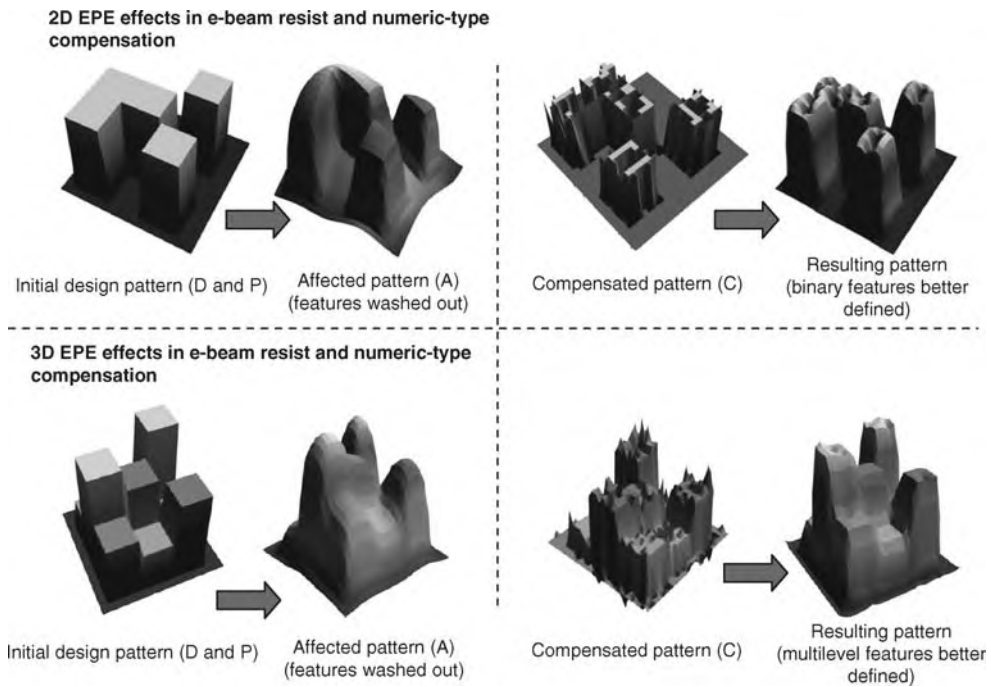


**Figure 13.15** The results of 2D EPE and 3D EPE compensation on numeric-type elements

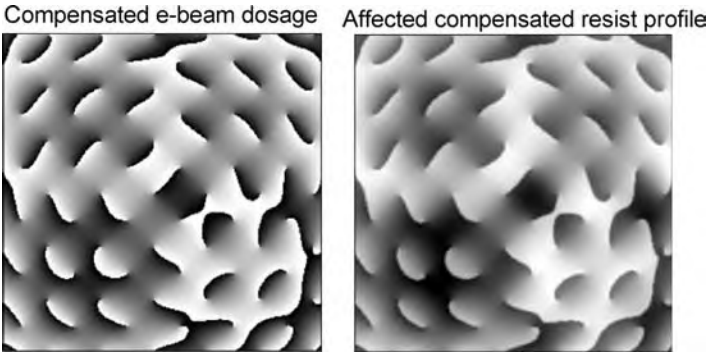


**Figure 13.16** Three-dimensional EPE effects compensation on an analog CGH profile

technique (consider, for example, the increase in data size when adding serifs to the shapes in rule-of-thumb compensation in optical proximity effects).

However, one needs to remember that even though the fracture grid is reduced, these smaller shapes do not need to be resolved. In fact, they should not be resolved at all, and therefore they create the compensation required to produce the larger feature with the desired resulting geometry.

Finally, Figure 13.16 shows an entire analog surface-relief CGH compensation example for 3D EPE effects. Compare Figure 13.16 with the uncompensated pattern and related effects presented in Figure 13.7.

An analytic-type element compensation process is now presented, with the use of an iterative DBS algorithm as described earlier in this section.

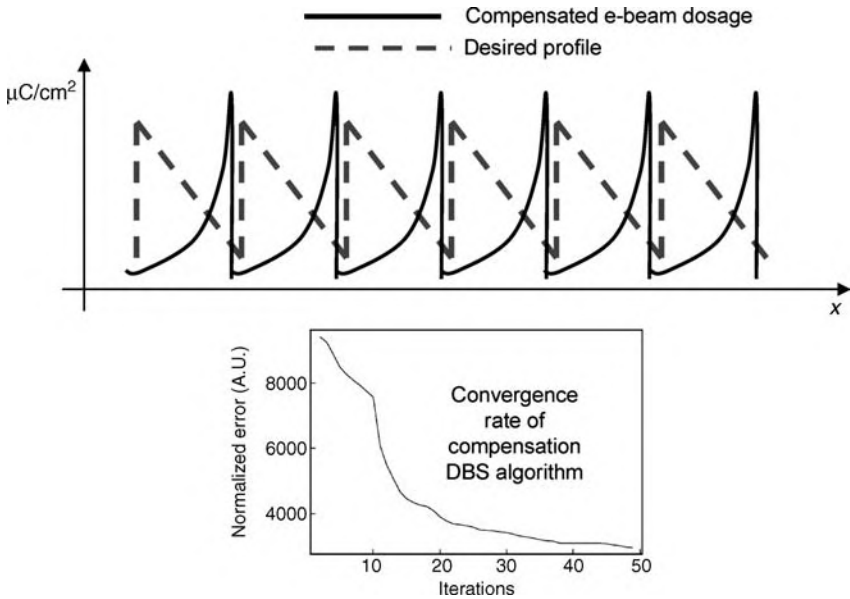


**Figure 13.17** An example of analytic-type element 3D EPE compensation by DBS algorithm

Figure 13.17 shows a blazed grating example and the optimized dosage that compensates the 3D EPE effects. The convergence rate of the DBS algorithm used for compensation is also shown in the same figure.

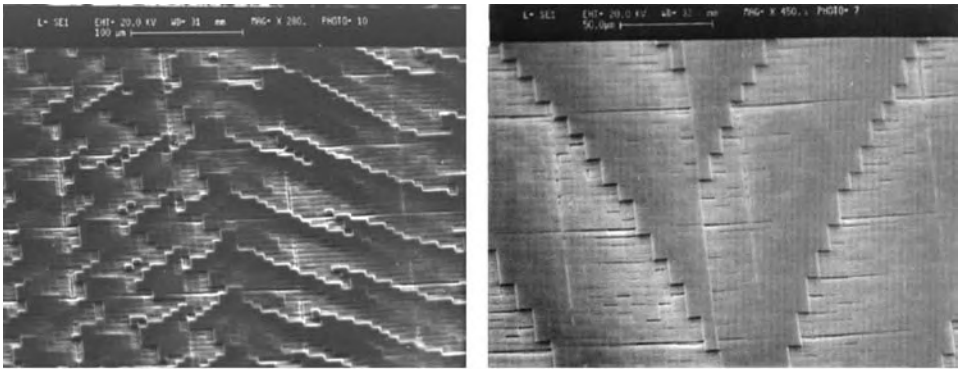
Such nonlinear spikes are characteristic of a 3D EPE effect compensation, and produce a much better side wall geometry, which is especially useful in yielding a high efficiency in blazed gratings or blazed Fresnel lenses.

Figure 13.18 shows some SEM photographs and interference surface profilometry plots of compensated numeric and analytic elements.

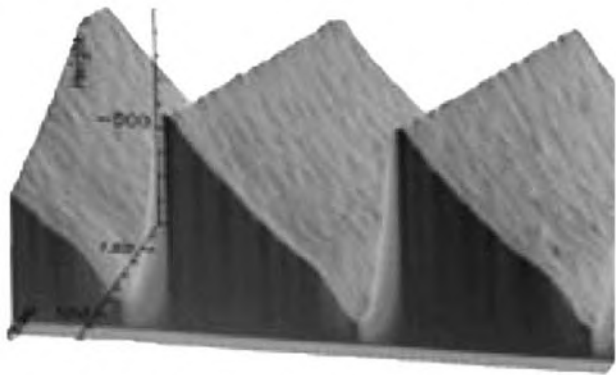
Figure 13.18 shows that although the resulting profile is decent (a blazed grating profile and quantized numeric CGH phase profiles), some roughness remains in the surface, and has not been cancelled by the compensation algorithm.

**13.2.1.2 Laser Beam Proximity Compensation**

Similarly to e-beam proximity compensation, one can implement 2D and 3D proximity effect compensations for other writing technologies; for example, laser beam writing of e-beam resist or even direct laser beam ablation (the Heidelberg Instruments direct Laser Beam Ablation (LBA) system).



SEMs of numeric-type compensated elements over 16 phase levels in e-beam resist



Interferometric topology plot of compensated blazed grating in e-beam resist

**Figure 13.18** Compensated numeric and analytic elements in e-beam resist



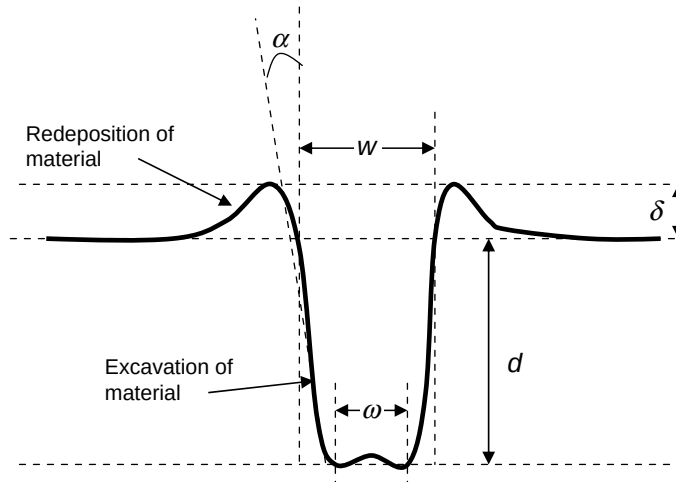


**Figure 13.19** Laser ablation patterning system PSF

In a laser beam writing in photoresist, the PSF function would be a Gaussian function or any other function that can be generated by a beam-shaping element introduced in the laser patterning confocal column, as indicated in Section 12.4.4. Such a PSF could thus be a top hat, a hexagonal shape and so on.

In a laser beam ablation system, the PSF function to replace the EPSF used in the previous section is a little more complex. This PSF also describes the laser beam profile at the mask (or wafer) plane, which can also be shaped by a beam-shaper element, but which also includes the material redeposited at the edges of the hole produced by this beam, as described in Figure 13.19. In the figure,  $d$  is the total material excavation depth,  $w$  is the outer hole width,  $\omega$  is the inner hole width,  $\delta$  is the material re-deposition height and  $\alpha$  is the side wall angle.

In the next section, the effects of optical proximity will be analyzed and a compensation algorithm will be proposed, similar to the one described for e-beam proximity effects in resist. An optimization technique for the generation of digital optics reticles will also be described, which is directly inserted in the digital optics design process as described in Chapters 5 and 6 (for analytic- and numeric-type digital optics). These methods are different from the ones used in the IC industry, as digital optics features are very different in essence from IC features.

### 13.2.2 Optical Proximity Effects in Projection Lithography

The previous section has analyzed the various proximity effects generated by the writing tools (either e-beam or laser, by exposure of resist or laser ablation). These effects are thus e-beam or optical beam proximity effects.

There are various imaging-related optical proximity effects, that are mainly due to the finite aperture of the lens. In a general way, a proximity effect is the change in image size depending on the environment of an object shape [4, 7–9].

#### 13.2.2.1 Effects on an Aerial Image

Due to the finite NA of the lithographic projection lens and due to the aberrations of that lens, one can model the imaging effects of such a system. Add to that a modeling of the polymerization process in the

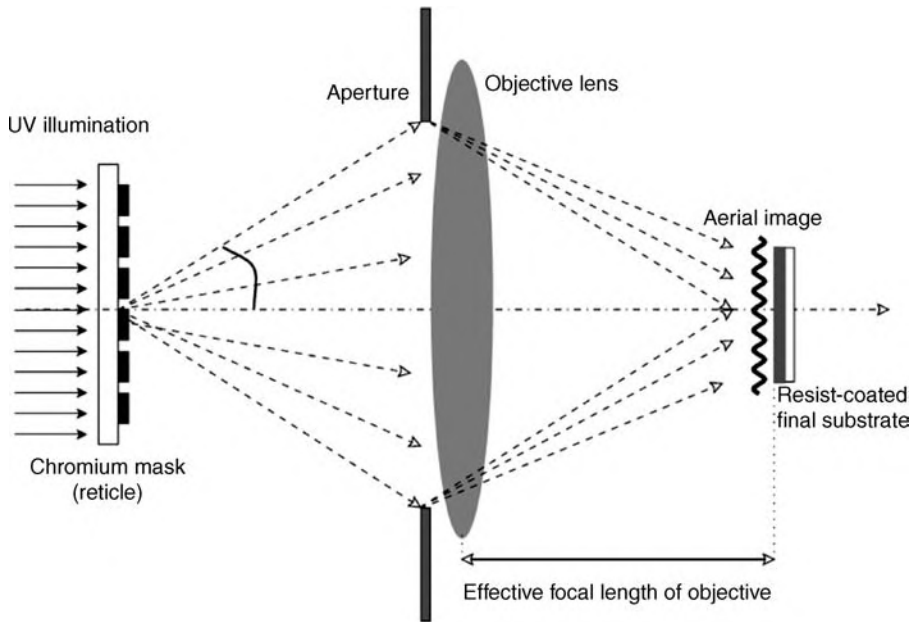


**Figure 13.20** The projection of an aerial sinusoidal image from a binary Cr reticle

photoresists due to the projected aerial image in the photoresist layer, and one can model exactly how the resist pattern will appear with a particular lithographic tool.

Figure 13.20 shows a typical lithographic projection lens and the various diffraction orders generated, as well as the image formed by the diffraction orders that are actually imaged through the projection lens.

Similar to the proximity effects in resist during writing, if one can model the effects of the projection tool and the aerial image formation within the resist layer, one can compensate for such effects. There are mainly two methods used in industry for Optical Proximity Compensation (OPC):

- the rule-of-thumb method; and
- the analytic compensation method.

The rule-of-thumb method is the most widely used in EDA software including OPC, and consists of some fundamental rules that are based on analysis of the vicinity of the shape and of the shape itself (see Figure 13.21(a)).

As one can see in Figure 13.21(a), most of the rule-of-thumb techniques consist of adding serifs and subtracting shapes in convex nodes, reducing line widths when similar shapes are close to each other and so on. Scattering lines (also called scatterers) introduce high diffraction angles, and therefore push light outside the acceptance cone of the projection lens, therefore reducing the light in these areas.

Such rules can often be complex and nonintuitive. However, as they are very well adapted to IC-type structures, they have very a limited effect for digital optics, and especially analytic-type elements (diffractive lenses) or numeric-type elements (CGHs). As the shapes required to produce digital optics are often customized for each element (no standard logic gate shape as in integrated transistors for example), one has rather to run an algorithm-based compensation process by using the exact model of the transfer function defined by the lithographic lens and the aerial image within the photoresist.

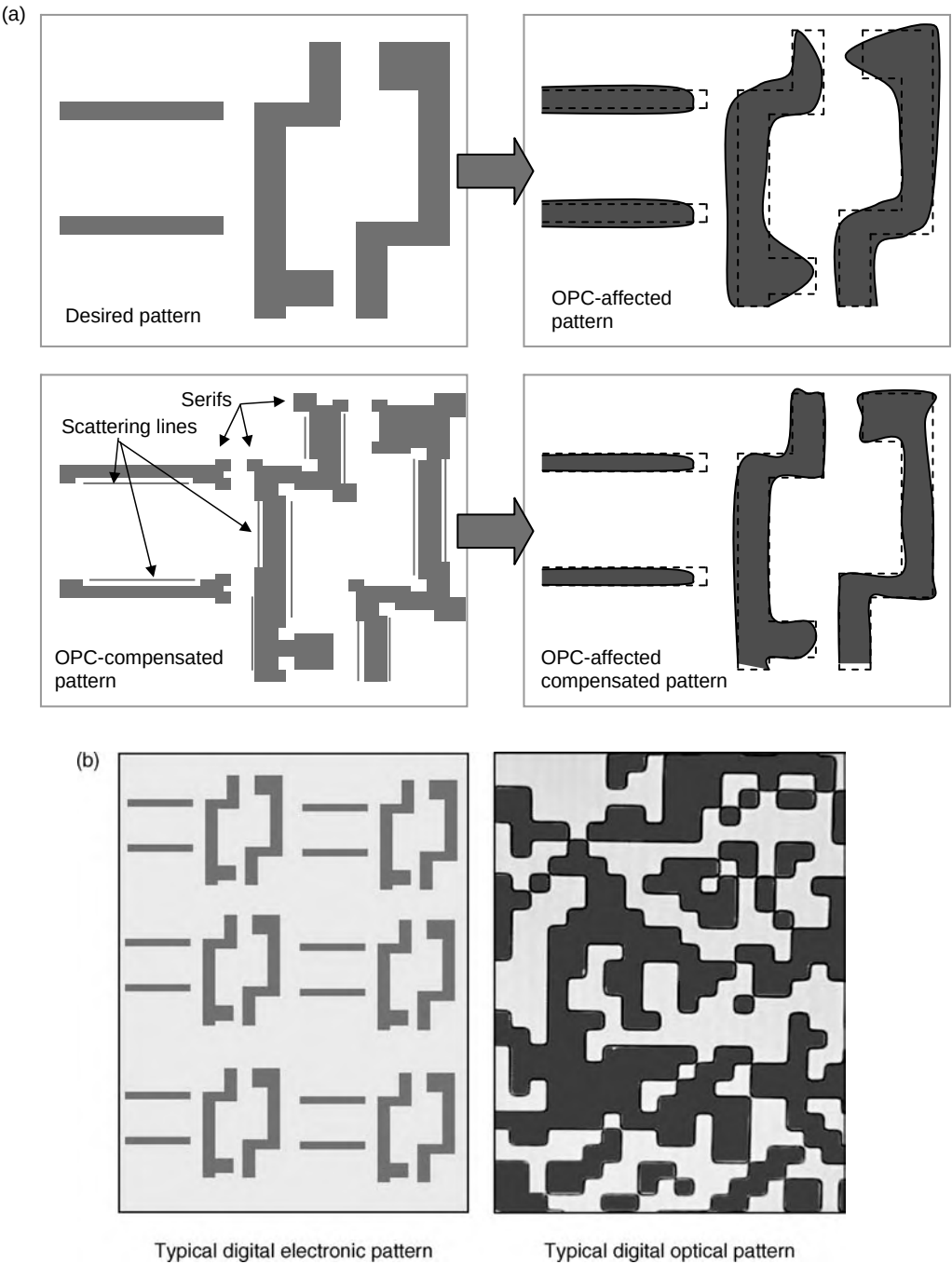


**Figure 13.21** (a) OPC rules of thumb in EDA software including RET techniques. (b) Typical micro-electronic and micro-optic patterns to be OPC compensated

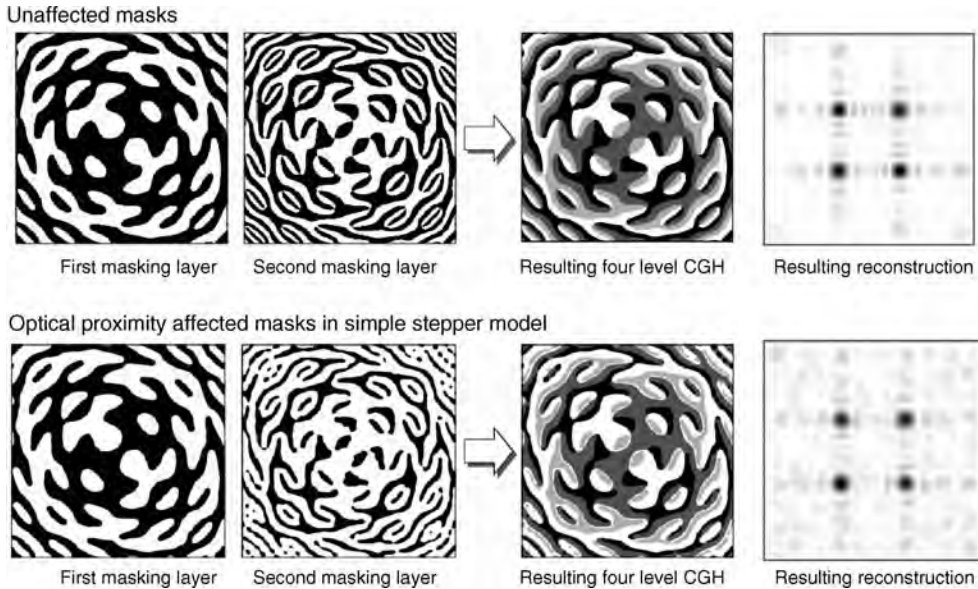


**Figure 13.22** Simple binary optical proximity effects in a simple stepper system

The problem with digital optics is that the related patterns do not look at all like digital electronic patterns to which one can apply rule-of-thumb OPC (such as serifs and scatterers). Figure 13.21(b) shows digital electronic and digital optical micro-patterns side by side. The figure highlights the need for another OPC compensation technique (other than those optimized for IC patterns) for typical digital optics micro-patterns.

Figure 13.22 shows a couple of binary photomasks (from the same CGH as in the previous section), affected by optical proximity effects in a simple stepper model.

It can be seen that the optical proximity effects on the binary masks in Figure 13.22 are quite severe, as noted in the IC fabrication industry (e.g. the reduction or even the vanishing of the small features on the second mask, which bears the smaller features). However, the optical reconstruction in the same figure is very close to the desired optical reconstruction for unaffected masks. This is due to the fact that the reconstruction (i.e. the four focused spots) is not located at the edges of the reconstruction window (the fundamental reconstruction window, as described in Chapter 11) but, rather, within two-thirds of that window, in the center. Therefore, the high spatial frequencies that were washed away from the optical proximity effects in the second mask (the mask with the highest frequencies) are not really contributing to the intensity reconstruction but, rather, take care of the high SNR in the vicinity of that reconstruction (at the edges of the window). This yields a drop in the SNR but not a drop in the uniformity of the reconstruction. A drop in the SNR is usually also related to a drop in the overall diffraction efficiency.

In order to evaluate the effects of optical proximity more precisely, a scan over the central part of the four phase level affected CGH is shown in Figure 13.23.

Figure 13.23 shows that there are high spatial frequencies that have disappeared totally and that other high spatial frequencies have appeared. Most of the smaller features have also narrowed. Again, this does not really change the reconstruction uniformity of position, but it has created high-frequency noise in the reconstruction window, which is shown in Figure 13.22.

Figure 13.24 shows an example of optical proximity effects on a gray-scale mask that has been used in the same stepper projection system. Here, the increase of noise (reduction of the SNR) is clearly

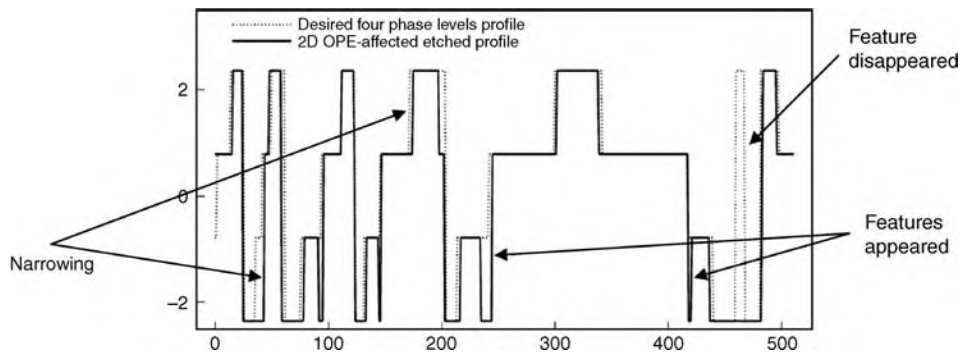


**Figure 13.23** A 1D scan of the central part of a four phase levels CGH affected by optical proximity

evident in the numeric reconstructions, but the position and the uniformity of the reconstruction are not altered.

As a practical example of optical proximity effects, Figure 13.25 shows an SEM photograph of a section of a binary diffractive lens in which some high spatial frequencies have disappeared in specific directions.

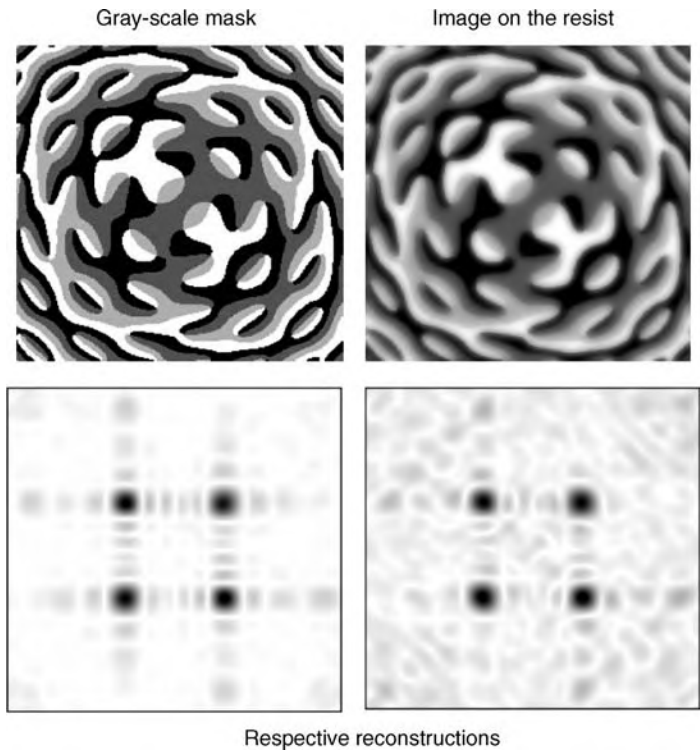


**Figure 13.24** A gray-scale mask used in stepper model optical proximity effects

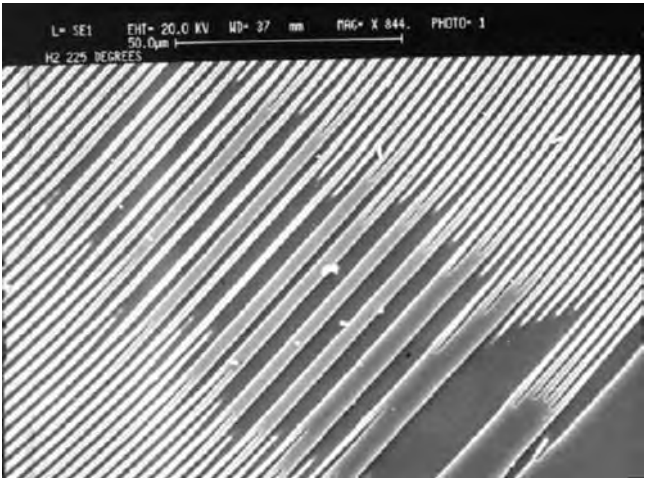


**Figure 13.25** An SEM photograph of an off-axis binary diffractive lens

This might also be an effect of the low-level fracture process (see Chapter 12), for which specific angles are fractured in a better way than others. When considering a simple SEM photograph, it is difficult to analyze where the problems arise from. The fringes that are stuck together might also be an effect of optical proximity, following the rule-of-thumb analysis in Figure 13.21.

The SEM photographs in Figure 13.26 show binary CGH structures with decreasing sizes, and increasing optical proximity effects. The CGH on the left-hand side has square cell of  $2.0\text{ }\mu\text{m}$ , in the middle  $1.5\text{ }\mu\text{m}$  and on the right-hand side the basic CGH cell reduces to  $1.0\text{ }\mu\text{m}$ , therefore increasing the proximity effects in a  $1\times$  projection mask aligner from left to right. First, the square cell edges become rounded up, then the cell becomes round, and then finally it almost disappears.

Finally, Figure 13.27 shows multilevel CGHs produced by field-to-field alignment in a stepper tool and a binary CGH, with decreasing cell sizes (the cells here are also square).

The rounding effect (an optical proximity effect) can be seen in the right-hand side element, where the individual pixels are  $0.8\text{ }\mu\text{m}$  wide. Actually, the proximity effects are quite severe, since individual pixels that should not touch are actually touching.

However, as discussed earlier, digital optics are much less sensitive to proximity effects than IC chips are. The light sees the pattern in a parallel way, whereas the electronic signal in an IC chip sees the pattern in a serial way. In a Fourier CGH, if 25% or even 50% of the pattern is missing, the reconstruction still appears, with a decreased SNR of course (much like when tearing a hologram into two and still being able

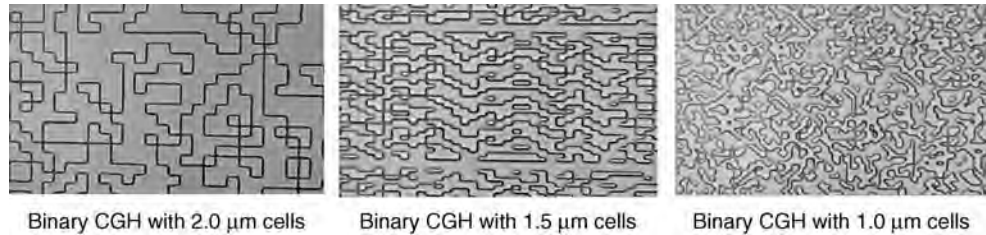


**Figure 13.26** Binary CGHs with decreasing feature sizes ( $1\times$  projection mask aligner)

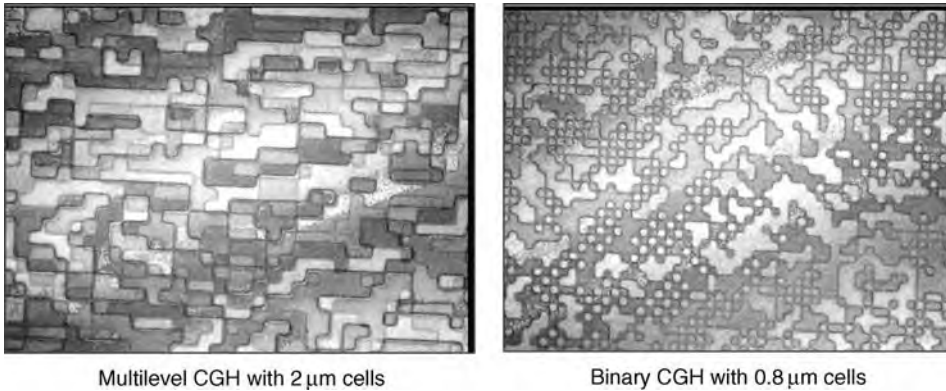


**Figure 13.27** Multilevel CGHs with decreasing cell sizes, fabricated via a G-line stepper

to view the entire 3D image). This is different if one considers a linear grating the efficiency of which is a function of the duty cycle factor within one period. The enlargement or reduction of the grating lines will directly affect the efficiency, but not the diffraction angle.

### 13.2.2.2 Compensation Algorithms

In order to compensate for optical proximity effects in a projection lithographic system, an algorithm similar to the one described for 2D or 3D EPE compensation can be used, by filtering the masks to be transferred into the final substrate one by one. The filtering process is performed after inverse numeric propagation to the reconstruction plane. The affected uncompensated and compensated CGH patterns are shown in Figure 13.28, along with the initial uncompensated pattern.

The affected compensated pattern is much closer to the original uncompensated unaffected pattern than the affected uncompensated pattern. But here again, the pattern fidelity is not an important issue, unlike in analytic-type element compensations; the accent is again rather put on reconstruction fidelity.

The optical reconstruction simulations show that there is a gain in efficiency of nearly 10% and a decrease of uniformity of about 4% for the compensated design. Add to this the effects of electron beam proximity effects and the anisotropic etching effects (see the next section), and one can easily see that the theoretical diffraction efficiency of the digital optical element is reduced by almost 20%.

### 13.2.3 Proportional Anisotropic Etching Effects

Chapter 12 has shown that proportional analog RIE transfer from resist into the underlying substrate is an important step in the fabrication of highly efficient digital optics [10–12].

In a similar way to the optical proximity effect in a stepper projection system, the adverse effects that are present in an anisotropic etch will now be analyzed. Etching is the step following the lithography step, for quasi-analog surface-relief elements, which form the vast majority of digital optics (either fabricated by multilevel lithography, direct analog write or gray-scale lithography).

The experimental characterization of proportional etching shows that the etching rate is actually dependent on the local slope of the shape to be transferred. In traditional binary etching, one does not bother about anisotropic etching effects, since the structures are mostly binary. This is why the IC industry and the main EDA software providers have not (yet) integrated any compensation techniques for anisotropic etching effects. Figure 13.29 reports the variation of the etch factor as a function of the local slope of the structure to be transferred into the substrate.

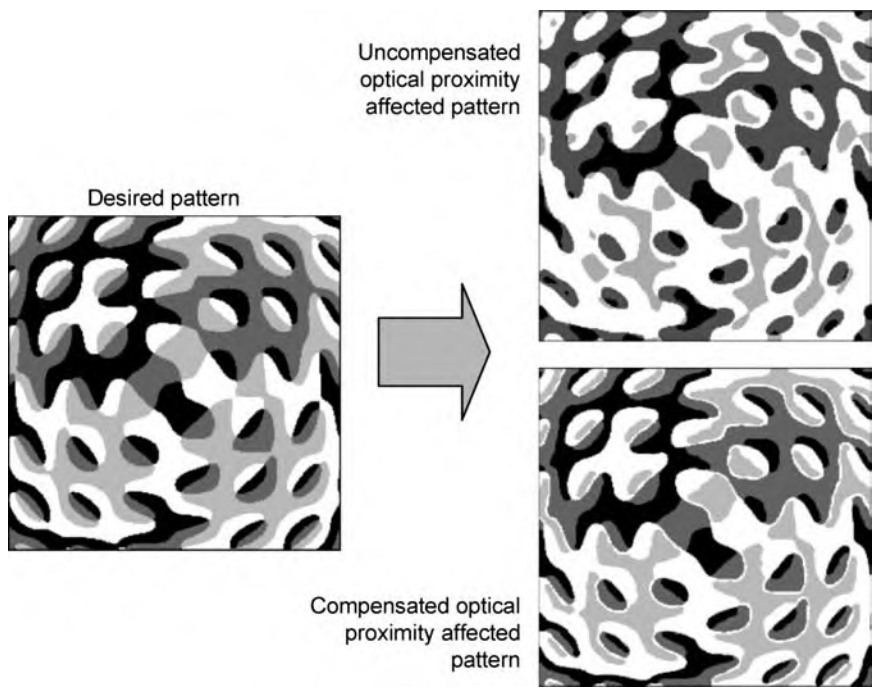


**Figure 13.28** The compensation of optical proximity effects in lithographic projection systems

The samples have been etched via a VEECO Instruments Micro Etch 301 system. The ion-milling system was modified to accommodate the introduction of reactive gases to enable the CAIBE process (see Chapter 12). The resist used was a Novolac-based resist OeBR-514 from Olin Ciba Geigy, which possesses a high resistivity against the CAIBE process.

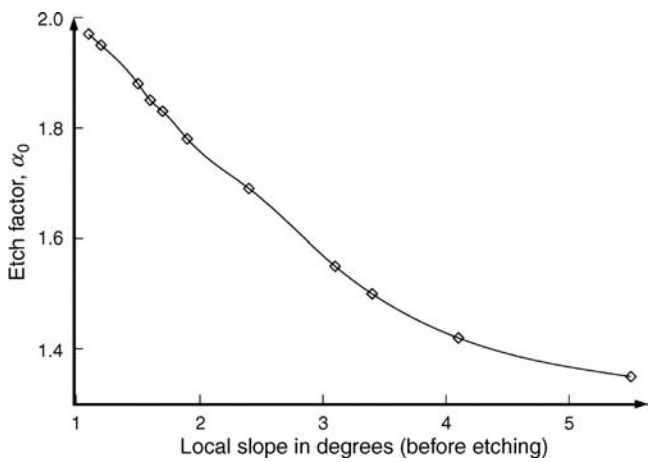


**Figure 13.29** Characterization of the anisotropic proportional etching process from resist to substrate



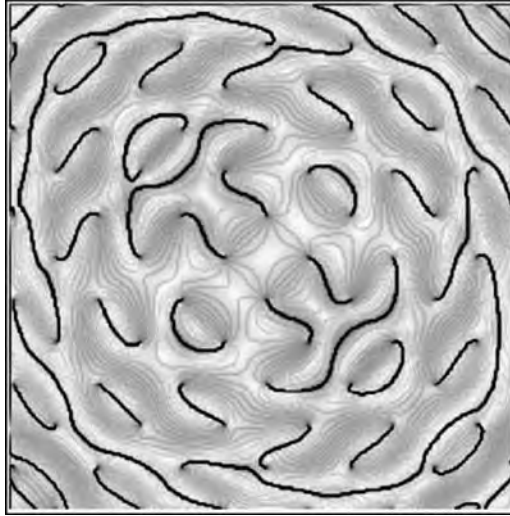


**Figure 13.30** A gradient map of the CGH profile in resist to be etched down the substrate

The experimental data set shows that the etch factor decreases with the local slope: therefore, either the resist etch rate increases and/or the substrate etch rate decreases with the local slope. Hence, a look-up table can be created to correct for the anisotropic etch, but only once the exact local slopes are being defined (with combined 3D EPE and optical proximity effect considerations).

### 13.2.3.1 Quantification of the Effects of Anisotropic RIBE Etching on CGH Performance

As the anisotropic etching rate is proportional to the local slope over the resist surface-relief profile, a gradient mapping of the CGH aperture is produced (with the same CGH as the one used in the previous sections). This gradient mapping is shown in Figure 13.30.

Based on the look-up table characterized experimentally (see Figure 13.29) and the previous gradient map (Figure 13.30), a numeric simulation of the anisotropic RIBE etching performed is presented in Figure 13.31.

The etched surface produces a ‘detail enhanced’ profile (best seen over the central fringe). This is due to the nonlinear behavior of the local etching rate versus the local resist profile slope. For example, the high structure on the left-hand side is perfectly transferred from the resist into the substrate, whereas the almost flat tops are deformed, and exaggerate the small defects in the resist profile due to the 3D e-beam proximity effects (see the previous section).

Based on Figure 13.31, one can conclude that a rule-of-thumb model is not sufficient for the wide range of effects that one can see, which include:

- large shallow structures are too shallow;
- large deep structures are too deep; and
- high-frequency structures have vanished.

There is therefore a need for an analytic compensation method that uses the exact look-up table measured in Figure 13.29.

As an indication, the numeric reconstruction is also computed in Figure 13.32.

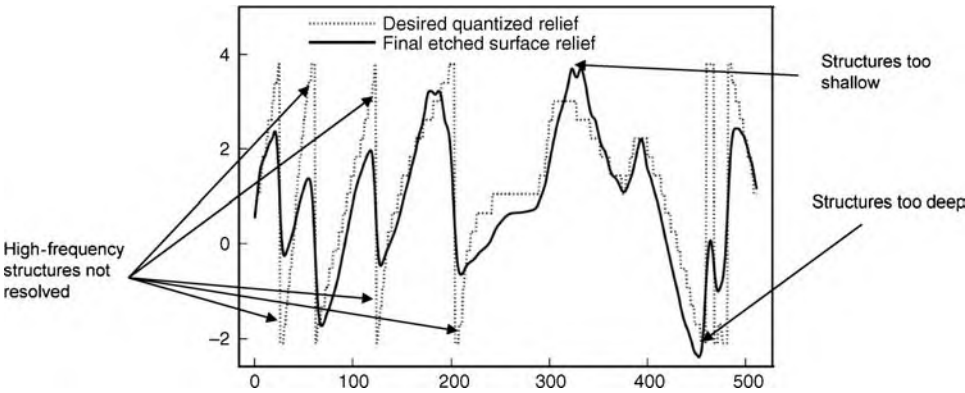


**Figure 13.31** Simulation of the final etched profile in the substrate

It can be seen that a large amount of noise is created in that reconstruction, on top of a deformation of the focused spots, which adds to the need to provide an adequate compensation method for the combined optical proximity errors in the lithographic stepper projection and the anisotropic proportional etching. This will be described in the next section.

As far as analog-type elements are concerned, the effects are the same, which leads to reduction of the side wall verticality and a change in the blaze slope, as shown in Figure 13.33 (a local section of a blazed diffractive lens).

On top of the anisotropic etching effects, there are other adverse etching effects that can also be modeled to a certain extent (and thus compensated for in the initial design). Some of these parasitic etching effects are described in Figure 13.34.

Figure 13.35 shows such an adverse effect, using a simple grating example.

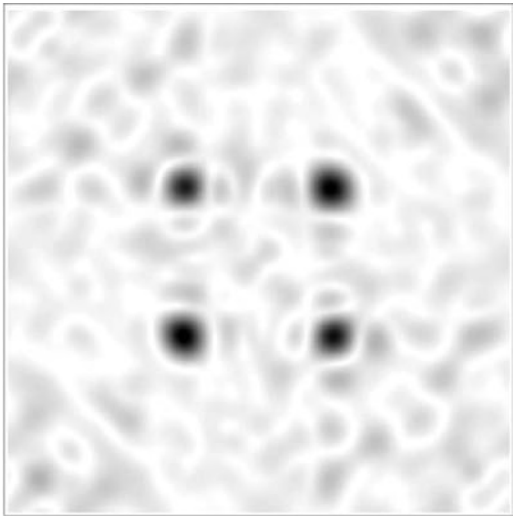


**Figure 13.32** Numeric reconstruction from an anisotropic etched CGH (noncompensated)

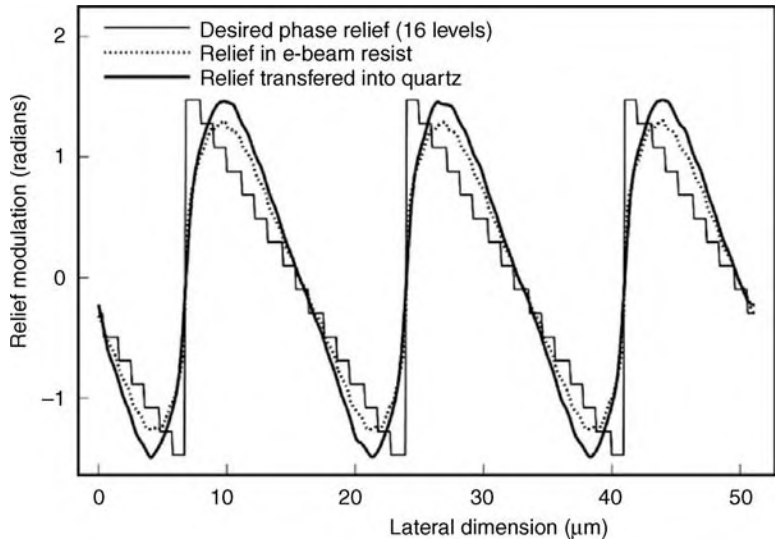


**Figure 13.33** Combined optical proximity effects in lithographic projection and anisotropic proportional etching for blazed structures

13.2.4 An Iterative Compensation Algorithm

This section proposes a general compensation method that aims to decrease the digital optical element’s sensitivity to the fabricating constraints and errors described previously (including e-beam proximity effects in resist, optical proximity effects in projection lithography and the anisotropic etching effect in

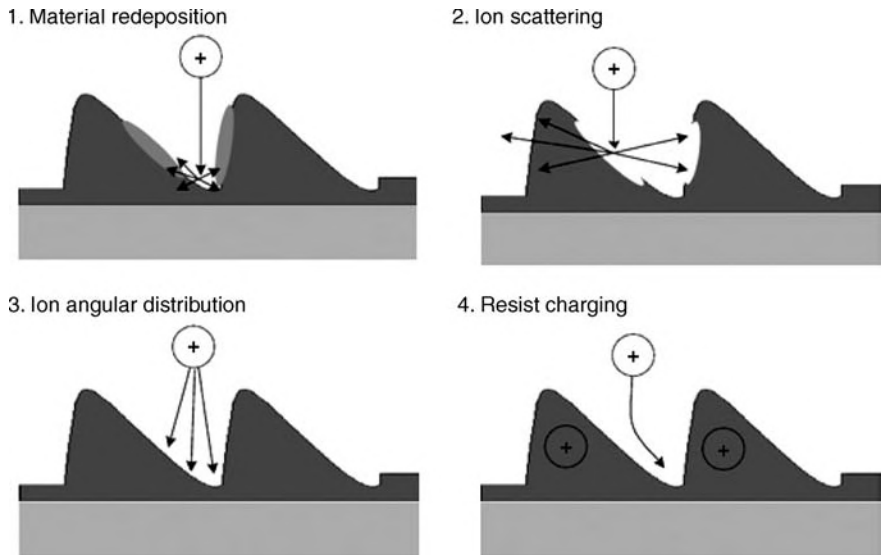


**Figure 13.34** Additional adverse etching effects in proportional etching

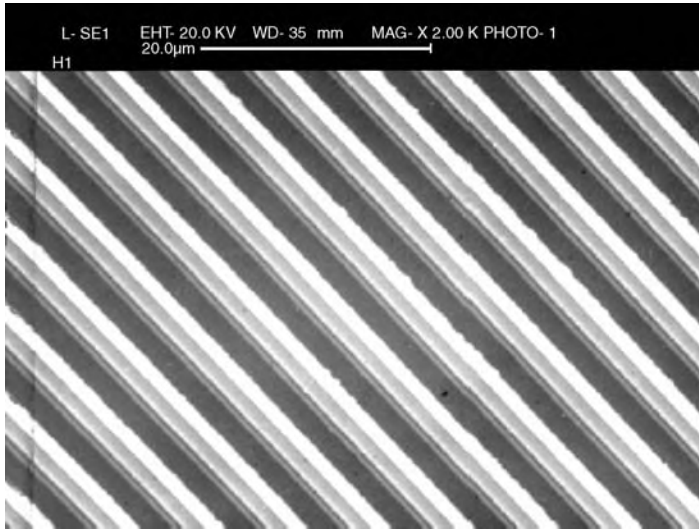


**Figure 13.35** An SEM of a grating with parasitic etching effects

proportional RIE transfer). This numeric technique is directly inserted within the element's design procedure, as has been described in Chapter 6 for numeric-type elements.

This technique is applied to the compensation of random errors and especially to anisotropic etching depth errors over the entire element's aperture. Due to the lack of *in-situ* etch depth characterization techniques (for more insight on these techniques, see Chapter 15) and to the nonlinear etching rate, the exact etch depth is hard to obtain. Typical tolerances of standard etching processes in mask shops are as low as 5 nm, which is still 25% of the height of a single step for a 16 phase levels DOE in reflection for HeNe laser light.

The technique alters the core numeric propagator that is used in either the IFTA-based or the steepest-descent optimization algorithm (see Chapter 5). The aim is to decrease the final element's sensitivity to the following reconstruction-related factor:

$$\rho = z\lambda \quad (13.19)$$

where  $z$  is the reconstruction distance (ideally, the focal length  $f$  of the DOE) and  $\lambda$  is the design wavelength.

This sensitivity factor is derived since  $f$  and  $\lambda$  never appear independently in the various numeric propagators described in Chapter 11, but always in the form of the factor  $\rho$ . Hence, if the element's response does not vary strongly with  $\rho$ , an element with extended depth of focus that would work correctly over a specific wavelength range is produced. Therefore, the following phase expression can be used instead of the regular quadratic phase factor for the Fresnel transform based propagator (see Chapter 11):

$$\phi(x, y) = \frac{\pi(x^2 + y^2)}{\lambda(f_0 + f_d \cdot \cos(\arctan(\frac{y}{x})))} \quad (13.20)$$

which actually resembles the phase profile of a 'Daisy lens' or 'sword lens' (see Chapter 5) that focuses in a line segment perpendicular to the diffractive lens plane instead of focusing into a single point (i.e. it increases the depth of focus). The line segment of the equivalent 'sword lens' is focused from  $f_0 - f_d$  to  $f_0 + f_d$ .

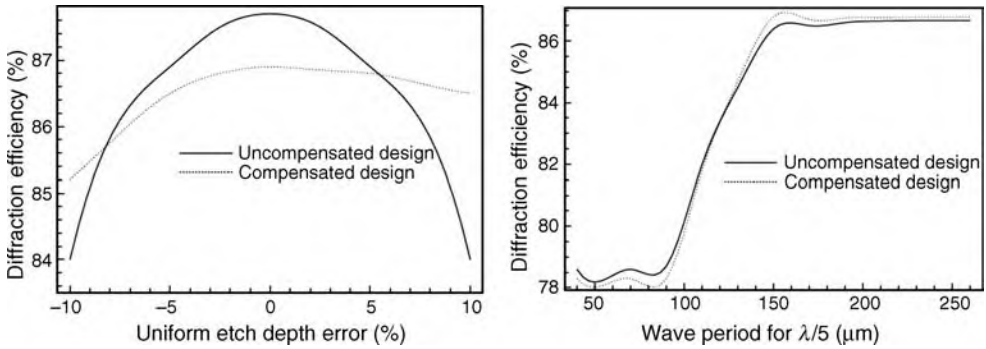


**Figure 13.36** The sensitivity of compensated and uncompensated elements to etch depth errors and substrate waviness

Since the digital optical element becomes less sensitive to the variations of the factor  $\rho$ , it also becomes less sensitive to wavelength variations. Besides, the only design parameter (apart from the substrate's material characteristics) that influences the desired optimal etch depth is the wavelength to be used for reconstruction. Therefore, by lowering the effects of wavelength drifts over the digital optical element's performance, the effects of anisotropic etching effects (directly related to equivalent wavelength drifts) are also reduced.

The effects of uniform etching errors over the CGH's efficiency (the previous CGH, but this time over 16 phase relief levels) are reported below, both for uncompensated and compensated DOE designs. The modeling of the effects of this compensation method when substrate waviness appears (here a substrate waviness amplitude of  $\lambda/5$  – see also the standard wafer specifications in Chapter 12) is also shown. Figure 13.36 reports these numeric results.

Although the uncompensated element yields the best results for perfect etching, the performance of the compensated element does not vary as quickly as before when etching errors occur, and gives the best results below  $-10\%$  with over  $5\%$  uniform etching errors.

On the other hand, the compensation algorithm does not improve the element's performance when the substrate shows waviness. Actually, its behavior does not change very much. It is, however, interesting to note that there seems to be a threshold of waviness frequency over which the effects of waviness decrease considerably (by  $10\%$ ). It is also noteworthy to see that below this frequency threshold the efficiency drops to a stable level.

### 13.2.5 Etch Bias Compensation

Direct biasing of structures to be etched is a linear straightforward compensation for the reduction of the features in a lithography/etching process, even for larger features, as described in Figure 13.37.

Such biasing is usually done through rule-of-thumb techniques, and inserted automatically in EDA DFM software.

### 13.2.6 The Phase Shift Masking (PSM) Technique

The previous sections have reviewed pure software techniques to increase the resolution of the mask (electron beam proximity effects compensation), decreasing the  $k_1$  factor (optical proximity effects compensation) as well as increasing the etching quality (anisotropic proportional etching effects and linear biasing). This section reviews another technique used to enhance the reticle design, and thus falls

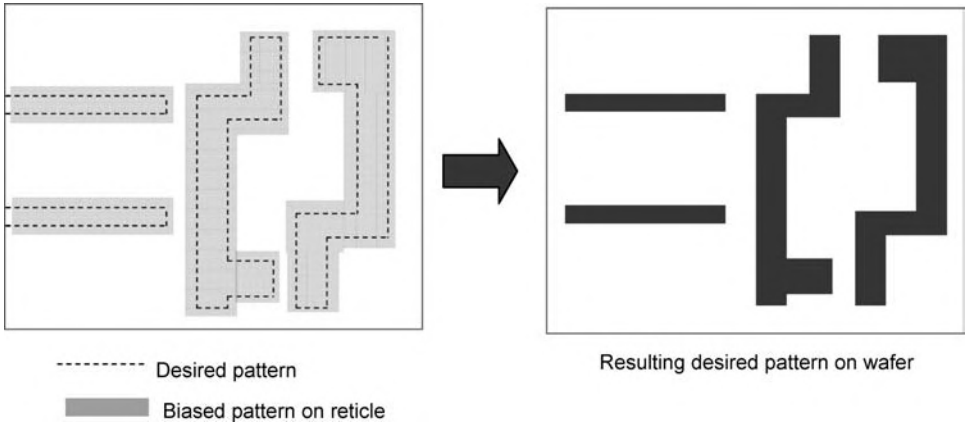


**Figure 13.37** Direct etch bias compensation

under the RET techniques. Phase-shifting masking techniques are a hybrid software/hardware solution to increasing the resolution of optical lithographic projection systems [13–15].

The IC fab industry and the EDA industry are using two main techniques today to implement phase-shifting RET techniques:

- the alternating phase-shifting mask (alternating PSM); and
- the attenuated phase-shifting mask (attenuated PSM).

**13.2.6.1 The Alternating Phase-shifting Mask**

Many diffraction orders are produced from a binary amplitude mask, which pushes light into the left and right side regions of the resulting aerial image. In order to reduce such effects, it was proposed to use alternating phase shifters from one feature to the other, which would produce a  $\pi$  phase shift, in order to produce locally complete destructive interferences, very similar to the design of binary digital optics (see Chapters 5 and 6). Such a phase shift is thus introduced by etching the right depth into the reticle for the considered projection wavelength in order to produce the  $180^\circ$  phase shift (see Figure 13.38).

As seen in Chapter 11, the depth for a  $\pi$  phase shift is

$$h = \frac{\lambda}{2(n - 1)} \tag{13.21}$$

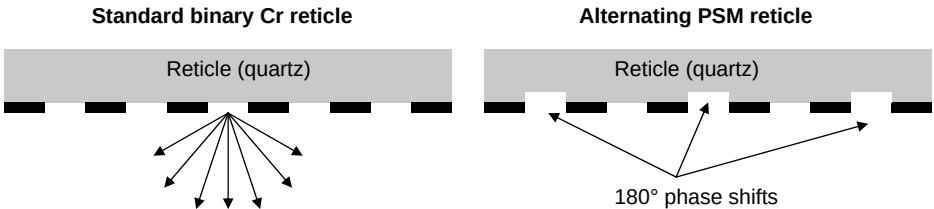


**Figure 13.38** An alternating phase-shifting mask technique

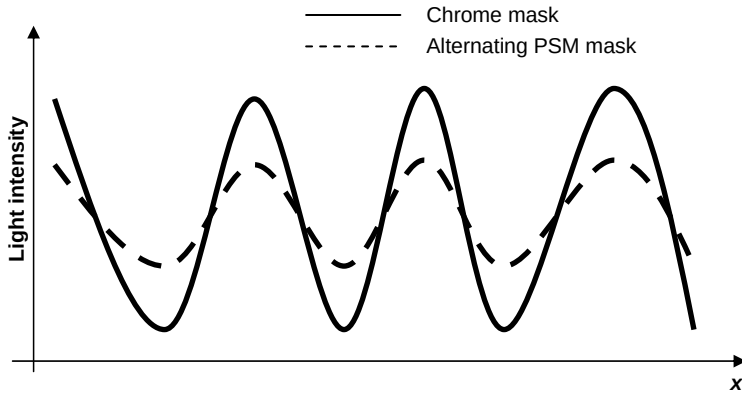


**Figure 13.39** Increasing the dynamic range of the aerial image in resist by an alternating PSM

Chapter 11 also showed that a  $\pi$  phase shift dramatically reduces the amount of light in the zero order, and pushes much more light into the fundamental orders, therefore increasing the dynamic range of the aerial image produced in the resist layer (see Figure 13.39). The alternating PSM is also called the Levenson Phase Shift Mask, after the name of its inventor.

In order to increase the resolution by a factor of two, one has to insert the phase shifters ranging from half a binary Cr pattern to the neighboring half of the Cr pattern. In this case, one cannot etch the substrate and pattern the Cr, since the Cr needs to reach through the etched grooves. In this case, phase shifters have to be inserted on the mask, by either  $\text{SiO}_2$  deposition and etching after the Cr patterning or simple resist patterning. Two techniques are used, the rim phase-shifting technique and the outrigger phase-shifting technique (see Figure 13.40).

Figure 13.40 also shows the amplitude of the rim PSM compared to that of the simple alternating PSM. The resolution here is multiplied by two. One can note the full amplitude swing of the modulation, whereas in a simple binary chrome mask the amplitude swing is limited to positive values.

The outrigger PSM uses an overlay principle, which is not resolved by the lithographic system; however, the phase shift still introduces the amplitude swing in the reconstructed pattern in the resist.

### 13.2.6.2 The Attenuated Phase-shifting Mask

A very desirable feature in PSM is attenuating the otherwise opaque regions of the reticle, by inserting a material in which optical transmission produces about 50% absorption (such as CrO, CrON, MoSiO or MoSiON), while also producing a phase shift of  $180^\circ$ . An attenuating PSM can be created in a single exposure, since there are no overlaps as in rim or outrigger PSM techniques. It is thus a cheaper method than the previous PSM techniques. Such a method is also known as halftone phase shifting.

## 13.3 Hardware Solutions

Hardware solutions to decrease the resolution include the numerous techniques described below. Usually, such techniques are much more costly to implement than the pure software RET techniques, which mainly require adequate EDA software development and higher resolution in the reticle plane. However, an EDA software including RET techniques is not cheap either. Lately, it has been the EDA software companies that have made most profits from the rush to decrease the resolution factor  $k_1$ , more so than the hardware companies.

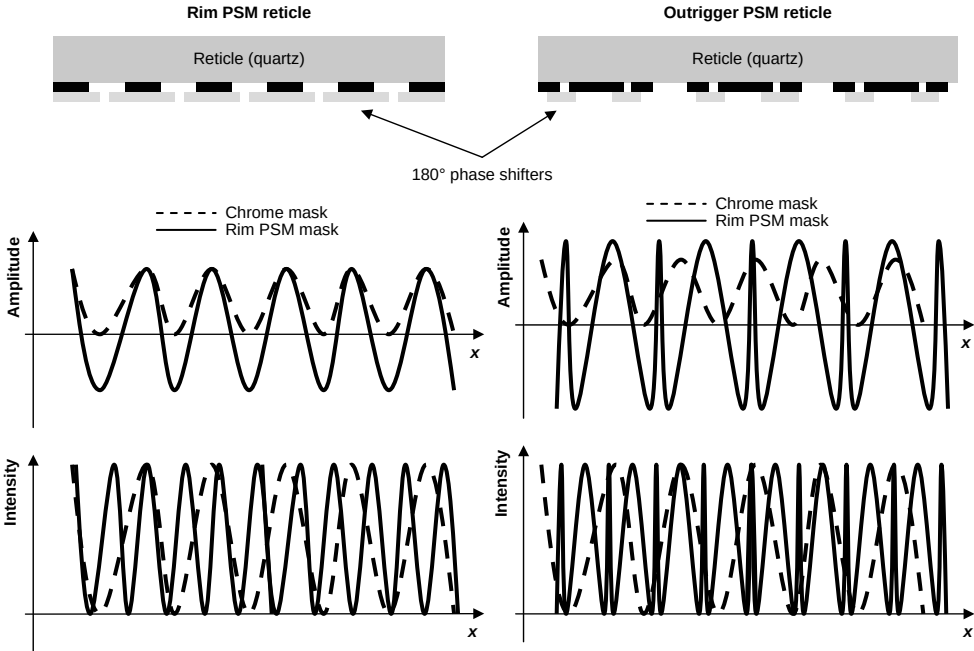


**Figure 13.40** A rim and outrigger PSM mask and its complex amplitude image

### 13.3.1 Reducing Wavelengths

Reducing the wavelength has been discussed at the beginning of this chapter. This is by far the most costly option in the IC fab industry (see Table 13.1). A major leap was achieved when migrating from mercury lamps to the use of UV and Deep UV (DUV) lasers (KrF, ArF and then F2), accompanied by modifying the projection lens train material from fused silica to calcium fluoride. There is another major leap from the DUV laser down to EUV sources, which require reflective optics (in off-axis mode) and operation in a vacuum (EUV). Other potential sources include X-ray sources and ion-beam lithography (IBL).

Due to the costs involved in changing the sources, in parallel with the RET software solutions presented in the previous sections, the IC fab industry has tried out other techniques, which are more or less costly. One of the most used techniques is the off-axis illumination technique. Other techniques use immersion lithography to increase the NA and soft lithography.

### 13.3.2 Off-axis Illumination

To produce an image, at least two diffraction orders are necessary (an interference pattern). For on-axis illumination, the fundamental diffraction orders are located symmetrically to the optical axis of the lens. In off-axis illumination, these orders are located in an asymmetric position, therefore pushing one fundamental order close to the optical axis of the lens (see Figure 13.41).

The direction of the illumination onto the reticle can be engineered by the use of an off-axis aperture placed between the condenser lens and the reticle. Such off-axis apertures can be either quadripole or annular, or may consist of other complex off-axis apertures, as described in Figure 13.42.



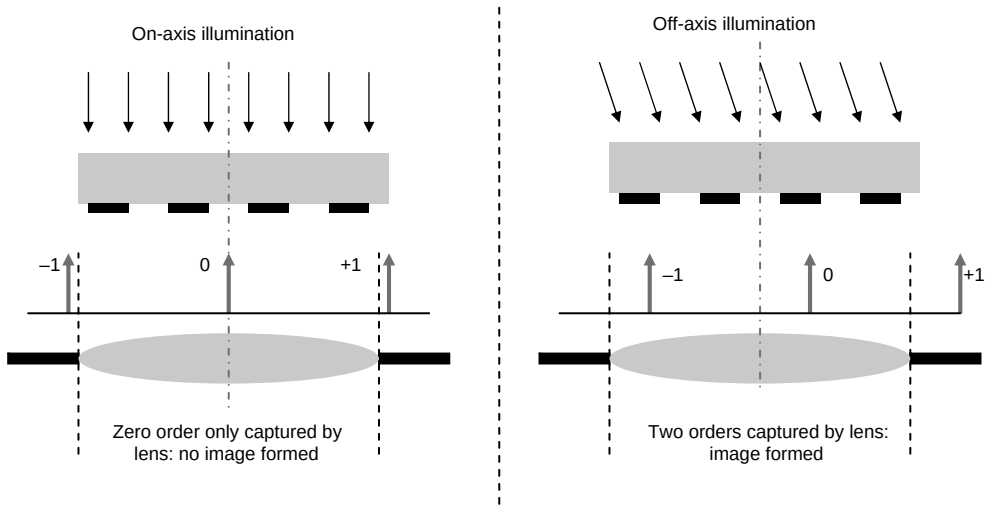


Figure 13.41 On-axis and off-axis illumination

Recently, the concept of using a Fourier CGH instead of a binary amplitude aperture has been introduced. The Fourier CGH generates a prescribed illumination, which is optimized for a specific pattern to be transferred to the wafer. Such an element produces in the far field a specific angular spectrum of plane waves, which is exactly what is needed for complex off-axis illumination. Figure 13.43 shows an example of a quasar illumination CGH, and its optical reconstruction.

Much more complicated illumination patterns can be created through a CGH (Fourier or Fresnel), as shown in Chapter 6. Such complex illumination tasks are a strong application sector for digital optics (see also Chapter 16).

13.3.3 Immersion Lithography

Immersion lithography is a complex hardware change that occurs at the wafer level, between the projection lens and the wafer. Here, the wafer is immersed in a liquid with an index of refraction  $n$ . Therefore, the NA of the lens seen from the wafer is enlarged by a factor  $n$ , without modifying the existing projection lens. Therefore, it is a very desirable feature to reduce the resolution factor  $k_1$ . Figure 13.44 shows the basic principle of immersion lithography.

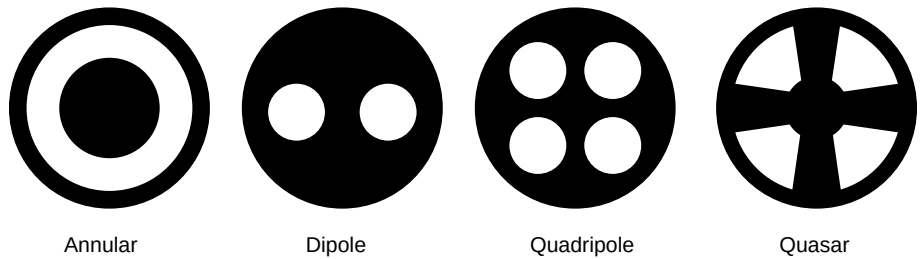


Figure 13.42 Off-axis illumination apertures

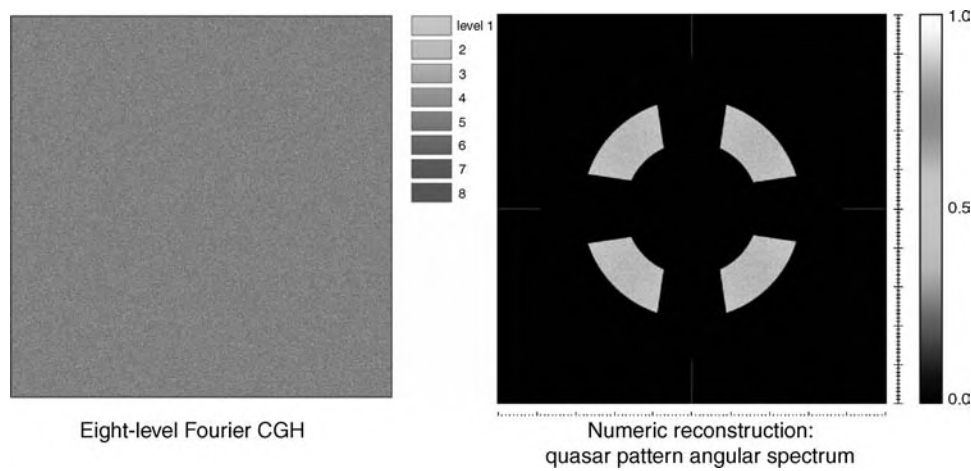


**Figure 13.43** Angular spectrum shaping for off-axis illumination in a stepper through a CGH

However, there are several problems that arise when implementing immersion lithography. The first one is that the wafer is now in a fluid, and therefore the resist is also in contact with the fluid. Photoresist is usually pre-backed to remove all the residual moisture before lithography, to prevent resist swelling. Here, the resist is placed directly into a fluid. Second, any turbulence of the fluid flow around the wafer would produce aberrations and thus reduce the resolution instead of increasing it. Finally, a fluid also has to be determined that has a matching index. There is no complex UV laser source involved in nanolithographic techniques, no complex stepper projection lens, no developer, no complex resists, no standing waves produced, no aerial image and so on. Chapter 14 will review four different techniques – Step and Flash Nano-Imprint (SNIL), thermoplastic NIL, roll nano-imprint and soft lithography using PDMS stamps – in detail.

*13.3.4 Maskless Lithography*

One can implement RET techniques to reduce the  $k_1$  factor by increasing the resolution of the reticle (mask). However, such high-resolution reticles are very costly. For example, in the 45 nm fabrication

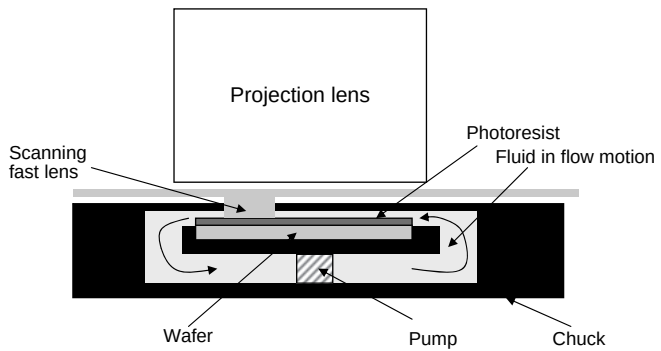


**Figure 13.44** Immersion lithography

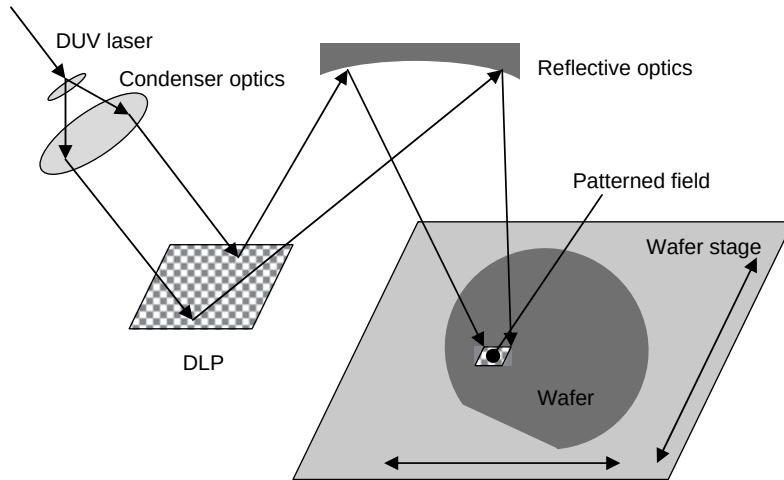


**Figure 13.45** Maskless optical projection lithography

node, the cost of an ‘easy’ mask including OPC is \$15 000, the cost of medium layers in the same node is around \$100 000 and the cost for a hard layer mask in that node can be as high as \$350 000.

One solution to this problem is to use an imaging device in place of the mask: this is the basis of maskless lithography [16, 17].

#### 13.3.4.1 Maskless Optical Projection Lithography

In maskless optical projection lithography, an imaging device is used, such as a simple programmable aperture plate system (APS) or a more complex DMD- or DLP-type microdisplay element – or a Grating Light Valve (GLV) (see Figure 13.45).

Such an architecture can be used in step mode, where the imaging device is a linear dynamic aperture stop, or in a stepper mode, where the imaging device is 2D.

#### 13.3.4.2 Zone Plate Array Lithography

Zone plate array lithography is similar to the previous maskless example; however, here one uses an array of high-NA diffractive microlenses (however, the old name ‘zone plate’ is used here), in order to focus an array of spots onto the photoresist (instead of the high-NA lithographic lens).

The source used in maskless lithography can be electron beams, DUV lasers or EUV sources. Although maskless lithography seems to make lots of sense economically, it is, however, still at the R&D stage, mainly due to problems linked to the reflective optics, field-to-field stitching errors and so on.

### 13.4 Process Solutions

Process solutions are cheaper to implement than hardware changes, especially when they are as severe as source wavelength change or immersion lithography. The process solutions discussed here are related to resist technology and exposure processes.

#### 13.4.1 Photoresist Technology

Chapter 12 shows that there are many different resists used in industry, from organic to inorganic. They can be positive or negative, but in most cases they rely on a polymerization process.

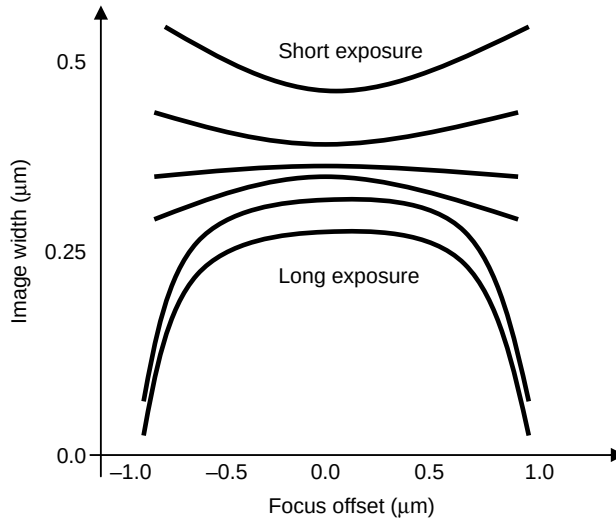


**Figure 13.46** The focus exposure matrix

The most commonly used photoresist today is a two-component system that consists of Novolac resin and the diazonaphthoquinone photoactive compound (or PAC, a dissolution inhibitor). The factors controlling the generation of an aerial image in the resist rely on many factors, including the absorption coefficient (the optical density of the resist), the thickness of the resist and the underlying substrate. The reflections from the underlying substrate produce standing waves, as described for Lippmann photography and holography (see Chapter 8). Such standing waves produce surface roughness within the side walls of the structures imaged.

In order to test the resist response in a lithographic system, it is usual to run a focus/exposure matrix job deck as depicted in Figure 13.46 (for the definition of a job deck, see Chapter 15). Chemically amplified resists can also help to produce higher resolutions.

The main problem with Deep UV (DUV) resist and Extreme UV (EUV) resists is the strong absorption of such radiations by almost all materials (resist included). Thus, the absorption depth when using DUV or EUV resist is less than 100 nm, therefore challenging the production of structures with high aspect ratios. Multilayer resists are being developed to find solutions to this problem. Alternative techniques include the use of X-ray lithography and deep proton irradiation.

### 13.4.2 Double-exposure Lithography

In double-exposure lithography, the task is to produce a smaller resolution than a single lithographic step would produce. The double-exposure technique is depicted in Figure 13.47.

The first exposure produces an intentionally underexposed (for positive resist) or overexposed (for negative resist) pattern. This pattern is then developed and transferred onto a hard metal layer underneath the original resist layer (Cr).

Then a new resist layer is spun on the processed wafer, and the wafer is again placed in the lithography tool with the same (or another) reticle. The alignment of the reticle to the wafer has to be very accurate, since the pattern (here a linear grating) is shifted by half a period in order to produce another under- (or over-) exposed pattern which, after second-step processing, leaves structures that can be half the size of the conventional structures that would arise from the use of a single-exposure process.

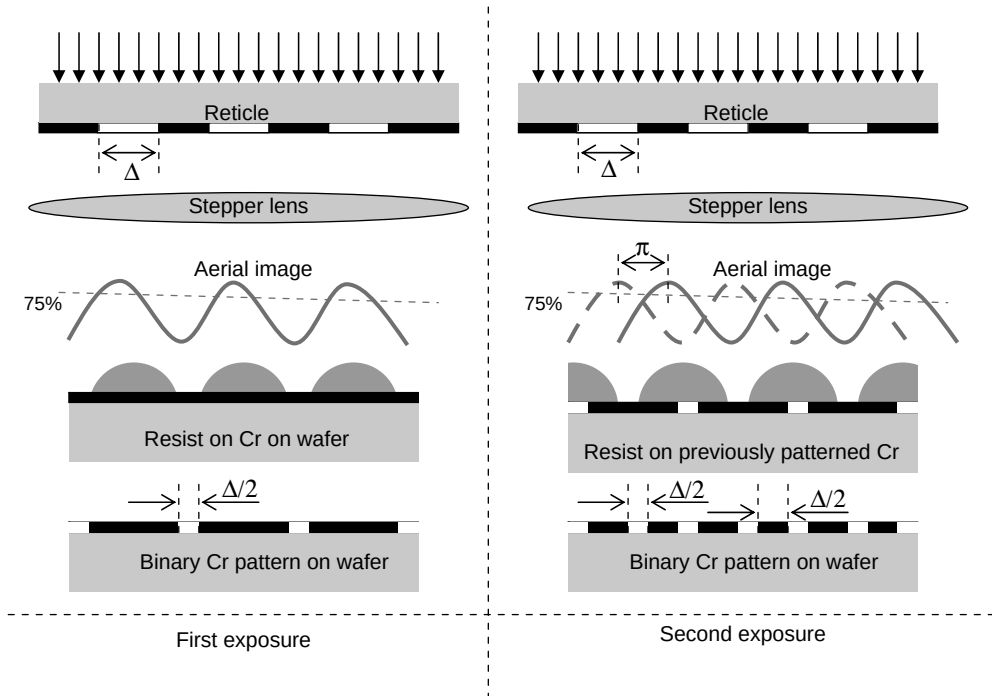


**Figure 13.47** Double-exposure lithography

The double-exposure process is not an easy task, since the over- (or under-) exposure has to be calibrated accurately, and the field alignment (with a  $\pi$  shift) puts an increased burden on the alignment tolerances for the stepper.

### 13.4.3 Holographic Interferometric Lithography

In holographic interferometric lithography, the photoresist is exposed through two or more collimated UV beams at very precise angles, to produce periodic structures in 2D, or even 3D, which are ideal for the fabrication of high aspect ratio sub-wavelength structures for photonic crystals or metamaterials. This process can be produced at wafer scale, or can be produced in a holographic fringe writer. Such a local interferometric holographic exposure process can actually produce 3D sets of fringes in a thick resist, voxel by voxel, making sure that there is no phase shift from one voxel to the other by using a phase-locking device (see Section 8.7.5).

If the 3D structure is to be replicated over a large area, traditional holographic interferometric lithography can be used (with two, three or even four beams), and if the 3D pattern has to vary smoothly in one or two directions (such as slanted Bragg gratings to form a holographic lens), a holographic interference fringe writer with phase-locking control can be used.

This chapter has reviewed the various software and hardware tools and related process methods employed in order to increase the resolution of the lithographic systems used to produce digital optics today. Several compensation techniques have been described that can alleviate systematic fabrication errors, and integrate them directly in the digital optical element design process. Such systematic errors include e-beam proximity effects in resist, optical proximity effects in projection tools, anisotropic proportional etching and so on.

## References

- [1] M.D. Levenson, 'Extending optical lithography to the gigabit era', *Microolithography World*, Autumn 1994, 5–13.
- [2] B.J. Lin, 'Where is the lost resolution?'; In H.L. Stover (ed.), *Proc. SPIE*, Vol. 633, 1986, 44–50.
- [3] A.K.-K. Wong, '*Optical Imaging in Projection Microlithography*', SPIE Press, Bellingham, WA, 2005.
- [4] J. Braat and P. Rennspies, 'Effects of lens distortion in optical step-and-scan lithography', *Applied Optics*, **35**(4), 1996, 690–700.
- [5] T.E. Zavec, 'Machine models for lithography', *Microolithography World*, Winter 1995, 14–18.
- [6] N.K. Eib, 'The art and science of lithography simulation', *Microolithography World*, Winter 1996, 11–15.
- [7] R. Socha, 'Resolution Enhancement Techniques (RET) for Optical Lithography: Course 2. Mask Resolution Enhancement Techniques (RET)', Short Course CO2067, ASML, Semi Zone, 2005.
- [8] N. Cobb and A. Zakhor, 'Fast sparse aerial-image calculation for OPC', in '*15th Annual BACUS Symposium on Photomask Technology*', G.V. Shelden and J.N. Wiley (eds), SPIE Vol. 2621, 1995, 534–545.
- [9] T. Brunner, 'Impact of lens aberrations on optical lithography', *IBM Journal of Research and Development*, **41**(1), 1997, 57–67.
- [10] S.C. Jackson and T.J. Dalton, 'A profile evolution model with redeposition', in '*Dry Processing for Submicrometer Lithography*', SPIE Vol. 1185, 1989, 225–233 (1989).
- [11] W. Daschner, M. Larsson and S.H. Lee, 'Fabrication of monolithic diffractive optical elements by the use of e-beam direct write on an analog resist and a single chemical assisted ion-beam etching step', *Applied Optics*, **34**(14), 1995, 2534–2539.
- [12] S. Somekh, 'Introduction to ion and plasma etching', *Journal of Vacuum Science Technology*, **13**, 1976, 1003–1007.
- [13] M.D. Levenson, 'The phase-shifting mask II: imaging simulations and submicrometer resist exposures', *IEEE Transactions on Electron Devices*, **31**(6), 1984, 753–763.
- [14] R. Kostelak, C. Pierrat, J. Garofalo and S. Vaidya, 'Exposure characteristics of alternate aperture phase-shifting masks fabricated using a subtractive process', *Journal of Vacuum Science Technology B*, **10**(6), 1992, 3055–3061.
- [15] C. Pierrat, A. Wong and S. Vaidya, 'Phase shifting mask holography effects on lithographic image quality', in *Technical Digest – International Electron Devices Meeting*, 53–56 (1992).
- [16] Ch. Brandstätter, H.-J. Doering and K. Reimer, 'MaskLessLithography', Sematech Litho Forum, Los Angeles, January 27–29, 2004.
- [17] R. Leachman, 'Relative Economics of Masked and Maskless Lithography', Short Course, Department of Industrial Engineering and Operations Research, University of California at Berkeley, November 3, 2006.

# 14

## Replication Techniques for Digital Optics

The fabrication of diffractive optics makes sense in an industrial setting only if these elements can be cheaply mass-produced. This is the *de facto* condition for diffractive optics to emerge from the laboratory and address industrial needs in consumer electronics products and other fields of application, such as the automotive industry, factory automation, biomedical and telecom applications, and so on. In the previous chapters, we have emphasized the generation of master elements, by holographic recording, diamond machining or microlithography.

There are several methods available to the optical engineer to replicate these master elements in volume. These include roll embossing, hot embossing, UV casting, injection molding and the sol-gel process. These various techniques will be reviewed in the present chapter.

### 14.1 The LIGA Process

Historically, the LIGA process was the first replication process used for mass-production of microstructures. Originally, it was intended for the replication of high aspect ratio microstructures used in particle detectors for atomic energy research [1,2]. The term LIGA comes from the German ‘Lithografie’, ‘Galvanoformung’ and ‘Abformung’ (lithography + electroforming + molding). Most of this early research was done at the German Kern Forschungs Zenter (KFZ) (the German Atomic Agency).

This technology allows high aspect ratio structures to be defined in nickel. The process consists of exposing a sheet of PMMA bonded to a wafer using X-ray lithography. The PMMA is then developed and the exposed material is removed (see Figure 14.1). Nickel is then electroplated up in the open areas of the PMMA. The nickel over-plate is removed by polishing, leaving high aspect ratio nickel parts. The PMMA is removed, and the nickel parts may remain anchored to the substrate or be released.

X-ray lithography is a very costly process (the source is usually a Synchrotron!); however, one can achieve very high aspect ratio structures due to the very small wavelength used during image projection (X-rays). The photomask used in such an X-ray lithographic device is very different from the standard Cr on quartz or even PSM masks discussed previously (Chapters 12 and 13). Figure 14.2 shows the structure of such an X-ray mask.

The X-ray photomask is deposited as a gold (Au) pattern on a thin Si membrane, mounted on a structural rigid frame (usually a silicon wafer). Glass and graphite membranes are also sometimes used). The membrane must be thin enough to let enough X-rays through and strong enough to withstand the radiation

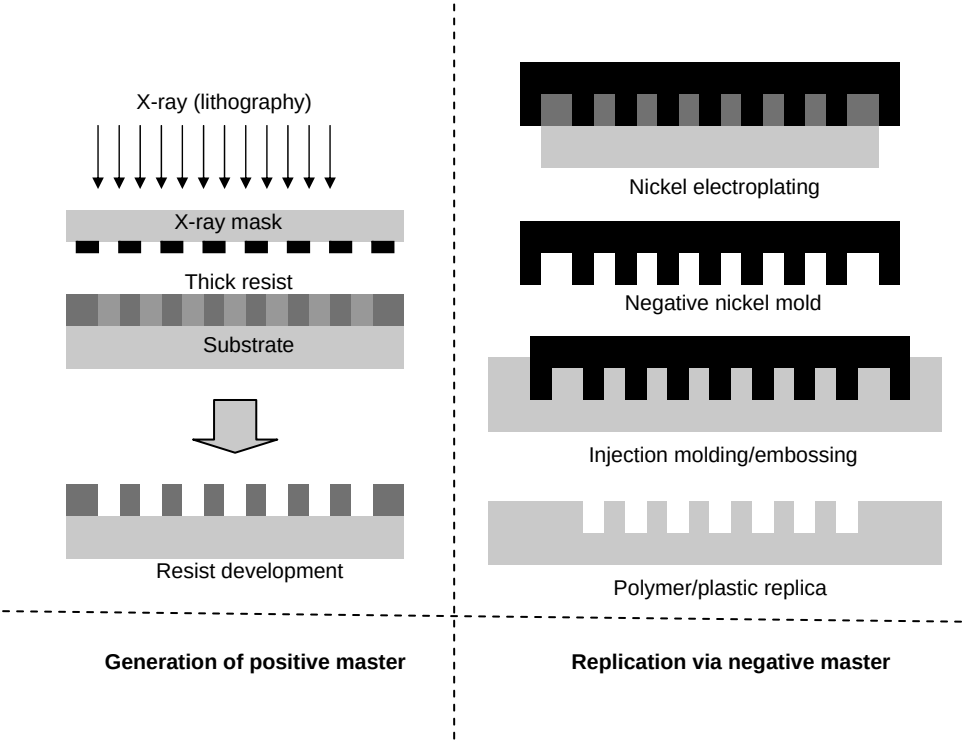


**Figure 14.1** The LIGA process

damage. X-ray resists are similar to electron-beam resists used in photomask patterning. Figure 14.3 shows typical X-ray high aspect ratio microstructures.

Such high aspect ratio microstructures can be adapted for the fabrication of special digital optics, especially in the vertical dimension (e.g. the *échelette* grating in Figure 14.3, which can be part of a waveguide grating router or WGR).

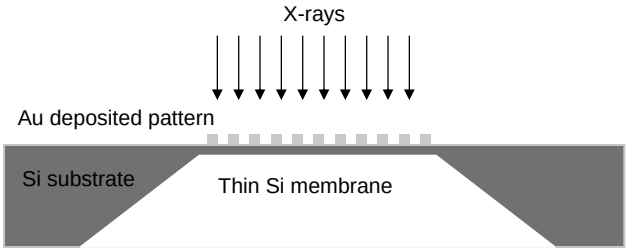
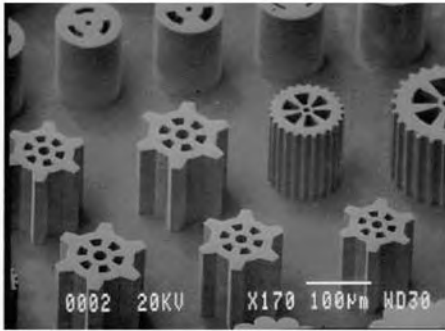
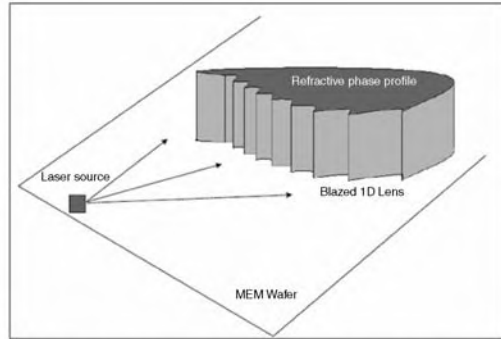


**Figure 14.2** The structure of an X-ray photomask





Typical LIGA structures



LIGA structures adapted to digital optics

**Figure 14.3** Typical LIGA high aspect ratio microstructures

## 14.2 Mold Generation Techniques

Mass-replication of a positive master element includes three successive processes:

1. Generation of a positive master through microlithography, diamond turning and so on.
2. Nickel electroplating of a negative master (mold on nickel film).
3. Generation of a final mold and replication (embossing or injection molding).

### 14.2.1 Diamond Machining

Diamond turning and diamond ruling (see Chapter 12) can produce high-quality master molds either directly, as the final hard material negative mold (nickel), or as an intermediate positive mold in plastic or other soft material. The typical elements that can be diamond turned directly into a final mold are mostly blazed or échelette linear gratings, and blazed diffractive and micro-refractive Fresnel lenses [13] with circular symmetric features.

### 14.2.2 Electroforming

Electroforming is one of the main techniques used to produce molds for digital optics [3]. The process requires the prior fabrication of a positive master (see Chapter 12) – identical to the final element – such as a photoresist layer, or a plastic or glass substrate. Figure 14.4 shows the electroforming of a nickel shim from a photoresist profile on a Si wafer. Sometimes a pre-coating of another metal is used to optimize the formation of the nickel shim on the surface (e.g. Ag).

As electroforming is used to produce the negative mold, it is better to use a soft material such as photoresist, which can be dissolved after electroforming. However, this also has its disadvantages, since the resist tends to swell in the electrolytic bath. Usually, when the resist profile is used to produce the final element (by etching, for example), the resist is post-baked and dried in order to remove all the moisture. Here, the resist is plunged into a liquid. The underlying substrate should be conductive, or at least partially conductive, in order to proceed to the electroforming in between an anode and a cathode.

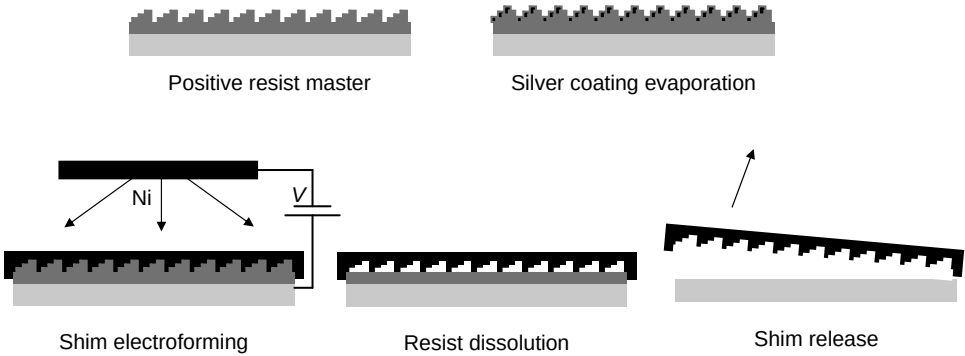


**Figure 14.4** The electroforming process for a negative nickel shim

When using an etched quartz substrate, two problems can arise during the electroforming process:

- the quartz is not conductive; and
- the quartz wafer can easily break during electroforming, due to the surface tensions of nickel on quartz.

When using an etched quartz substrate, it is better to use an extra thick quartz substrate, which has a thickness of about 500  $\mu\text{m}$  for a 4 inch wafer and 700  $\mu\text{m}$  for a 6 inch wafer (see Chapter 12), rather than the standard quartz wafers.

Figure 14.5 shows a nickel shim (on the right) produced by the electroforming of resist structures lithographically patterned on a silicon wafer master (on the left). The nickel shim shown in the figure is placed on a larger metallic plate for added rigidity. The initial wafer was a 500  $\mu\text{m}$  thick 4 inch silicon wafer with a resist coating thickness of 0.9  $\mu\text{m}$ .

Nickel shims are very thin, and need to be placed on a stronger base, which is also made out of nickel. This base can be mounted on another base for added rigidity – for hot embossing, for example. The microstructures in nickel are very hard, and conform well with the original resist microstructures (see Figure 14.6).

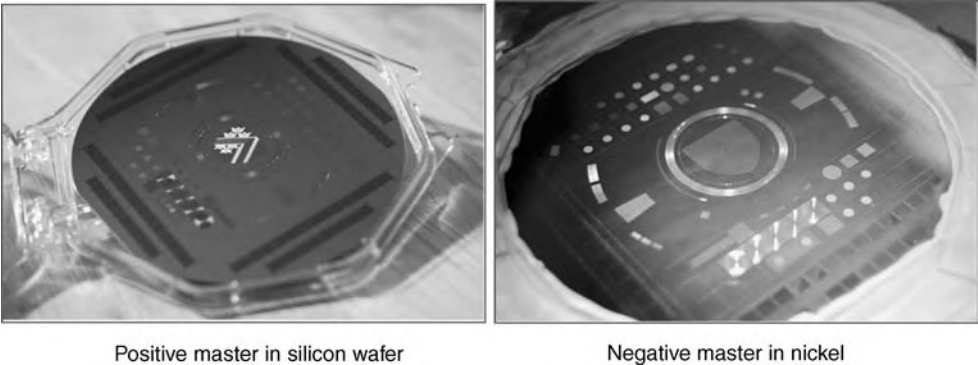


**Figure 14.5** A resist-on-silicon master and a replicated shim (binary diffractive optics)

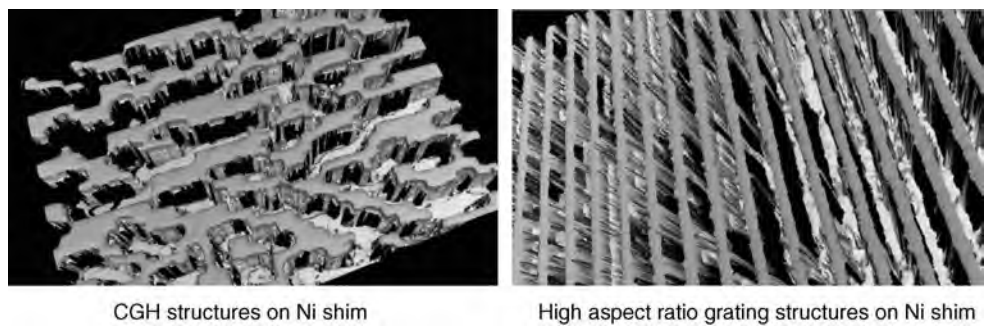


**Figure 14.6** An example of micro-profilometry over nickel shim microstructures

14.2.3 Shim Replication and Recombination

Once the nickel shim has been produced, shim recombination is performed in order to produce a large shim that can be replicated in order to produce a set of shims, depending on the number of replicas (see Figure 14.7).

The master nickel shim is the first nickel master grown as a result of electroforming and the image is wrong-reading (a negative master). That master, and potentially others grown from other positive masters, is replicated to form a ‘mother’ (a right-reading image, or positive master) and those ‘mothers’ are replicated to form ‘daughters’ (wrong-reading images, again negative masters). The master nickel shim is

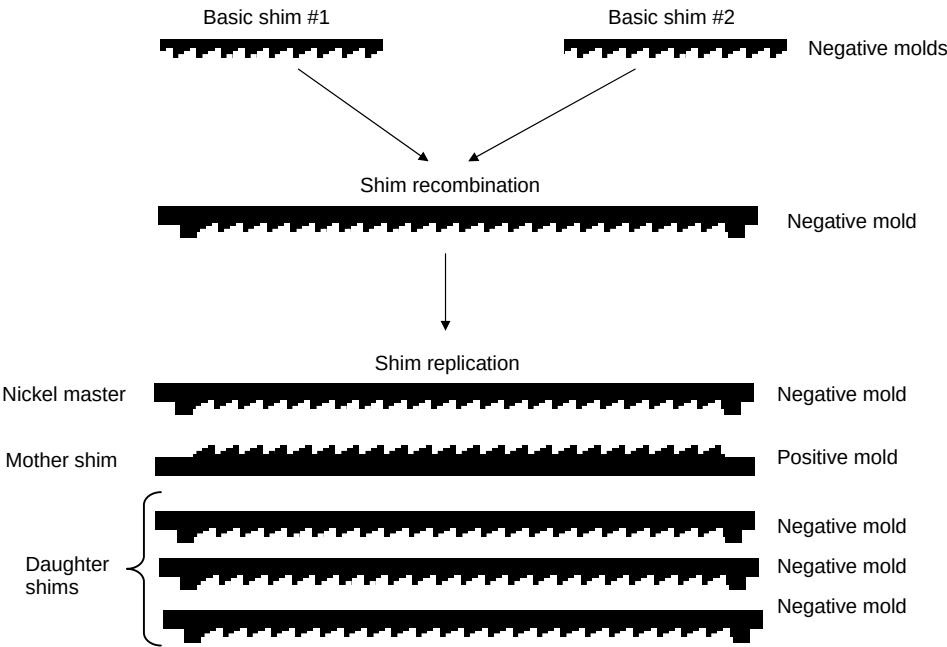


**Figure 14.7** Shim recombination and the generation of daughter shims

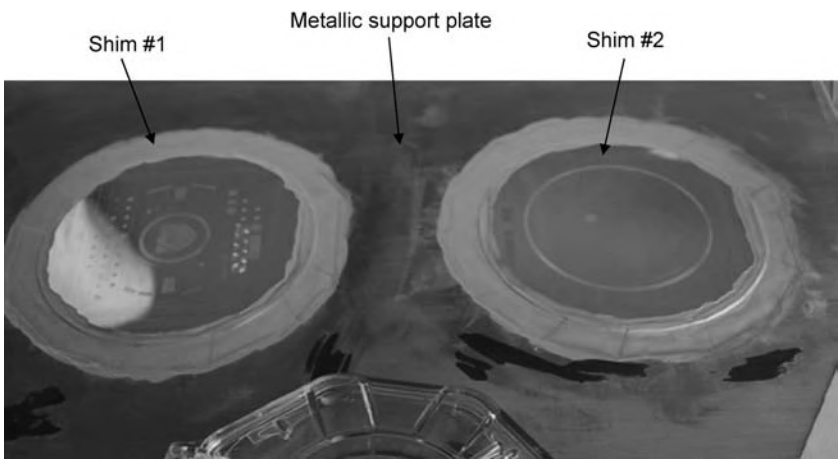


**Figure 14.8** The recombination of two shims in a single embossing shim plate

archived and the mother is stored and inventoried to grow more daughters as needed. The ‘daughters’ are then sheared to the customer’s mounting specifications and are shipped to the customer’s production facility.

Figure 14.8 shows an example of a recombined shim, including two different masters from two different resist on silicon wafers. The resulting shim is to be used in hot embossing. After this process, the shim is prepared for injection molding, casting, hot embossing or roll embossing (Figure 14.9).

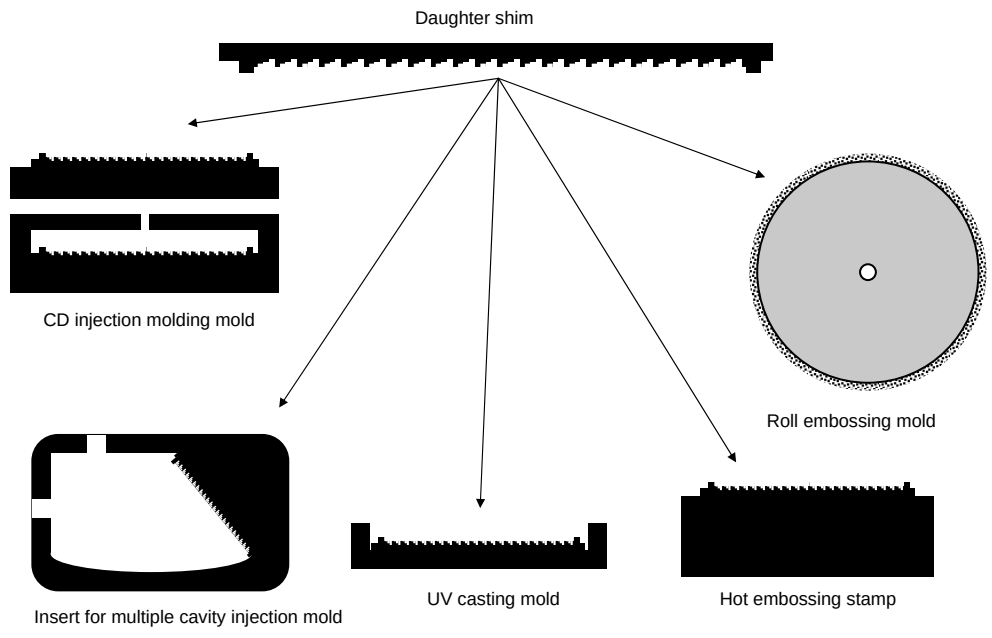


**Figure 14.9** Preparation of the mold for injection molding, casting, hot embossing or roll embossing

## 14.3 Embossing Techniques

Embossing techniques are the fastest and cheapest techniques for the replication of microstructures in a plastic material [4–8]. Various embossing techniques are available for the replication of microstructures, ranging from hot embossing, to roll embossing to UV embossing.

### 14.3.1 Hot Embossing

In hot embossing, the Ni mold is used to stamp a soft polymer. Before the stamping, the polymer is raised just above its glass temperature. This method has been successfully applied to create grating microstructures and optical waveguides in polymer layers.

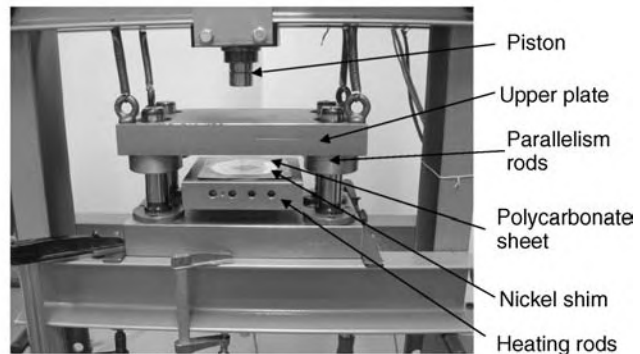
There are three phases during hot embossing:

- Phase 1:
  - closing and evacuation of the process chamber;
  - heating of the mold insert and substrate;
  - plastification of the polymer sheet.
- Phase 2:
  - elevation to maximum temperature;
  - pressing of the Ni shim mold into the polymer;
  - application of the maximum molding pressure.
- Phase 3:
  - cooling of the mold insert and substrate;
  - venting of the chamber;
  - opening the mold and removing the molded micro-parts.

Hot embossing is a very simple and precise process (see Figure 14.10).



Hydraulic press



Embossing head

**Figure 14.10** The hot embossing process

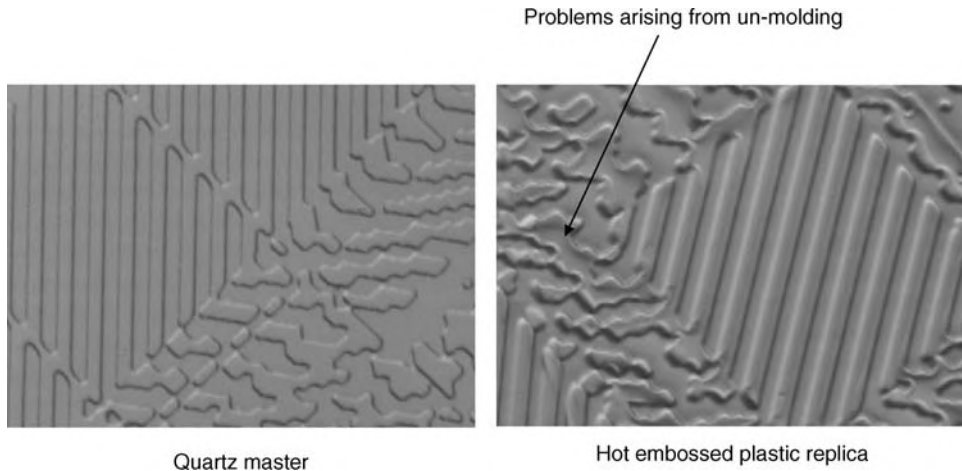


**Figure 14.11** A comparison of microstructures between an initial quartz master and an embossed plastic replica

A hot embossing press can be fabricated easily on the basis of a standard hydraulic press, by integrating heating pads and vertical parallel sliding rods to ensure an even pressure application. It is also a gentle process due to a low embossing and un-molding velocity. Therefore, the material remains homogeneous and produces low internal stresses, such as birefringence. Hot embossing is compatible with other technologies and allows integration of complex systems with reduced assembly efforts.

However, there can be problems arising from the un-molding, such as side-wall imperfections and top surface modulations, such as those shown in Figure 14.11. The photograph on the left shows the initial quartz master, and the photograph on the right is the plastic replica made using the hot embossing process.

Figure 14.12 shows some examples of diffractive elements replicated by embossing on polycarbonate (PC) plastic sheets.

Hot embossing remains a reproducible, cost-effective process for small-scale series, for high-performance digital optical microstructures. Binary grating structures with periods down to the sub-micron level have been replicated in PC sheets. Figure 14.13 shows the surface topology, using a white light confocal interferometer microscope, of an embossed replicated grating in plastic. The binary grating structures have been rounded by the process. In some cases this might be a disadvantage, but in others it is an advantage (when quasi-sinusoidal gratings are desired as a final product). Figure 14.13 also shows various artifacts present on such embossed plastics, such as scratches, dust particles, defects, tears and so on.

### 14.3.2 Roll Embossing

When the number of desired replicas is very large and the price has to be very low, the solution might be to use roll-to-roll embossing, in which the shim is directly inserted on a roll and plastic is fed through (Figure 14.14). In the figure, the final patterned roll is also coated with a protective laminated sheet, and a back-coating film that can act as an adhesive layer. Figure 14.14 also shows the problem of foil stitching, which is one of the main problems with roll embossing. The stitching is on the roll and translates into a repetitive stitching error line on the roll.

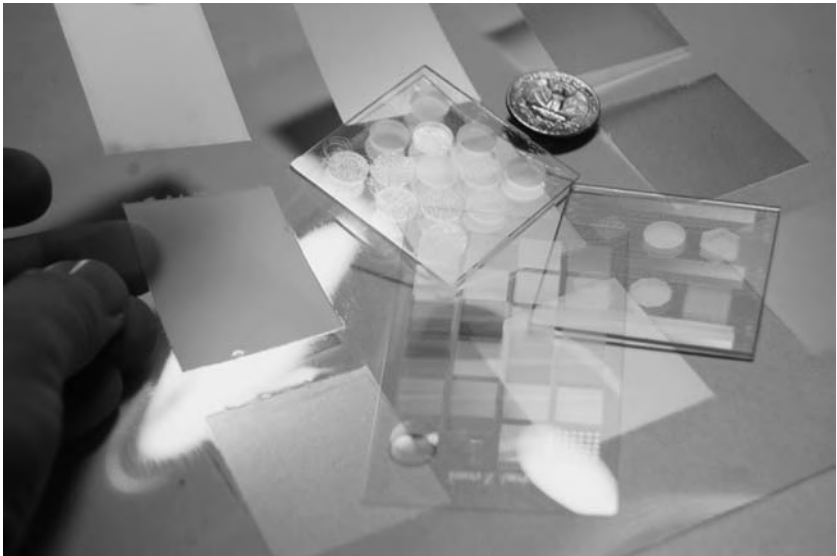


**Figure 14.12** Embossed digital diffractive optics in polycarbonate plastic sheets

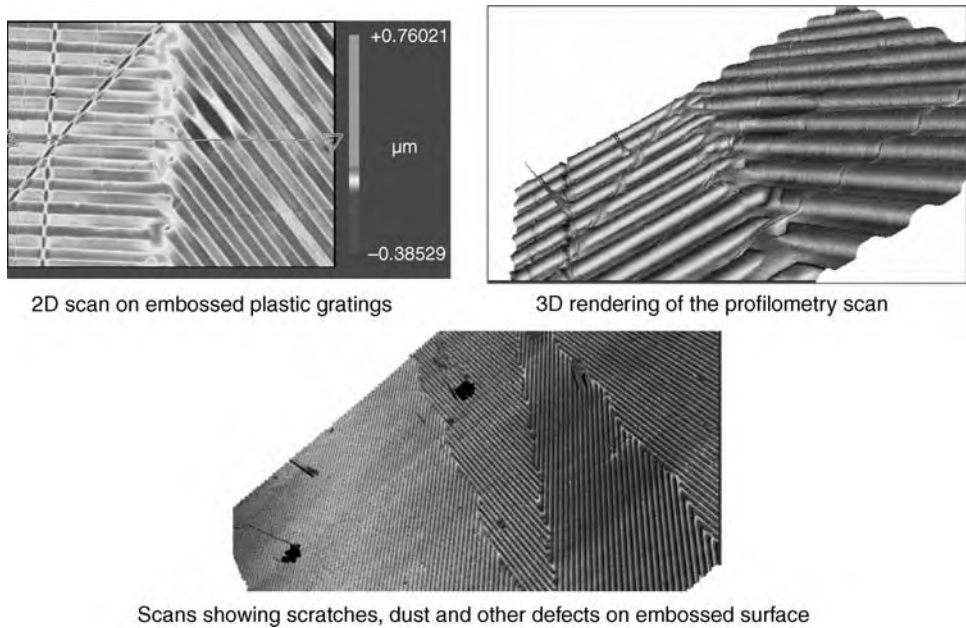


**Figure 14.13** The surface topology of a grating replicated by plastic embossing

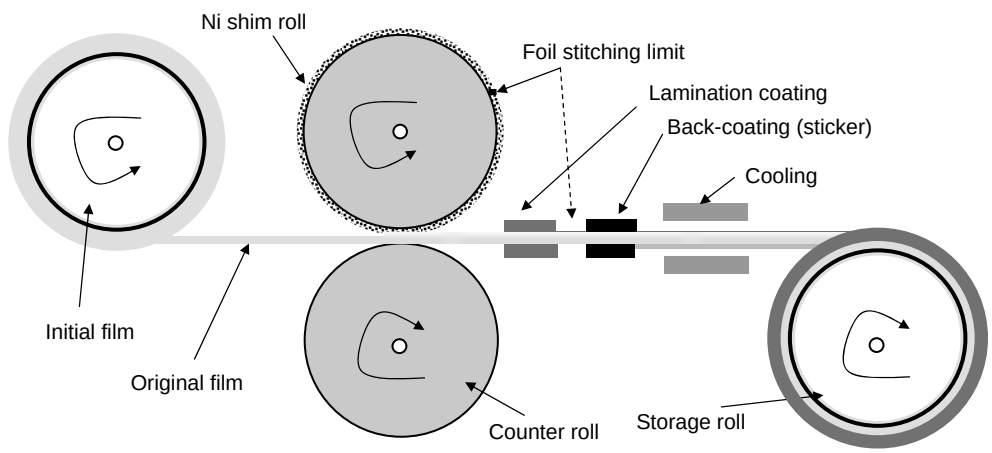
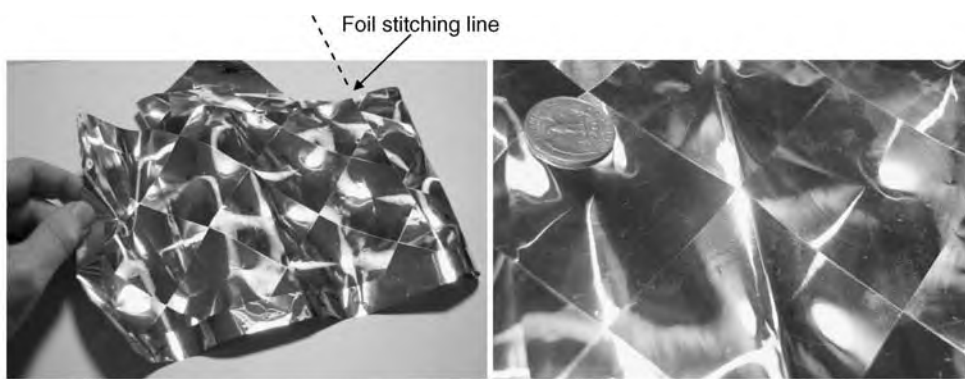


**Figure 14.14** The roll embossing process

Figure 14.15 shows an example of reflective Mylar™ roll embossing, of a square-aperture Fresnel lens array that has been fabricated by diamond turning. The stitching line is clearly visible in the left-hand photograph.

Figure 14.16 shows another photograph of a transparent plastic embossed diffractive lens array (the imaging qualities are shown). The images produced in Figure 14.16 are from a US quarter coin. The on-axis imaging quality is quite good, but the off-axis imaging is worse, due to the fact that these lenses are simple spherical lenses on a flexible film.

Figure 14.17 shows some surface topology of roll-embossed grating arrays (confocal white light interferometric microscope photographs). It is very difficult to perform surface topology on roll-embossed plastics, since such sheets are not planar (they are flexible and wavy), and this gives rise



Mylar™ roll-embossed square-shaped Fresnel lenses      Close view of roll-embossed elements

**Figure 14.15** An example of a roll-embossed array of Fresnel lenses in Mylar™



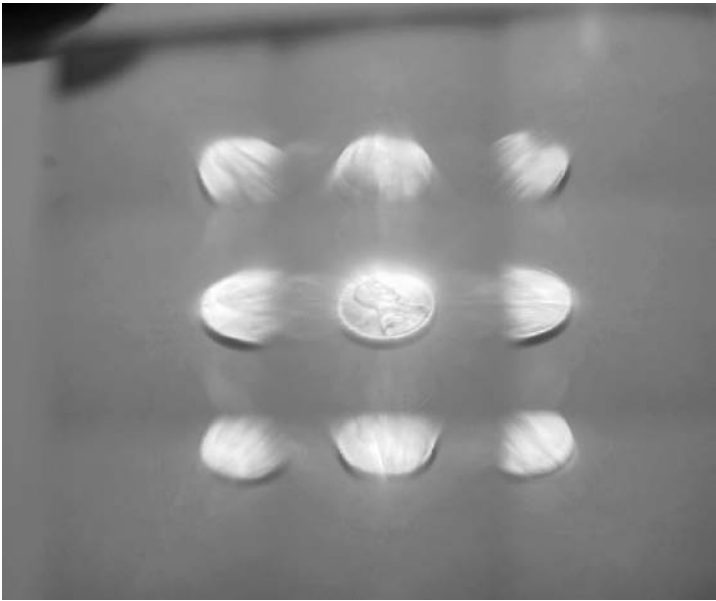


**Figure 14.16** An example of replication of a transmission diffractive lens array by roll embossing

to the waviness and surface wobbles seen in the plots in Figure 14.17. However, one can scan micro-profiles such as gratings by correcting for the wobbles using software (see the linear grating scan in Figure 14.17).

Materials for roll embossing can be Mylar™, PET, BOPP or polycarbonate, and the film thickness are usually between around 25 μm and 100 μm. The thickness of the roll can be as large as 1 m or more.

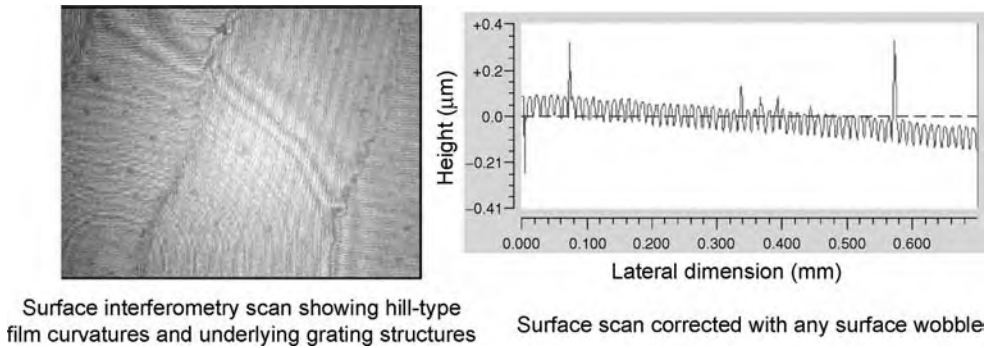


**Figure 14.17** The surface profilometry of roll-embossed grating arrays

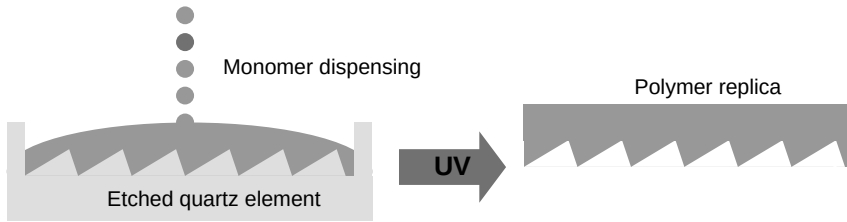


**Figure 14.18** The UV casting process

### 14.3.3 UV Embossing

As an alternative to hot embossing and roll-to-roll embossing, one can also use ultraviolet (UV) embossing, where UV light exposure is used to polymerize a liquid polymer. Ultraviolet embossing requires one side of the mold structure to be UV-transparent [9].

UV embossing is very similar to hot embossing. However, instead of heating the polymer beyond the glass transition temperature, the monomer is liquid. The hardening in hot embossing is performed by cooling, and in UV embossing by polymerization through UV exposure. However, hot or UV embossing can produce parasitic shrinkage problems.

## 14.4 The UV Casting Process

UV casting does not require any embossing or pressure [9]. A liquid monomer solution is poured on top of a micro-surface, and cured by UV exposure or simple air exposure curing (see Figure 14.18).

Here, the initial surface does not need to be a nickel shim, and can be any etched surface. However, the initial surface cannot be a photoresist, since the monomer would eat away the photoresist. A good positive master for the UV casting process is an etched quartz element. It is transparent to UV and is solid enough to be used over and over.

This might be the least expensive process to replicate micro-surface-relief elements such as diffractive gratings and micro-refractive lenses. Such UV curing liquids can be simply polymer liquids that are sold on the market for lens cleaning. They do not require any special UV exposure – 5 minutes in air is sufficient to produce the replica, which can be peeled off using a tape applied to the quartz wafer surface.

The problem with this technique is that when the polymer is peeled off, the mechanical properties of the film limit its usage in industrial applications. The films are very fragile, and cannot be used without a proper substrate (glass or plastic). However, it is a very interesting table-top experiment that anyone can do for a couple of dollars with any diffractive optical element (even with laser pointer diffractives replicated in plastic).

## 14.5 Injection Molding Techniques

Injection molding is one of the most popular techniques for the replication of macroscopic, mini- and microscopic optical elements in transparent plastics. All scales can be produced in one replication process and then mixed to generate final optical sub-assemblies [10,11].

There are two main types of plastic injection molding techniques available today:

- conventional injection molding; and
- compact disk (CD) type injection molding.

In the injection molding process, a thermoplastic polymer is injected at high pressure into a mold. There are a variety of optical plastics that can be used for plastic injection molding. Table 14.1 summarizes some of them.

**Table 14.1** Optical plastics used in injection molding processes (www.gsoptics.com)

| Properties                                           | Acrylic<br>(PMMA)                                                        | Polycarbonate<br>(PC)                     | Polystyrene<br>(PS)    | Cyclic olefin<br>copolymer                                          | Cyclic olefin<br>polymer                                         | Ultem 1010<br>(PEI)                                                |
|------------------------------------------------------|--------------------------------------------------------------------------|-------------------------------------------|------------------------|---------------------------------------------------------------------|------------------------------------------------------------------|--------------------------------------------------------------------|
| Refractive index                                     |                                                                          |                                           |                        |                                                                     |                                                                  |                                                                    |
| $N_F$ (486.1 nm)                                     | 1.497                                                                    | 1.599                                     | 1.604                  | 1.540                                                               | 1.537                                                            | 1.689                                                              |
| $N_D$ (589.3 nm)                                     | 1.491                                                                    | 1.585                                     | 1.590                  | 1.530                                                               | 1.530                                                            | 1.682                                                              |
| $N_C$ (656.3 nm)                                     | 1.489                                                                    | 1.579                                     | 1.584                  | 1.526                                                               | 1.527                                                            | 1.653                                                              |
| Abbe value                                           | 57.2                                                                     | 34.0                                      | 30.8                   | 58.0                                                                | 55.8                                                             | 18.94                                                              |
| Transmission % visible spectrum 3.174 nm thickness   | 92                                                                       | 85–91                                     | 87–92                  | 92                                                                  | 92                                                               | 36–82                                                              |
| Deflection temperature                               |                                                                          |                                           |                        |                                                                     |                                                                  |                                                                    |
| 3.6 °F/min @ 66 psi                                  | 214 °F/101 °C                                                            | 295 °F/146 °C                             | 230 °F/110 °C          | 266 °F/130 °C                                                       | 266 °F/130 °C                                                    | 410 °F/210 °C                                                      |
| 3.6 °F/min @ 264 psi                                 | 198 °F/92 °C                                                             | 288 °F/142 °C                             | 180 °F/82 °C           | 253 °F/123 °C                                                       | 263 °F/123 °C                                                    | 394 °F/201 °C                                                      |
| Max. continuous                                      | 198 °F                                                                   | 255 °F                                    | 180 °F                 | 266 °F                                                              | 266 °F                                                           | 338 °F                                                             |
| Service temperature                                  | 92 °C                                                                    | 124 °C                                    | 82 °C                  | 130 °C                                                              | 130 °C                                                           | 170 °C                                                             |
| Water absorption % (in water, 73 °F for 24 h.        | 0.3                                                                      | 0.15                                      | 0.2                    | <0.01                                                               | <0.01                                                            | 0.25                                                               |
| Specific Gravity                                     | 1.19                                                                     | 1.20                                      | 1.06                   | 1.03                                                                | 1.01                                                             | 1.27                                                               |
| Hardness                                             | M97                                                                      | M70                                       | M90                    | M89                                                                 | M89                                                              | M109                                                               |
| Haze (%)                                             | 1 to 2                                                                   | 1 to 2                                    | 2 to 3                 | 1 to 2                                                              | 1 to 2                                                           | —                                                                  |
| Coeff of linear expansion (cm × 10 <sup>−5</sup> °C) | 6.74                                                                     | 6.6–7.0                                   | 6.0–8.0                | 6.0–7.0                                                             | 6.0–7.0                                                          | 4.7–5.6                                                            |
| $dN/dT$ × 10 <sup>−5</sup> °C                        | −8.5                                                                     | −11.8 to −14.3                            | −12.0                  | −10.1                                                               | −8.0                                                             | —                                                                  |
| Impact strength (ft-lb/in) (Izod notoh)              | 0.3–0.5                                                                  | 12–17                                     | 0.35                   | 0.5                                                                 | 0.5                                                              | 0.60                                                               |
| Key advantages                                       | Scratch resistance<br>Chemical resistance<br>High Abbe<br>Low dispersion | Impact strength<br>Temperature resistance | Clarity<br>Lowest cost | High moisture barrier<br>High modulus<br>Good electrical properties | Low birefringence<br>Chemical resistance<br>Completely amorphous | Impact resistance<br>Thermal and chemical resistance<br>High index |

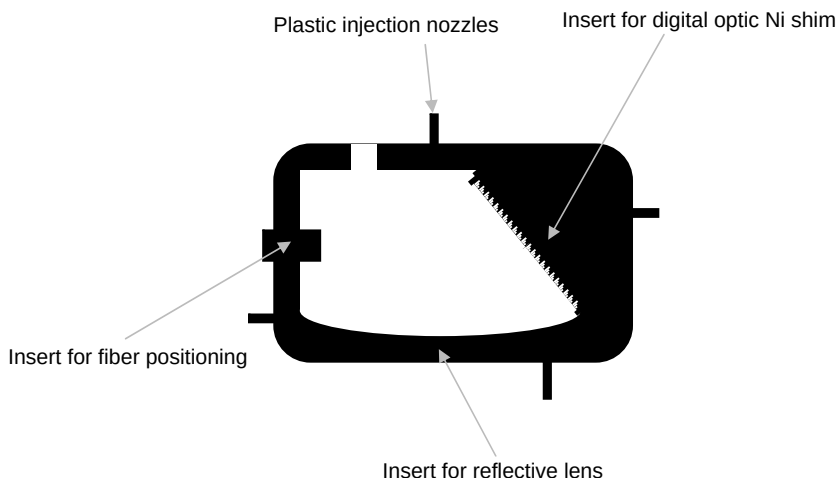


**Figure 14.19** Multi-cavity plastic injection molding, including digital optics Ni shim inserts

Optical plastics can be chosen for their refractive index, their transparency, their melting point, their birefringence, their linear expansion coefficient, or simply their availability and price.

### 14.5.1 Conventional Molding

Conventional plastic molding machines can be used to replicate optical plastic systems with inserts that are produced by Ni shim electroforming [12–15]. The resist of the multiple cavity mold can be generated by conventional mold shaping (see Figure 14.19).

The various inserts are linked to critical optical surfaces, such as refractives, reflectives, diffractives and so on [16]. The system depicted in Figure 14.19 is a spectroscopic device based on a linear blazed grating and a reflective lens. Another very popular example is the optical subsystem of an optical mouse, which comprises a couple of lenses, prisms and so on (see Figure 14.20).

An optical mouse is a very precise optical sensor, relying on speckle movement analysis on a  $32 \times 32$  CMOS sensor (with a red or IR laser) and very fast analysis electronics. The optical sub-assembly consists of a prism and two lenses. The production price point is about \$1.5 today: considering the price of the mechanics, the laser source, and the PCB, DSP and CMOS sensors, the optical sub-assembly must only cost a few cents.

Figure 14.21 shows the replication in plastic of planar refractive lens arrays, which are used as fly's eye arrays (see Chapter 4).

Figure 14.22 shows mini-refractive elements replicated by plastic injection molding (used in LED collimation, micro-optical sensors and disposable biomedical devices). Note that the quality of the surface (insert) is much smoother for the actual optical element than for the assembly and package.

### 14.5.2 CD Injection Molding

Compact disk injection molding was developed for the audio CD industry at the end of the 1970s, to replace vinyl disks. In the early 1980s, the replication techniques routinely produced replicas with sub-micron structures that offered unprecedented quality for the growing audio and video industry, both for CD media and for OPU objective lenses and gratings. Today, CD injection molding is capable of meeting

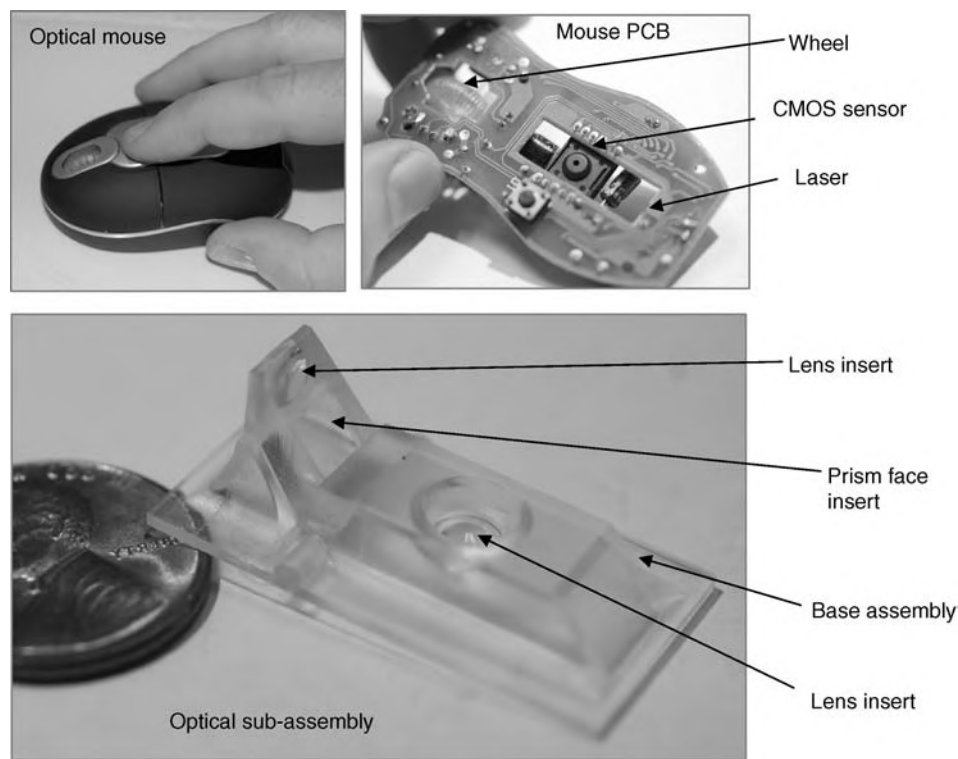


**Figure 14.20** An injection molded optical mouse sub-assembly

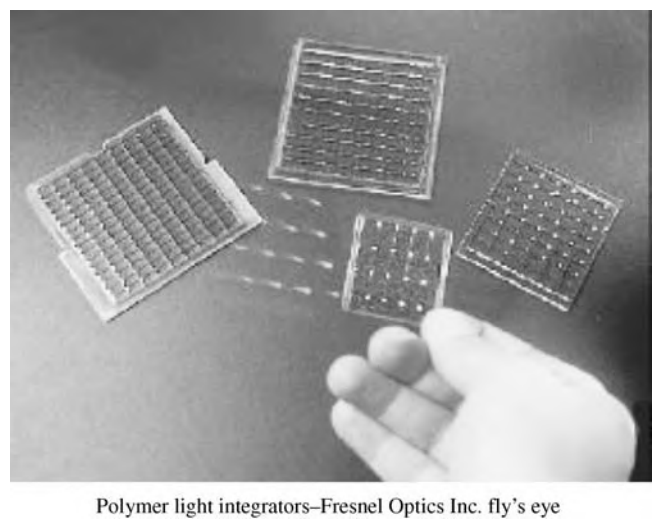


**Figure 14.21** The replication of planar fly's eye lenslet arrays by the injection molding process



**Figure 14.22** The replication of mini- and micro-optics by plastic injection molding

the specifications needed for DVD and Blu Ray, which require replication of 300 nm structures over 120 mm disks with multiple layers, each being 100  $\mu\text{m}$  thick. However, CD and DVD replication is not perfect, and any imperfection over the disk must be compensated by a correction algorithm.

CD injection molding is a hybrid injection molding/embossing process, performed in highly automated machines that can produce several hundred disks per minute (see Figure 14.23).

In a standard CD replication process, there are numerous steps that are not involved in replication of a diffractive element, such as the application of the reflective coating, the application of the lacquer, the lamination process, the protective coating process and so on.

The replication of digital diffractive structures using CD injection molding has been used for more than a decade, and has proved to be an efficient way to replicate sub-micron optical structures for either transmission or reflection mode (Figure 14.24) at a very low cost (around \$20 per disk).



**Figure 14.23** The CD injection molding replication process

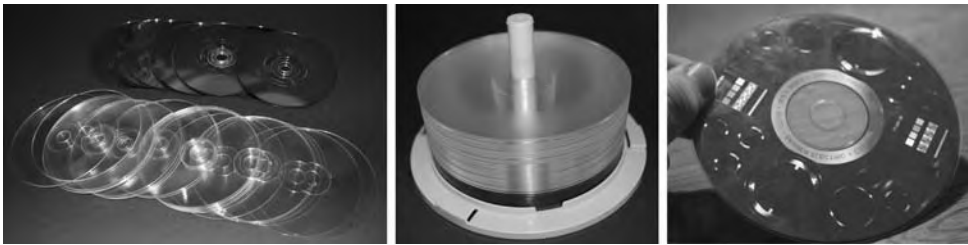


**Figure 14.24** Sets of reflective and transmission diffractive elements replicated in 120 mm CD disks

Many disk formats are available (see Figure 14.25). The conventional CD diameter is 120 mm, but there are also the mini-CD format (80 mm), the business card format and the nano-CD format (32 mm).

Two thicknesses are available (1.2 mm for the CD type and 600  $\mu\text{m}$  for the DVD type). Figure 14.26 shows the replication of reflective and transmission nano-CDs (32 mm format) with diffractive optical



**Figure 14.25** Different formats of CD replication of digital diffractive optics

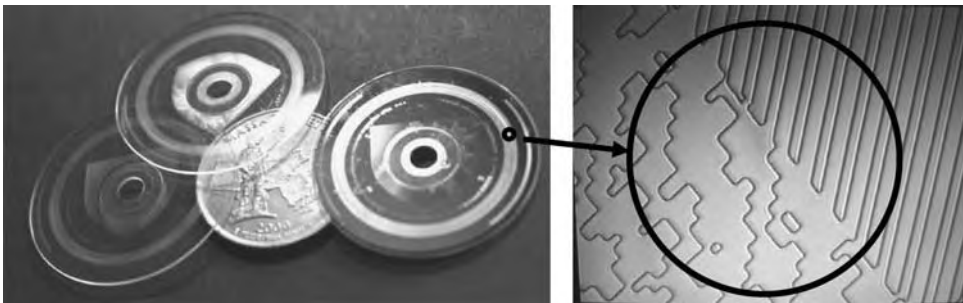


**Figure 14.26** Nano-CD injection molding and a microscope photograph of the replicated microstructures

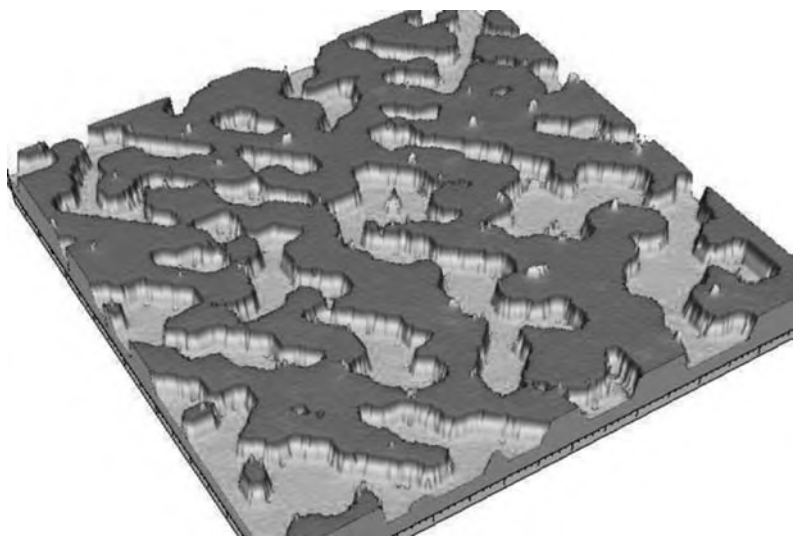


**Figure 14.27** The micro-topography of a CGH replicated by CD injection molding in polycarbonate encoders, and a microscope photograph of the microstructures on the diffractive encoder channel. Typical materials for CD injection molding are PMMA or polycarbonate.

The topographic scan in Figure 14.27 shows a CD injection molded replica of a CGH. The side walls are very straight and there do not seem to be any adverse effects of un-molding, as typically observed for hot embossing (see the previous section).

AFM measurement techniques can be used to assess the quality of the replica. AFM images of the CD replicated structures are presented in Figure 14.28, and show a high surface quality, free of roughness.

In order to estimate the fidelity of replication by CD injection molding, we have taken photographs of the original mold and the injected plastic structures (see Figure 14.29). The mold has been grown by nickel

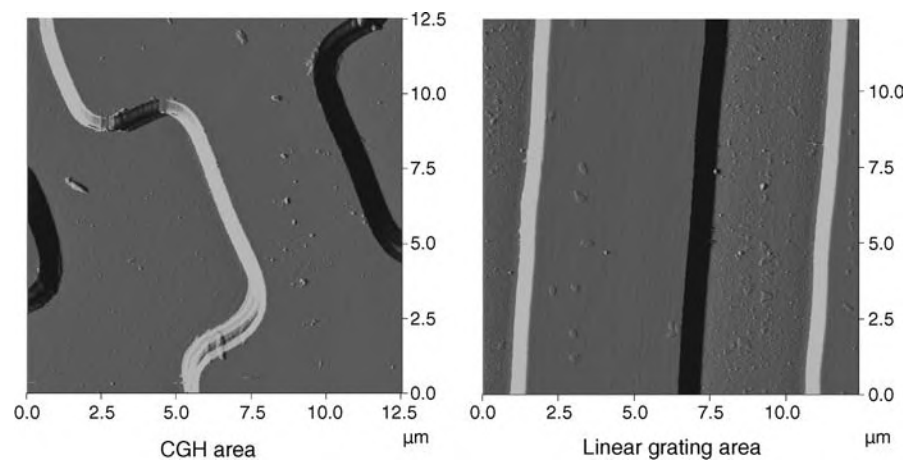


**Figure 14.28** AFM plots of a CD injection molded CGH and grating



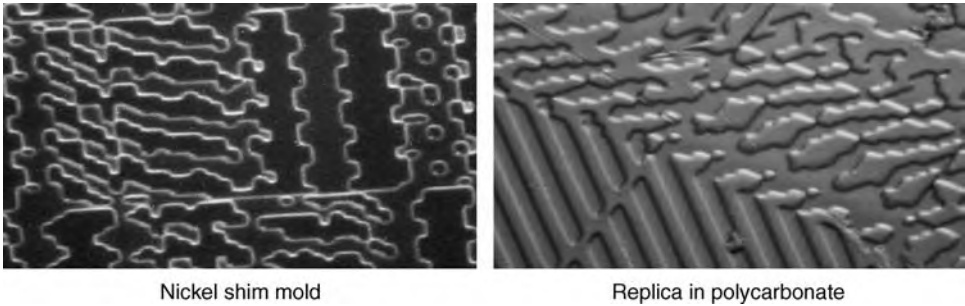


**Figure 14.29** A comparison between a nickel mold and structures replicated in plastic by CD injection molding

electroforming, and has been recombined on a thicker plate to be inserted in the CD injection mold. Figure 14.29 shows that the mold and the replica are almost identical, and that very little deformation has occurred during and after molding.

CD injection molding thus seems to be a high quality and cheap replication technology for digital diffractive optical elements. However, the main disadvantages are the following:

- the size is limited to 120 mm disks (with a 110 mm active area);
- the technique is limited to flat substrates; and
- the central hole is usually an inconvenient feature – however, it can be a desirable feature for some applications where the diffractive has to turn around a shaft (such as in a diffractive optical encoder).

14.6 The Sol-Gel Process

The sol-gel process is a desirable replication process since, unlike the previous plastic replication process [17–19], this process can yield high-quality, high-purity elements in glass-like structures (i.e. silica-loaded replicas) (see Figure 14.30). Therefore, refractive optics are best suited for sol-gel replication, but analog phase-relief diffractives can also be produced by this method. Silica optical elements that are produced and replicated by this method have low thermal expansion, good mechanical

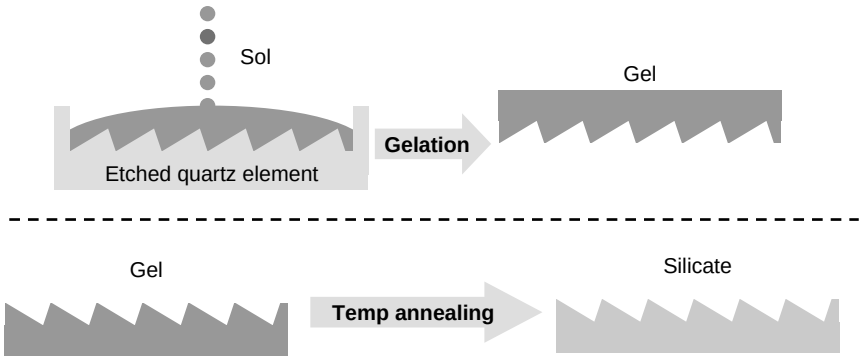


**Figure 14.30** The sol-gel process

strength, hardness and chemical durability, and high transmission over the spectral range from 0.2  $\mu\text{m}$  to 3.2  $\mu\text{m}$ .

The sol-gel process can be divided into three parts: sol and gel formation, drying and densification. The chemical ingredients are mixed to produce a sol with a viscosity only slightly greater than that of water. The sol is poured into a mold. Hydrolysis and poly-condensation of the sol form the three-dimensional network of the future glass, and the solution stiffens into a 'wet gel'. The gel is then left in the mold to age. After drying, the gel is heat treated to change the porous solid into a dense homogeneous glass. During aging, drying and densification, the gel shrinks by a factor that can be as large as 50%, as the silica matrix builds up. Since this shrinkage is controlled in all three dimensions, the shrinkage has an additional advantage in the way that the master prototype can be fabricated at a greater scale. In addition, imperfections of the mold surface decrease in size during aging, improving the surface finish of the sol-gel optics over that of the mold.

A widely used type of silica base material (precursor) is tetraethylorthosilicate ( $\text{Si}(\text{OC}_2\text{H}_5)_4$ ). DI water is added to the material to enter the 'sol' phase in a 'hydrolysis' reaction.

The sol-gel method can also be used directly in a photolithographic step and repeat process, on top of a wafer coated with resist. Such a technique has been used for the production of waveguide structures and diffractive optics.

However, the inevitable shrinkage issues have slowed down the introduction of sol-gel as a micro-structure replication technology in industrial sectors, even in the previously discussed wafer-based process.

## 14.7 The Nano-replication Process

Nano-imprint and soft lithography are two replication processes that have been developed to alleviate the costly problem of having to reduce the  $k_1$  factor in optical projection lithography by decreasing the wavelength, increasing the NA or using costly masks and RET techniques such as OPC.

Four techniques are presented below:

- step and scan nano-imprint lithography;
- thermoplastic nano-imprint;
- roll nano-imprint; and
- soft lithography.

In Nano-Imprint Lithography (NIL) and soft lithography (ink-based), there is no complex UV laser source involved (only flood exposure), no complex stepper projection lens, no developer, no complex resists, no standing waves produced, no aerial image and so on. And yet NIL can produce ultra high resolution lines in resist (the resist does not even need to be high-resolution resist, with a very good edge roughness). Finally, and perhaps most importantly, there are no successive alignment errors. If the initial mold is multilevel (etched fused silica for NIL or a PDMS stamp for soft lithography), it behaves like a gray-scale mask and performs the multilevel lithographic transfer without any etch depth or lateral field misalignment errors.

### 14.7.1 Nano-imprint Processes

#### 14.7.1.1 Nano Imprint Lithography (NIL)

The nano-imprint stamp can be made out of nickel (as for the previous molds presented) or using an etched fused silica wafer. The mold wafer is put into contact alignment with the final resist-coated wafer in a slightly transformed mask aligner (e.g. a Karl Suss MA 6).

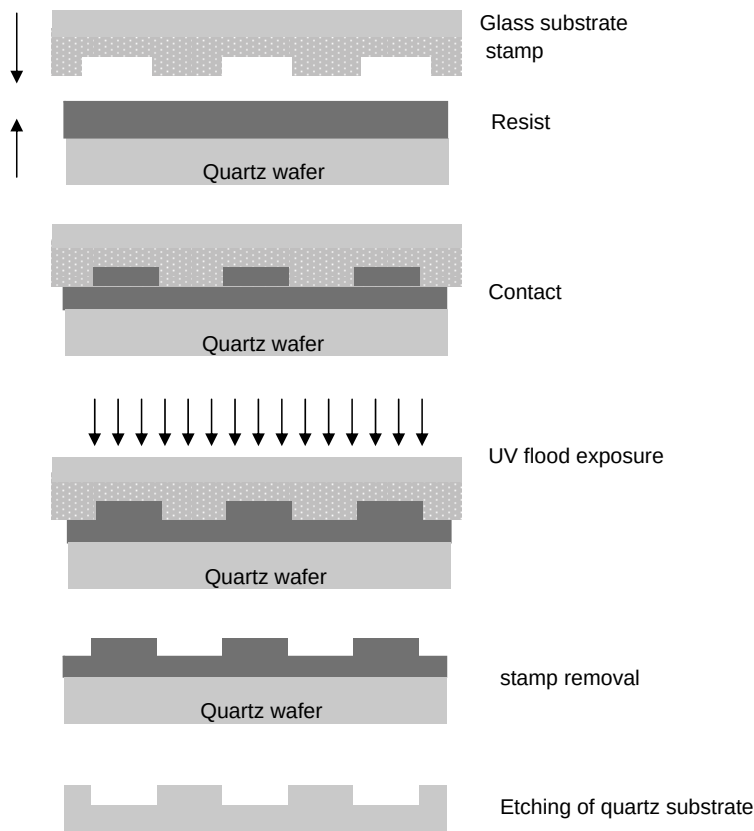


**Figure 14.31** The step and scan nano-imprint lithography process

In step and scan NIL (SNIL; see Figure 14.31), the resist is not exposed as in traditional lithography [20]. Instead, the resist (or other) layer is deformed by the contact of the stamp, and then flood exposed by a UV lamp to polymerize the entire resist layer.

Such nano-imprint techniques are used to replicate photonic crystal structures on large wafer areas.

**14.7.1.2 Thermoplastic Nano-imprint**

In thermoplastic NIL, the stamp is heated before contacting the resist-coated wafer. The resist-coated wafer is then deformed by pressure and heat, and is subsequently etched (temperature curing rather than UV curing).

**14.7.1.3 Roll Nano-imprint**

In roll nano-imprint technology, the stamp is used as a nickel shim in roll-to-roll embossing (see the previous section). There are two main ways to implement roll nano-imprint (see Figure 14.32):

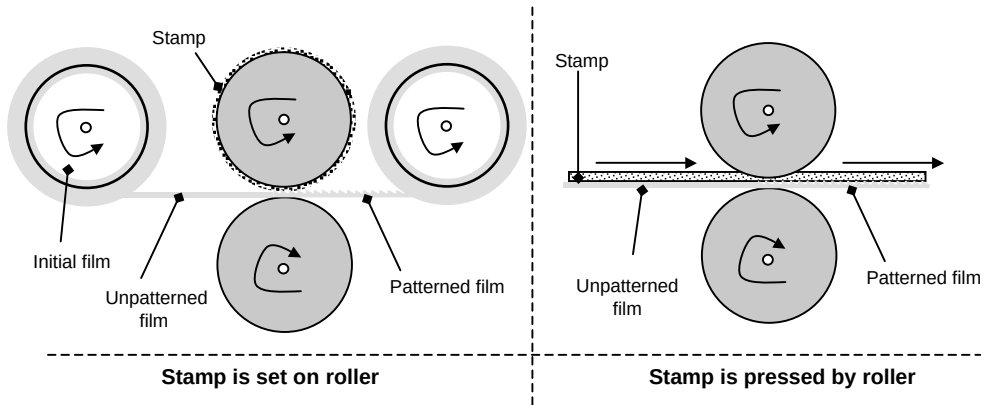


**Figure 14.32** Roll nano-imprint techniques

- One technique is to use a cylindrical PDMS mold. This is done by imprinting onto the PDMS and then bending it into a cylindrical shape, mounting it around a roller and rolling the roller on the resist-coated substrate.
- The other technique is to imprint over a flat mold by putting the mold directly on the substrate, and rotating the roller on top of the mold.

Such nano-imprint roll embossing can be an efficient tool to implement the replication of sub-wavelength slanted structures.

### 14.7.2 The Soft Lithography Process

Soft lithography is a lithographic operation that is very close to the very first type of lithography operated by Johannes Gutenberg in the 15th century, namely by the use of an ink stamp.

Here, the term ‘soft’ relates to the use of elastomers or organic materials for stamps or molds. It includes, among others, transferring lithographically created patterns in polydimethylsiloxane (PDMS), and low-cost elastomeric vacuum casting with silicone molds and rubber stamping, which can structure even highly curved surfaces.

Polydimethylsiloxane (PDMS) is a common stamp material. PDMS is a stiff elastomer that is nontoxic and readily available. Once the PDMS is coated on a resist pattern, a UV curing process stiffens the layer. The UV-cured PDMS stamp is then lifted off the original silicon wafer substrate. The lift-off process can be improved by dissolution of the underlying resist. The stamp is then coated with an alkanethiolate ink solution, and used as a stamp on a wafer coated with either Ag or Au. The ink stays on the wafer after stamp removal, and the underlying mask can then be wet etched. Further processing can be done by, for example, dry etching the underlying quartz substrate for the generation of sub-wavelength binary gratings.

As there is no imaging involved, there is no need to use fancy RET techniques or even any stepper. However, in order to implement the soft lithography, a mask aligner is often used in order to provide the contact alignment as well as the UV exposure for the PDMS curing.

The soft lithography process is depicted in Figure 14.33.

Such a PDMS film can thus be used to produce lithography on a flat or curved substrate, in order to replicate, for example, digital diffractive surfaces on a refractive surface to produce high-quality hybrid

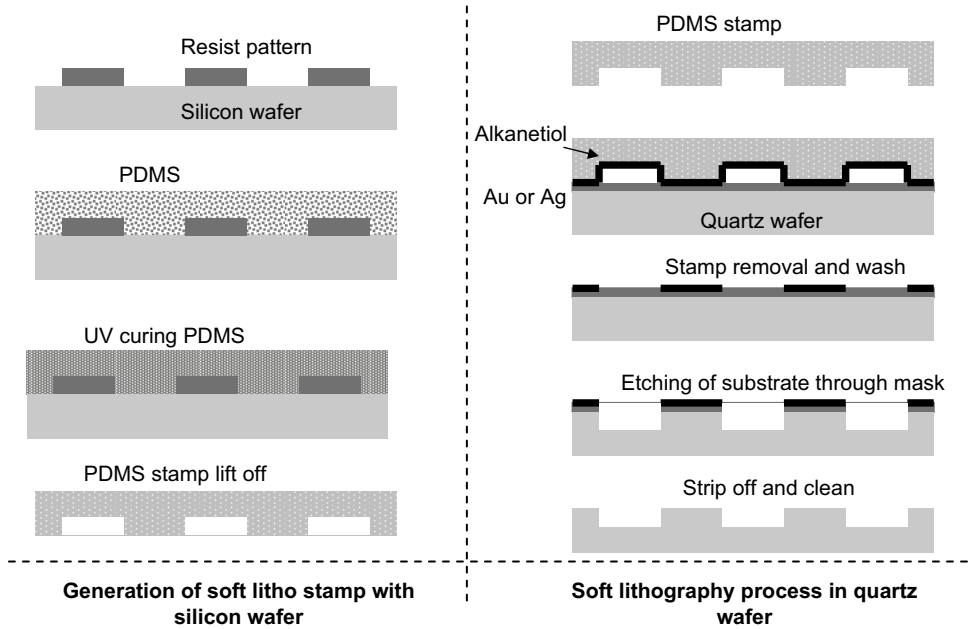


Figure 14.33 The soft lithography process

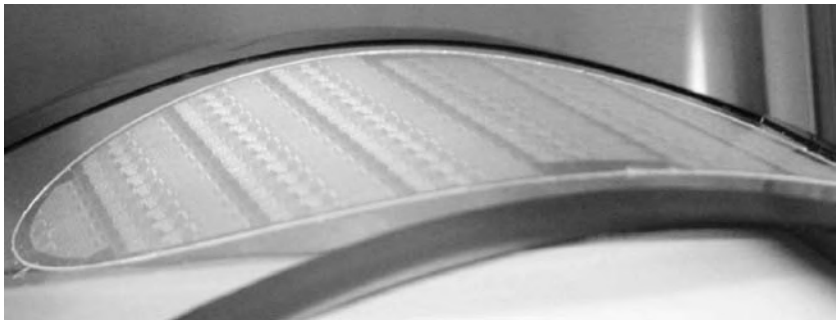


Figure 14.34 A 900 μm thick 6 inch PDMS soft lithography stamp from Philips

optical elements. Figure 14.34 shows such a flexible PDMS stamp (a 900 μm thick stamp, with a 6 inch diameter).

14.8 A Summary of Replication Technologies

Table 14.2 summarizes the main digital optics replication technologies used in industry today.

**Table 14.2** A summary of the major diffractive optics replication technologies used in industry today

| Replication technology           | Element type                          | Typical lot size      | Price per replica     | Resulting substrate                                    | Advantages                                                  | Limitations                                               |
|----------------------------------|---------------------------------------|-----------------------|-----------------------|--------------------------------------------------------|-------------------------------------------------------------|-----------------------------------------------------------|
| Conventional optical lithography | Binary/multilevel surface relief      | 25 wafer cassette lot | High (\$0.1K to \$1K) | Any wafer (quartz, fused silica, glass, Silicon, etc.) | Very high precision via optical lithography and dry etching | Pricing, lot size, dicing                                 |
| Gray-scale lithography           | Analog surface relief                 | Individual wafers     | Very costly (>\$1K)   | Selected wafers                                        | Very high efficiency                                        | Costs and availability                                    |
| CD/DVD injection molding         | Thin surface-relief profiles          | >1000s to Millions    | Cheap < \$1           | Polycarbonate                                          | Rigid substrate, proven technology                          | Need for specialized equipment and nickel shim generation |
| Planar hot embossing             | Thick surface-relief profiles         | 1000s                 | Cheap \$1–\$10        | Any plastic                                            | Rigid substrate, no need for specialized equipment          | Nickel shim generation needed                             |
| Roll embossing                   | Thin surface-relief profiles          | Millions              | Very cheap < \$0.1    | Mylar, thin plastics, etc.                             | Cheap, proven technology                                    | Substrate not rigid, nickel shim generation               |
| UV casting                       | Thin to thick surface-relief profiles | 1000s                 | Cheap \$0.1–\$10      | Polymer                                                | No need for special equipment                               | No rigid substrate, shrinkage, tear, wear                 |
| Sol-gel                          | Thick surface-relief profiles         | 100s                  | Average \$10–\$100    | Glass                                                  | Resulting element is silica based (glass)                   | Shrinkage and geometrical control                         |
| Step and flash NIL               | Thin surface-relief profiles          | Individual wafers     | High                  | Resist on Silicon/glass                                | Replica on curved surfaces                                  | Replica is in resist, needs to be etched                  |
| Roller NIL                       | Thin surface-relief profiles          | 1000s                 | Medium to cheap       | Resist on film of substrate                            | Replicas on rolls                                           | Replica remains in resist, no etching possible            |

Such replication technologies, which in most cases have been developed and optimized for other technologies, such as CD and DVD disk replication, or nano-imprint or soft IC lithography, provide an invaluable asset to diffractive optics to allow it to finally come out of the academic/research arena, and to propose concrete industrial solutions to real market needs in many consumer electronics, factory automation and telecom applications, as we will see in Chapter 16.

## References

- [1] A. Heuberger, 'X-ray lithography', *Solid State Technology*, **28**, 1986, 93–101.
- [2] E. di Fabrizio, M. Gentili, L. Grella *et al.*, 'High-performance multilevel blazed X-ray microscopy Fresnel zone plates fabricated using X-ray lithography', Presented at the 38th International Symposium on Electron, Ion and Photon Beams, New Orleans, May 31 – June 3, 1994.
- [3] K.-H. Brenner, M. Kufner, S. Kufner *et al.*, 'Application of three-dimensional micro-optical components formed by lithography, electroforming and plastic molding', *Applied Optics*, **32**(32), 6464–6469.
- [4] F.P. Schwartzman, 'SURPHEX: new dry photopolymers for replication of surface relief diffractive optics', in '*Holographics International*'92', Y.N. Denisyuk and F. Wyrowski (eds), SPIE Vol. **1732**, 121–130, (1993).
- [5] F.P. Schwartzman, 'Holographic optical elements by dry photopolymer embossing', in '*Practical Holography*', V. Stephen and A. Benton (eds), Proc. SPIE Vol. **1461**, 1991, 313–320.
- [6] W.C. Sweatt, M.R. Descour, A.K. Ray-Chaudhuri *et al.*, 'Mass-producible microholographic tags', in '*Diffractive and Holographic Optics Technology III*', I. Cindrich and S.H. Lee (eds), Proc. SPIE No. **2689**, SPIE Press, Bellingham, WA, 1995, 170–174.
- [7] S. Traut, M. Rossi and H.P. Herzig, 'Replicated arrays of hybrid elements for application in a low-cost micro spectrometer array', *Journal of Modern Optics*, **47**(13), 2000, 2391–2397.
- [8] J. Anagnostis, S. Payette and D. Rowe, 'Replication of high fidelity surface relief structures', ASPE Spring Topical Meeting, 1999.
- [9] P. Dannberg, R. Bierbaum, L. Erdmann and A.H. Braeuer, 'Wafer scale integration of micro-optic and optoelectronic elements by polymer UV reaction moulding', in '*Optoelectronic Integrated Circuits and Packaging III*', M.R. Feldman, J.G. Grote and M.K. Hibbs-Brenner (eds), Proc. SPIE Vol. **3631**, 1999, 244–251.
- [10] F. Nikolajeff, S. Jacobsson, S. Hård *et al.*, 'Replication of continuous-relief diffractive optical elements by conventional compact disc injection-molding techniques', *Applied Optics*, **36**(20), 4655–4659.
- [11] M. Tanigami, S. Ogata, S. Oayama, T. Yamashita and K. Imanaka, 'Low-wavefront aberration and high temperature stability molded micro-Fresnel lenses', *IEEE Photonics Technology Letters*, **1**, 1989, 384–385.
- [12] M. Tanigami, S. Ogata, S. Aoyama, T. Yamashita and K. Imanaka, 'Low-wavefront aberration and high-temperature stability moulded micro Fresnel lens', *IEEE Photonics Technology Letters*, **1**, 1989, 384–385.
- [13] U. Kohler, A.E. Guber, W. Bier, M. Hekele and T. Schaller, 'Fabrication of microlenses by combining silicon technology, mechanical micromachining and plastic molding', *Proceedings of SPIE*, **2687**, 1996, 18–22.
- [14] Y.S. Liu, R.J. Wojnarowski, W.A. Hennessy *et al.*, 'Polymer optical interconnect technology (POINT): optoelectronic packaging and interconnect for board and backplane applications', in '*Optoelectronic Interconnects and Packaging*', R.T. Chen and P.S. Guilfoyle (eds), Proc. SPIE Vol. **CR62**, 1996, 405–414.
- [15] Y. Ishii, S. Koike, Y. Arai and Y. Ando, 'Ink-jet fabrication of polymer microlens for optical I/O chip packaging', *Japanese Journal of Applied Physics*, **39**, 2000, 1490–1493.
- [16] G. Kritchewsky and J. Schäfer, 'Plastic optics offer unique design freedom', *Laser Focus World*, 10/1997; [http://www.laserfocusworld.com/display\\_article/314432/12/none/none/Feat/MOLDED-OPTICS:-Molded-glass-aspheric-optics-hit-the-target-for-precision-and-cos](http://www.laserfocusworld.com/display_article/314432/12/none/none/Feat/MOLDED-OPTICS:-Molded-glass-aspheric-optics-hit-the-target-for-precision-and-cos).
- [17] W.V. Moreshead, J.-L.R. Nogues, R.L. Howell and B.F. Zhu, 'Replication of diffractive optics in silica glass', Proceedings Vol. **2689**, in '*Diffractive and Holographic Optics Technology III*', I. Cindrich and S.H. Lee (eds), Proc. SPIE Vol. **2689**, SPIE Press, Bellingham, WA, 1996, 142–152.
- [18] B.E. Bernacki, A.C. Miller, L.C. Maxey *et al.*, 'Hybrid optics for the visible produced by bulk casting of sol-gel glass using diamond turned molds', in '*Optical Manufacturing and Testing*', V.J. Doherty and H.P. Stahl (eds), SPIE Vol. **2536**, 1995, 463–474.

- 
- [19] S. Biehl, R. Danzebrink, P. Oliveira and M.A. Aegerter, 'Refractive microlens fabrication by ink-jet process', *Journal of Sol-Gel Science and Technology*, **13**, 1998, 177–182.
  - [20] S.Y. Chou, P.R. Krauss, W. Zhang *et al.*, 'Sub-10 nm imprint lithography and applications', *Journal of Vacuum Science and Technology B*, **15**, 1997, 2897–2904.
  - [21] P. Nussbaum, I. Philpoussis, A. Husser and H.P. Herzig, 'Simple technique for replication of micro-optical elements', *Optical Engineering*, **37**, 1998, 1804–1808.



# 15

## Specifying and Testing Digital Optics

The past 14 chapters have reviewed the various design, modeling and fabrication tools available to the optics engineer to produce digital optics. The next chapter will show practical examples of products, available on the market today, that include such digital optics.

As the number of applications requiring digital optics grows every day, especially in the consumer electronics sector, numerous industries are showing strong interest in such elements, and in the lithographic fabrication and replication of such elements, which are similar to processes in the IC industry. However, for many industries, digital optics remains a small part of their overall final product, even though it might be a key component. Other constituting elements of the final product might include standard macroscopic optics, digital and analog electronics, mechanics, fluidics, software and so on.

Therefore, it is very seldom and also very unlikely that an industry might invest in the various design and various fabrication tools for producing digital optics, especially when lithographic techniques are to be used for the production of such elements (even a small microlithographic cleanroom fab can be very costly).

Thus, more and more industries are turning fabless, or partially fabless, and specify digital optics for external design houses and foundries. This leads to the following questions:

1. How do you specify your digital optics fabrication processes?
2. How do you make sure the fabrication has been done correctly?
3. How do you make sure that what you specified is actually going to give the desired result in your final product?

This chapter attempts to answer these three questions.

### 15.1 Fabless Lithographic Fabrication Management

This first section summarizes what a fabless company needs to know about specifying the digital optics fabrication process and creating a Process Definition – Manufacturing Instructions (PDMI) document for the fabrication of such optics.

Emphasis will be put on conventional multilevel free-space digital diffractive optics, since this is the most widely available process. For more specific analog direct write techniques and gray-scale processes (see Chapter 12), the PDMI is usually custom tailored for each fab.

## 15.2 Specifying the Fabrication Process

As the microlithographic foundry is usually not dedicated solely to digital optics fabrication, but rather to standard IC fab (CMOS, Bi-CMOS, etc.), a PDMI has to be created to specify the process for the target technique to be used.

The various lithographic fabrication processes that are used in industry today have been described in the previous chapters. The focus will now be on the main five processes available to the optical engineer today, namely:

- diamond turning;
- binary lithography;
- multilevel lithography;
- gray-scale lithography; and
- resist reflow lithography.

Emphasis will be placed on the multilevel lithography technique, which accounts for most of the digital optics fabrication today in industry.

### 15.2.1 Specifying the Mask Patterning Process

When specifying a photomask patterning process (see Chapter 12), several specifications have to be sent to the mask shop, in the Mask Order Form (MOF), which is written in a particular lingo that is usually not comprehensible by an optical engineer. The standard machine-related lingo has already been defined in Chapter 12. The specifications that are required in the MOF are described below.

#### 15.2.1.1 The Fracture Grid

The fracture grid is the resolution of the writing grid during e-beam or laser beam exposure. Chapter 12 has shown that the digital optical element can be fractured into a high-level format, such as GDSII, CIF, DXF or OASIS, which is not dependent on the patterning machine to be used. The fracture processed by the mask shop is related to the specific mask patterning machine to be used, fracturing the GDSII data into low-level trapezoids or raster scans as illustrated in Chapter 12, and also – and most of all – to:

- the fidelity of the low-level fractured shapes to the high-level fractured shapes in GDSII; and
- the price of the mask patterning process.

Indeed, one can fracture in the digital optics EDA tool a digital optical element in GDSII format over a 1 nm grid, and send it to the mask shop. However, mask shops usually fracture regular masks over a 0.5 or 0.25  $\mu\text{m}$  grid, but can also, at additional cost, fracture the polygons over a 100 nm or even 50 nm or lower grid. As depicted in Figure 15.1, the mask patterning price usually grows exponentially with the fracture grid and the size of the GDSII file.

Figure 15.1 also shows also the price structure for gray-scale masks (such as HEBS) and PSM masks (very rarely used for digital optics). There is a plateau price up to a defined flat GDSII data file size, beyond which the price increases.

A flat GDSII file does not have any structured data hierarchy, and is basically equivalent to a standard image file. A GDSII file is a complex hierarchized data system that includes libraries, structures and arrays, as defined in Figure 15.2.

Although any GDSII shape has to be defined using a library and a structure, flat data is usually composed of a single library and a single structure. So, when designing a diffractive lens with a small phase error over a 1 nm grid (for phase errors during the GDSII fracture process, see Chapter 12), the file to

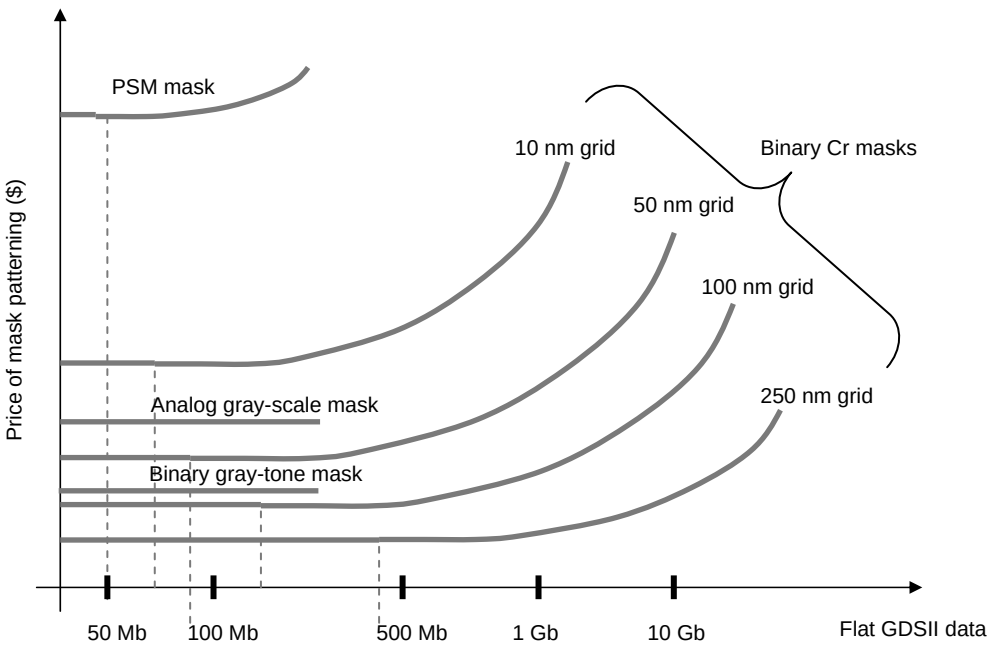


**Figure 15.1** Mask patterning pricing as a function of the fracture grid and data size

be patterned by the specific machine has to be re-fractured over a 500 nm, 250 nm, 100 nm, 50 nm or even a nonisotropic fracture grid (e.g. 200 nm in the  $x$  direction and 50 nm in the  $y$  direction), depending on the budget that one has for such fabrication. Figure 15.3 shows such fracture examples over a diffractive lens, which has been fractured in GDSII format, with a very tight phase error of  $\lambda/50$ .

As one can see in Figure 15.3, a cheap mask can be patterned quickly using a 500 nm grid for a lens fractured in GDSII over a  $\lambda/50$  maximum phase error, but the result is going to be very bad in terms of optical performance. The same  $\lambda/50$  fractured lens can be patterned over a 100 nm or 50 nm grid, and the result will be very good (but more expensive to pattern, of course).

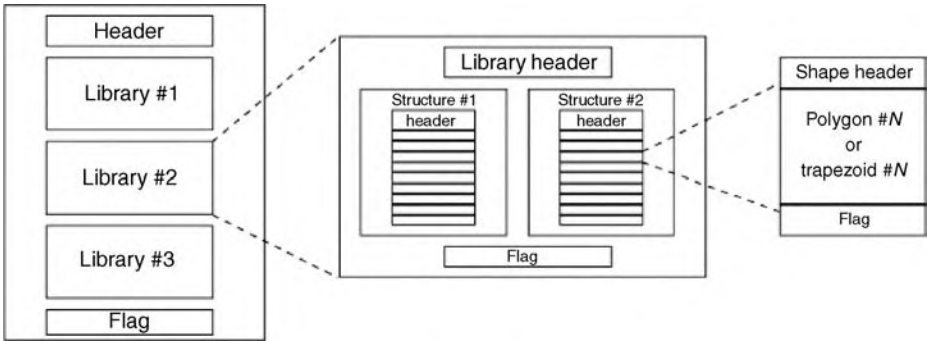


**Figure 15.2** The hierarchy of a GDSII file structure

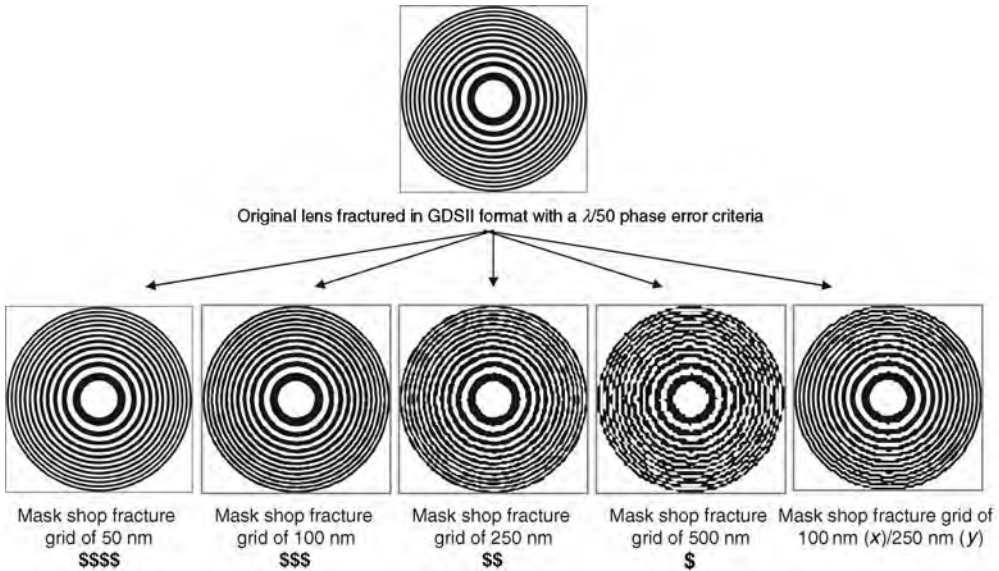


**Figure 15.3** Results of mask shop fracture grids over diffractive lens fractures with a phase error of  $\lambda/50$

There is very often a misunderstanding on the part of optical engineers and product development managers for optical devices who see a beautifully fractured lens on their digital optics CAD tool and a poor result on the photomask, and who think that a mask is just a mask, without analyzing the effects of low-level grid fracture accompanying cheap mask tickets.

### 15.2.1.2 The Die, Array and Field

The ‘die’ is the single element that is to be diced out after lithographic processing. It can be a single Fresnel lens, a single CGH or digital element, or a combination of such digital optical elements. The die is used in the final application; therefore, the dicing of the die is of great importance (see also wafer dicing and wafer scribing in Chapter 12). The array is the area made up of the rows and columns of dies on  $1\times$  masters,  $5\times$  reticles and other reduction reticles.

The ‘field’ is the area that can actually be stepped on a stepper (limited by the NA of the lithography lens) or the field of the mask aligner, which can be transferred to the wafer. A field size for a typical  $5\times$  stepper is about  $20\times 20$  mm, and for a mask aligner, it is about 85% of the wafer size: see Figure 15.4, which shows the dies composing the field on the reticle, the stepping scheme during lithography and the actual usable array on the final wafer.

### 15.2.1.3 The Critical Dimension (CD)

The CD is the smallest feature written on the mask. The CD is usually a part of the Process Control Monitor (PCM) or part of the device itself (e.g. the smallest fringe width in a Fresnel lens, or the width of a waveguide). The price of photomask patterning is closely related to the CD. The CD usually helps in the choice of the final fracture grid size (see Figures 15.1 and 15.2).

When using a  $5\times$  reticle, the resulting mask CD is five times larger than the CD on the wafer, and thus reduces the price of the mask patterning. A photomask used in a mask aligner is directly patterned at  $1\times$ , so such a mask can become pricey since the feature sizes are very small.

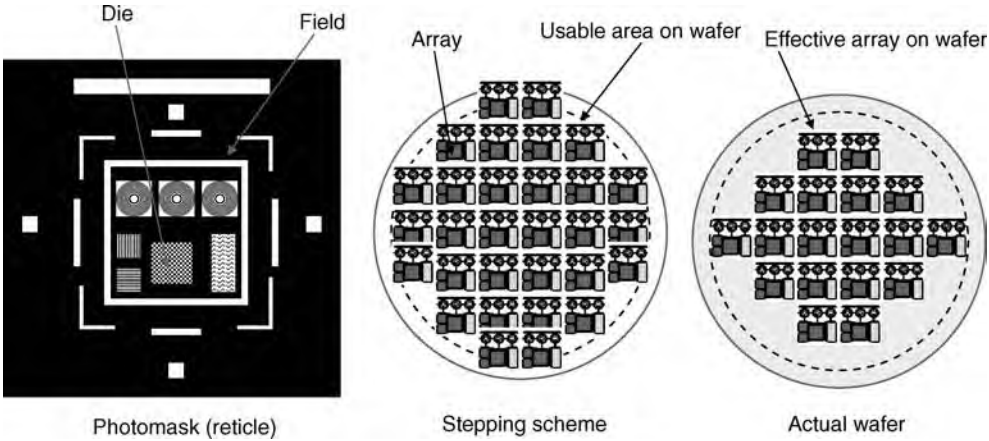


Figure 15.4 The die, array and field

15.2.1.4 The Field Tone

The field tone is the way the mask is physically written. The field tone is also referred to as the polarity of the mask. A dark tone would be produced by a positive photoresist during mask patterning, and a clear tone would be produced by a negative photoresist. Figure 15.5 shows the same mask written in a clear tone (left) and in a dark tone (right). Actually, the pattern on the mask in Figure 15.5 is a hybrid clear/dark tone (the central rectangle being of opposite tone than the rest).

In order to make sure that the tone is right for the client, mask shops double check by providing the tone of the field and the tone of the pattern (digitized patterns). Thus, a positive resist writing would be called ‘Dark Field Digitized Patterns Clear’, or a negative resist writing would be ‘Clear Field Digitized Patterns Dark’. It is always good to have the alignment marks in a clear tone, even if the rest of the mask is in a dark tone (see Figure 15.5), since it is good to see through the alignment marks to align two fields one on top of the other.

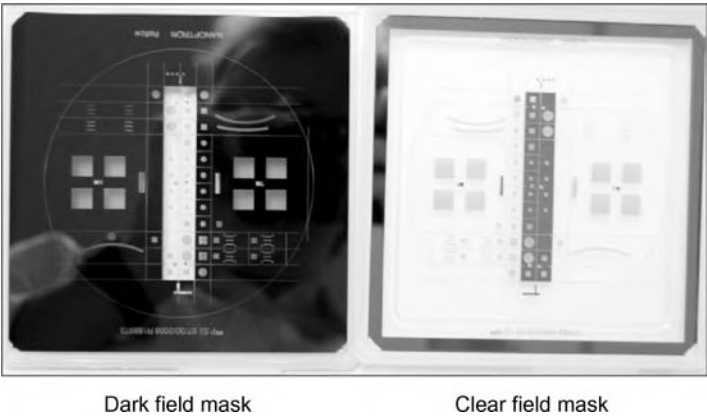


Figure 15.5 Clear tone and dark tone photomasks

### 15.2.1.5 Mask Parity

The parity of the mask is important, since it is the way the pattern is written in the Cr layer. Usually, the pattern needs to be mirror imaged in order to produce the right image on the wafer, since in contact, proximity and projection alignment, the mask pattern is mirror imaged. If the pattern is symmetric and does not contain any text, one need not consider mirror imaging. The parity specifications are given in terms of reading and Cr position, as follows:

- a '*Cr up right reading*' means that when one is looking at the Cr layer (on top), one can see the pattern as it should appear on the wafer; and
- a '*Cr up wrong reading*' means that the pattern has been mirror imaged.

Things also depend on whether the photomask is to be set Cr up on the lithography system or Cr down. Therefore, a '*Cr down right reading*' is equivalent to a '*Cr up wrong reading*'.

### 15.2.1.6 Defect Inspection

Defect inspection is very important when fabricating an IC mask. Therefore, it is also important to incorporate a pellicle to prevent dust imaging on the wafer (see Chapter 12). However, this is less important when fabricating a digital optics mask, especially an optical element for use in free space, such as a diffractive or refractive optical element.

Unlike in an IC element, while a device cannot work due to a transistor malfunction as an effect of a dust particle in the photomask (and thus a skewed image of the transistor gate on the wafer), a digital optical element will work perfectly well even with some dust on the mask and wafer, since the process is a parallel process (the light beam sees the whole element in parallel, and not in a serial way as in IC). Therefore, defect analysis does not need to be specified for a mask patterning that includes free-space digital optics, thus reducing the price tag a bit and speeding up delivery. This, however, is not true for a digital waveguide optical mask, where the defects have to be analyzed, since a small defect on a waveguide structure can disable the whole system.

### 15.2.1.7 Inserting the Process Control Monitors (PCMs)

Process control monitors are standard features that are inserted in the layout in order to be able to measure the quality of the fabrication at each process step in the fabrication flow. Usually, such PCMs can be located within the same area in the successive field alignment fiducials. Such PCMs are standard in the IC process (CMOS or Bi-CMOS, for example). PCMs that are optimized for the digital optics realm are described below.

In a digital optical element used as a diffractive element, one is mainly interested in one final specification: the resulting efficiency. The resulting diffraction efficiency is related to numerous fabrication parameters, the quality of which can be assessed by a set of customized PCMs, which are different from the PCMs used in the IC industry. Such PCMs need to include mainly four features:

- the quality of the lithography (resolution);
- the quality of the field-to-field alignment;
- the quality of the etch depth and the side walls (and uniformity over the wafer); and
- the quality of the surface (roughness etc.).

There are two ways to define a PCM for digital optics:

- the traditional way, by inserting verniers, resolution targets and so on, and using a microscope to read them after fabrication; or

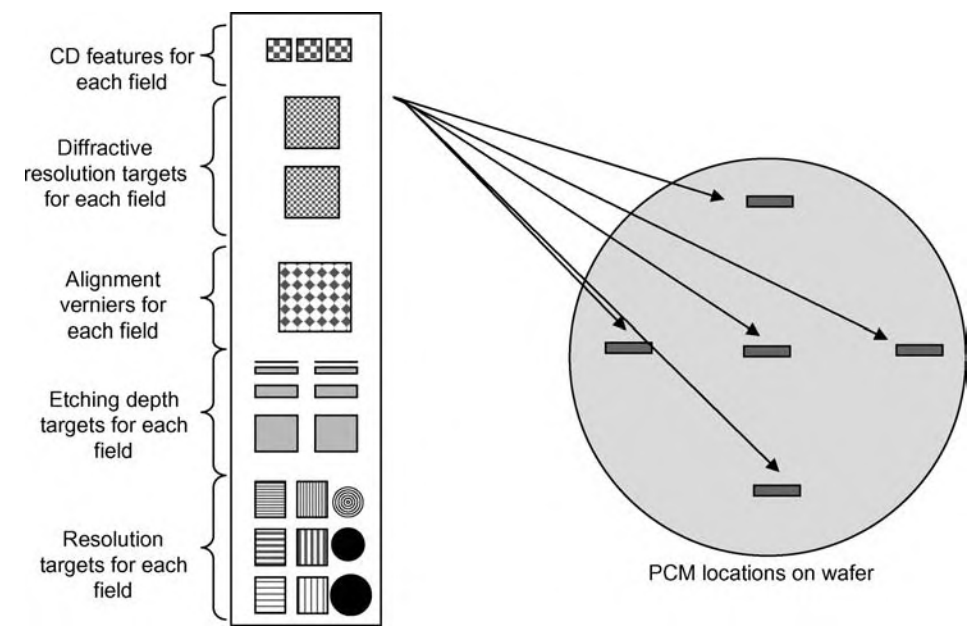


**Figure 15.6** An example of a conventional and diffractive PCM array

- the diffractive optical way, by simply using a laser beam and shooting it at the PCMs, and then visualizing and analyzing the far-field diffraction pattern.

Both sets of PCMs are introduced and analyzed in more detail in Section 15.3.2. Figure 16.6 shows such a PCM array.

The alignment verniers in the PCMs are not alignment marks; they are not used to align the fields. The alignment marks are used to align the fields. The alignment verniers are used to assess the quality of the alignment once the fields are aligned and etched.

Alignment marks can be introduced in PCMs. However, it is best to keep the alignment marks in separate fields, in order not to confuse the alignment algorithm (in the case where the alignment is automatic, as in a stepper) or the alignment technician (in the case where the alignment has to be done manually, as in some mask aligners).

The next section analyzes such standard PCMs on fabricated examples, and the last section describes how to use such custom diffractive PCMs, which are a much better fit for assessing the fabrication of digital optics, especially digital diffractive optics.

### 15.2.1.8 Creating the Job-deck File

Chapter 12 has reviewed the individual element fracture process and GDSII layout generation process, as well as the alignment marks, the PCMs and the various texts to be inserted. The number of individual GDSII files describing the various elements in a field and their relative positions can be high. The job-deck file is a file describing the various individual GDSII files, and how to pattern the photomask or reticle. The information in a job-deck file includes:

- the type of mask that is to be patterned ( $5\times$  reticle,  $1\times$  photomask etc.);
- the lithography machine that is to be used (in order to include standard alignment mark libraries linked to a particular lithography machine);





- the alignment features and PCMs;
- the PDMI instructions;
- the Mask Order Form (MOF); and
- the desired defect analysis.

Most mask shops use FTP servers to receive the data. However, as the amount of data for digital optics can be very large (several Gb of data), FTP is not recommended due to the prohibitive transfer time and the limited bit error rate associated with FTP servers. When the final data size is larger than 500 Mb, it is best to burn the data on a CD or DVD and send it out to the mask shop.

It is best to zip all the files into a single compressed file and document the contents in a separate file, which can be sent to the mask shop by e-mail, so that the mask shop can keep track of all the documents associated with the project, especially when it comes to assessing problems in fabrication after the returned mask has been received.

15.2.2 *Creating the PDMI Sheet*

In IC fab lingo, the process of ‘taking a picture’ means performing an entire lithographic step. Such a step involves many sub-steps, which are defined in Figure 15.8.

15.2.2.1 **Wafer Cleaning**

Wafer cleaning is a crucial part of lithography, since if this step is not performed correctly, the entire following process is useless. Original quartz or silicon wafers can be cleaned in a solution of heated (around 60 °C) sulphuric acid mixed with a reactant (hydrogen peroxide 30%). The wafers can also be cleaned with pre-mixed solutions of Nano-Strip™. Heating is a key parameter, but one has to be careful, since too much heating can have a dangerous effect (acid explosion). Therefore, this process should always be performed with acid gloves and protective goggles, under a laminar flow hood, in closed baths and with a protective window (see Figure 15.9). The acid bath should also be stirred continuously for better cleaning results.

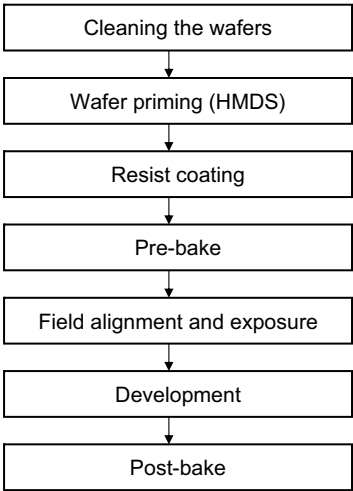


**Figure 15.8** The various steps involved in ‘taking a picture’

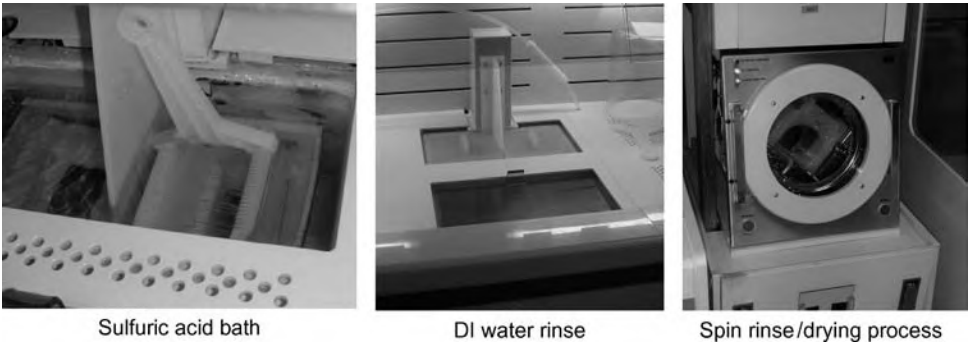


**Figure 15.9** The wafer-cleaning process

Following the acid bath, the wafer cassette is placed into a deionized (DI) water bath (using sprinklers and stirrers, in an ultrasonic bath). Following this DI step, a rinsing and drying step has to be performed in a centrifugal heated rinse/dryer (see Figure 15.8) or some other pressurized air dryer. Such dryers can rotate the cassette very quickly, at around 1000 rpm, without breaking them.

**15.2.2.2 Wafer Priming**

Wafer priming is an important step, since without HMDS (the adhesion promoter hexamethyldisiloxane) applied to the wafer surface, the resist will not stick to the wafers, especially during the development process. The HMDS can be spun onto the individual wafers in a spinner (similar to the resist spinner) or applied in a vapor phase (100–160 °C) directly onto an entire wafer cassette (see Figure 15.10).

**15.2.2.3 Resist Coating**

Resist coating is the process of applying a resist coat on the wafer. There are many resists available today in the industry. Positive resist is used if the exposed area is to be removed after development. Negative resist is used if the exposed area is to remain on the wafer after development. By changing the tone of the mask (see the previous section), one can change the polarity of the resist to give similar results. Where possible, negative resist should be avoided, since its chemical composition is harmful and the solvents used to strip the resist are also more harmful than for positive resists (acetone is used to dissolve positive resists, but



**Figure 15.10** The resist coating process

does not dissolve negative resists). Photoresist is applied by letting drops of resist fall onto a spinning wafer (see Figure 15.10).

The spinning speed is tuned to give the target resist thickness. Usually a resist coating of 1.0–1.5  $\mu\text{m}$  thickness is used for standard resolutions around the micron, and thinner resist coats are used for increased resolution. Typical spinning speeds are between 4000 and 800 rpm.

**15.2.2.4 Resist Pre-bake**

A pre-bake is necessary to prepare the photoresists for exposure. It evaporates any remaining humidity in the resist, to avoid any unwanted swelling. A coated resist that has been pre-baked should be exposed within the following hour in order to prevent swelling (Figure 15.10).

**15.2.2.5 Wafer Exposure**

The exposure process is the central part of the lithographic process. The mask is set in the mask holder and inserted in the lithographic tool (see Figure 15.11). The wafer cassette is inserted in the robotic arm of the aligner or stepper, and the process can begin.

If the field is the first one, the only alignment is the rough wafer alignment to the photomask (aligning to the wafer flat). If the field is to be aligned to a pre-existing field on the wafer, the operators can either use an automatic alignment process (steppers and some mask aligners) or a manual process (mask aligners). One has to remember that the alignment is always only as good as the alignment marks.

**15.2.2.6 Wafer Development**

Wafer development is usually performed cassette by cassette in a mixture of 66% DI water and 33% developer, with an agitator (Figure 15.11). The development time is a critical factor in the resulting feature sizes (e.g. 90 seconds in the development bath).

**15.2.2.7 Post-bake**

Prior to post-bake, the wafer is examined under a microscope in order to evaluate the quality of the lithography. If the lithography is not satisfactory (e.g. by evaluating the PCMs with the naked eye – see the previous section), the resist is stripped off by means of a dry asher, or a heated sulphuric acid or acetone bath, cleaned and sent back to lithography (Figure 15.11). Resist post-bake is usually carried out at 120–145 °C for about 30–45 minutes.

The lithography process is now complete, and the wafers are now ready to be etched.

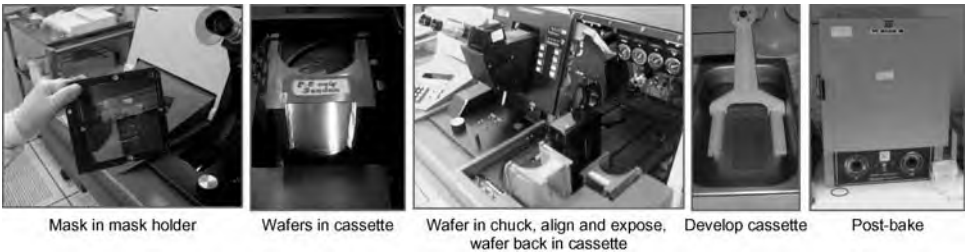


**Figure 15.11** The wafer exposure process (‘taking the picture’)

### 15.2.3 Specifying the Etching Process

Following the optical lithography step specifications, the etching step now has to be specified. In multilevel lithography, there are successive lithography and etching steps ( $N$  steps for  $2^N$  levels). Chapter 12 has reviewed some of the etching processes and related problems. Wet etching is fine for the removal of intermediate Cr layers, but not for etching the substrate. Dry RIE etching should be used if the wafer is made of quartz, fused silica, ZnSe, SiO<sub>2</sub> and so on.

The specification of the etching process is focused not on the process itself, since it is usually proprietary to each fab, but on the geometrical results:

- the etch depth;
- the etched floor surface quality; and
- the side walls.

In the case of refractive profiles, the key criteria are the resulting surface roughness as well as the total etch depth. The etch depth is not related to efficiency problems, but only to lens specification variations (focal length etc.).

In case of diffractive profiles, the key criteria are the side-wall quality and the etch depth. The etch depth is directly related to the efficiency. The quality of the side walls is also related to the efficiency, since they change the duty cycle of the grating or DOE. Such criteria can be measured via the PCMs (see the following sections).

Finally, these criteria should hold on the whole wafer, and not just in one particular area. That is why PCMs should be inserted at all five main locations on the wafer (left, right, top, bottom and center). The alignment marks are usually also placed at these locations. Figure 15.12 shows the three criteria usually specified for etching.

The etch depths should be as close as possible to the desired depths. However, it is important to remember that the estimated depth is only as good as the estimate of the index of the material to be etched at that wavelength:

$$\begin{cases} h_{trans}(p) = \frac{\lambda}{2^p(n(\lambda) - 1)} \\ h_{refl}(p) = \frac{\lambda}{2^{p+1}} \end{cases} \quad 1 < p < N \quad (15.1)$$

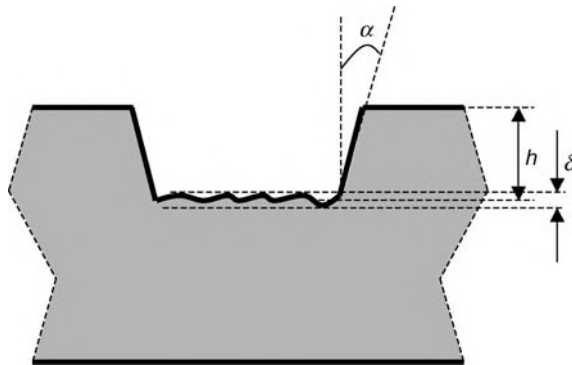


**Figure 15.12** Dry etching parameters for digital diffractive optics

For reflective elements, the index of the material does not appear in the depth specification (see also Chapter 5). Equation (15.1) refers to the etch depth of the individual layer  $p$  for multilevel element fabrication. The maximum etch depth when all  $N$  levels are etched is therefore as follows:

$$\begin{cases} h_{trans\_max} = \frac{(2^N - 1)}{2^N} \cdot \frac{\lambda}{(n(\lambda) - 1)} \\ h_{refl\_max} = \frac{(2^N - 1)}{2^N + 1} \cdot \lambda \end{cases} \quad (15.2)$$

It is therefore very important to know (or to measure if unknown) the exact index of refraction of the material for the wavelength at which the element is going to be used. An error of 7% in the index (e.g. 1.45 instead of 1.55) can reduce the efficiency slightly – a decrease in efficiency of about 5% efficiency is 100% correct (correct with regard to the erroneous depth expression). Now, if one adds an additional 5% error during the etching process, the total etching error becomes then  $7\% + 5\% = 12\%$ , which can take a high toll on the efficiency (more than 10%). This is one of the main errors that is made when specifying an etch depth for a digital diffractive optical element (one takes  $n = 1.5$  for any glass at any wavelength).

The etched side walls should be within  $2^\circ$ . Side walls are much worse (standing waves, resist swell etc.) when the element is still in resist, and is actually used as a resist element. This is why resist does not usually provide a good efficiency.

Generally, in dry RIE etching, there is a pre-asher and a post-asher, which takes care of the roughness in the resist and the remaining resist on the floor. Such pre- and post-ashers (with  $O_2$  gas only) produce better results than just an etching step (e.g.  $CHF_3$  for fused silica etching – ‘oxide etching’).

#### 15.2.4 What Are the Key Parameters to Watch in Multilevel Lithography?

The key elements in the specification of a digital element fabrication process are at least threefold:

- *Resolutions*: the CD has to be reached (the CD is located in the PCM group).
- *Field-to-field alignment*: this is a critical issue, and it has to be measured after fabrication. The measurement verniers are also located in the PCMs. Field misregistrations produce loss of efficiency due to high-frequency noise generation, ghost images and overall reduction of the SNR. Usually, the zero order is not seriously affected.
- *Etch depth*: etch depth errors can almost always be measured by the amount of zero order appearing for a diffractive element. Such etch depth measurement structures (topographical and optical) are also located within the PCMs.

#### 15.2.5 Specifying the Dicing and Die Sorting Process

As there is nothing closer to a diced-out die than another diced-out die, one has to specify how the individual dies are to be numbered and sorted in individual packages. When arrays of slightly different elements are to be fabricated, it is very difficult to sort out the dies after dicing when they are not referenced out on the wafers. Therefore, it is always a good idea to include within a die a reference number or even some text describing the element, so that in the event of shuffling, one can still sort out the dies.

#### 15.2.6 An Example of a Multilevel Lithography/etching Process Specification

Below are the specifications for a simple fabrication process, which is for a four-level digital diffractive element. The element is to be used at 850 nm (VCSEL laser) with a fused silica wafer:

*Wafer specification*

- Fused silica with a flat
- inch diameter, 500  $\mu\text{m}$  thick
- 20  $\mu\text{m}$  TTV
- Scratch/dig: 10/5
- Batch size: 20
- Refractive index  $n(\lambda)$  to be used at 850 nm

*Mask patterning specification*

- Mask specifications:
  - Two reticles
  - 5 inch square, 125 mil thick
  - High-reflectivity Cr on quartz
  - No pellicle
- GDSII data file specifications:
  - File fractured at  $1\times$  scale right reading
  - first masking layer in layer #1 and second masking layer in layer #2
  - Data are flat, user units are nm
  - Library name is 'Library'
  - Structure name is 'Structure'
  - Alignment marks library to be inserted: Canon FPA 2500i stepper
- Mask patterning specifications:
  - CD is 1.05  $\mu\text{m}$  at  $1\times$ , located
  - Fracture grid to be used: 100 nm
  - Cr up wrong reading
  - Field is dark, digitized patterns are clear
- Post-processing:
  - No defect analysis
  - CD to be measured in PCM at 22 °C
  - No pellicle needed

*Optical lithography specification*

- Resolution is 1.05  $\mu\text{m}$  at  $1\times$
- Field-to-field misregistrations:  $< \pm 150 \text{ nm}$

*Etching specifications*

- Etch depth is:
  - $h_1 = 850 \text{ nm}/(n(\lambda) - 1)$ , where  $n$  is the refractive index of the wafer at  $\lambda = 850 \text{ nm}$
  - $h_2 = h_1/2$
- Etch depth errors:  $< \pm 5\%$  for  $h_1$  and  $h_2$
- Side walls:  $< 2^\circ$  for  $h_1$  and  $h_2$
- Etched surface roughness amplitude:  $< 2 \text{ nm}$  for  $h_1$  and  $h_2$

The specifications of the fabrication have thus been described. The PDMI for this fabrication step is described in Figure 15.13.

Although the process seems to be simple (four phase levels with two masks), if a single step is not defined or is ill defined within the process (e.g. no Cr layer—or, worse, there is no mention of the polarity of the resist during lithography), one can end up with a diverging diffractive lens instead of a converging diffractive lens. This, of course, would be a little uncomfortable for the end user, since it would be the responsibility of the end user not to have précised the polarity of the resist during lithography. However, due to the Babinet principle (inversion of binary pulses), the polarity of the resist is not important for binary (two-level) diffractive optics.

When something goes wrong, the error has to be located and the step that went wrong has to be repeated (for free). This is why it is important to produce a specification list for the various fabs involved, rather than

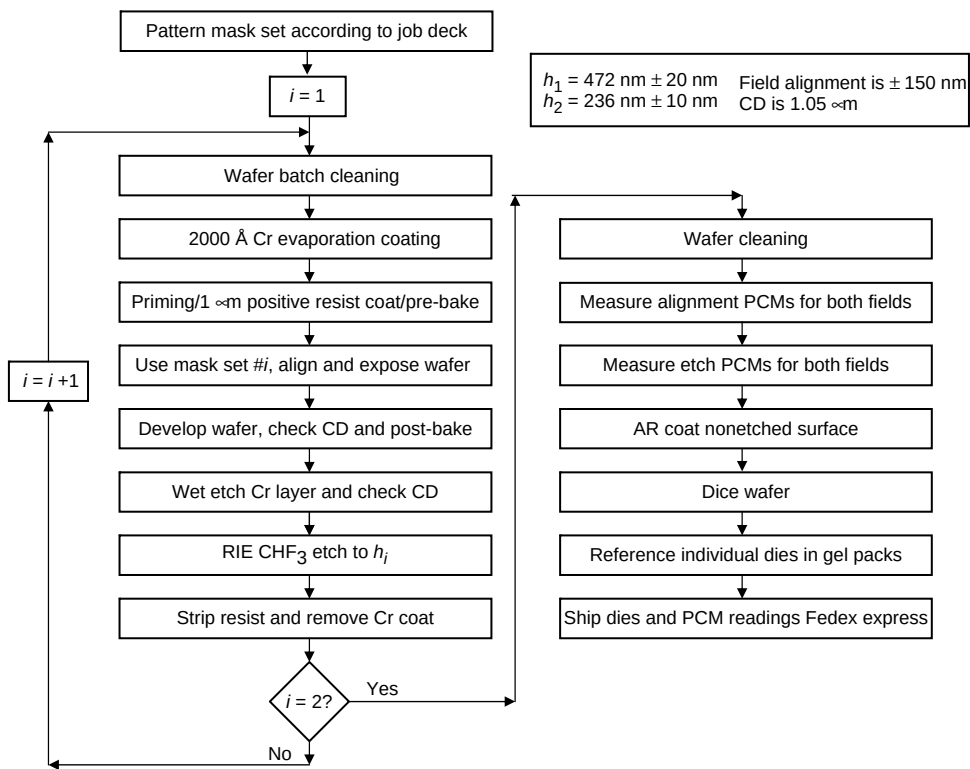


Figure 15.13 A PDMI for four phase level fabrication

a specification list which might run like this: ‘please fabricate a four phase level element for 850 nm’ (such a minimalist specification list is very often used today, and the resulting elements are usually very surprising – but not in a good way, though).

As an example, Figure 15.14 shows the various processes during the lithographic fabrication of such a four phase level element, through mechanical surface profilometry (the Tencor Alpha Step Profilometer).

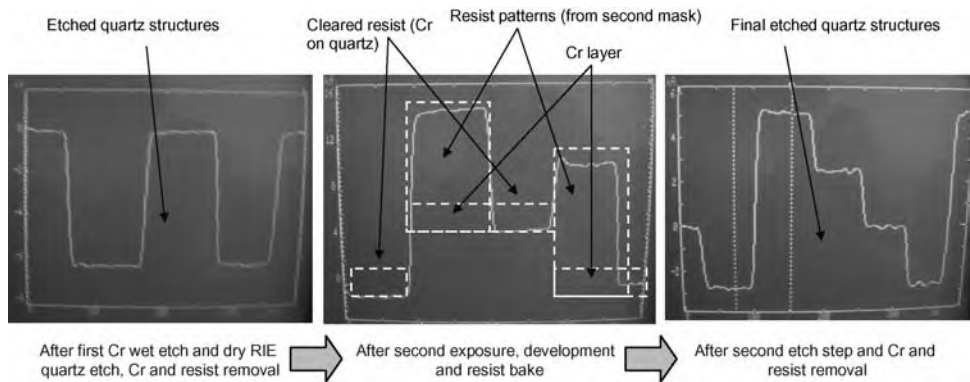


Figure 15.14 The fabrication of a four phase level element in quartz

Figure 15.14 shows a Cr layer in between the quartz and the resist, in order better to form the shape to be etched (lateral shape production). The additional Cr layer is also used so that the wafer handling robot can actually see the wafer during lithography (since it is usually designed to handle only opaque – silicon – wafers).

### 15.3 Fabrication Evaluation

After fabrication, and upon receipt of the final (or partial) elements, one has to assess the quality of the fabrication that has been performed [1–3]. There are two ways to assess the fabrication quality:

- use metrology over the fabricated structures and compare the target specs; or
- use optical PCMs in order to evaluate the quality of the fabrication.

When one has access to optical profilometry and microscopic measurement tools, one can assess the fabrication quality using such tools. However, each tool has its own resolution and its own range of operation. The various tools used in industry today are reviewed below, along with their advantages and drawbacks.

If one does not have access to the previously mentioned costly micro-profilometry measurement tools (which can cost up to several hundreds of thousands of dollars), one can still assess the quality of the fabrication by using inverse optical metrology on the various optical PCMs described in the previous section. This is effectively the ‘poor man’s’ metrology laboratory, since it only requires a \$1 laser pointer and some smart gray cells.

Finally, in many digital optics applications, unlike in VLSI fabrication, the aim is a desired optical functionality and not so much a microstructure pattern fabricated to exact micrometric specs. Very often, systematic fabrication errors can actually improve the optical performance of digital optics, rather than reducing them. That concept is very far from the quality concepts in the traditional semiconductor fabrication industry, which have made the fortunes of companies such as KLA Tencor and others.

#### 15.3.1 Microstructure Surface Metrology

Surface quality analysis and metrology of microstructure surfaces can be performed using the following tools:

- microscopes (lateral measurements only);
- contact surface profilometers (lateral and depth measurements); and
- optical surface profilometers (lateral and depth measurements).

Figure 15.15 shows two contact surface metrology tools: the mechanical stylus profilometer and the Atomic Force Microscope (AFM).

Figure 15.16 shows two optical metrology tools: a phase contrast microscope and a white light interferometric confocal microscope.

These four tools will be used in the following sections.

##### 15.3.1.1 Microscopy

Microscopy is the first tool used to assess the quality of a microlithographic fabrication. Microscopes are usually used right after a lithographic projection, in order to decide if the profile within the resist is good





Stylus profilometer (Alpha Step)                      Atomic Force Microscope (AFM)

**Figure 15.15**    Contact surface metrology tools

enough to be etched down the substrate, or should be stripped down and reprocessed in optical lithography. This is a good way to increase the yield.

*The Traditional Optical Microscope*

Figure 15.17 shows a series of optical photographs of a digital diffractive element after the first etching step, the second and the third (for two, four and eight phase levels).



Phase contrast microscope                      Confocal white light interferometric microscope

**Figure 15.16**    Optical surface metrology tools

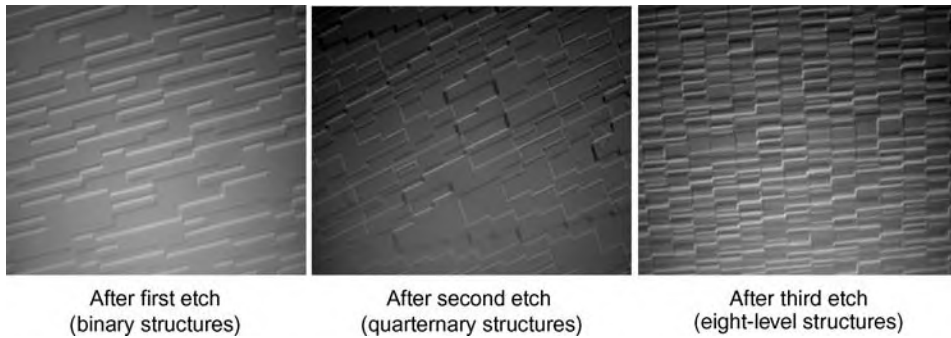


**Figure 15.17** Examples of optical microscope shots of successive steps to produce an eight phase level element

The CGH cells in Figure 15.17 are rectangular. On the four and eight phase level photographs, one can see some of the remaining field-to-field misregistrations. However, such a microscope cannot tell precisely how well the fields are aligned or how deeply the etching has been done.

#### *The Phase Contrast Microscope*

Usually, one ends up with a transparent wafer to measure (fused silica, BK7, quartz, soda lime etc.). Transparent microstructures are often a challenge to see with traditional microscope, especially in mask aligners, when the alignment has to be done manually. This is why an additional amplitude layer (Cr) was used in Figure 15.13 (multilevel lithography).

However, when the fabrication is done, the intermediate optional Cr layer is removed and one has a complete transparent wafer to be analyzed. Phase contrast microscopy is an optical microscopy illumination technique in which the small phase shifts produced by the various etching steps are converted into amplitude or contrast changes in the image. Therefore, the various etching steps appear as increasing gray level steps under a phase contrast microscope.

#### *Electron Microscopy*

In Scanning Electron Microscopy (SEM), the sample has to be prepared and coated with a gold layer prior to being inserted in the SEM vacuum chamber, so usually the sample is destroyed. Also, the sample needs to be quite small in order to be able to be inserted on the vacuum chamber holder. So it is not possible to analyze an entire 4 or 6 inch wafer without breaking it and coating it with gold. The process of coating changes the surface quality and the surface height, as well as the side-wall angles. Figure 15.18 shows some SEM photographs of the same element at different magnifications.

However, SEM remains a practical tool that allows a much higher magnification than optical microscopes. Depth measurements are possible by cutting the wafer and taking a picture on the edge of the wafer. However, this is very tricky, since the cut has to be made properly and located very accurately on the wafer. Finally, an SEM photo has a tremendous depth of focus, since electrons are used to form the image rather than photons. Such SEM photos are usually very attractive for advertising, but less so for accurate metrology, especially in the third dimension.

#### *Near Field Scanning Microscopy*

Near Field Scanning Optical Microscopy (NSOM) is a microscopic technique for nanostructure investigation that breaks the far-field resolution limit by exploiting the properties of evanescent waves.

This is done by placing the detector very close ( $d \ll \lambda$ ) to the specimen surface. This allows for surface inspection with a high spatial, spectral and temporal resolving power. With this technique, the resolution

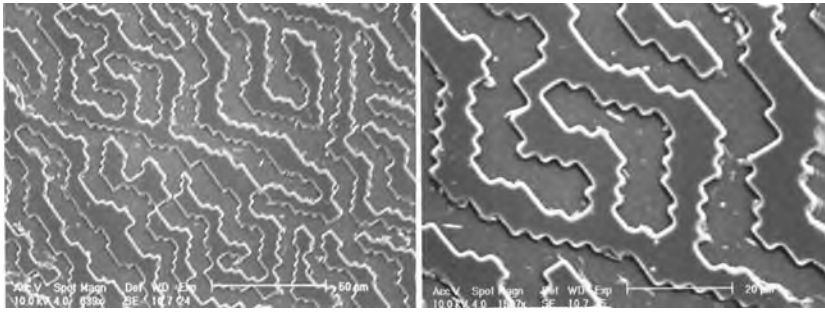


**Figure 15.18** SEM photographs of a Fresnel beam shaper under various magnifications

of the image is limited by the size of the detector aperture and not by the wavelength of the illuminating light. As in optical microscopy, the contrast mechanism can be easily adapted to study different properties, such as the refractive index, the chemical structure and local stress.

#### ***Fourier Tomography Microscopy***

Fourier tomography microscopy, or diffraction tomography, is a variation of optical coherent tomography, and is very closely related to digital holography as presented in Chapter 8.

Fourier tomography microscopy is used to image and analyze transparent elements, especially in biotechnology. However, such elements are also very helpful for the metrology of index-modulated holograms. Holograms having index modulations cannot be measured by the various profilometry techniques used for surface-relief (transparent) elements. Optical Fourier tomography is a tool that allows the optical engineer to ‘see’ the Bragg planes within the holographic material.

### **15.3.1.2 Contact Surface Profilometers**

Contact surface profilometers were the first tools available to measure accurately the surface profiles of microsystems such as multilevel digital optics. Two types of contact profilometers are available for different resolution ranges (see Figure 15.19).

#### ***Stylus Profilometers***

A stylus profilometer is a simple tip on a spring-loaded system, which is linked to a scale (see Figure 15.19). Such a profilometer is a cheap way to measure surfaces directly in a cleanroom environment. It does not require any anti-vibration table. It does, however, have to be calibrated every now and then, due to spring tension variation and tip wear-out. It is a truly mechanical device, and such scans have been reported in Figure 15.14.

Note that in Figure 15.14, there is no mention of the underlying structure, which was resist on chrome on quartz. The only information one gets is the surface profile. If the surface profile has remaining resist or Cr, one does not measure the etch depth correctly. This is also true for most of the following measurement tools. Thus, the etch depth, no matter how it is measured, involves either knowledge of the exact resist height (which changes with temperature and humidity) or complete resist removal (which is easy using a cotton tip with acetone).

#### ***The Atomic Force Microscope***

The Atomic Force Microscope (AFM) or Scanning Force Microscope (SFM) is a very high resolution type of scanning probe microscope, with a demonstrated resolution of fractions of a nanometer. The precursor to the AFM, the Scanning Tunneling Microscope (STM), was developed by G. Binnig and H. Rohrer in the

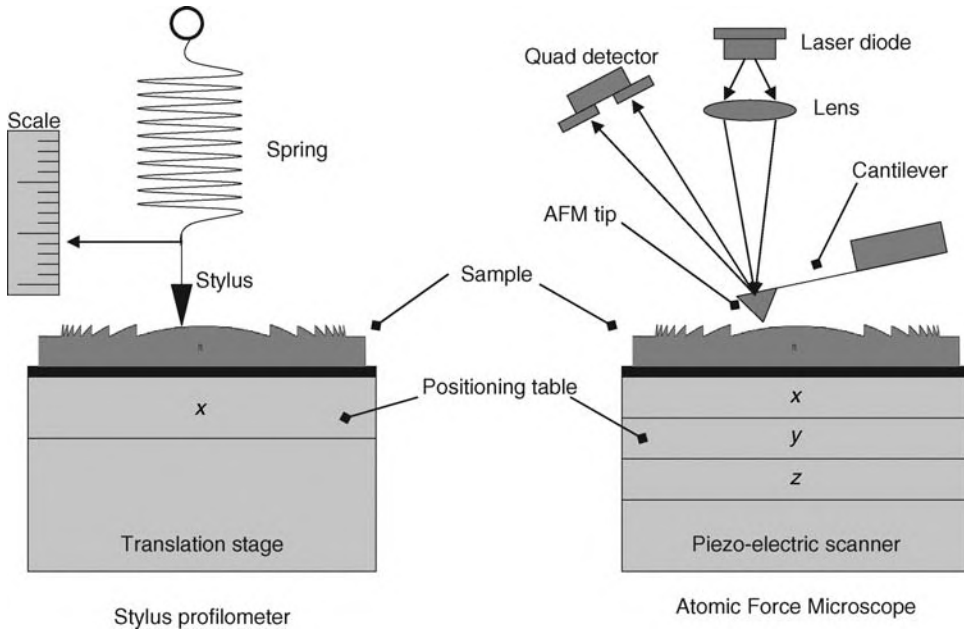


**Figure 15.19** Contact profilometers: the stylus profilometer and the AFM

early 1980s, a development that earned them the Nobel Prize for Physics in 1986. G. Binnig, C.F. Quate and C. Gerber invented the first AFM in 1986. The term 'microscope' in the name is actually a misnomer, because it implies looking, while in fact the information is gathered by 'feeling' the surface with a mechanical probe (see Figure 15.19). Piezoelectric elements that facilitate tiny but accurate and precise movements on (electronic) command enable the very precise scanning.

AFM measurements can only be performed over a very small area, usually 5 or 10  $\mu\text{m}$  wide, and the scanning is done in 1D over a 2D surface, thus it can take a long time to scan a surface with a small scanning step. The scanning steps can be increased on the piezo-table, but the resolution is thus decreased. It has been reported that carbon nano-tubes have been placed under AFM tips to increase the resolution of these systems ever further. AFM plots are compared to interferometric white light confocal microscope plots in the next section.

### 15.3.1.3 Noncontact Surface Profilometers

Noncontact profilometers (see Figure 15.15) have numerous advantages over scanning tip profilometers. First, there are noncontact metrology tools, which can scan a 2D or 3D surface topology in a single shot. The vertical resolution of these devices is a fraction of the wavelength, and can be as small as the nanometer!

One of the disadvantages of such profilometers is that they measure surface profiles and phase changes equally well, and do not discriminate between them. However, if the surface to be measured is a clean surface with a large index variation (such as air to glass, or air to plastic), the measurement (in reflected light) is very accurate. However, the lateral resolution is limited to the optical resolution of the microscope device. AFMs have thus much better lateral resolution, since they are not limited by any optical resolution.

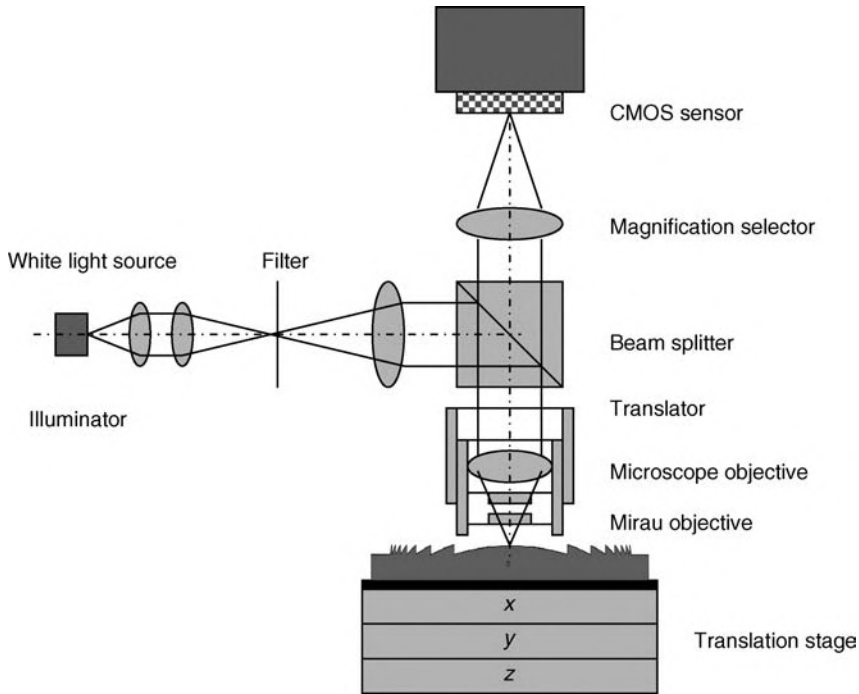


**Figure 15.20** A diagrammatic view of an interference microscope

### ***Microscope Interferometers***

The interference microscope is shown diagrammatically in Figure 15.20. There are many variations of the interferometric head, which can be inserted in the microscope objective to produce the interference objective [4–6]. The three major ones are listed below:

- the Mirau interferometer;
- the Michelson interferometer; and
- the Linnik interferometer [7].

These three interference objectives are also described in Figure 15.21.

A Mirau objective provides medium magnification, a central obscuration and a limited numeric aperture. A Michelson interference objective provides low magnification and a large field of view, but the beam splitter limits the working distance. There is no central obscuration. A Linnik interference objective has a large numeric aperture and a large magnification, and the beam splitter does not limit the working distance; however, they are expensive and require two matched objectives.

### ***The Heterodyne Two-wavelength Interferometric Microscope***

In a two-wavelength interferometric microscope, the equivalent wavelength (which is the wavelength corresponding to the beat frequency between the two wavelengths) provides the measuring wavelength [8]. The equivalent wavelength is as follows:

$$\lambda_{eq} = \frac{\lambda_1 \lambda_2}{|\lambda_1 - \lambda_2|} \quad (15.3)$$

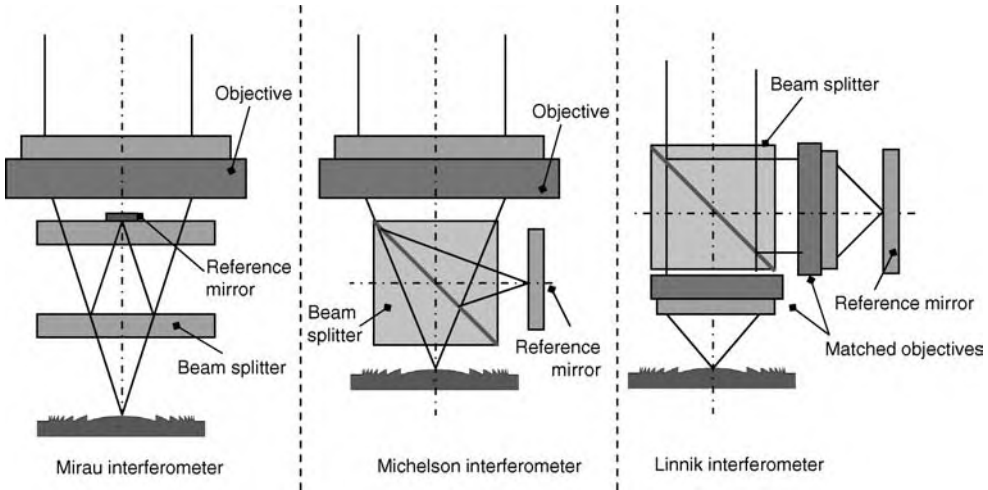


**Figure 15.21** Interference objectives for white light confocal interferometric microscopes

The resulting equivalent phase is as follows:

$$\varphi_{eq} = \varphi_1 - \varphi_2 \quad (15.4)$$

### ***The White Light Confocal Interferometric Microscope***

The optical sources for optical interferometers can be laser or even white light [9–11]. The advantage of white light over laser light is that it creates less noise and no spurious fringes, multiple-wavelength operation is inherently present (for the measurement of large steps) and the focus is easy to determine.

Figure 15.22 shows some scans performed by a confocal white light interferometric microscope, with a Mirau objective. The sample is a four phase level diffractive lens etched in fused silica (Zygo system).

Although there was no reflective coating on the surface, the 4% Fresnel reflections were sufficient to acquire a good profilometry scan of the surface, showing the four phase levels. The interferometric scan is displayed in the lower part of the right-hand side. The reconstructed 3D data are reproduced in the upper right-hand side. A linear scan over the surface is reproduced in the lower part of the left-hand side.

A similar scan of an array of high aspect ratio binary lenses is shown in Figure 15.23. Although the (relatively) high aspect ratio of the structures in Figure 15.23 could be a challenge for interferometric measurements, the results are still valid. However, the limitations of this technology will be described below.

Figure 15.24 shows two scans performed with two different objectives on the same four phase level vortex lens.

Figure 15.23 shows that the results of white light interferometric plots are better when the structures are larger, even with a smaller objective. The lateral resolution is dependent on the microscope, so it is not very accurate. However, the longitudinal resolution can be accurate to several nanometers.

### ***Surface Roughness Analyses***

Rough surfaces reduce the fringe contrast of interferometric plots. Figure 15.25 shows a scan over fine structures with fairly rough surfaces, fabricated over four phase levels (the off-axis region of the same lens). One can see clearly in Figure 15.25 that the fringe contrast is reduced, mainly due to surface roughness.

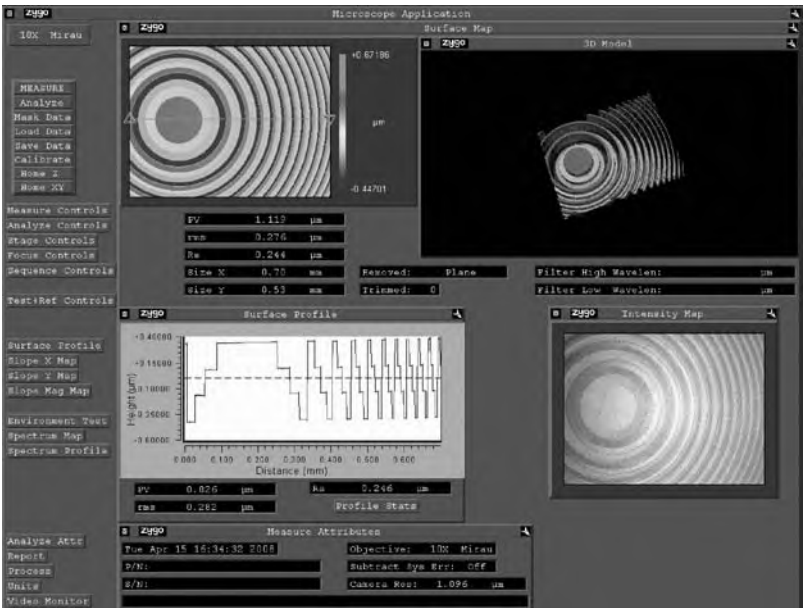


**Figure 15.22** An example of an interferometric plot through a Mirau confocal interferometric white light microscope

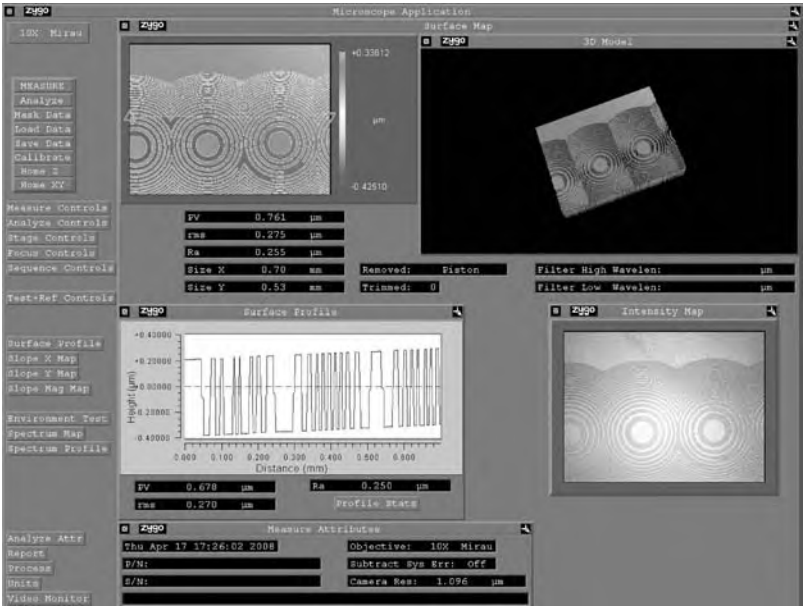


**Figure 15.23** An interferometric scan over an array of high aspect ratio binary lenses

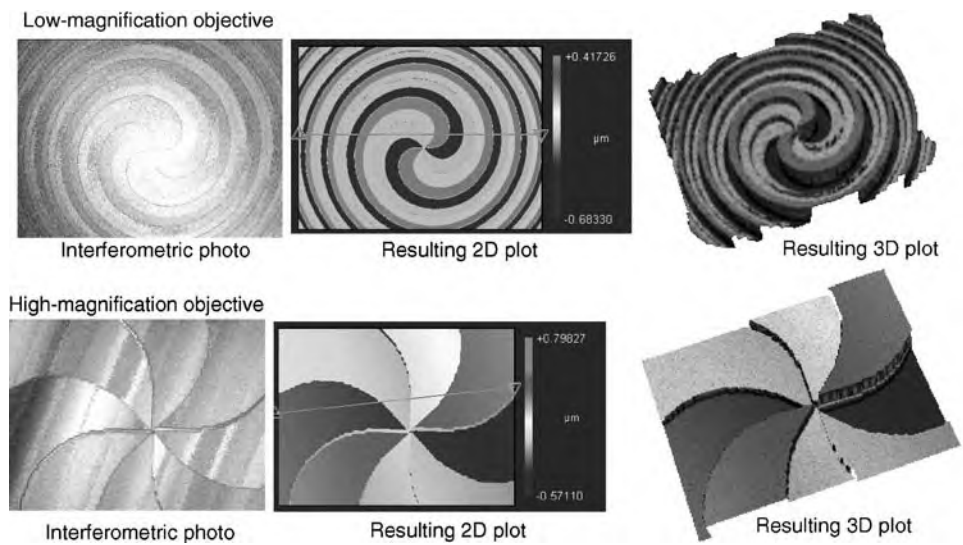


**Figure 15.24** Successive scans over a four phase level vortex lens

If the surface height distribution of the etched phase elements is Gaussian, and the standard deviation is  $\sigma$ , the normal probability distribution for the height  $d$  is as follows:

$$P(\delta) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(\frac{\delta^2}{2\sigma^2}\right)} \tag{15.5}$$

The fringe contrast reduction  $C$  due to surface roughness is then as follows:

$$C = e^{-8\left(\frac{\pi\sigma}{\lambda}\right)^2} \tag{15.6}$$

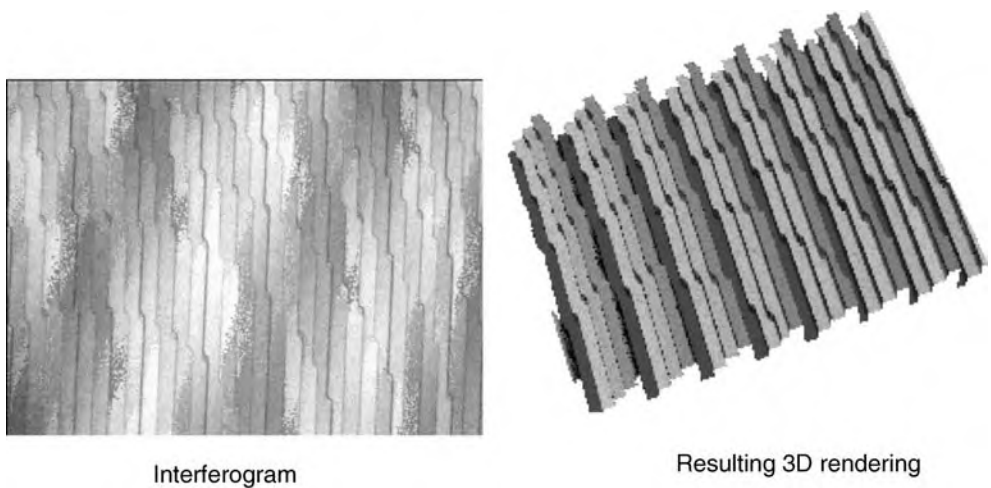


**Figure 15.25** The off-axis region of the lens shown in Figure 15.24



### ***Comparing Interferometric and AFM Plots***

Interferometric plots and AFM plots are compared below, in order to be able to decide when one should use the first or second technique to analyze the surface profile of surface-relief digital optics. To do this, a surface-relief element is chosen that shows continuously varying periods, such as a chirped grating, an EMT element or a diffractive microlens. Figure 15.26 shows an interferometric plot of an EMT-type element (see Chapter 10), which is basically a blazed linear grating sub-fringe, and a plot of two binary microlenses.

It can be seen in Figure 15.26 that the depths of the structures decrease with the local period. One can then conclude that the fabrication (e.g. the etching) is not isotropic and produces shallower grooves when the structure is smaller. On the other hand, one can also conclude that the interferometric confocal microscope system is not ideal for such scans. Figure 15.27 shows a 3D rendering of high- and low-NA binary microlenses, and shows a very good and clean profile for the larger structures (the center of the high-NA lens and the entire low-NA lens) but a very bad profile for the outer region of the high-NA lens.

Figure 15.28 shows the same part of the lens, scanned with an AFM and with the white light Mirau objective confocal microscope. The figure shows that there are actually no depth variations, and that the depth variations seen previously are actually an artifact of the interferometric measurement tool, related to the limited pixel-to-pixel resolution.

Figure 15.29 shows another scan over the same region of a CGH with a confocal interferometric microscope and an AFM.

The CGH cell size here was  $1.5\text{ }\mu\text{m}$ . Both scans show good results. However, the AFM took about 10 minutes, whereas the confocal microscope plots took only 5 seconds. The AFM plots, however, could provide information about the remaining resist lips over the unetched parts of the wafer.

Figure 15.30 shows a scan over a CGH with larger cells of  $2.5\text{ }\mu\text{m}$ , and by using the same objective, one can see from the simple interferometric scans the same remaining resist lips on the unetched parts of the wafer.

One can thus conclude as follows:

- use a confocal as much as you can, since it is fast and very accurate for structures down to about twice the wavelength (around  $1.5\text{ }\mu\text{m}$ ); and
- use an AFM below  $1.5\text{ }\mu\text{m}$ .

An AFM has to be calibrated much more often than a confocal microscope.

There are other artifacts that are specific to the AFM, and these can be seen in Figure 15.31. The effects in Figure 15.31 are very often present in AFM plots. They do not provide information about the precise geometry of the etched side walls. This is due to the fact that a mechanical tip is used. Such effects are not encountered in confocal interferometric plots.

## ***15.3.2 Process Control Monitors***

The previous sections have reviewed contact and noncontact profilometry techniques to assess the quality of the fabrication. One can also use metrology patterns to assess the same qualities without using a microscope or any other device apart from a laser pointer.

### **15.3.2.1 The Use of Conventional PCMs**

When aligning and measuring the accuracy of the alignment, the IC industry uses alignment marks and alignment verniers. The former are used to align the field and the latter to assess how well it is aligned. Alignment marks and PCMs can be joined in a single pattern set that is replicated over the wafer at

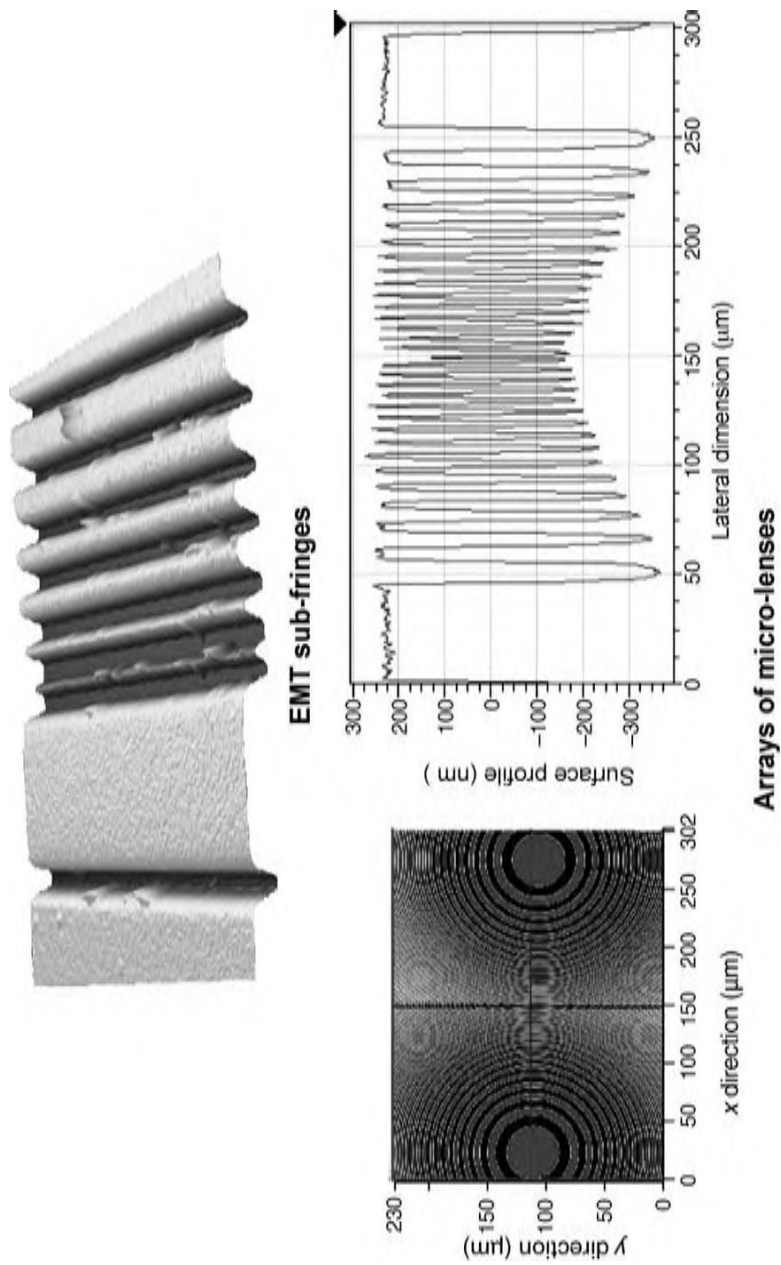


**Figure 15.26** An interferometric plot of EMT sub-fringes and binary lenses

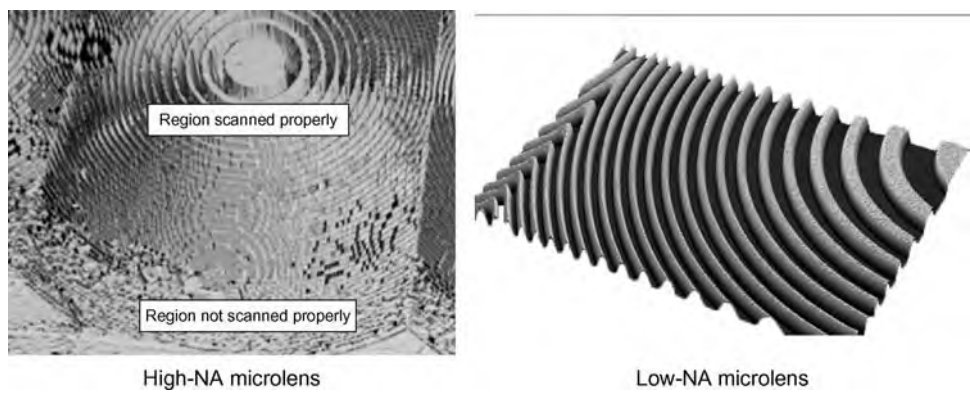


**Figure 15.27** A 3D plot of large and small binary lens structures

strategic locations, especially when using a mask aligner. Figure 15.32 shows the positioning of such PCM/alignment mark patterns. Different positions for the alignment marks and the PCMs have been used in the figure. The PCMs are also located close to the most important features of the mask, in order to see how well one part of the wafer or another has been fabricated.

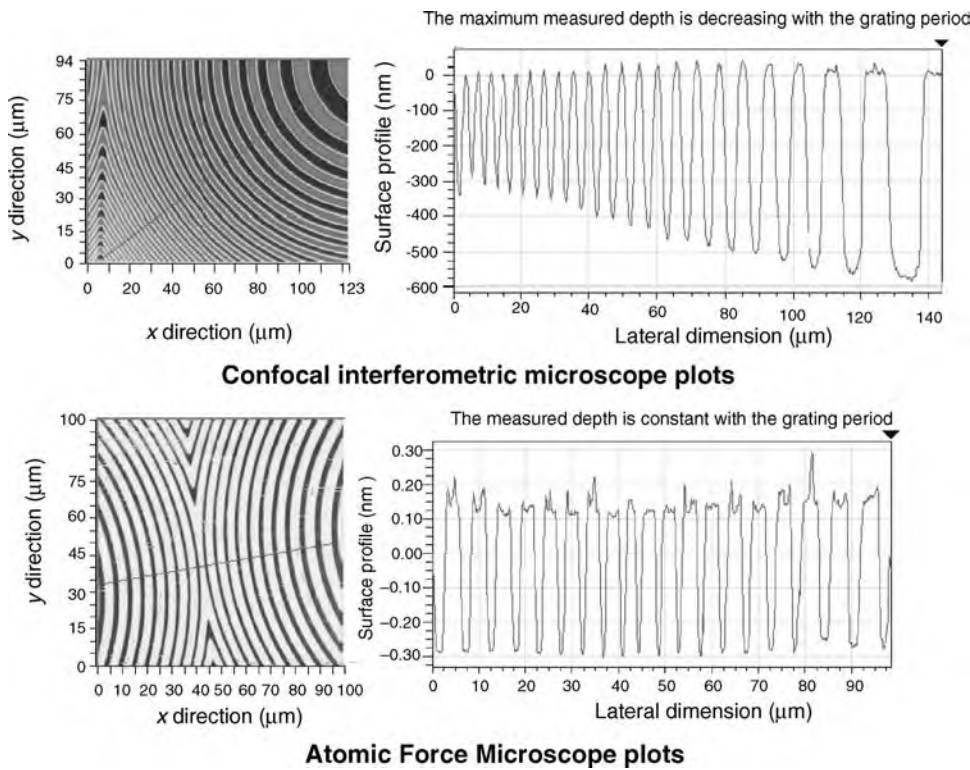


**Figure 15.28** A comparison between AFL and interferometric plots

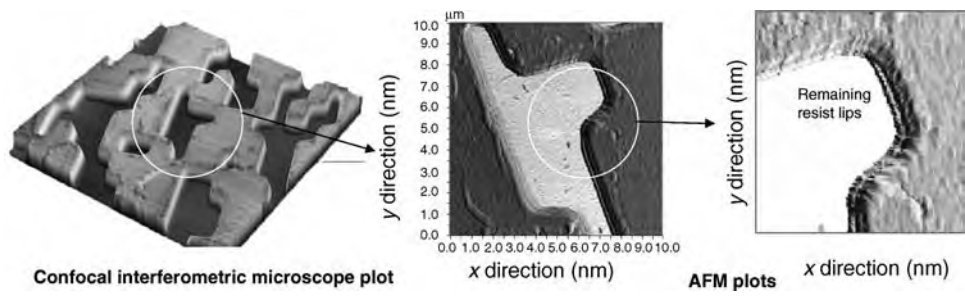


**Figure 15.29** A comparison between an interferometric microscope and AFM scans over an area of a CGH

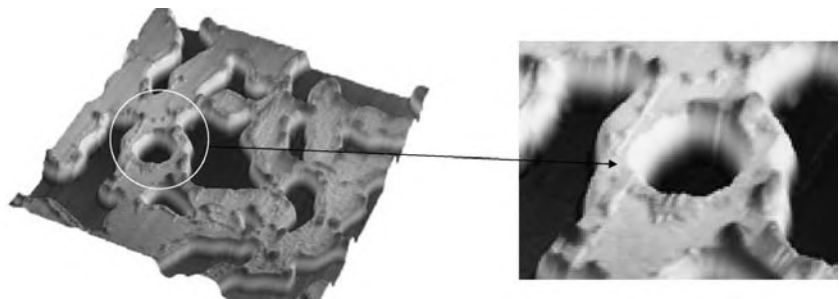


**Figure 15.30** Scans over a 2.5  $\mu\text{m}$  cell size CGH

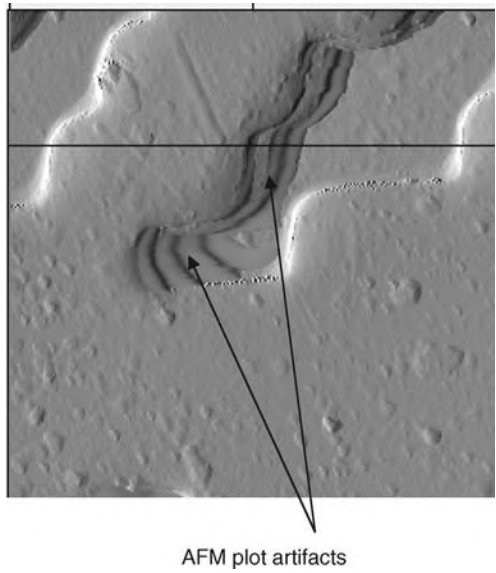


**Figure 15.31** AFM scans, showing problems with side-wall resolution

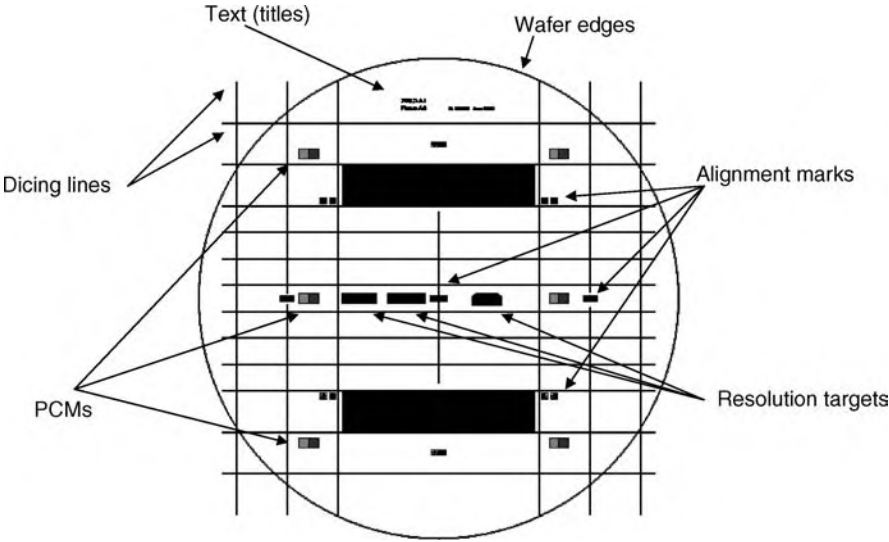


**Figure 15.32** The positioning of PCM/alignment mark patterns on a wafer for a mask aligner

Figure 15.33 shows a typical hybrid PCM/alignment mark die. In the figure there are three types of alignment marks:

- boxes in boxes,
- crosses in crosses, and
- crosses in boxes

and two types of alignment PCMs:

- linear verniers, and
- right-angle gratings.

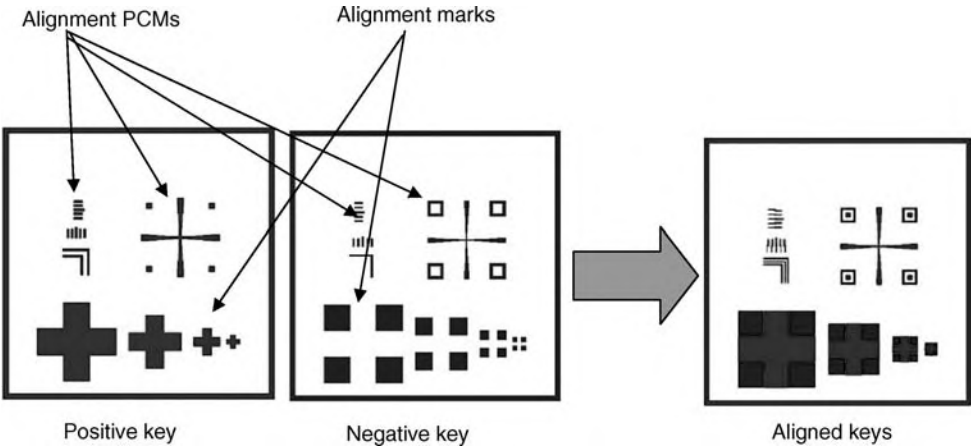


**Figure 15.33** A hybrid PCM/alignment mark die

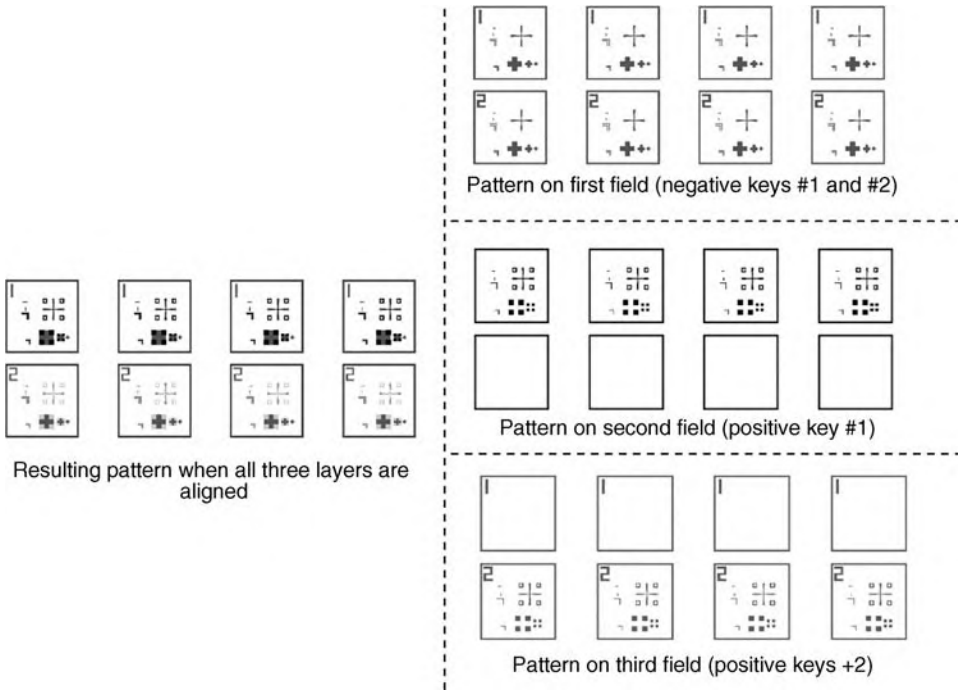


**Figure 15.34** Repetition of the alignment marks/PCMs for alignment of three layers

Other PCMs that are valuable are sets of orthogonal gratings with decreasing periods, and circular gratings.

The alignment marks have to be easy to evaluate by the naked eye of the technician operating the aligner. Verniers and moiré effects (for wafer-to-wafer alignment, see Chapter 12) are not easy to assess, and moreover they are not standard in the IC industry. Verniers have to be set in both directions in order to analyze the quality of the alignment in both directions.

In order to align more than two layers, such marks have to be repeated, as shown in Figure 15.34.

Figure 15.35 shows another type of alignment mark system, where the same pattern is replicated at smaller and smaller sizes, so that the aligning technician can first focus on the larger marks, and then gradually align more and more finely.

Figure 15.35 also shows another type of PCM: the etch depth measurement features. These features resemble a chirped grating, with smaller and smaller, but very tall, structures. Therefore, one can find these structures easily, align the stylus or AFM tip over them and scan from the larger ones to the smaller ones. As these structures are individual and do not have any other layers around them, one can assess after multilevel fabrication how the first level was fabricated, and then how the second level was fabricated, and so on, while checking the final wafer with all the levels etched. Such individual layered etching structures are very important if one does not have access to the wafer in between successive lithographic steps.

Figures 15.36 and 15.37 present actual results of field-to-field alignments on fused silica wafers, using the previously described alignment marks and PCMs. It is interesting to note that in these figures the first and second layers are easily recognizable, since the first one is etched more deeply than the second (actually to half the depth). The results for both PCMs show a better than 0.25  $\mu\text{m}$  alignment accuracy in both directions.

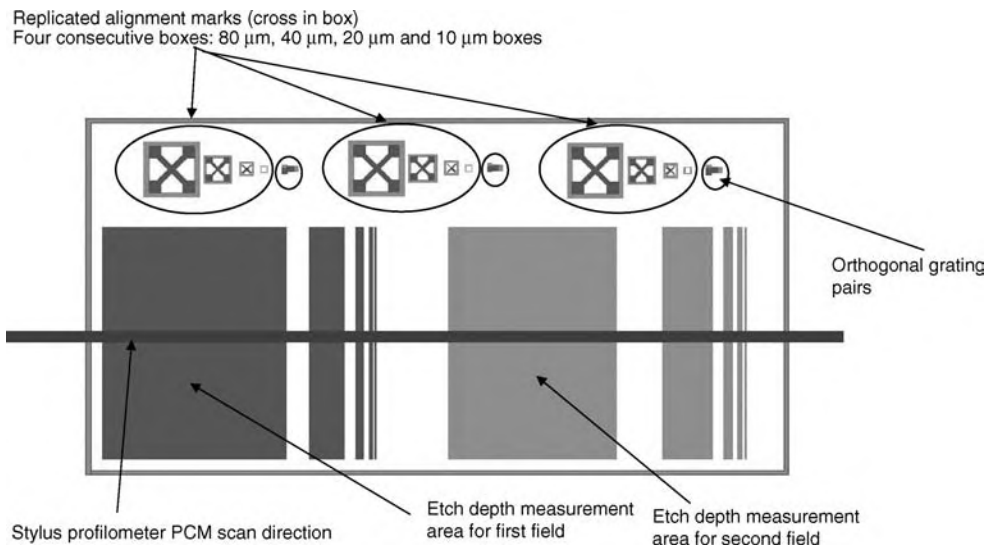


**Figure 15.35** Successively smaller alignment marks for a mask aligner

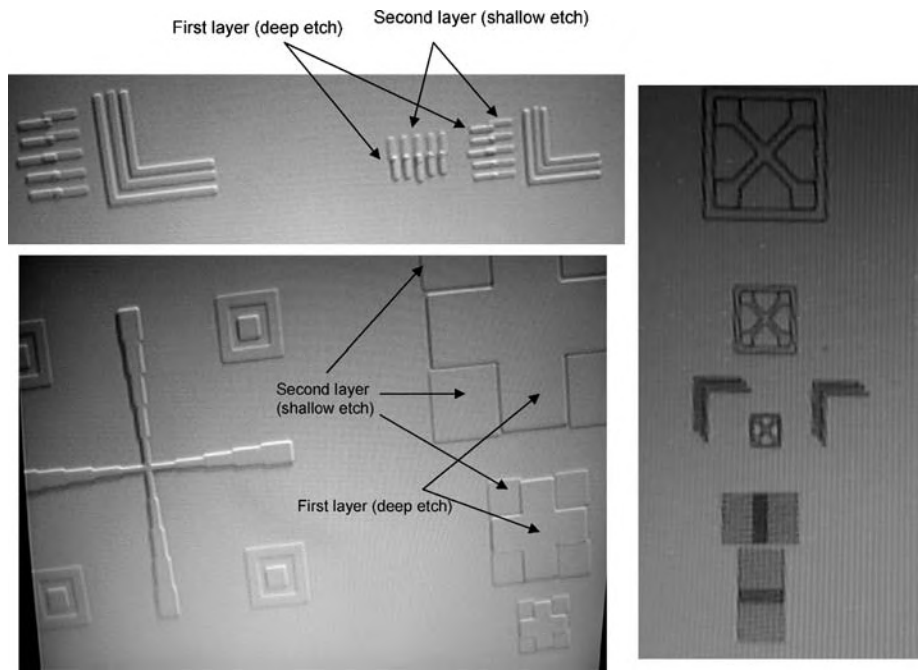


**Figure 15.36** Examples of etched field-to-field alignment marks

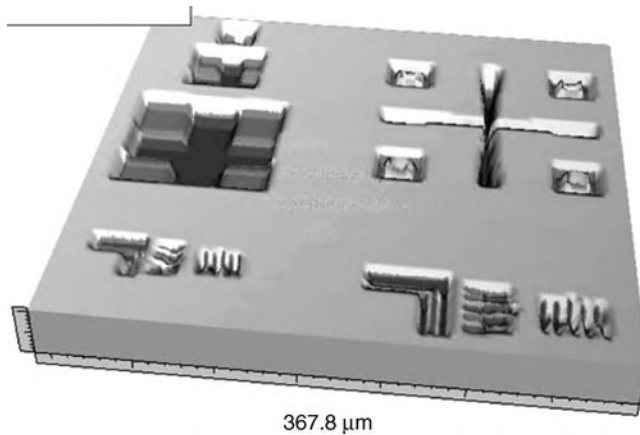


**Figure 15.37** A confocal microscope scan over aligned PCMs

### 15.3.2.2 Diffractive PCMs or Inverse Lithography

Diffractive PCMs are a new kind of PCM, which can be applied to either IC manufacturing or digital optics manufacturing. Basically, the task is to use a laser pointer, point it at the diffractive PCM and watch the reconstruction in the far field, and from the analysis of that reconstruction get an idea of how well the structures have been fabricated. This is also called ‘inverse lithography’, in that the quality of the lithography and of the 3D structures fabricated is computed from the diffraction pattern, similarly to how an X-ray diffraction provides information about 3D crystal structure and crystal orientation.

Figure 15.38 shows how a CGH producing an on-axis array of concentric circles can produce a very good analysis of the lateral lithographic resolution and the etch depth on a fabricated wafer. Furthermore, one can assess the directional resolutions achieved by the lithographic exposure tool by analyzing the amount of light in particular angular directions in the reconstructed light circles.

Figure 15.39 shows the CGH PCM used in the previous example. The fundamental window reconstruction and the higher orders are also shown on the same figure. A CGH PCM that is used for resolution evaluation should always be binary, even if the fabrication is a four-, eight- or 16-level fabrication. A CGH PCM should be inserted for every field, so that the resulting PCM window shows a CGH PCM that can refer to the individual resolution of the first masking step, the second step and so on, and is not linked to either alignment errors or multiple exposures with various resolution results. Therefore, there should be  $N$  CGH PCMs for  $N$  masking sets in a single PCM window.

## 15.4 Optical Functionality Evaluation

The previous section reviewed the various tools (direct metrology and indirect optical PCM measurements) to assess the quality of the fabrication, and compare it to the target specs [12].

However, one might be more interested in how the final digital optical element behaves in the desired product rather than how well the fabrication has been done. A decent digital optical element can actually be based on lower-performance fabrication tools [13–16]. In many cases, in digital optics, the final product works better with a lower fabrication standard [17,18].

### 15.4.1 An Optical Test Bed for PCM

A typical optical test bed for PCM analysis is described in Figure 15.40. Such a test bed can be used to produce the diffractive PCM analysis presented in Figures 15.38 and 15.39 [19]. The reconstruction plane



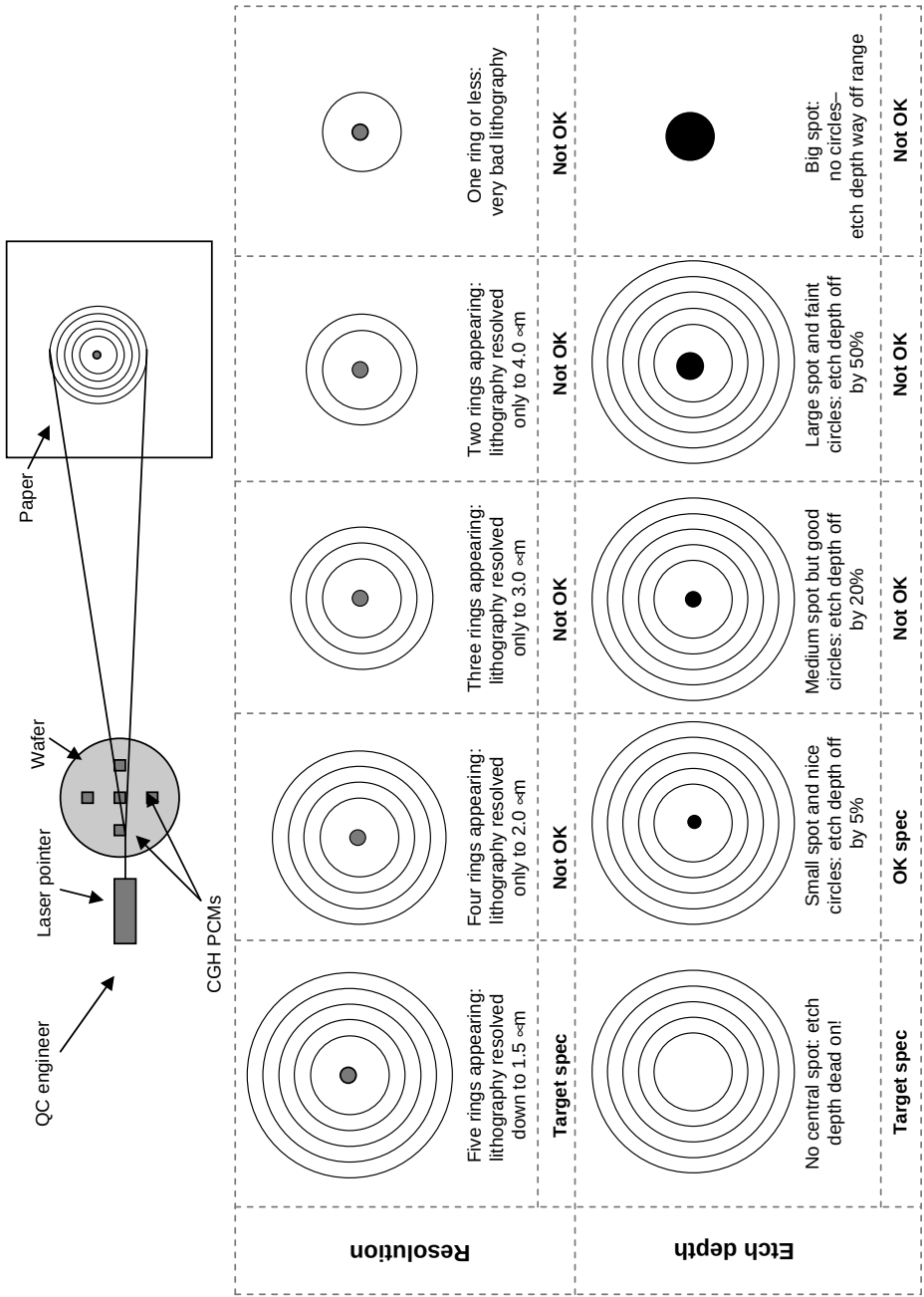


Figure 15.38 The use of diffractive PCMs to analyze the quality of the lithographic resolution and the etch depth

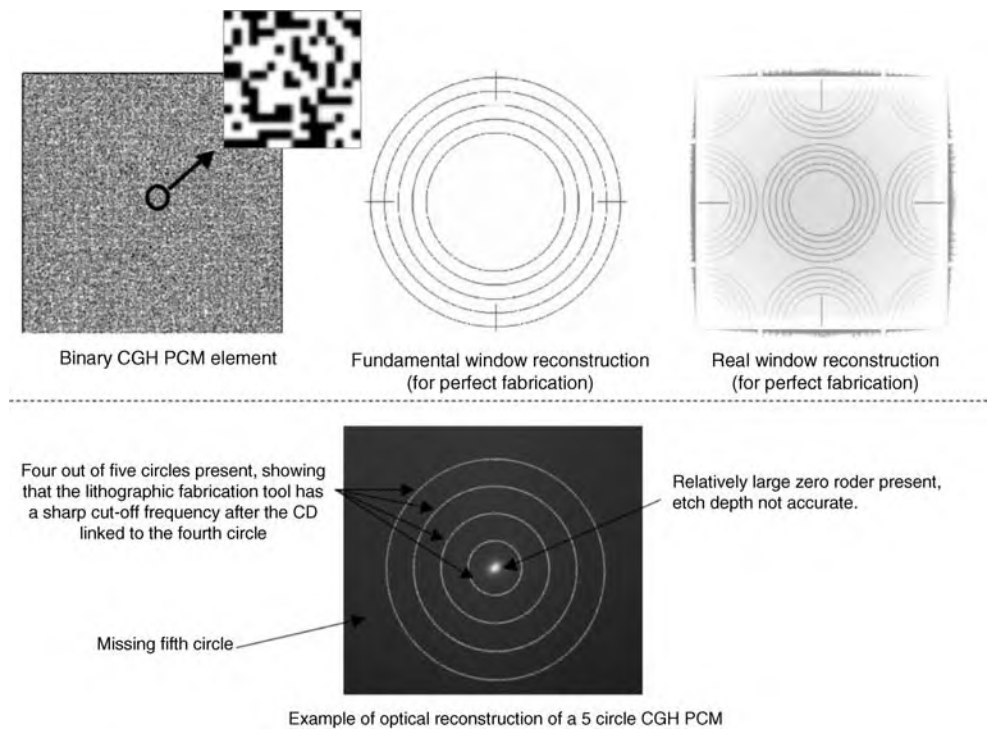


**Figure 15.39** An example of a CGH PCM for resolution analysis

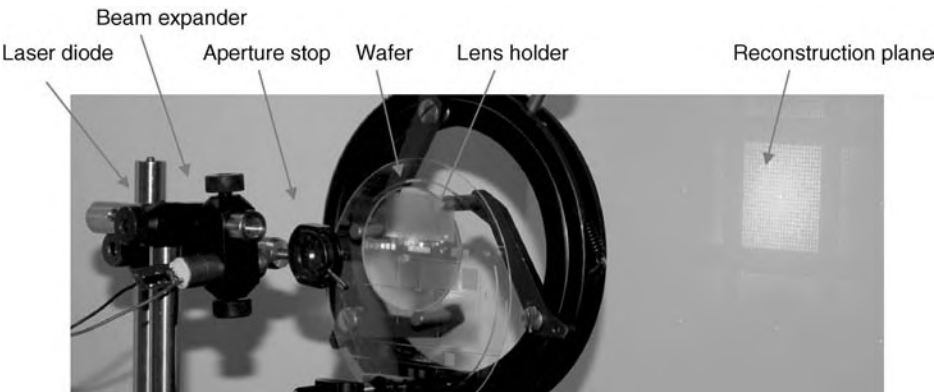


**Figure 15.40** A typical test bed for PCM analysis

is the Fourier plane of the lens. The laser beam is collimated. The far-field reconstruction therefore appears on that plane.

This test bed is not only usable for Fourier CGHs and gratings, but also for Fresnel CGHs and diffractive microlenses, since one can analyze the quality of the lenses much better when taking a look at the far field pattern, especially if these are arrays of microlenses. The beam expander and aperture stop are used to shape the incoming beam to the size of the die on the wafer, since if any laser light falls onto unetched parts of the wafer, they compensate for the zero-order intensity.

15.4.2 Digital Optics Are Not Digital ICs

Digital optics are not ICs, and therefore there are different criteria to be taken into consideration during fabrication [20,21]. One of these is the etch depth, and the other is the surface roughness. The resolution and the field overlay accuracy are two criteria that are of equal importance for ICs and digital optics. The PCMs described earlier (standard and diffractive) inform the engineer about these criteria for a processed wafer.

The next section will review the effects of etch depth errors and the effects of field-to-field misalignments.

15.4.3 Effects of Etch Depth Errors

Etch depth errors usually translate into an increase of intensity in the zero order, and thus a reduction in the overall diffraction efficiency. Figure 15.41 shows the reduction of the efficiency in the fundamental orders (positive and negative combined) for a binary lens as a function of the etch depth error. The results in Figure 15.41 are only valid for single mask etch depth errors. For consecutive additive errors, the results are more complex, since the errors are no longer uniform across the range of phase levels.

Etch depth errors, however, do not alter the geometry of the reconstruction. The various orders are where they should be and produce the effect desired, but with a very different intensity distribution.

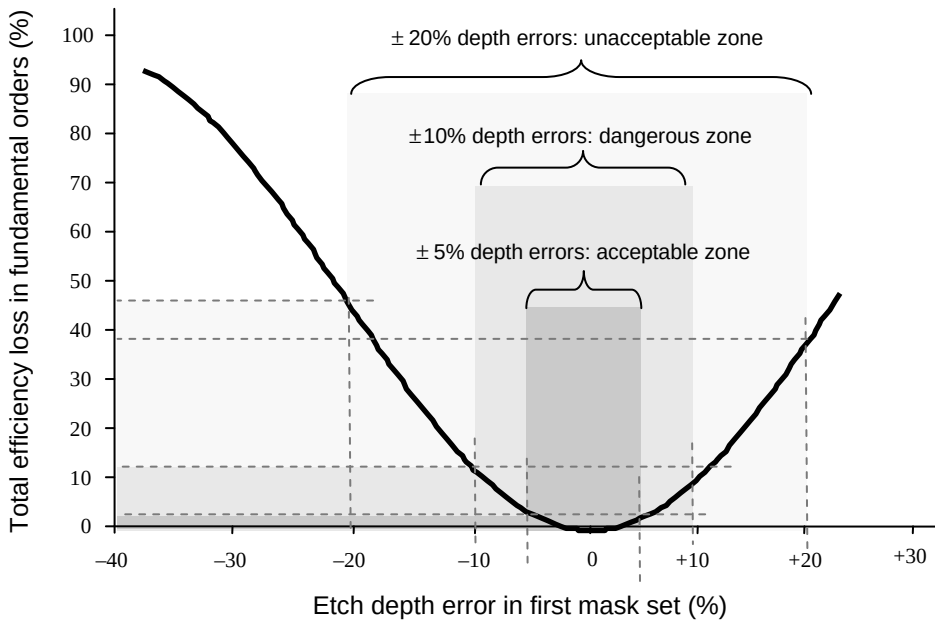


Figure 15.41 The decrease in efficiency as a function of the etch depth errors

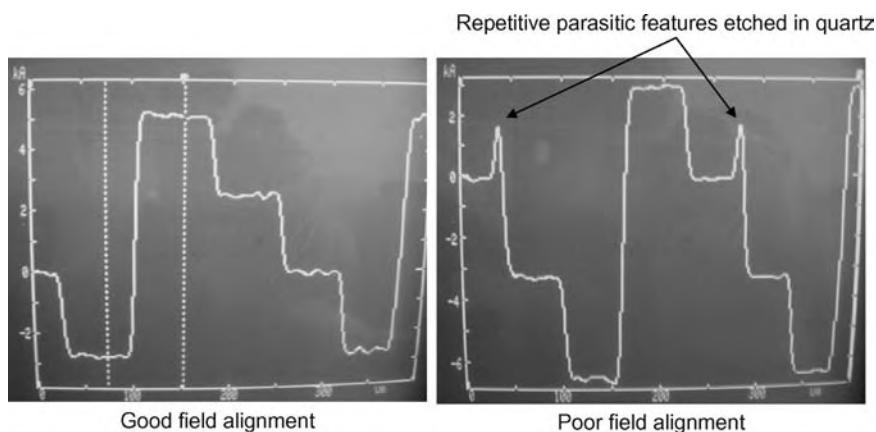


**Figure 15.42** Good and poor alignment for a four phase level element

#### 15.4.4 Effects of Field-to-field Misalignments

When a field is misaligned to a previously transferred field on the wafer, high-frequency artifacts appear on the structures [20–22]. These artifacts can be much smaller than the resolution of the lithographic tool. This is an effect similar to double exposure to reduce the  $k_1$  resolution factor in DUV lithography (see Chapter 13), and it proves that one can produce much smaller features by simply aligning carefully fields with a precision much smaller than the optical resolution of the lithographic device. However, here it is a negative effect, which produces high-frequency noise, reduces the overall SNR and produces ghost images in CGHs. Figure 15.42 shows a four phase level element with good alignment and with alignment errors producing the parasitic effects described here. In this case, the parasitic features are etched into the quartz substrate.

As the parasitic features are repetitive (if the pattern is repetitive), they can create high-order grating effects, superimposed on the lower-frequency gratings, which are constrained within lower diffraction cones. The effects on the resulting optical functionality or intensity pattern can, however, be reduced since the diffraction angles of such high-frequency optical noise are relatively high and can therefore be filtered out without affecting the fundamental orders. However, parasitic phase structures on the phase (wavefront) are more tricky, since they are imprinted over the wavefront propagating through the element, and thus perturbate this wavefront, especially when this element is to be used as a wavefront corrector (null CGH) or in an interferometric device.

Figure 15.43 shows the optical reconstructions of off-axis Fourier pattern generators fabricated as binary and four phase level phase elements. The center reconstruction is affected by a combination of alignment errors and etch depth inaccuracies, which yields light in the conjugate fundamental order. Such light should not appear for high near perfect four-level fabrication (see the right-hand side reconstruction). In the left-hand reconstruction, the element is binary, and therefore might only have etch depth errors (and potentially other errors linked to the lithographic tool used), since no field alignment is required (only one masking field layer). The apparition of light in the conjugate fundamental order can be explained by both the etch depth and the alignment errors. In addition, the etch depth errors produce stronger zero order.

In order to decipher whether the zero order light arises from either etch depth errors or field alignment errors, we perform the following test: a set of binary and four-level elements as depicted before, but this time with perfect etch depth accuracy (see Figure 15.44). Therefore, the only fabrication error that remains is the field alignment in the center reconstruction (poor field-to-field alignment). As one can see in

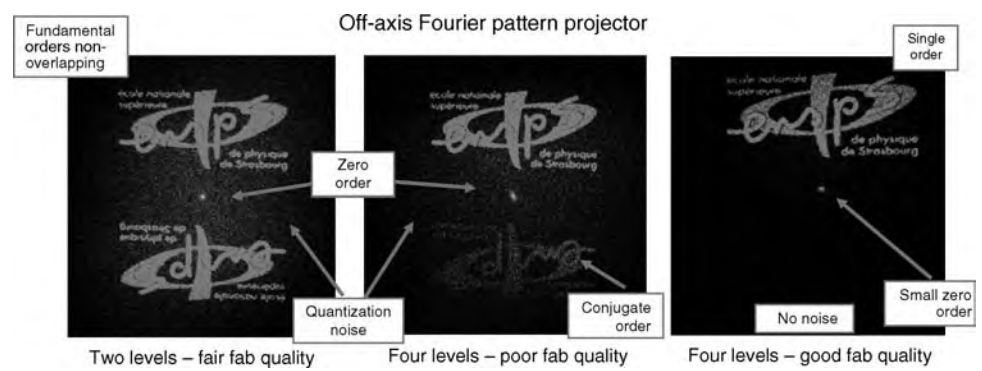


**Figure 15.43** Optical results for various etch depth and field alignment errors on four phase level elements

Figure 15.44, the field alignment inaccuracies produce the apparition of light in the conjugate order, but no zero order.

The effects of field-to-field misalignments are more complex to predict than the effects of etch depth errors, which are mainly efficiency related. We have seen in Figure 15.43 and 15.44 that field misalignments can create produce light in the fundamental conjugate order for Fourier elements, which is not easy to grasp at first. Field alignments also create higher-order optical noise, quantization noise, ghost images or functionalities and so on, and are also very different for Fresnel and Fourier elements.

In Fresnel elements such as diffractive lenses, field alignment errors can produce a wide variety of effects, such as aberrations in the lens (especially coma), and can also create multiple focus along the optical axis of the lens. If the field-to-field alignment errors are severe, an off-axis beam with weak efficiency can be created, or even multiple off-axis beams.

Figure 15.45 shows alignment errors in an eight phase level element. The confocal interferometric scans are compared using an optical microscope in order to see if the spikes are artifacts of the measurement tools or if they really exist.

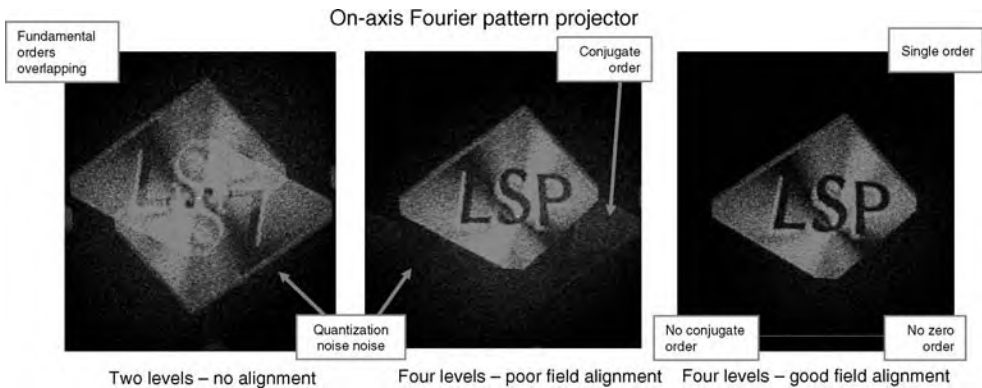


**Figure 15.44** Optical results for good and poor field alignment on four phase level elements with perfect etch depths

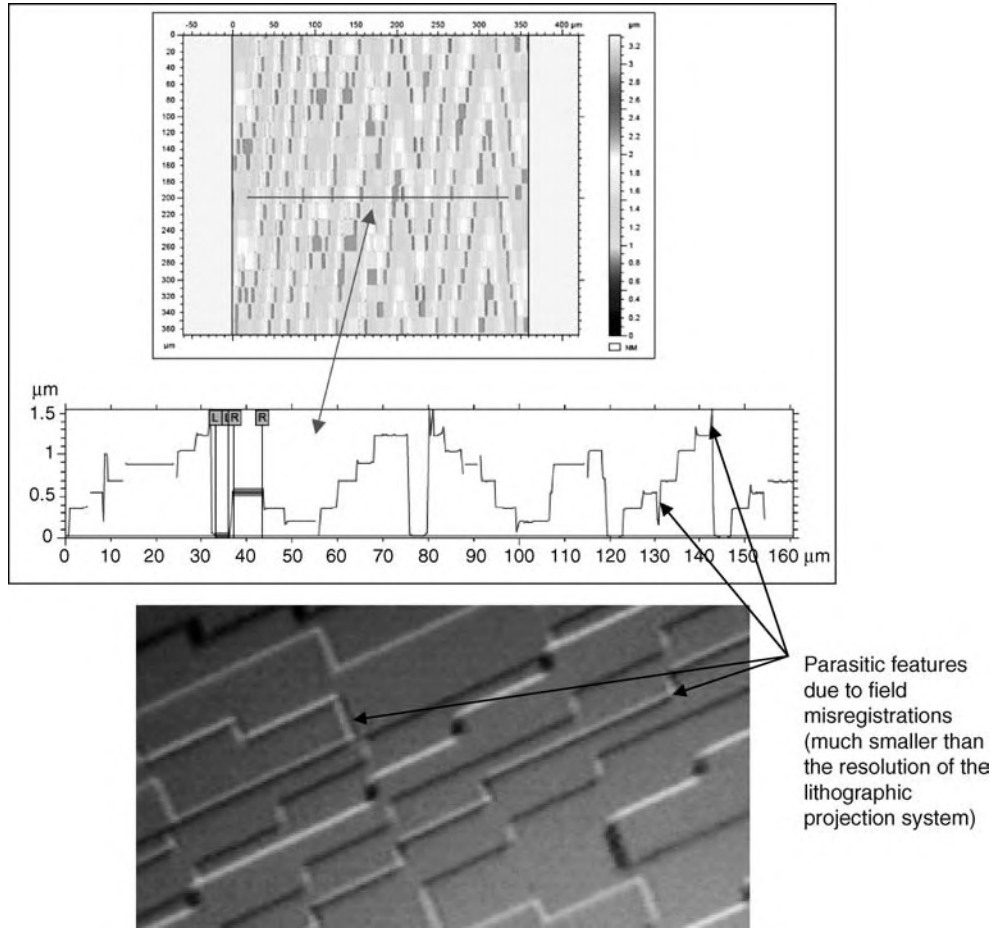


**Figure 15.45** Confocal interferometric microscope plots and an optical microscope photograph of an eight phase level CGH with alignment errors

As an example of modeling multiple field misalignments and resulting effects on the final multilevel element, an eight phase level  $3 \times 3$  fan-out grating is considered, designed by the IFTA algorithm described in Chapter 6. The resulting eight phase level element and the three mask sets are shown in Figure 15.46.

Note how the features are reduced in size when the mask index increases. It is therefore important to actually reduce the alignment errors when aligning higher and higher mask sets. If the alignment tolerances between the first and second mask are  $\pm\delta$ , the alignment between the first and third should be  $\pm\delta/2$  (see Figure 15.47). Note that one does not align two higher-index fields to each other, but always a higher field to the fundamental field (first mask). This is also shown in the alignment keys presented in Figure 15.34, where the first field had both positive alignment keys, and the higher indices 2 and 3 had only one set of negative alignment keys. The super-gratings (rounded grating fringes) that appear in both directions create noise in these directions, and therefore reduce the efficiency as well as the SNR.

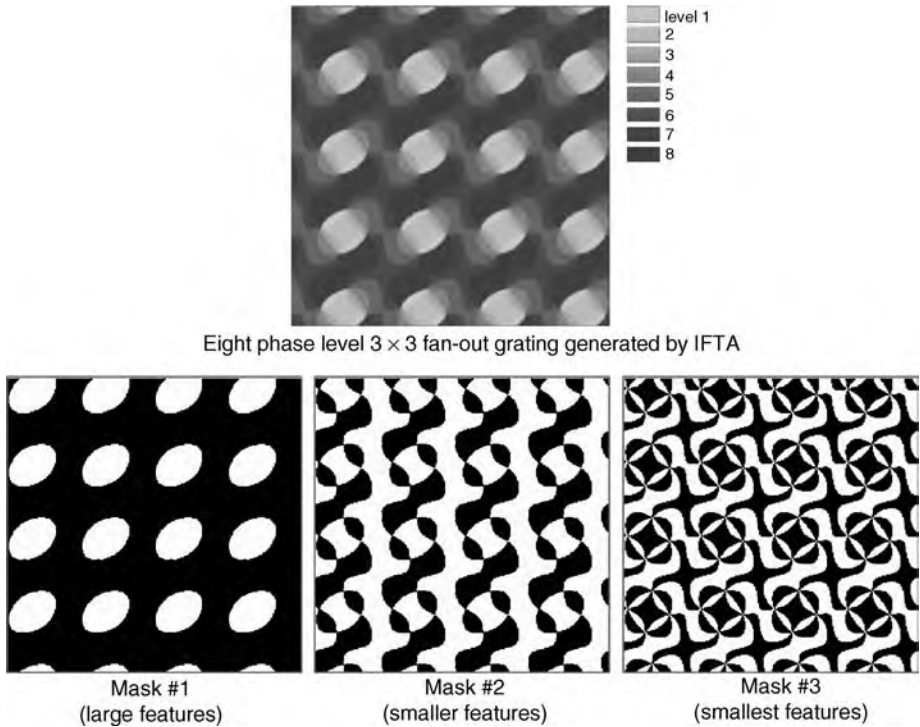


**Figure 15.46** The eight phase level element and mask set for a  $3 \times 3$  fan-out grating

15.4.5 *Effects of Edge Rounding*

Effects of edge rounding have already been described in Section 13.2.2. The edge rounding effect is mainly due to the limitations of the resolution of the optical projection tool. This is very important for IC-type elements (transistor gates etc.), but less important for digital optics. Optical proximity effects can be compensated by either rule-of-thumb techniques (adding serifs and scatterer bars) or by an iterative compensation algorithm that is directly inserted in the digital optics design process (see Chapter 13).

15.4.6 *E-beam and Optical Proximity Effects*

Chapter 13 has also analyzed the effects of e-beam proximity effects in e-beam resists on the optical performance of the final element. Such direct write proximity effects in resist can also be compensated by an adapted algorithm.

15.4.7 *Other Negative Effects*

Many other negative effects can be analyzed by using a simple optical microscope. Resist lift-off and other effects are depicted in Figure 15.48.

Resist lift-off can be produced by either a lack of priming of uneven resist thicknesses due to dust particles on the wafer surface prior to resist spin coating, or nonuniform resist exposures and development. Once the resist structures are lifted off, the remaining structures (or lack of structures) can be transferred into the underlying substrate by ion-beam etching, as seen in the example in Figure 15.48. Thus it is important to detect such resist lift-offs right after development.

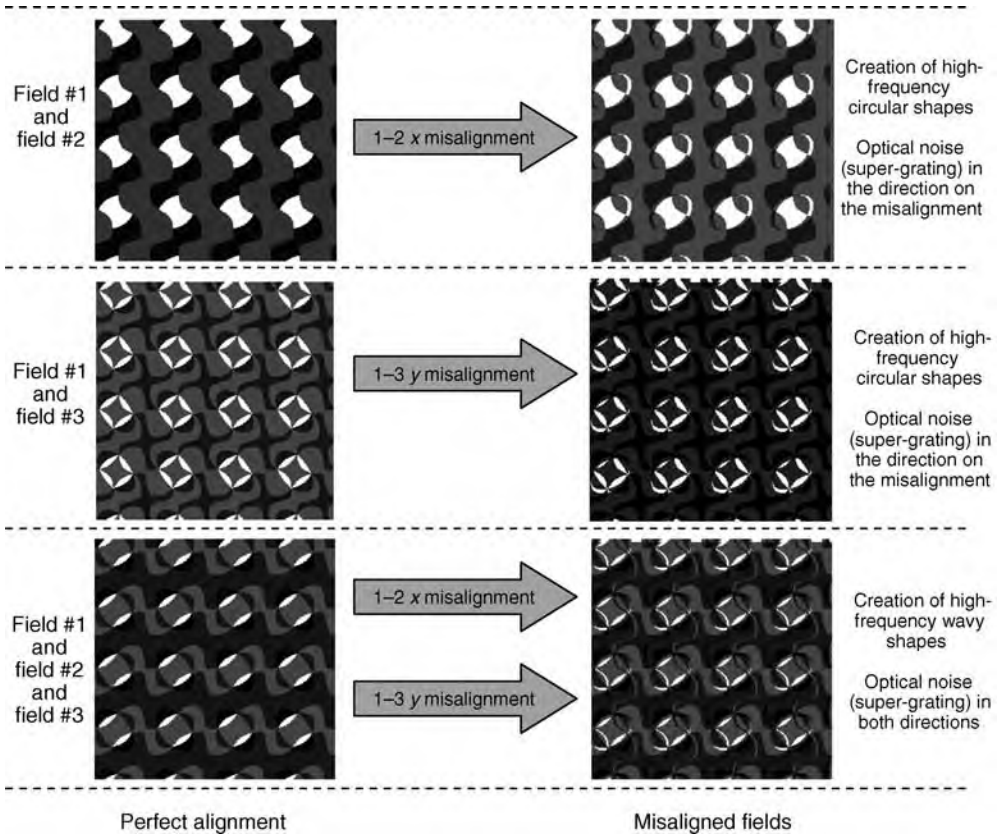


**Figure 15.47** Field misregistrations between first-order and higher-order fields

### 15.4.8 *Relaxing the Fabrication Specs to Increase the Optical Functionality Can Be Good*

Finally, on a positive note, negative effects in optical lithography can produce positive effects on the optical reconstruction (end user functionality, which is what is important after all).

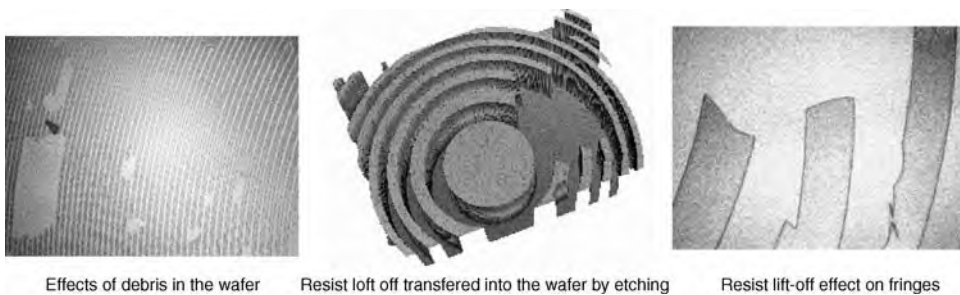


**Figure 15.48** Resist lift-off and effects of debris and dust on the resist, transferred into the wafer after etching



The rounding effects of the CGH cells described in Chapter 13 can have beneficial effects as they reduce the high-frequency noise present in the CGH. This is particularly true when using a lithographic tool that can resolve much smaller features than the ones to be resolved for the current job. For example, a CGH with a  $10\mu\text{m}$  cell size for YAG laser operation will provide better results if the lithography tool has a cutoff frequency around  $5\mu\text{m}$ , rather than if the tool has a cutoff frequency around  $0.5\mu\text{m}$ . This is true since the CGH is approximated in its design (see Chapter 6) by square or rectangular cells. The cells in the design process are actually not square or rectangular; they have no dimension at all. They are positioned on a square or rectangular grid, and it is that grid that dictates their geometry and size. Such effects can be easily modeled by using a scalar propagator and the oversampling and embedding factors described in Chapter 11. A simple reconstruction via an FFT propagator (within an IFTA design iteration, for example) does not take into consideration the geometry of the basic CGH cell.

The rounding of the cell edges thus has two consecutive positive effects:

- the reduction of high-frequency noise, without compromising the optical reconstruction in the fundamental negative order; and
- reduction of the cost of the lithography, since the resolution of the projection system is intentionally decreased.

However, both the field-to-field alignment accuracy and the etch depths have to be controlled very accurately, and only the resolution can be reduced.

This chapter has reviewed the various specification lists that one has to provide to a fab when operating in a fabless mode, and the various techniques used to assess the quality of the fabrication after the wafer is sent back.

## References

- [1] J.P. Allebach, 'Representation related errors in binary digital holograms: a unified analysis', *Applied Optics*, **20**(2), 1981, 290–299.
- [2] M.B. Stern, M. Holz, S.S. Medeiros and R.E. Knowlden, 'Fabricating binary optics: process variables critical to optical efficiency', *Journal of Vacuum Science Technology*, **B9**, 1991, 3117–3121.
- [3] B.K. Jennison and J.P. Allebach, 'Analysis of the leakage from computer-generated holograms synthesized by Direct Binary Search', *Journal of the Optical Society of America A*, **6**(2), 1989, 234–243.
- [4] W. Krug, J. Rienitz and G. Schulz, '*Contributions to Interference Microscopy*', Adam Hilger, Bristol, 1967.
- [5] W. Beyer, '*Interferenzmikroskopie*', Ambrosius Barth, Leipzig, 1974.
- [6] P. Caber, 'Interferometric profiler for rough surfaces', *Applied Optics*, **32**, 1993, 3438–3441.
- [7] A. Pförtner and J. Schwider, 'The dispersion error in white-light-Linnik interferometers and its implications for the evaluation procedures', *Applied Optics*, **40**(34), 2001, 6223–6228.
- [8] J. Schmit and A. Olszak, 'High-precision shape measurement by white-light interferometry with real-time scanner error correction', *Applied Optics*, **41**(28), 2002, 5943–5950.
- [9] P. deGroot and L. Deck, 'Three-dimensional imaging by sub-Nyquist sampling of white light interferograms', *Optics Letters*, **18**, 1993, 1462–1464.
- [10] L. Deck and P. deGroot, 'High speed noncontact profiler based on scanning white light interferometry', *Applied Optics*, **33**, 1994, 7334–7338.
- [11] P. deGroot and L. Deck, 'Surface profiling by analysis of white-light interferograms in the spatial frequency domain', *Journal of Modern Optics*, **42**, 1995, 389–401.
- [12] Honeywell Technology Center, ARPA-sponsored CO-OP DOE Foundry Run 'Specifications on process errors and analysis', 1995.
- [13] J.M. Miller, M.R. Taghizadeh, J. Turunen and N. Ross, 'Multilevel-grating array generators: fabrication error analysis and experiments', *Applied Optics*, **32**(14), 1993, 2519–2525.

- [14] J.A. Cox, B.S. Fritz and T.R. Werner, 'Process-dependent kinoform performances', in '*Holographic Optics III: Principles and Applications*', G.M. Morris (ed.), Proc. SPIE Vol. 1507, 1991, 100–109.
- [15] J.A. Cox, B.S. Fritz and T.R. Werner, 'Process error limitations on binary optics performances', in '*Computer and Optically Generated Holographic Optics; 4th in a Series*', I. Cindrich and S.H. Lee (eds), Proc. SPIE Vol. 1555, 1991, 80–88.
- [16] M.W. Farn and J.W. Goodman, 'Effects of VLSI fabrication errors on kinoform efficiency', in '*Computer and Optically Formed Holographic Optics*', I. Cindrich and S.H. Lee (eds), Proc. SPIE Vol. 1211, 1991, 125–136.
- [17] D.W. Ricks, 'Scattering from diffractive optics', in '*Diffractive and Miniaturized Optics*', S.H. Lee (ed.), Proc. SPIE Vol. CR49, 1993, 187–211.
- [18] P.D. Hillman, 'How manufacturing errors in binary optic arrays affect far field pattern', in '*Micro-optics/ Micromechanics and Laser Scanning and Shaping*', M.E. Motamedi and L. Beiser (eds), Proc. SPIE Vol. 2383, 1992, 298–308.
- [19] T.R. Jay, M.B. Stern and R.E. Knowlden, 'Effect of microlens array fabrication parameters on optical quality', in '*Miniature and Micro-optics: Fabrication and System Applications II*', C. Roychoudhuri and W.B. Veldkamp (eds), Proc. SPIE Vol. 1751, 1993, 236–245.
- [20] H. Andersson, M. Ekberg, S. Hård *et al.*, 'Single photomask multilevel kinoforms in quartz and photoresist: manufacture and evaluation', *Applied Optics*, **29**, 1990, 4259–4267.
- [21] R.W. Gruhlke, K. Kanzler, L. Giammona and C. Langhorn, 'Diffractive optics for industrial lasers: effects of fabrication error', in '*Miniature and Micro-optics: Fabrication and System Applications II*', C. Roychoudhuri and W.B. Veldkamp (eds), Proc. SPIE Vol. 1751, 1993, 118–127.
- [22] S.M. Shank, F.T. Chen and M. Skvarla, 'Fabrication of multi-level phase gratings using focused ion beam milling and electron beam lithography', *Optical Society of America Technical Digest*, **11**, 1994, 302–306.

# Digital Optics Application Pools

When a new technology becomes integrated into consumer electronic devices, the industry generally agrees that this technology has entered the realm of mainstream technology. This has a double-edged sword effect: first, the technology becomes democratized and thus massively developed, but then it also becomes a commodity, and thus there is tremendous pressure to cut down the production and integration costs without sacrificing any performance.

Such a leap from high-technology research to mainstream industry can only be achieved if this technology has been developed through extensive and expensive academic and governmental projects, such as the Defense Advanced Research Projects Agency (DARPA), the National Science Foundation (NSF) and the Small Business Innovation Research (SBIR) program.

Digital diffractive optics was introduced into research and academia as early as the 1960s, and the first industrial applications were seen as early as the 1960s (especially in spectroscopic applications and optical data storage). However, digital optics have long suffered from a lack of adequate fabrication tools, until the recent development of IC fabrication tools and replication technologies (CD replication and embossing), and thus have spent a long time as high tech curiosities, with no real industrial applications [1].

Digital holograms have been around a little longer; however, they have suffered mainly from materials problems (costs, MTBF, volume replication etc.). This is why holographic optics have taken a relatively long time leap from military applications and museum displays, where costs are usually not a major factor (holographic HUDs, 3D holographic displays etc.), down to industrial markets [2, 3].

Micro-refractives have been introduced to consumer products more recently, especially with the ever-growing need for better LCD displays and digital cameras.

Digital waveguide optics (integrated PLCs) had a tremendous boost during the optical telecom heyday (1998–2002), before the bubble burst in 2003. However, this boost was sufficient to thrust digital PLCs into mainstream telecom applications, which are now integrated everywhere from CATV to DWDM to 10 Gb/s Ethernet.

Although the realm of photonic crystals is no longer in its infancy, it still remains a research challenge and it has very few industrial applications today – although there is great potential for the future. This is mainly due to the lack of available fabrication tools, with the notable exception of photonic crystal fibers, which are less challenging to fabricate.

The field of metamaterials has been projected into the public eye quite recently, but will remain a research endeavor for some time (and especially a defense and security research effort), before entering the mainstream. Here also, the lack of adequate fabrication techniques as well as characterization methods is slowing its expansion.

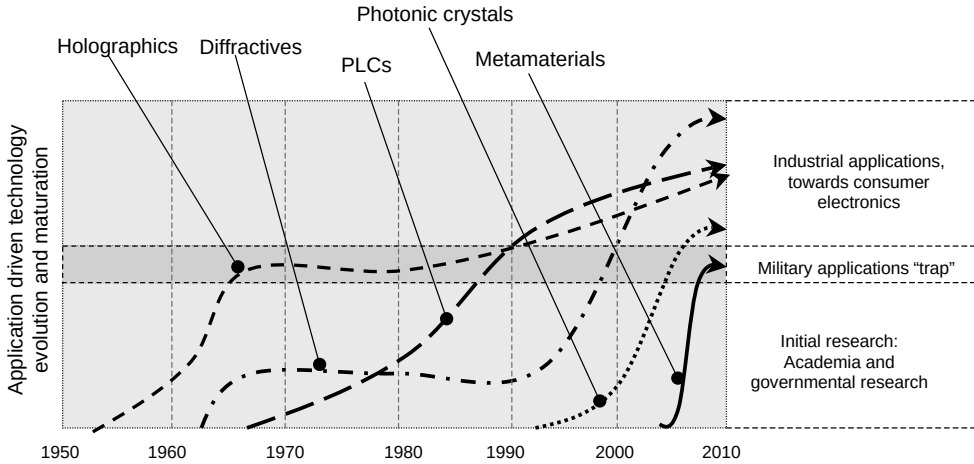


**Figure 16.1** The transfer from research to industry for the various digital optics realms

Figure 16.1 shows these five areas of digital optics and how they have been transferred from research into industrial applications, through the military gap. This figure shows that holographics may have been dominated by military and security applications for nearly three decades, before starting to reach the industrial sector (other than as spectroscopic applications and security holograms). Recent breakthroughs in holography are in clean tech (solar cells concentrators), telecom applications, LCD screen displays and optical page data storage.

Diffractives had a long infancy period in research and academia, mainly due to the lack of adequate fabrication tools (IC fabs). However, now that this has been solved and cheap plastic replication techniques have been proposed, they have been pushing through the military trap and into the industrial sector much faster than holographics.

It is interesting to note that PLCs have had a head start in industry after a fast germination period in academia, especially due to the optical telecom bubble of the late 1990s and early 2000s.

Today, photonic crystals have taken on the role that diffractives had about 30 years ago, and metamaterials have taken the same route as holographics did 40 years ago.

## 16.1 Heavy Industry

Heavy industry is a long-time user of diffractive optics such as gratings and diffractive lenses, in reflection or transmission, for low-power visible lasers or high-power IR lasers. This section reviews some of the related applications, in a chronological way, beginning with spectroscopic gratings, nondestructive testing and metrology, down to laser material processing, lighting and finally automotive applications.

### 16.1.1 Spectroscopy

References: Chapters 5 and 6

Historically, spectroscopy was the first application realm for diffractives, and more precisely for reflective linear and curved ruled gratings (for more details, see Chapter 5). For many people, including optical

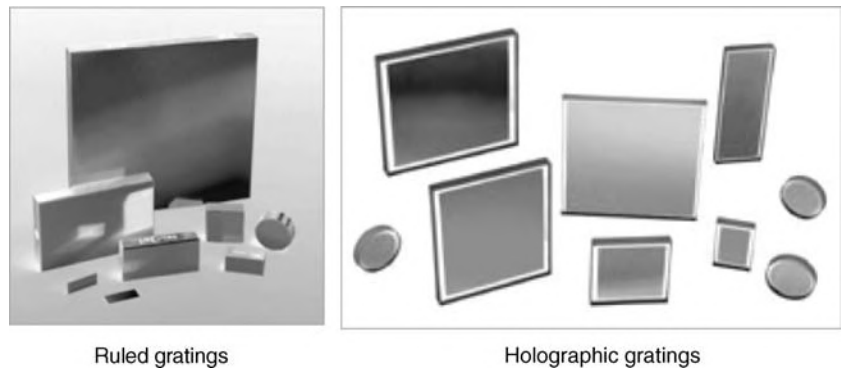


**Figure 16.2** Ruled and holographically generated spectroscopic gratings

engineers, spectroscopic gratings remain the one and only application of diffractive optics today, a notion that needs to be reconsidered seriously, as we will see throughout this chapter.

Figure 16.2 shows two physical implementations of spectroscopic gratings: on the left-hand side diamond-ruled gratings and on the right-hand side holographically exposed gratings in photoresist (with subsequent reflective gold coating).

Such reflective gratings can be tuned in order to obtain the optimal efficiency over a broad wavelength range. Figure 16.3 shows a typical free-space reflective grating spectrometer.

For more insights on the spectral dispersion characteristics of reflective and transmission gratings, see Chapter 5.

*16.1.2 Industrial Metrology*

Industrial metrology includes holographic nondestructive testing (a very early application of holography), 3D sensing through structured laser illumination and digital holography.

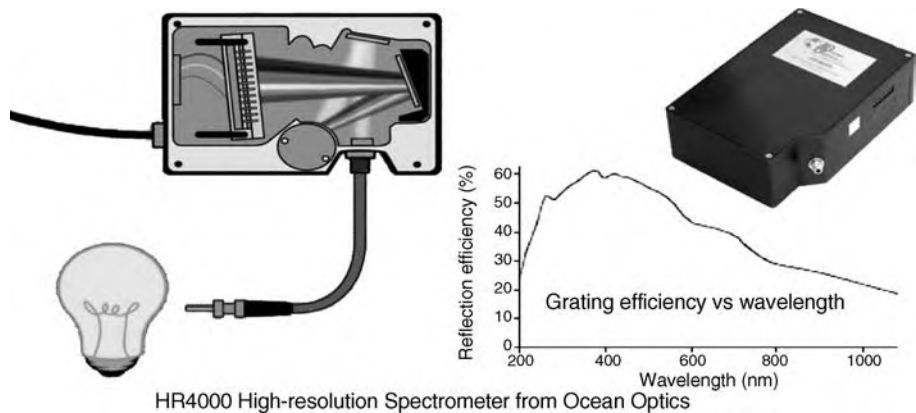


**Figure 16.3** A reflective grating spectrometer ([www.oceanoptics.com](http://www.oceanoptics.com))

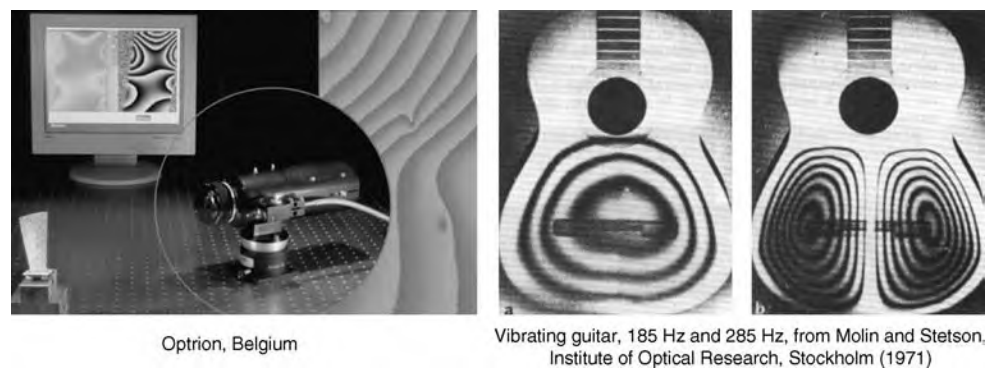


**Figure 16.4** Holographic nondestructive testing

**16.1.2.1 Holographic Nondestructive Testing**

Reference: Chapter 8

The principles of holographic nondestructive testing [4] have already been described in Chapter 8. Figure 16.4 shows an industrial testing apparatus based on holographic nondestructive testing, as well as some of the first tests. Most of the applications today are linked to vibration mode analysis in automotive and avionics, as well as stress analysis in various mechanical components.

**16.1.2.2 Three-dimensional Sensing through Structured Illumination**

Reference: Chapter 6

Diffraction optics are being used increasingly in infrared image sensors for remote 3D sensing in automotive and factory automation. A CGH is designed to project in the far field a set of geometrical shapes or grids through a visible or IR laser diode. This scene is then captured by a digital camera. A software algorithm analyzes the deformation of the projected geometrical patterns and computes back the 3D topology of the scene. Figure 16.5 shows such a fringe projection scheme and the automotive section will show some more applications. The sets of horizontal fringes in Figure 16.5 are projected one set after

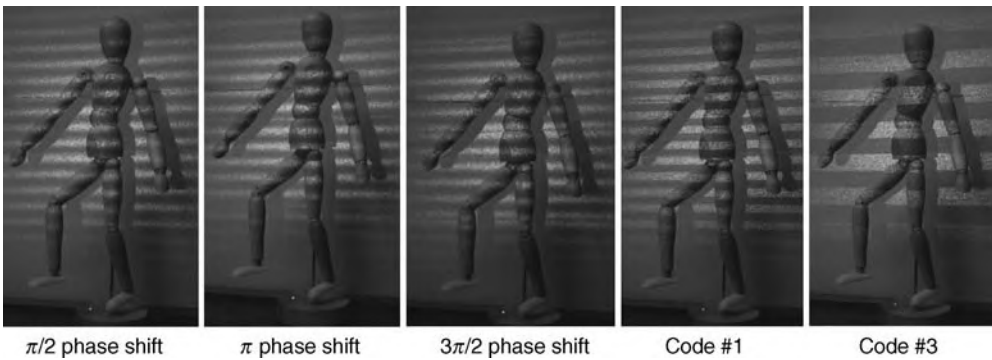


**Figure 16.5** Diffraction fringe projections for 3D shape analysis

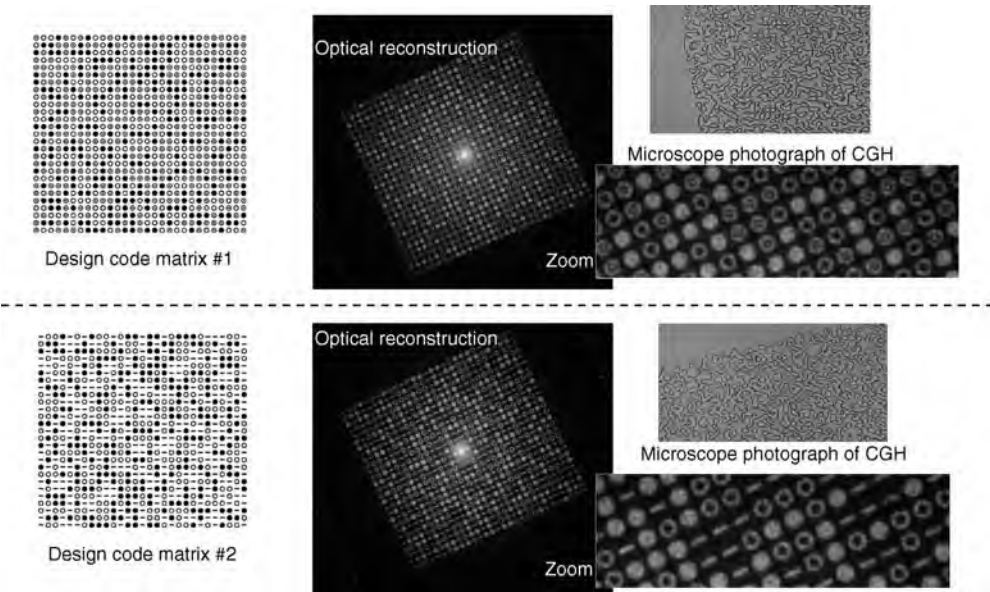


**Figure 16.6** Nonrepetitive code projection via diffractive elements, for 3D shape extraction and absolute positioning

the other. The first four sets are sinusoidal fringe patterns with phase shifts of  $\pi/2$  between each set, followed by sets of binary Gray code fringes.

Figure 16.6 shows some other diffractive projected patterns, which bear more information than the previous linear fringes. These 2D codes includes absolute position information. For example, on top of the 3D shape extraction, one can locate precisely the position of that shape in 2D (or even 3D). Such structured illuminations can be used for large objects in free space, or at the end of an endoscopic device for military or medical *in-vivo* 3D imaging applications.

**16.1.2.3 Digital Holography for 3D Shape Acquisition**

Reference: Chapter 8

Digital holography for 3D sensing has been described in Chapter 8. It is a method that relies heavily on numeric computation and high-resolution image sensors, and therefore could not be introduced into industry until just recently, when these technologies became available, although the principles of digital holography have been known since the dawn of holography. Digital holography is a powerful tool for the remote measurement of 3D shapes by recording a hologram on a high-resolution CMOS sensor and back-propagating the field using a numeric algorithm, similar to those described in Chapter 11. There are considerable potential applications of this technique for industrial 3D sensing and robot 3D vision. This technique can be applied to amplitude objects as well as phase objects (similar to Fourier tomography).

**16.1.3 Industrial Laser Material Processing**

Laser cutting, laser welding, laser marking, laser engraving and laser surface treatment are some of the functionalities used in the laser material processing industry, a market that has been growing at a steady rate for more than two decades. Laser beam shaping for high-power CO<sub>2</sub> and YAG lasers are now common

elements, and are usually fabricated by diamond turning in substrates that are transparent at a wavelength of  $10\text{ }\mu\text{m}$ , such as ZnSe, ZnS or Ge.

More complex CGHs, fabricated as reflective elements etched in quartz with a reflective gold coating, can incorporate specific beam engineering functions (such as complex logo engraving or accurate intensity redistribution for welding applications). These are promising solutions for fast but accurate laser material processing tasks, without having any moving parts in the optical path (no moving mirrors, shutters etc.).

#### 16.1.3.1 Laser Mode Selection

Digital optics (phase plates or diffractive optics) can be inserted directly into a laser cavity in order to discriminate between the various modes that can actually get amplified in the laser cavity. It is thus possible to trigger a specific mode; for example, a fundamental  $\text{TEM}_{00}$  mode.

#### 16.1.3.2 Amplitude Mode Selection

High-power lasers (YAG,  $\text{CO}_2$  or excimer lasers) usually produce a large number of modes, and therefore reduce the quality of the focused spot used for welding, cutting or marking. High-power lasers were early adopters of diffractive optics for amplitude mode selection devices. Such diffractive mode selectors are usually placed inside the laser cavity, and filter out the unwanted modes, limiting the output of the laser to either the fundamental  $\text{TEM}_{00}$  mode or other fundamental modes, excluding higher-order modes that could cause the Strehl ratio of the focused spot to deteriorate.

#### 16.1.3.3 Polarization Mode Selection

In a similar way to amplitude mode selectors, polarization mode selectors can be implemented as circular gratings etched on top of VCSEL lasers [5]. Such polarization mode selectors can force the laser to output a specific polarization state, if the application is highly polarization sensitive (such as display applications using polarization selective microdisplays, like HTPP or LCoS devices).

#### 16.1.3.4 Beam Delivery

##### *Beam Samplers*

References: Chapters 5 and 6

Beam samplers are diffractive gratings or off-axis diffractive lenses with a diffraction efficiency that has intentionally been kept very low (e.g. a few percent). The main signal propagates through the diffractive without noticing anything. A few percent of the signal is diffracted onto either a detector (a high-power beam sampler) or a fiber (a fiber tap), or to monitor laser power for datacom applications (in closed loop operation).

##### *IR Hybrid Lenses*

Reference: Chapter 7

Thermal IR optics are very expensive elements, and are used for industrial and military applications. Fabricating surface-relief aspheric profiles in IR materials such as Ge, ZnS or ZnSe is a difficult and expensive task. Off-axis reflective optics are also difficult to produce. Diffractive and hybrid refractive/diffractive optics (see Chapter 7) are a good alternative to reduce the number of optical elements required for a specific focusing or beam-shaping task, and to reduce the production costs at the same time.

#### 16.1.3.5 Laser Material Processing

References: Chapters 5–7

Laser material processing is a traditional application pool for diffractive optics.



***Laser Marking/Engraving***

Reference: Chapter 6

Diffractive pattern generators are useful elements when the task is laser marking of repetitive structures into a workpiece. Such elements can work both as far-field or near-field elements. In a case in which a Fourier element is being considered, an additional focusing lens should be used. If a Fresnel element is being used, nothing else is required, and the element is usually referred to as an infrared ‘focuser’. Such ‘focusors’ can focus an incoming high-power YAG or CO<sub>2</sub> laser beam directly into the pattern to be engraved, without having to move either the lens or the workpiece. They are usually designed as interferogram-type diffractive lenses (see Section 5.5).

***Laser Surface Treatment***

Reference: Chapter 6

Laser surface treatment and laser heat processing can be implemented in the same way as the previous engraving applications; however, here the element behaves more like an anisotropic diffuser rather than as a ‘focuser’ (i.e. as a Fourier element rather than a Fresnel element). Here, the intensity profile of the projected pattern has to be tightly controlled in order to produce the desired surface heat treatment effect, whereas in focusors the intensity is mainly binary along the engraving or cutting line.

***Laser Cutting and Welding***

References: Chapters 6 and 7

Similarly to laser engraving, laser cutting and laser welding are tailored jobs for diffractive optics. Such digital diffractive optical elements can shape the beam into the tool required for the cutting or welding tasks (including beam shaping, beam homogenizing and beam focusing in the same element).

***Chip-on-board Integration***

References: Chapters 5 and 6

Direct chip-on-board integration usually requires a lot of hole drilling and subsequent soldering of the multiple legs of each chip onto the PCB board. Parallel soldering can be performed by the use of fan-out digital optical elements such as Fresnel CGHs. The soldering of several hundred points on a PCB can thus be carried out in a single step, without moving the laser or the PCB.

Figure 16.7 shows some examples of the implementation of digital diffractive optics for engraving, surface treatment, beam shaping or chip-on-board integration.

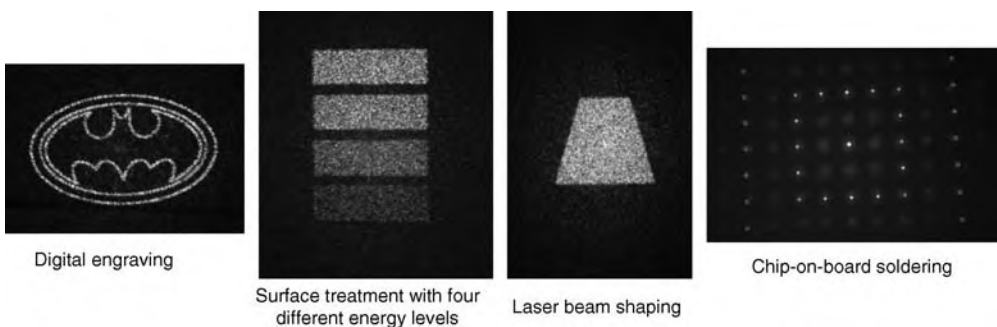


**Figure 16.7** Example of laser material processing tasks taken on by digital optics



**Figure 16.8** Holographic wrapping paper replicated by roll embossing in Mylar™

### *16.1.4 Industrial Packaging*

References: Chapters 5 and 14

#### **16.1.4.1 Diffractive Wrapping Paper**

The holographic wrapping paper industry (e.g. Christmas wrapping paper) is a large user of roll-embossed diffractive gratings (see Figure 16.8). Such gratings can be quite exotic, and can have all kinds of curvatures and chirps in order to produce a specific visual effect, or eye candy (similar to the OVID elements in optical security; see Section 16.2.3.5).

#### **16.1.4.2 Holographic Tags**

References: Chapters 8 and 14

Holographic tags and holographic bar codes are part of the packaging industry but, instead, we will review them in the optical security section below, since their main task is to authenticate the packaging material of the product (or the product itself) and therefore reduce potential counterfeiting.

### *16.1.5 The Semiconductor Fabrication Industry*

The semiconductor fabrication industry is not only a provider of fabrication tools for digital optics, but it is also a large user of digital optics.

#### **16.1.5.1 Phase-shift Masks**

References: Chapters 10, 11 and 13

Alternating phase-shift masks and attenuating phase-shift masks are sub-wavelength diffractive optics encoding complex amplitude and phase information in a reticle or photomask. They are carefully optimized by either rule-of-thumb techniques or vector electromagnetic modeling tools in order to yield smaller and smaller features (CD) on the wafer. For more details, see Section 13.2.6.

#### **16.1.5.2 Complex Illumination in Steppers**

References: Chapters 12 and 13

The off-axis and complex illumination used to reduce the  $k_1$  resolution factor in optical projection lithography can be achieved by the use of Fourier beam-shaper CGHs. For more details of this technique, see Section 13.3.2. This is a fast-growing market for digital optics today.

### **16.1.5.3 *In-situ* Etch Depth Monitoring**

References: Chapters 12 and 13

The *in-situ* real-time monitoring of the etching process in a plasma chamber is a very desirable feature, since a high etch depth accuracy is required to produce a high diffraction efficiency in digital optics. Such a monitoring technique can be implemented by launching a laser beam onto PCMs, which can be composed of simple gratings, by measuring the reflected zero-order intensity. The amount of light in the zero order provides information about the depth of the grating structure (for an analytic analysis of the diffraction efficiency versus phase shift or etch depth, see Chapter 5).

### **16.1.5.4 Maskless Imaging via Microlens Arrays**

Reference: Chapter 13

Chapter 13 has described maskless lithography, which uses arrays of high-NA diffractive microlens arrays.

### **16.1.5.5 Beam Shaping in Direct Laser Write**

Reference: Chapter 12

When using a laser beam writer for mask patterning or direct writing on photoresist, it is often the case that a CGH Fourier beam shaper is used in between the beam splitter and the objective lens, in order to ‘write’ the structures with an optimal beam footprint (e.g. a uniform square beam ‘stamp’). This technique not only reduces proximity effects in the resist, but also speeds up the patterning process.

Beam splitters are also used to produce multiple beams that can write patterns in parallel on the photoresist, each of them modulated individually by an acousto-optic modulator or a MEMS mirror deflector.

### ***Direct Laser Ablation***

Reference: Chapter 12

The previous technique can also be used in laser ablation patterning systems to shape the beam with which the material is to be ablated, to reduce ablation proximity effects and produce a cleaner and more uniform ablated surface: see, for example, the Laser Beam Writer (LBW) and Laser Beam Ablation (LBA) machines from Heidelberg Systems.

## **16.1.6 *Solid State Lighting***

### **16.1.6.1 Photonic Crystal LEDs (PC-LEDs)**

Reference: Chapter 10

In an LED, a thin planar slab serves as a waveguide. At some frequencies, spontaneous emitted light can be coupled into the waveguide. A photonic crystal (PC) on top of an LED can produce optical band gaps that prevent spontaneous emitted light from coupling into the waveguide, therefore enhancing the efficiency of the solid state light source (see left side of Figure 16.9).

### **16.1.6.2 LED Beam Extraction**

Reference: Chapter 10

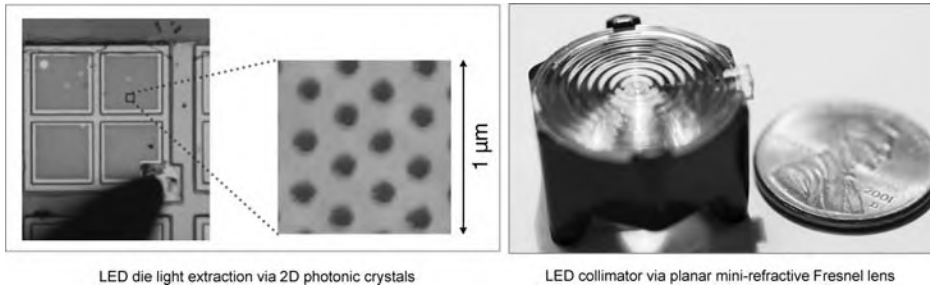


**Figure 16.9** Secondary LED optics as a planar Fresnel lens

Other diffractives can be etched on top of the LED window in order to extract light that would otherwise be reflected by TIR. Diffractive Fresnel lenses and other beam-shaping diffractives can be mounted on top of the LED device in order to collimate the light and produce a more uniform beam of light.

### 16.1.6.3 LED Beam Collimation and Beam Shaping

References: Chapters 5 and 6

LED beam collimation and beam shaping can be performed by either micro-refractive profiles (Fresnel lenses) or diffractive beam shapers as secondary optics on an LED die (see Figure 16.9).

## 16.1.7 Automotive Applications

The automotive sector has recently become a new user of micro-optics and digital optics, for external lighting applications, sensors, displays, panel controls and even internal lighting.

### 16.1.7.1 HUD Systems

References: Chapters 5–10

Head-Up Display (HUD) systems have been applied in avionics for more than 40 years, through the use of holographic or dichroic combiner optics. The optical combiners combine a digital display with the field of view, and often produce a virtual image of that display at a position several feet in front of the windshield or cockpit.

The first integration of HUDs into automotive applications used large and cumbersome reflective/catadioptric optics buried under the dashboard (such as for the Chevrolet Corvette, the Pontiac Grand Prix etc.), where the optical combiner was simply the windshield.

New architectures of optical combiners promise to use mass-replicable digital diffractive combiners and digital optical see-through displays, based on dynamic holograms such as edge-lit elements. Figure 16.10 shows an edge-lit HUD produced by BAE Systems, composed of sandwiched waveguide holograms that route the edge-emitted laser beams in both directions, and that extract the light from the slab by Bragg coupling.

### 16.1.7.2 Virtual Optical Interfaces

Reference: Chapter 6

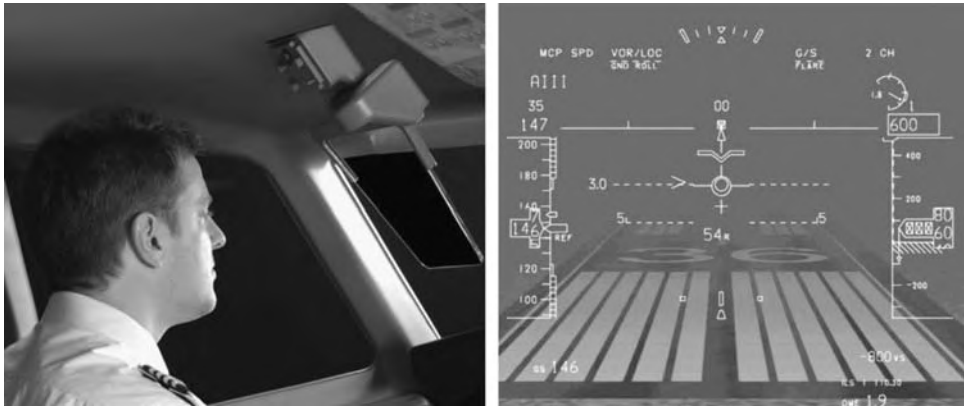


**Figure 16.10** An edge-lit waveguide hologram HUD by BAE Systems

Virtual optical interfaces similar to diffractive keyboard projectors have not yet been implemented in the automotive industry, but promise to be a valuable solution to demanding applications. Such virtual optical human-machine interfaces could also replace costly touch screen interfaces in applications subject to severe environmental conditions and even vandalism, such as ATMs, public booths and so on. In addition to virtual interface projection, there has to be a good 2D or 3D hand position sensor, such as triangulation through IR illumination, the use of fiber cloth or some other optical sensing technique, such as the generation of optical beamlet grids through diffractives.

#### 16.1.7.3 Contact-less Switches

Reference: Chapter 6

The first elements that are most likely to fail in a car are the mechanical switches (signals, wipers, lights, mirrors, windows, AC etc.). Replacing mechanical switches with noncontact optical switches is an on-going effort in the automotive industry, and can provide valuable solutions to this very old problem. Such contact-less switches can be implemented in a similar way to absolute position optical encoders (see Section 16.4.1.1).

#### 16.1.7.4 Three-dimensional Remote Sensing

Reference: Chapter 6

New IR laser automotive sensors use structured laser illumination to sense 3D structures around the car, in order to anticipate collisions or other problems. Such structured illumination can be produced by tiny integrated visible or IR lasers launched onto digital diffractive Fourier pattern generators (see Figure 16.11).

#### 16.1.7.5 Automotive LED Lighting

References: Chapters 6 and 8

In recent years, automotive lighting has been moving away from bulbs toward solid state lighting (LED). Bare LEDs produce visual hot spots, and thus need beam-shaping and beam-diffusing optics to produce illumination distributions that are compatible with today's automotive safety standards. Fourier CGHs can provide such beam homogenizing and isotropic or anisotropic beam diffusion (elliptic beam diffusion

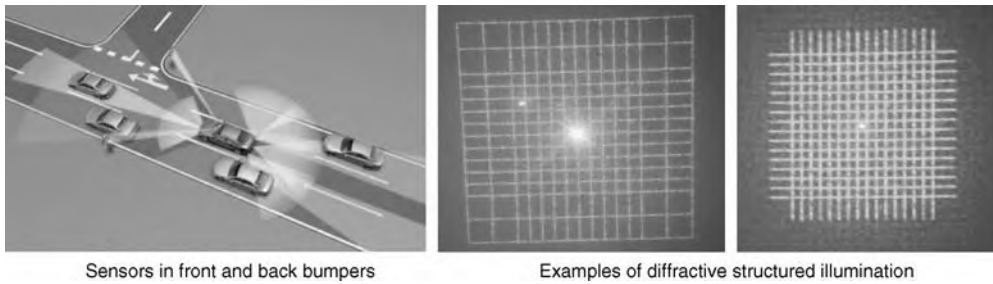


**Figure 16.11** Three-dimensional sensing in automotive applications

with the fast axis in the horizontal direction, for example, or beam shaping into uniform square or circular patches of light).

## 16.2 Defense, Security and Space

### 16.2.1 Weaponry

The military sector, and more specifically weaponry, are usually the first users of any new technology. This has been the case for holographic optics, diffractives, micro-optics and photonic crystals, and will certainly also be the case for metamaterials. The latter may, however, promise too much too quickly for the military sector (see, e.g., Section 10.9.2.2 on optical cloaking).

#### 16.2.1.1 Helmet-mounted and Head-mounted Displays

Reference: Chapters 5, 6 and 8

Helmet-mounted displays (HMDs), wearable displays and near-eye displays are applications that are heavy users of digital optics, especially off-axis holographic elements and sub-wavelength grating structures. An HMD is basically a wearable version of an HUD, which was described in the previous section. Such applications can be integrated in military helmets, or even motorcycling helmets, protective headgear for construction workers or headsets for surgeons.

#### 16.2.1.2 Holographic Gun Sights and Targets

References: Chapters 6 and 8

Holographic gun sight targets have been making use of digital diffractive optics mainly due to the fact that they are Fourier elements, and thus able to focus on any surface, similar to a simple laser pointer pattern generator.

#### 16.2.1.3 IR Missile Optics

References: Chapters 7 and 10

Imaging tasks in the mid-IR for laser-guided or other missile applications are strong application pools for hybrid optics, especially when it comes to athermalizing imaging optics (see Chapter 7). Wavefront coding is another military application for imaging in missiles, for the same reason (athermalization of IR imaging optics in missile heads).

#### 16.2.1.4 Laser Weapons

Reference: Chapter 9

Laser weapons for ‘Star Wars’ type ‘applications’ can make good use of diffractive optics, such as for controlling the propagation of beams in free space over long distances (compensating for divergence, the use of diffractive axicons etc.), or for correcting thermal and other atmospheric turbulences through the use of wavefront sensors coupled to adaptive optics (see the next section).

16.2.2 Astronomy

References: Chapters 4, 7 and 10

Astronomy, as well as the military and homeland security sectors, has been an early application pool for diffractive optics. Astronomy is one of those application sectors where high technology is used without the pressure of low-cost mass production, and without the pressure of development time or cost. Therefore, it has provided an ideal technology push for diffractives for very specific applications.

16.2.2.1 Shack-Hartmann Wavefront Sensors for Adaptive Optics

Reference: Chapter 4

The design of a Shack–Hartmann wavefront sensor based on arrays of micro-refractive or micro-diffractive elements has been discussed in Chapter 4 [4, 6–8]. Such wavefront sensors are used today in adaptive optics applications mainly to compensate for atmospheric turbulences (wavefront distortion). Adaptive optics feedback can be implemented using pistons beneath a primary telescope mirror, or as more complex MEMS devices for imaging tasks in military or biomedical applications, and even in telecom applications (free-space laser communications). Figure 16.12 shows a Shack–Hartmann sensor based on microlens arrays.

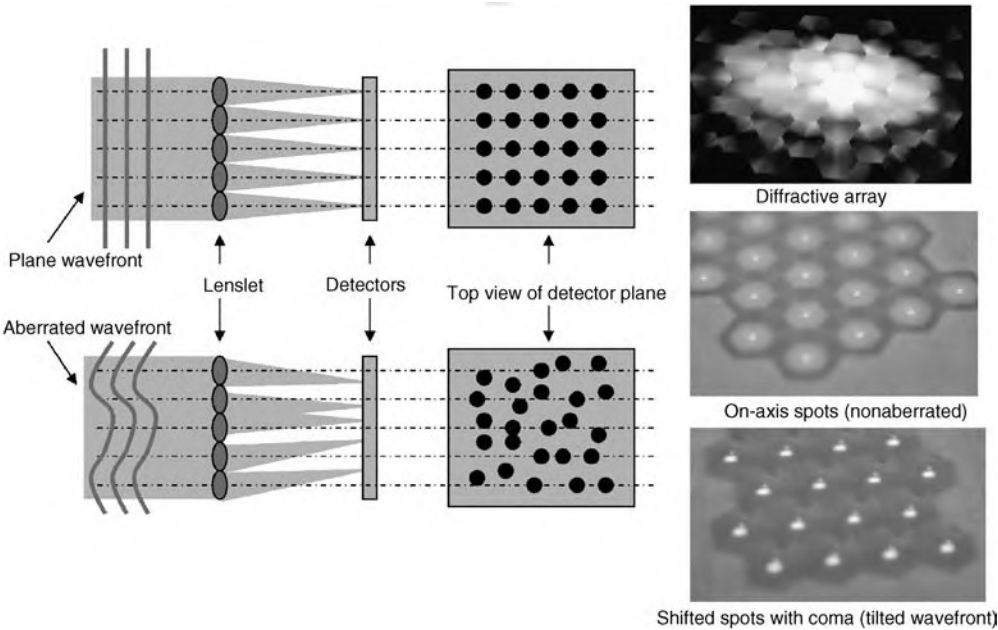


Figure 16.12 A Shack–Hartmann wavefront sensor based on a diffractive microlens array

### 16.2.2.2 Null CGH Lenses

Reference: Chapter 5

Null CGHs are mainly used to test large aspheric surfaces in astronomy. Such lenses are described in Chapter 5. Null CGHs implement a specific aspheric phase profile that impregnates an incoming wavefront generated by a refractive or reflective element (such as a telescope mirror or lens). The resulting wavefront, when interfering with a reference wave, shows fringes the deformation of which relates to the differences between the exact phase profile from the null CGH and the approximate lens or mirror profile from the telescope optics.

### 16.2.2.3 Alignment Elements for Telescope Mirrors

References: Chapters 5 and 6

Alignment elements such as Fourier target projectors can be etched into telescope mirrors in order to align them to other optical elements.

## 16.2.3 Homeland Security

Homeland security is becoming a strong user of digital optics, for both security and sensing applications.

### 16.2.3.1 Biometrics Using Structured Illumination

Reference: Chapter 6

The structured illumination scheme presented in Section 16.1.2.2. can be applied to biometrics. By projecting sets of structured laser illuminations of a face, one can retrieve the 3D facial geometries that make up a person's biometrics. As the projection can be performed by an IR laser, such a sensor can operate on a stealth basis.

### 16.2.3.2 Gas Sensors

Reference: Chapter 5

Gas sensors are making heavy use of diffractive optics for both their high spectral dispersion characteristics and their potential for miniaturization and mass-production (disposable sensors).

### 16.2.3.3 Chemical Sensors

Reference: Chapter 10

Chemical sensors can be implemented as surface plasmon elements, which are described in Chapter 10. Such elements use sub-wavelength metallic gratings patterned on top of a slab waveguide substrate.

### 16.2.3.4 Distributed Smart Camera Networks

References: Chapters 6 and 9

Networks of distributed CMOS cameras can include diffractive structured illumination schemes (in the IR or the visible) to be used as networks of 3D sensors rather than networks of video cameras. Such smart camera networks can be used in either an industrial environment or in security applications, by monitoring transportation flows (tracking specific 3D shapes). The processing (3D contour extraction) can be performed on board the CMOS sensor and thus such sensor arrays do not require any large data bandwidth,





**Figure 16.13** Security holograms

since they are not used for their imaging capability but, rather, for their special sensing functionalities. The window for such a sensor can comprise both the objective lens for the CMOS sensor as well as the various diffractive elements producing the structured illumination, thus reducing the overall size and price of the integrated optical camera sensor.

### 16.2.3.5 Optical Security

#### *Anti-counterfeiting Holograms*

References: Chapters 5, 6 and 9

Holograms have been used since the 1970s to implement anti-counterfeiting and optical security devices in products such as credit cards, banknotes, passports and other critical documents used for authentication purposes. Figure 16.13 shows typical banknote holograms and other security holograms.

#### *Optical Variable Imaging Devices (OVIDs)*

References: Chapters 5 and 14

Optical Variable Imaging Devices (OVIDs) are spatially multiplexed gratings and/or diffractive off-axis lenses. Such OVIDs and other synthetic 3D display holograms can either be originated via traditional holographic exposure, or can be generated by computer as arrays of DOEs or CGHs (see Chapters 5 and 6). In the latter case, they are mastered by conventional optical lithography and replicated by roll embossing (see Chapters 12 and 14).

OVIDs are variable devices, since the spectral pattern diffracted in a viewing direction varies with the angle between the grating and the source (the sun or a bulb). The diffractive angles can be carefully crafted in order to produce dynamic projections of spectral shapes so that the viewer has the feeling that the devices are actually dynamic, in both shape and color.

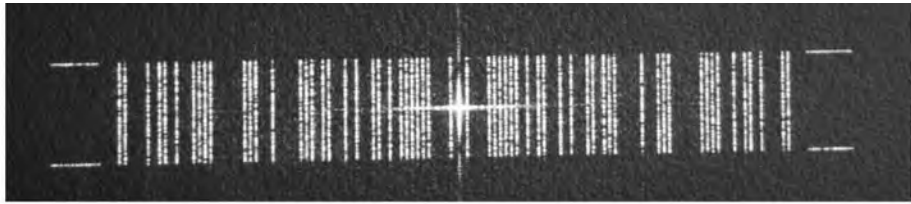
#### *Holographic Bar Codes*

References: Chapters 6 and 14

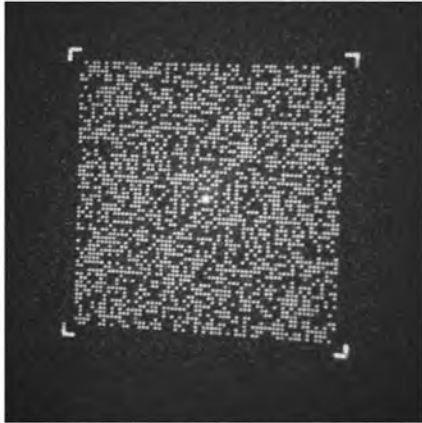
OVIDs and other holograms provide only a relative amount of security, since they can easily be replicated overnight by any holographic replication laboratory (legal or illegal). The human eye is not the perfect sensor; therefore, it should not be used to secure products such as medication or luxury goods.

Rather, one would prefer to use machine-readable optical information, such as bar codes. Holographic bar codes, in 1D and 2D, have thus been developed, and use the same mastering and replication technologies as OVIDs. Figure 16.14 shows the optical reconstruction of such 1D and 2D diffractive bar codes.

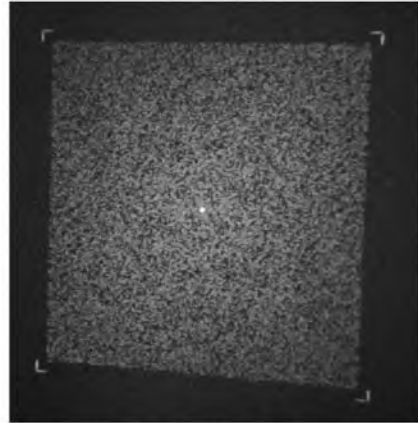
Such holographic bar codes, especially in 2D arrays, are machine-readable holograms (via a 1D laser scanner or a 2D CMOS sensor) that can include large amounts of digital data, much larger than can be integrated in RFID devices today. Such hybrid visual/machine-readable synthetic holograms still make use of Write Once Read Many (WORM) type holographic tags, but are much cheaper to mass-replicate than RFIDs.



*1D holographic bar codes (64 bits coding)*



*2D holographic bar code (256 Kb bits coding)*



*2D bar holographic code (1 Mb bits coding)*

**Figure 16.14** One- and two-dimensional diffractive bar codes

### 16.2.3.6 Holographic Scanners

References: Chapters 8 and 14

Holographic scanners (for conventional bar codes) were early adopters of holographic technology. Most of the scanners used in industry today are based on multiplexed holographic gratings, as depicted in Figure 16.15.

### 16.2.4 Optical Computing

The IC industry has been engaged in relentless efforts to try to keep up with the momentum of Moore's law, increasing computing power and reducing the size of the features on the IC chip. One of the directions is to reduce the size of the smallest transistor or gate printable on the wafer (see Chapters 12 and 13). Another direction is to reduce the heat dissipation and reduce the parasitic inductive or capacitive effects linked to higher clock frequencies. As a matter of fact, the modern CPU looks electrically more like a light bulb (power), an oven (current) or a flashlight (voltage), and will likely require advanced thermal management solutions. Eventually, the IC industry will run into a brick wall and will be unable to reduce the size and the heat dissipation of its circuits any more.

By replacing electronics with photons, it is theoretically possible to reduce the overall size of the chip while at the same time reducing the heat dissipation, driving the chip at much higher frequencies without any parasitic effects and producing chips in the third dimension, even using free space.

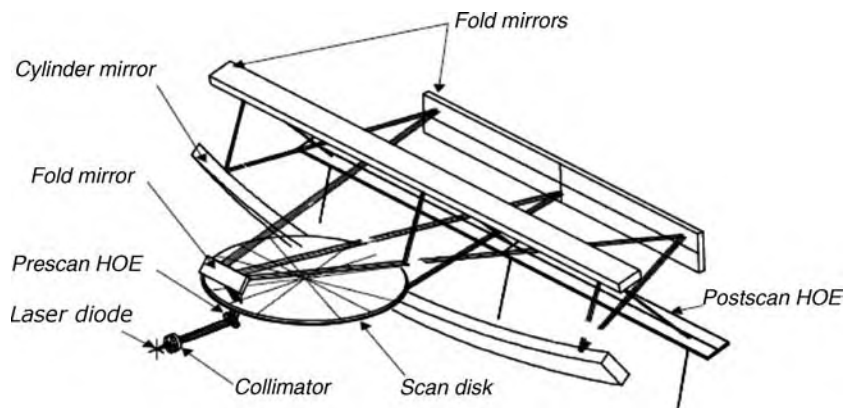


**Figure 16.15** A holographic product scanner

The first step in using photons rather than electronics is to produce densely photonic links between chips: optical interconnections, where light is emitted by a source (VCSEL or laser diode) and launched onto a detector. Such beams can be routed or modulated by integrated optical devices or split into many beams, but they remain passive links. The second step, which is much more complex, is to provide fully photonic logic gates on the chip where logic operations can be performed on the light without having to fall onto a detector and being regenerated by a laser source (much like a full optical switch in optical telecom – see Section 16.5).

Throughout the 1990s, optical computing [9, 10] became a hot topic of research and development, and used diffractive optics technology for both optical clock distribution in multi-chip modules (MCM) and optical interconnections in massively parallel computing architectures. However, this technology has not yet transferred to consumer electronics, and still remains in development. However, there has recently been renewed interest in optical interconnections. Texas Instruments Inc. and Sun Microsystems have recently issued statements that optical interconnections are one of the 10 main technologies for tomorrow's IC industry development.

#### 16.2.4.1 Optical Clock Distribution

References: Chapters 4, 5 and 6

Optical clock distribution [11–14] was one of the first implementations of optical interconnections in the computing realm, especially for MCMs, in which the optical clock has to be broadcast at high frequencies. Figure 16.16 shows two different approaches. The first one uses the H tree architecture with TIR and redirecting sub-wavelength gratings etched in a slab waveguide, and the second one uses a 1 to 16 multifocus Fresnel lens in a space-folded free-space thick slab architecture.

#### 16.2.4.2 Parallel Optical Interconnections

References: Chapters 4, 5 and 6

Complex interconnection architectures [15–24] can be implemented optically in three dimensions via 2D arrays of VCSELs and 2D arrays of detectors, with 2D arrays of diffractive optical elements located in between. Figure 16.17 shows the implementation of the twin butterfly interconnection architecture via an array of spatially multiplexed off-axis Fresnel lenses, and an array of phase-multiplexed Fresnel lenses.

Section C4.4 of Appendix C shows the interconnection scheme for the implementation of a 2D FFT. Such interconnection architectures can also be implemented optically.

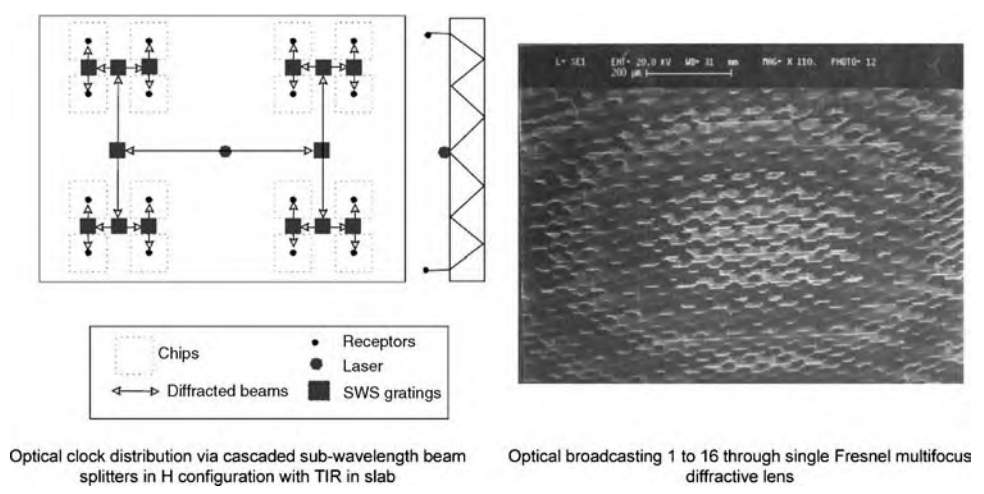


Figure 16.16 Optical clock distribution architectures (1 to 16)

Figure 16.18 shows an example of an array of diffractive lenses (single-focus and multiple-focus CGH lenses), which are used in a planar opto-electronic interconnection architecture. In the lower right-hand part of the figure, one can observe the focused spots (from the single lenses and the  $3 \times 3$  multifocus lenses).

16.2.4.3 Parallel Optical Image Processing

Reference: Chapter 6

Parallel image processing [25] can be done via Fourier filtering with the help of filtering CGHs, as described in Chapter 6. Such Fourier filtering typically includes 2D pattern detection (Vander Lugt filters), edge detection and so on.

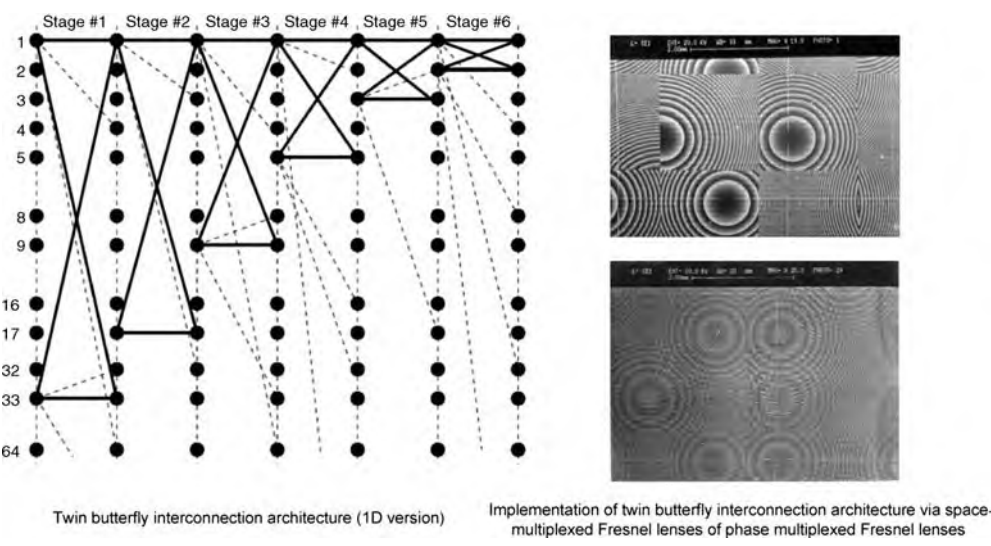


Figure 16.17 The implementation of an optical interconnection architecture via arrays of diffractive optics

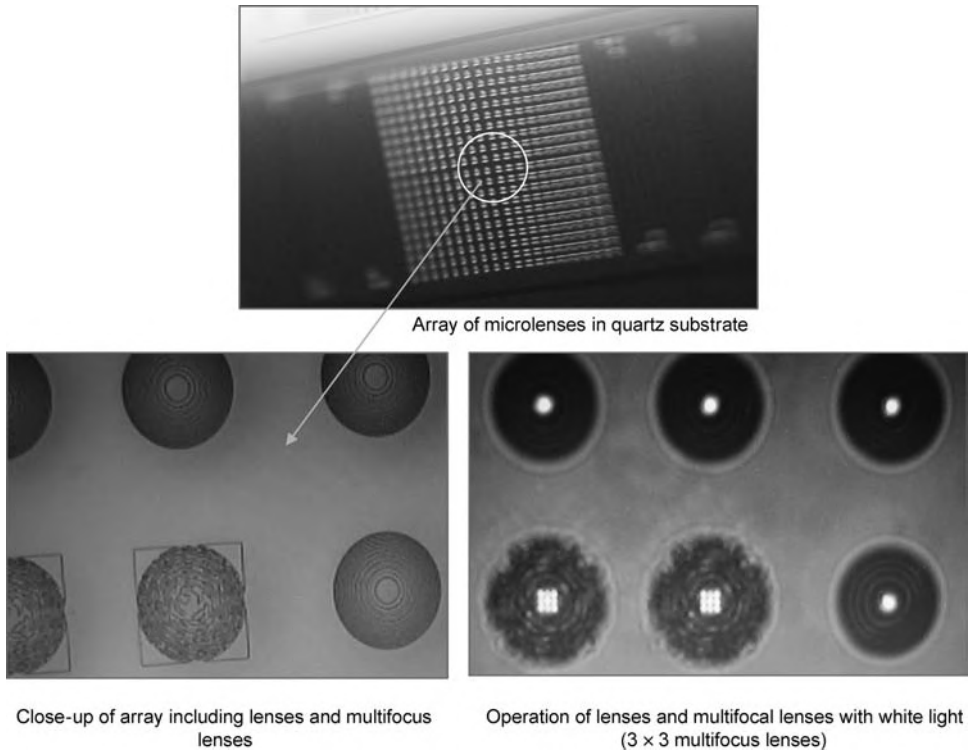


**Figure 16.18** An example of a planar diffractive opto-electronic interconnection architecture

### 16.2.5 Metamaterials

References: Chapters 10 and 11

Applications using metamaterials in the optical region are triggering considerable interest today from the military, industrial and biomedical sectors. Chapter 10 has described the various optical cloaking, perfect imaging and super-prism effects, which have promising applications in biomedical sensors and IC fabrication sectors. However, we are far from a broadband optical cloak that would work for a range of angles and that would not be sensitive to small perturbations.

## 16.3 Clean Energy

Clean energy has been the subject of considerable attention since the beginning of the last energy crisis. Solar energy is a perfect candidate to integrate planar digital optics for applications such as anti-reflection coatings, solar concentrators, optical tracking and photon trapping.

### 16.3.1 Solar Energy

#### 16.3.1.1 Anti-reflection Structures

Reference: Chapter 10

Anti-reflection (AR) structures have been described in Chapter 10, as sub-wavelength structures integrating effective medium theory in order to mimic the effects of multiple stacked thin films. Such AR structures can be directly etched within the top window of the cells, and replace costly thin films.

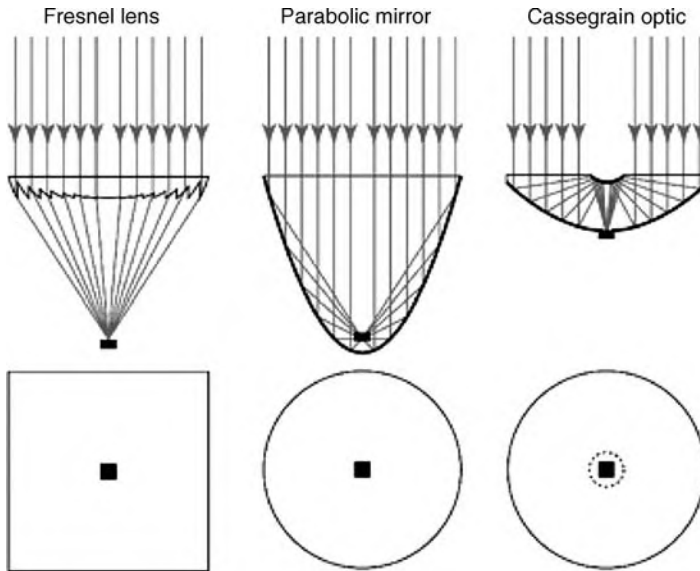


**Figure 16.19** Solar concentrators used in industry today

### 16.3.1.2 Solar Concentrators

Reference: Chapter 5

Solar concentrators have used refractive Fresnel lenses and other micro-optical elements for a long time. Diffractive broadband concentrators can be used today, since broadband design and fabrication tools are now available (see Chapter 5). Diffractive concentrators can reduce the size and price of conventional concentrators. Figure 16.19 shows some of the concentrators used in industry today.

### 16.3.1.3 Solar Tracking

Reference: Chapter 5

Solar tracking is a key to increasing the efficiency of a photovoltaic system (PV). Mechanical tracking is used today; however, static tracking is a more desirable feature.

### 16.3.1.4 Solar Trapping

References: Chapters 5 and 10

Solar trapping is a very desirable feature in which photons are trapped by a planar device that redirects these photons to the PV cells. The aim is to integrate, on a single or dual sheet of plastic microstructures, an AR layer, a solar concentrator and a solar tracking functionality over a single PV device. Companies such as Prim Solar Inc. and Holox have proposed solutions for solar light trapping (see Figure 16.20).

## 16.3.2 Wireless Laser Power Delivery

Reference: Chapter 6

Wireless laser power delivery is one of the new ways to redirect power in a wireless way through free space over small distances. Such systems have already been integrated (see [www.powerbeam.com](http://www.powerbeam.com)). For

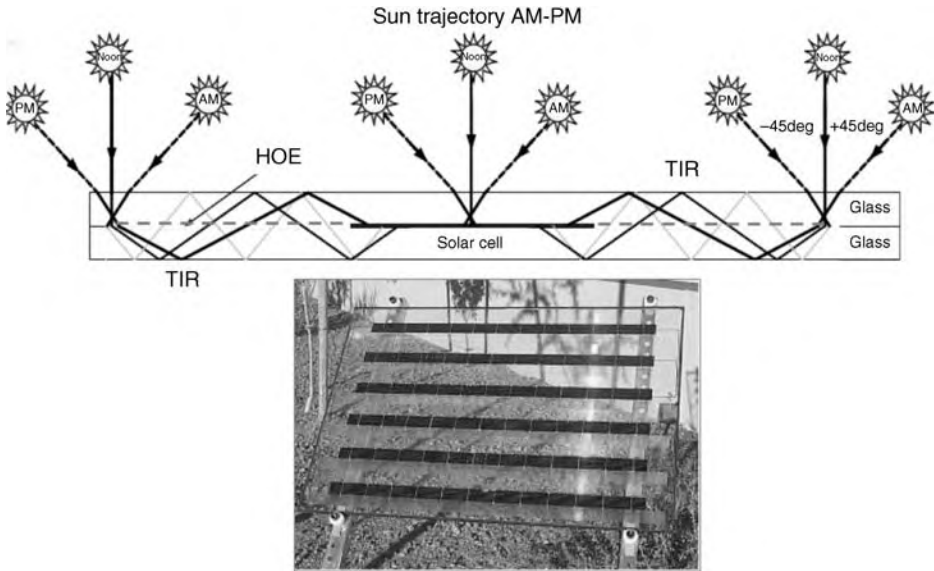


**Figure 16.20** A holographic PV system (Prism Solar Inc.)

example, a laser can produce a beam that is split by a dynamic diffractive element (such as an H-PDLC) onto devices that are selected by the user within a room, and thus optimize the direction of the beam in real time to generate maximum energy where it is needed.

### 16.3.3 The National Ignition Facility (NIF)

#### 16.3.3.1 Laser Damage Resistant Diffraction Ratings

Reference: Chapter 5

Beam sampling gratings have been fabricated to provide small samples of the NIF 351 nm high-power laser beams for monitoring purposes. The sampled fraction will be used to determine the laser beam energy on target and to achieve power balance on hundreds of NIF beams. Petawatt pulse compression gratings have also been fabricated.

#### 16.3.3.2 Large-aperture Diffractive Lenses

Reference: Chapter 6

Phase plates, beam-correcting optics and large-aperture segmented Fresnel lenses have also been fabricated for the NIF facility.

## 16.4 Factory Automation

### 16.4.1 Industrial Position Sensors

Industrial optical sensors have been a large pool of application for diffractive optics. Gas sensors, position, displacement and motion sensors, strain, torque and force sensors, spectral sensors and refractive index sensors, Doppler velocimeter sensors as well as diffractive microsensors in MEMS architectural platforms have been reported in the literature as well as in industry.



**Figure 16.21** Micro-e diffractive linear encoders from Micro-e Inc. ([www.microesys.com](http://www.microesys.com))

#### 16.4.1.1 Diffractive Optical Encoders

References: Chapters 5 and 6

Industrial optical sensors (especially motion and position encoders) are sustained by a large and steady market that is growing slowly and has not been affected by any of the technological bubbles – as have been, successively, the optical data storage, optical telecom and biomedical markets. The current encoder market (linear and rotational) today exceeds \$5 billion a year and is growing continuously.

##### *Linear Encoders*

References: Chapters 5 and 6

Linear interferometric incremental encoders based on dual gratings have been developed and are now available on the market. Figure 16.21 shows some diffractive linear encoders from Micro-e Inc.

Such incremental encoders are often integrated with two gratings, one being the ruling grating (long grating) and the other the read-out grating, which can have a period half that of the ruling grating. The various diffraction orders produce a traveling interference fringe pattern that is sensed by a linear detector array. However, it is a difficult task to produce long gratings which do not have any phase discontinuities. Such encoders have a very high accuracy but are also very sensitive to temperature drifts, shocks/vibrations and humidity.

Absolute linear encoders can be implemented as a linear succession of CGHs producing various binary codes (see Figure 16.22). Absolute diffractive linear encoders are very desirable, since they can implement a 2D encoding functionality while using a single direction on the detector areas. A CGH can redirect the diffracted beams in any direction, and is not limited to the direction of the scan. Figure 16.22 shows a 2D diffractive encoder strips over 4 bits in each direction (in reflective, transmission and hybrid mode).

##### *Rotational Encoders*

References: Chapters 5 and 6

##### *Incremental Encoders*

Conventional optical encoders are usually implemented as a periodical set of two openings in an opaque disk (A and B signals) located in phase quadrature over a circular channel (see Figure 16.23).

Such openings in an opaque disk (e.g. chrome on glass) can be replaced by gratings, which can redirect the light onto detectors that do not need to be placed exactly in the line of the openings in the disk.



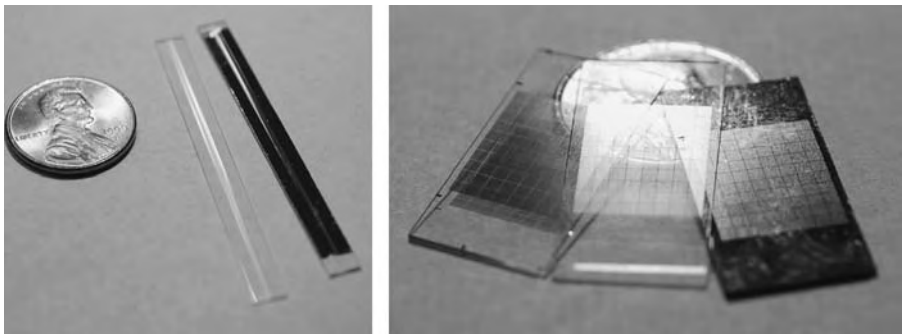


**Figure 16.22** One- and two-dimensional linear diffraction absolute encoders using arrays of CGHs

Similarly to linear incremental diffraction encoders, rotational encoders can be produced with sub-wavelength gratings. It is easier to implement such gratings in the rotational case, since the distance (radius) is finite and can be relatively small (see Figure 16.20).

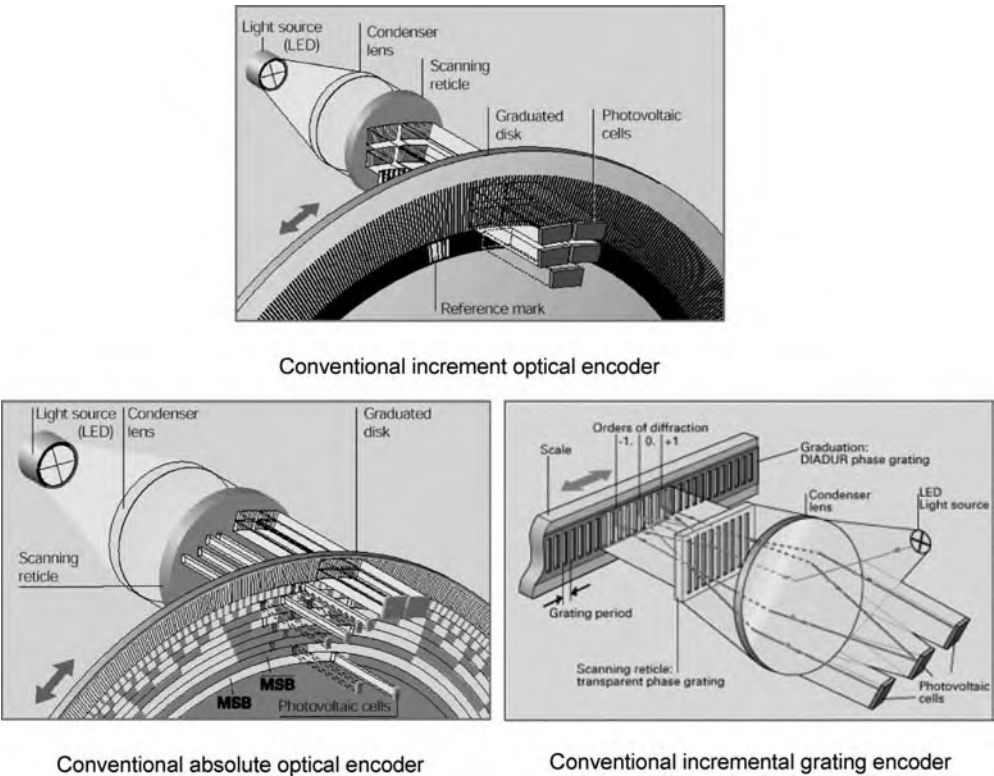


**Figure 16.23** Conventional optical encoders: amplitude incremental and absolute, and diffractive incremental (courtesy of Heidenhain)

### *Absolute Encoders*

Absolute optical encoders are usually made out of an amplitude disk that has  $N$  openings for  $N$  bits of resolution in line with  $N$  detectors. Such signals are often implemented in a Gray code architecture (see Figure 16.24).

Similarly to incremental encoders, such Gray codes can be generated by diffractive elements, this time not gratings, but series of more complex CGHs that diffract a specific binary pattern in the far field (or at the focal plane of a detector lens). These CGHs can be placed in a circular geometry to form a circular channel. Figure 16.24 shows such diffractive absolute encoders over 12 bits, mass-replicated as transmission or reflective disks, on 600  $\mu\text{m}$  thick polycarbonate small DVD-type disks of 32 mm diameter (see also Chapter 14).

### *Hybrid Encoders*

References: Chapters 5 and 6

Hybrid diffractive encoders integrate incremental and absolute channels on a very small physical channel, and can thus be used as a signal-on-demand encoder (incremental for high speeds and absolute for low speeds).

### *Multidimensional Encoders*

References: Chapters 5 and 6

Two-dimensional encoders can be implemented as combined 1D absolute encoders, as discussed earlier, using a single linear detector array. Similarly, radial/circular encoders can be implemented by similar CGHs, which can, for example, encode the rotation over 12 bits and the radial position of 4 or 8 bits, on the same linear detector array, by redirecting the diffracted beams adequately. Such a multidimensional encoder could monitor the wear of a shaft or the torque produced on the shaft, as well as the rotation angle (elliptical rather than circular motion control).

#### **16.4.1.2 Machine Tool Security Sensors**

Reference: Chapter 6

Machine tool security sensors can be implemented by either static or dynamic beam-splitter CGHs, which cover a much larger field than single beams would do. We have demonstrated beam splitting from one beam to as many as 100 000 beams in Chapter 6. The typical beam-splitting ratio in machine tool security sensors can be as high as 100 (in order to monitor up to 100 different areas around the machine tool).

## **16.5 Optical Telecoms**

The optical telecoms sector has seen a massive surge of interest in diffractive optics since the DWDM revolution at the end of the last decade, and more recently with the steady growth of CATV (Cable TV) and 10 Gb/s optical Ethernet lines.

### *16.5.1 DWDM Networks*

#### **16.5.1.1 DWDM Demux/Mux**

References: Chapters 3–5

#### *Free-space Grating Demux*

Diffractive optics have been applied extensively to spectral Demux and Mux applications, as reflective linear ruled gratings or transmission phase gratings (see Chapter 5). Mux assemblies can be integrated

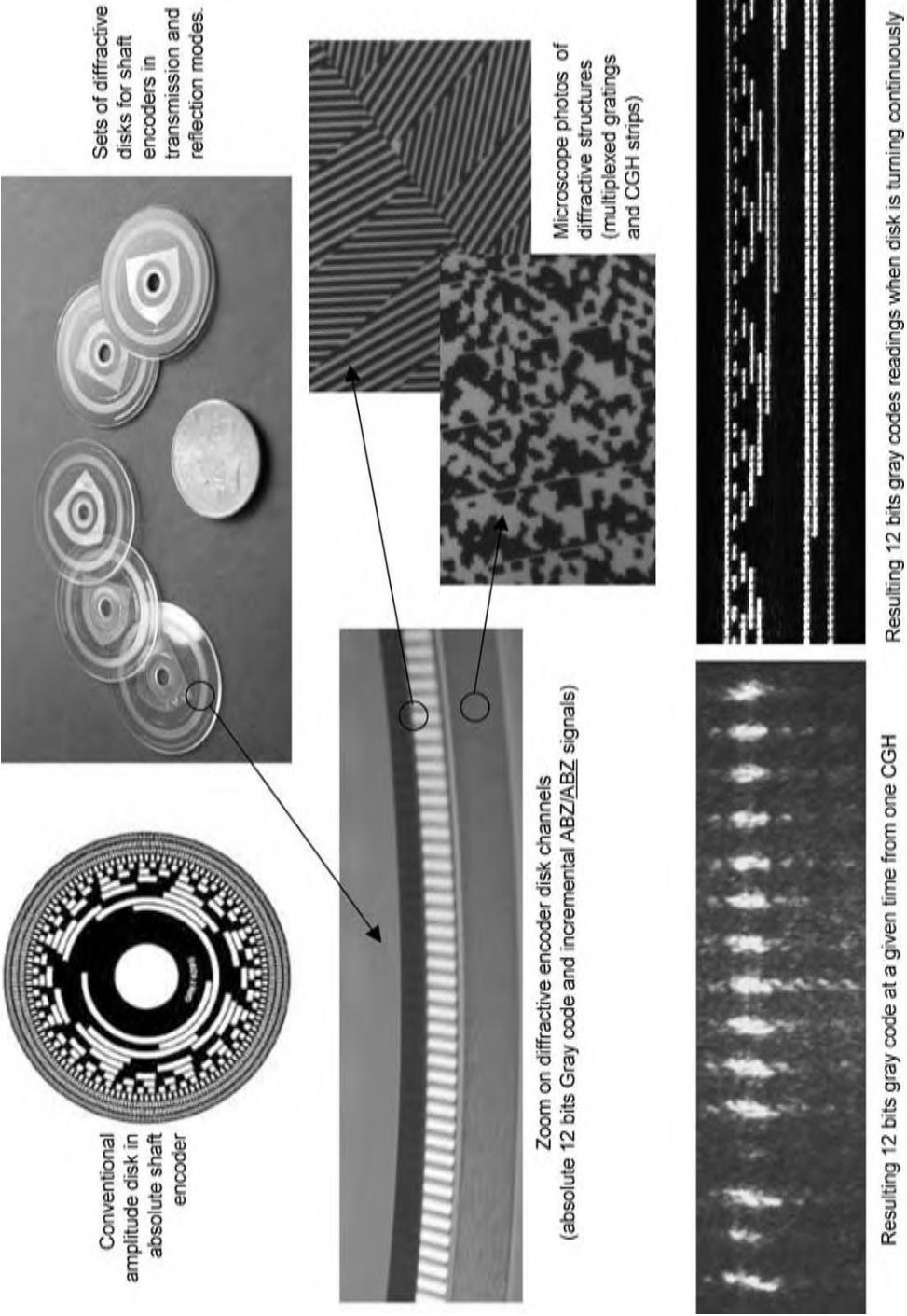


**Figure 16.24** Absolute diffractive rotational encoders: replicas made in small DVD disks. The signal is Gray code over 12 bits

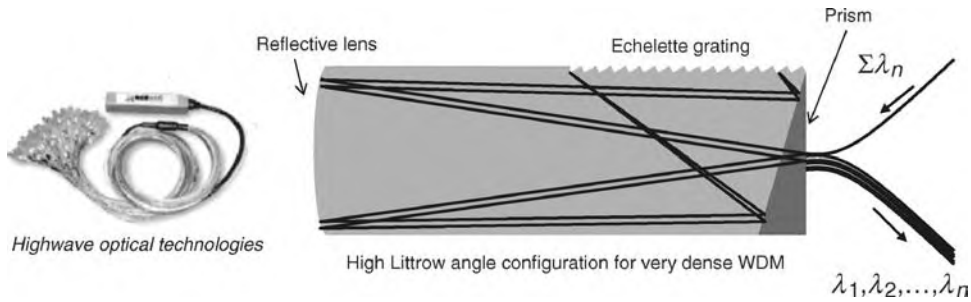


**Figure 16.25** A DWDM reflective grating Demux application (www.highwave-tech.com)

with other technologies to produce more complex functionalities, such as optical add-drop modules (see Chapter 3).

Figure 16.25 shows a reflective grating integrated in a hermetic package, which can perform wavelength demultiplexing of 192 channels in a 50 GHz spacing.

- *Cascaded dichroic Demux* based on thin film filters and arrays of GRIN lenses has been discussed in Chapter 4.
- *AWG grating Demux*: AWG- and GWR-based PLC Demux devices have been discussed in Chapter 3.
- *Integrated échelette grating Demux* devices have been discussed in Chapter 3.

### 16.5.1.2 Integrated Bragg Grating Devices

References: Chapters 3–5

- *Bragg reflectors* used in semiconductor laser diodes have been discussed in Chapter 3.
- *Wavelength lockers*: Bragg reflectors used for wavelength locking have been discussed in Chapter 3.
- *Add-drop devices*: reflectors based on multiplexed Bragg reflectors have also been discussed in Chapter 3.

### 16.5.1.3 EDFA Fiber Amplification

References: Chapters 3–5, 7, 9 and 10

- *Polarization splitters and combiners* have been discussed in Chapter 9 (sub-wavelength gratings).
- *Variable Optical Attenuators (VOAs)* have been discussed in Chapter 10, as dynamic MEMS gratings.
- *Dynamic Gain Equalizers (DGEs)* have been discussed in Chapters 7 and 10, as hybrid waveguide holographic chirped Bragg couplers.

## 16.5.2 CATV Networks

References: Chapters 3–5

The CATV signal is usually transported in the same fiber as DWDM for long haul, but over a different wavelength range (1310 nm instead of 1550 nm; see Chapter 3).

### 16.5.2.1 The 1310/1550 Splitter

CATV is a large user of wavelength splitters or broadcasters to split the DWDM bands around 1550 nm and the CATV at 1310 nm. Dichroic filters or diffractive elements can be used for this task.

### 16.5.2.2 Optical Fiber Broadcasting

Optical broadcasting is widely used in CATV networks, especially for FTTH (Fiber To The Home applications). A 1 to  $N$  pigtailed splitter can be implemented between two GRIN lenses and a Fourier fan-out grating, as described in Chapter 4.

## 16.5.3 10 Gb/s Optical Ethernet

### 16.5.3.1 Optical Transceiver Blocks

References: Chapters 5 and 6

In recent years, diffractives have been applied to 10 Gb/s optical Ethernet lines for fiber coupling and detector coupling, signal and other monitoring functionalities. Figure 16.26 shows a 12-line 10 Gb/s optical Ethernet optical assembly block for 850 nm VCSEL laser-to-fiber coupling and fiber-to-detector coupling using dual-side substrate patterning.

In Figure 16.23, the left-hand part shows a 6 inch quartz wafer with many individual optical block assemblies etched into it, and the right-hand part (as well as the central parts) show a single diced-out assembly ready to be integrated into a 12-array 10 Gb/s Ethernet fiber bundle.

### 16.5.3.2 Vortex Lens Fiber Couplers

Reference: Chapter 5

Chapter 3 discussed the coupling of a laser beam into a graded-index plastic fiber through the use of a diffractive vortex coupling lens. Such a vortex lens creates a helicoidal beam for an optimal coupling into

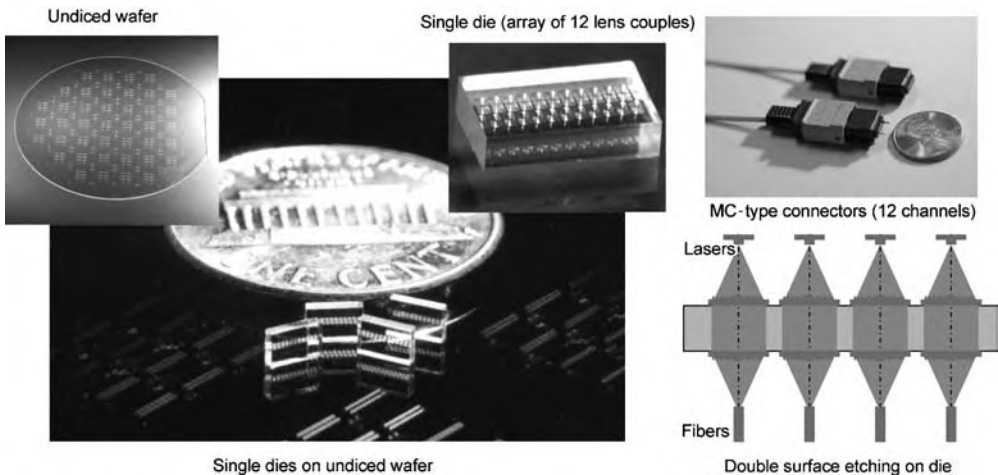


**Figure 16.26** A multichannel 10 Gb/s Ethernet optical block assembly for 850 nm VCSEL lasers/detectors

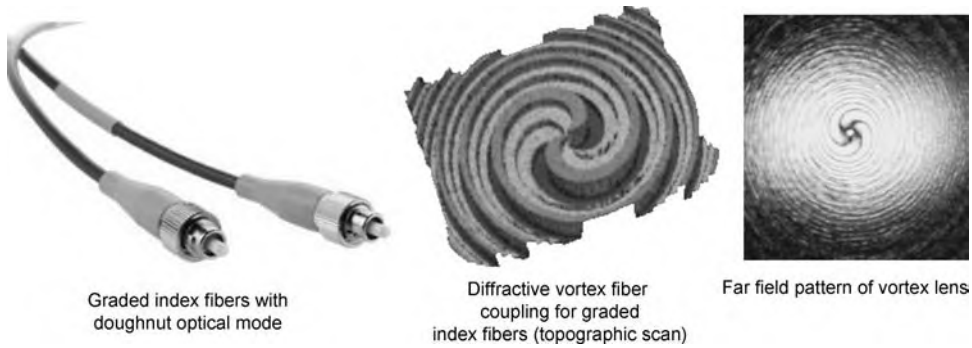


**Figure 16.27** A diffraction vortex lens far-field pattern and surface scan

such a fiber, which exhibits a phase discontinuity on its central core region (the discontinuity is linked to a systematic fabrication error of graded-index fibers). The fabrication of diffraction vortex lenses is also described in Chapter 12. Figure 16.27 shows the far-field pattern of a vortex lens and a profilometry scan of the lens surface.

#### 16.5.4 The Fiberless Optical Fiber

References: Chapters 5 and 9

The ‘fiberless optical fiber’ is the traditional name for free-space optical communications via laser. Such systems may use diffraction lenses for the collimation of the beams at the launching site and for the focusing of the beams onto detectors at the reception site. Furthermore, such systems may use dynamic optics linked to real-time wavefront analysis systems, as described previously in the telescope application (see Section 16.2.2.3).

### 16.6 Biomedical Applications

In recent years, the biomedical field has been showing great interest in diffraction optics, for applications in optical sensors, medical imaging, endoscopic imaging, genomics and proteomics, as well as individual cell processing.

Diffraction optics have been used to implement surface plasmon sensors and refractive index sensors by using the Bragg coupling effect in integrated waveguides. Parallel illumination of assays in genomics have used CGHs as laser beam splitters. Such diffraction splitters can illuminate large numbers of individual arrays of samples to read out the induced fluorescence. Laser tweezers have taken advantage of dynamic diffraction in confocal microscope architectures to move individual cells in a desired 2D pattern in real time.

More generally, diffraction can be used in any biomedical apparatus that requires homogenization and/or shaping of the Gaussian laser beam into a specific intensity mapping (e.g. fluorescence measurements in hematology by uniform laser illumination of a blood flow).

#### 16.6.1 Medical Diagnostics

##### 16.6.1.1 Ophthalmic Applications

References: Chapters 5 and 14

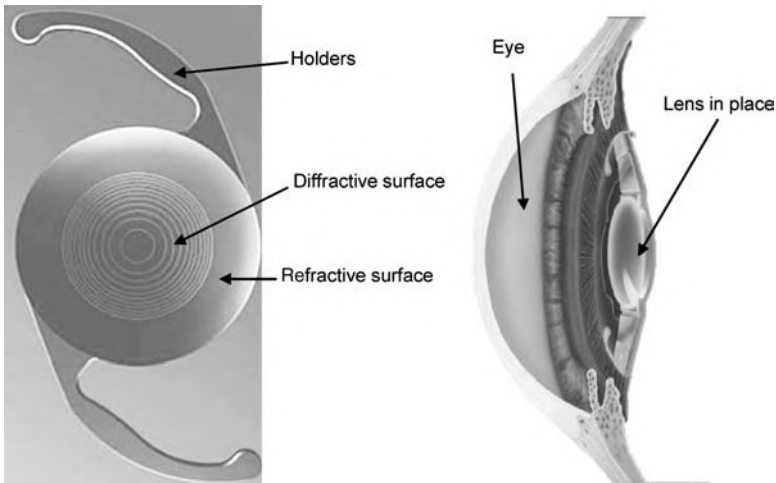


**Figure 16.28** A hybrid refractive diffractive intra-ocular lens

Hybrid refractive diffractive intra-ocular lenses have been used for some time now, 3M being one of the pioneers in this domain. We have seen how to achromatize such hybrid lenses in Chapter 7. Figure 16.28 shows such a hybrid bifocal intra-ocular lens in place in the eye.

In order to increase the depth of view of such lenses, one of the techniques described in Chapter 5 can be implemented. Figure 16.29 shows the implementation of an extended depth of focus diffractive lens, which can be used in an intra-ocular lens in order to increase the depth of focus on the resulting compound lens.

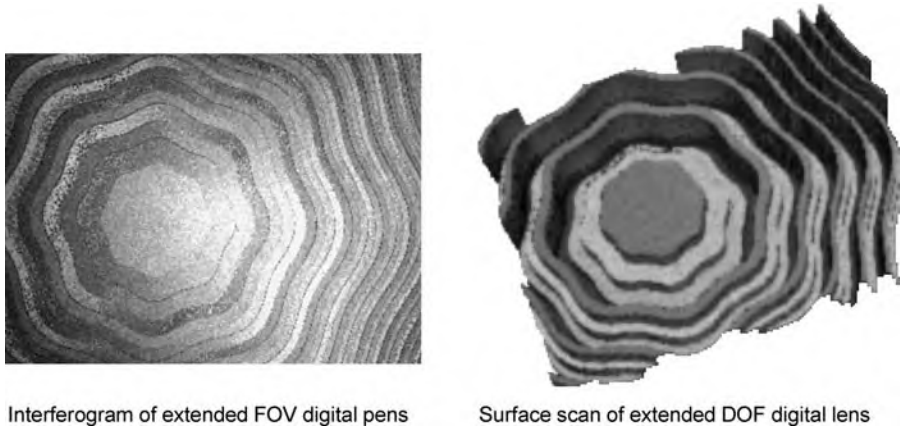


**Figure 16.29** An extended DOF diffractive Daisy lens for a hybrid intra-ocular lens

### 16.6.1.2 Endoscopic Systems

References: Chapters 4 and 5

Endoscopic systems are very well suited to use the potential of structured illumination via a confocal system, or to increase the imaging qualities of the end tip of the endoscope.

### 16.6.1.3 Flow Cytometry and Beam Shaping

References: Chapters 4 and 6

In flow cytometry applications, the illumination of the blood flow has to be uniform for optimal results. Therefore, the use of a beam-shaping diffractive or refractive micro-optical element has been proposed.

### 16.6.1.4 Diffractive Coherence Tomography

References: Chapters 8 and 15

Holographic confocal imaging (Fourier tomography imaging or diffraction tomography) is a relatively new technique that allows the imaging of small phase objects in 3D. This is especially interesting for biomedical research applications (for more details, see Chapters 8 and 15).

## 16.6.2 Medical Treatment

### 16.6.2.1 Laser Skin Treatment

References: Chapters 5 and 6

It has been demonstrated that multiple small laser beams focused under the skin can help reduce wrinkles, age spots, sun spots, roughness or acne scares. Such arrays of beamlets can be produced by refractive optics, but can also be produced more efficiently by fan-out gratings or multifocus lenses.

Figure 16.30 shows a commercial device (the Fraxel™) in operation on a patient's skin, and an example of beam array generation from a diffractive fan-out grating. The wavelength is the same as for the telecom C band (1550 nm), and the spot size is about 100  $\mu\text{m}$  for a depth of about 300  $\mu\text{m}$ .

### 16.6.2.2 Chirurgical Laser Treatment

Reference: Chapter 6

In chirurgical laser treatment, elements similar to the ones we discussed for laser material processing can be used, namely diffusers for heat treatment and focusators for cutting and suturing.



**Figure 16.30** Skin treatment via arrays of small focused laser beams



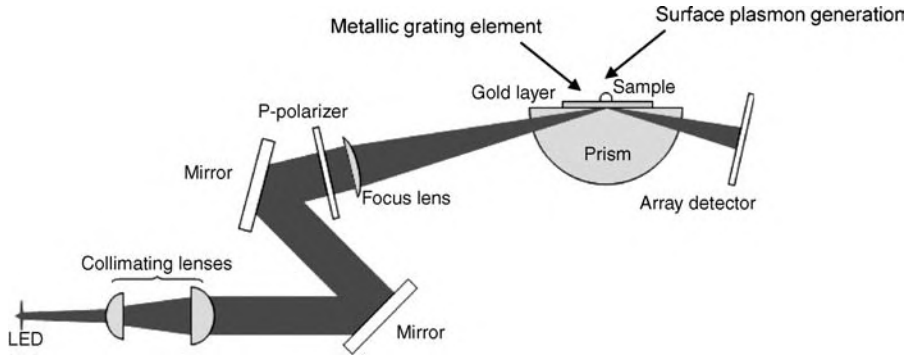


**Figure 16.31** Integration of a surface plasmon sensor

### 16.6.3 Medical Research

#### 16.6.3.1 Integrated Optical Sensors

Reference: Chapter 10

Integrated optical sensors are a large pool for digital optical elements integration. Examples include surface plasmon sensors and resonant sub-wavelength gratings (see Chapter 10). Figure 16.31 shows the implementation of a typical surface plasmon sensor.

#### 16.6.3.2 Immunoassay Sensors

##### *Surface Plasmon Immunoassay Sensors*

Reference: Chapter 10

The surface plasmon effect discussed in Chapter 10 is created over a metallic nanostructure such as a grating or an array of holes in a metallic structure. Such a surface plasmon extends only a few hundred nanometers over the surface of the metallic grating and is thus strongly affected by the local refractive index in this region (immunoassays, for example). Therefore, very small changes in the local permittivity of the assays (the refractive index) can be measured when launching light onto such a SPP device and measuring the intensity reflected back. SSP sensors can probe ultra fine layers (monolayers) and the samples can be extremely small (perfect for immunoassays).

##### *Immunoassay Illumination*

Reference: Chapter 6

Immunoassay or DNA assay illumination has been integrated by the use of fan-out gratings, as depicted in Figure 16.32. This reduces the complexity of the assay analysis by rendering the system static (laser beam deflection or assay movement no longer required).

### 16.6.4 Optical Tweezers

References: Chapters 6 and 9

Optical tweezers are used in biomedical research to move small particles or even molecules trapped within the focus of a laser beam via a microscope objective lens. Very often, several molecules have to be moved

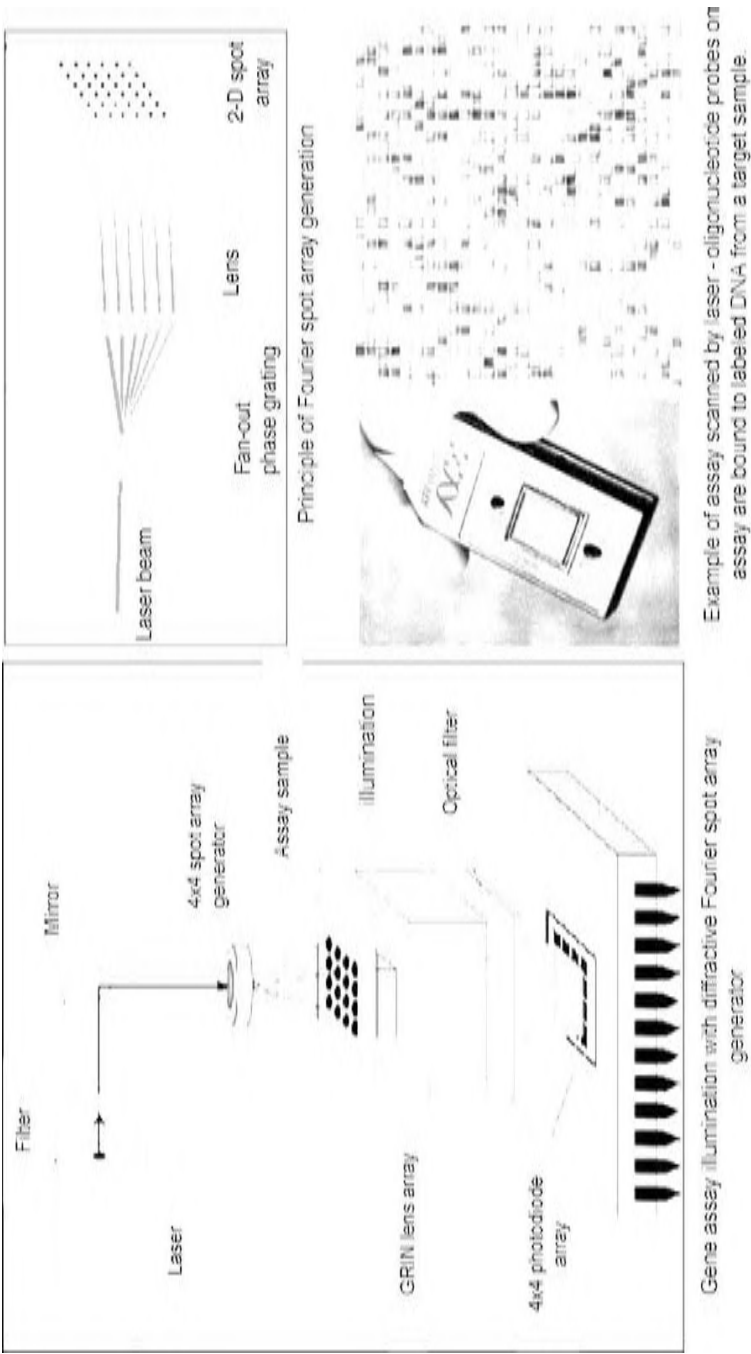


**Figure 16.32** The use of a fan-out grating for assay illumination (Affymetrix Inc.)

in a specific order; therefore, the use of dynamic holograms, which can produce a predetermined set of patterns (focused in a confocal microscopic system), has been proposed. Such elements can be dynamic reconfigurable optical elements or dynamic tunable optical elements such as the ones described in Chapter 9.

### 16.7 Entertainment and Marketing

The entertainment and marketing sectors are heavy users of digital optics and are helping to drive down the costs of mass-replication techniques for digital micro-optics.

#### 16.7.1 Laser Pointer Pattern Generators

Reference: Chapter 6

Laser pointer pattern generators are perhaps the best known application for digital diffractive optics and the one that reaches the most people directly today. Such elements are usually fabricated multilevel Fourier CGHs, calculated by an IFTA algorithm as presented in Chapter 6 and replicated in plastic by embossing. Figure 16.33 shows a red laser and many interchangeable heads, each bearing a different plastic replicated Fourier CGH.

Also depicted in Figure 16.33 are some far-field patterns projected by such laser pointers. The pattern on the left shows a binary intensity level image, which shows that a diffractive element can not only generate a precise far-field pattern but also control precisely the intensity levels over this pattern. The pattern in the center shows a 256 gray-scale image diffracted by a multilevel Fourier CGH. The zero order is right in the middle of the image, under the letter ‘S’, and only one diffraction order appears. By controlling the fabrication parameters, and especially the etch depth, one can almost completely remove the zero order as depicted in this pattern and push almost all the laser light into the

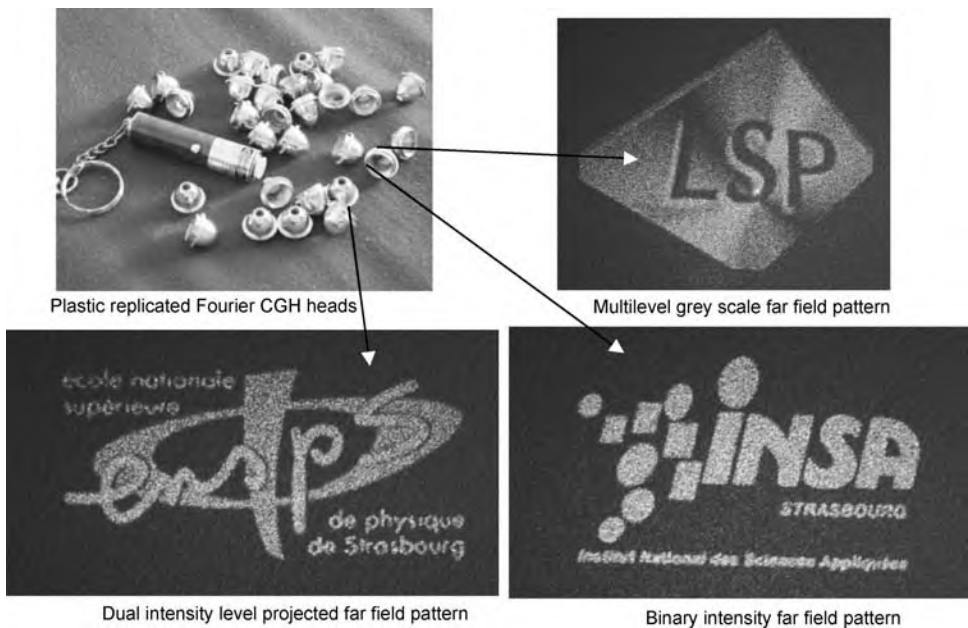


**Figure 16.33** Diffraction laser pointer pattern generators and projected pattern examples

desired pattern (the fundamental diffraction order). The pattern on the right-hand side is binary intensity, which includes both high- and low-frequency patterns. The intensity levels for both types of patterns are well controlled and produce a very uniform image.

These three types of laser intensity shaping are very useful in laser material processing (see Section 16.1.3). High-frequency diffracted patterns with binary intensity profiles can be useful to cut a complex shape in a workpiece, whereas a two-level intensity pattern can be used to cut through a metal sheet while leaving some chads cut through at only 50% or 25%, so that the cut shape remains hanging on the metal sheet for further processing. A complete analog intensity profile can be used to perform some surface heat treatment or hardening of a metal piece.

### *16.7.2 Diffractive Food*

References: Chapters 5 and 14

Food has also been an application pool for diffractives. There have been reports of gratings and other OVIDs embossed on the sugar coating of candies, to make them more appealing to children. Holographic candy wrappings are also a traditional market segment for holographic foil.

### *16.7.3 Optical Variable Imaging Devices (OVIDs)*

References: Chapters 5 and 14

We have already reviewed OVIDs in the previous optical security section. However, OVIDs are also widely used for their artistic beauty and optical effects in anti-counterfeiting labels and other marketing materials.

#### **16.7.3.1 Holographic Projector Screens**

References: Chapters 5, 8 and 14

Holographic projector screens are directional diffusers for digital projectors. Thus, they can be very bright for a viewer in the main diffusion angle. However, they remain expensive.

#### **16.7.3.2 Diffractive Static Projectors**

References: Chapter 6, 10 and 14

Diffractive static projectors are related to laser pattern projectors placed over a turning wheel, which projects a repetitive set of patterns. This can be very interesting, for example, in the case of structured illumination projectors for 3D contour extraction by successive pattern projection.

## **16.8 Consumer Electronics**

Consumer electronics is currently becoming one of the main users of digital optics, surpassing their use in industrial applications as mentioned before. Such consumer electronics includes digital imaging applications, data storage applications, LCD display applications and computer peripherals.

### *16.8.1 Digital Imaging*

References: Chapters 5 and 7

Developing a diffractive imaging system is usually considered a difficult or near impossible task when dealing with broadband illumination (white light). However, as seen in Chapter 5, diffractives can be designed as broadband or multi-order elements, or can help reduce chromatic and thermal aberrations and reduce the overall number of lenses needed, as seen in Chapter 7. It is therefore very desirable to use hybrid elements in optical systems, provided that the efficiency of the diffractive remains more or less constant over the whole spectrum considered. This latter task is more difficult than reducing lateral and longitudinal chromatic aberrations.

Figure 16.34 shows a simple example of broadband imaging through diffractive optics, with a binary and a multilevel lens. Figure 16.34 also shows a multiple imaging task through an hexagonal array of diffractive lenses.

It is noteworthy to point out that although the multilevel diffractive lens imaging task could be performed by a traditional spherical Fresnel lens (a single image generated in the center), the right-hand imaging task, which produces two conjugate images, could not be implemented by any kind of element other than a binary digital optical element. The two images produced are actually a converging wave and a diverging wave, producing a real and a virtual image (see Chapter 5).

16.8.1.1 The Hybrid SLR Camera Objective Lenses

References: Chapters 5 and 7

Some years ago, Canon Inc. of Japan (Figure 16.35) introduced a hybrid diffractive/refractive objective lens in their line of digital SLR cameras (telephoto and super telephoto zoom). Canon uses a set of sandwiched Fresnel lenses with different refractive indices in order to reduce the overall size and weight of

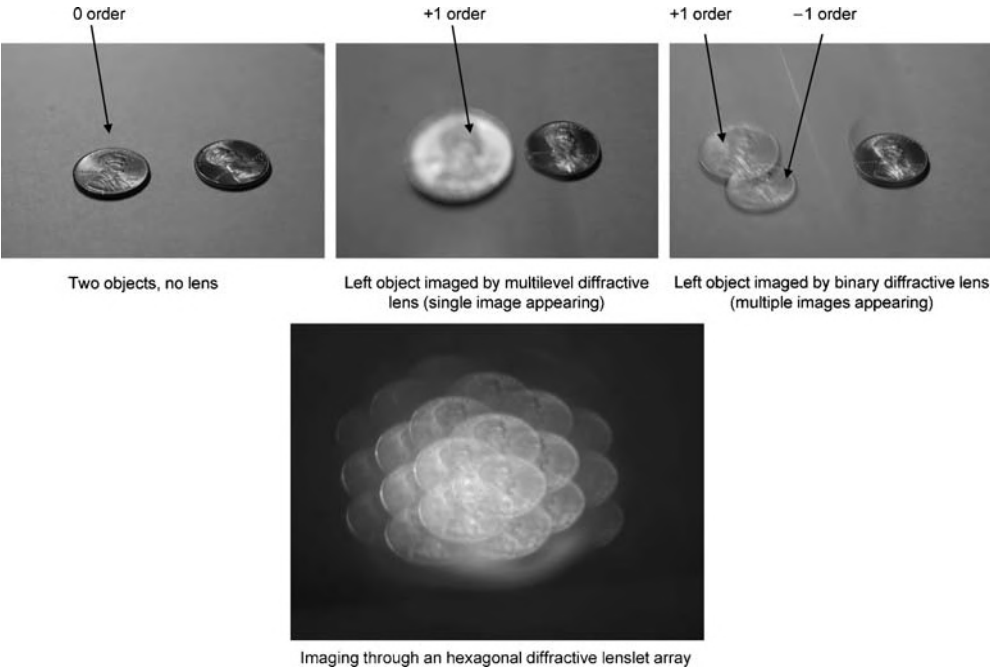


Figure 16.34 Broadband imaging through diffractive lenses



**Figure 16.35** The Canon diffractive telephoto and super telephoto zoom lenses

the objective. Not only is the resulting objective lens package shorter; it also incorporates fewer lenses for similar performances.

However, it is a slight abuse of terminology to speak about diffractive lenses when the grooves and the widths of the smallest zones in the Fresnel lenses are hundreds of times the wavelength. These elements are actually sandwiched micro-refractive Fresnel lenses, rather than sandwiched diffractive lenses. Such a Fresnel lens has very low diffractive power (see Chapters 1 and 5). This is why Canon had to use two different sandwiched refractive indices to correct chromatic aberrations, rather than using the technique described in Chapter 7, which uses a single index. If the technique described in Chapter 7 is used to produce a singlet achromat, the big challenge remains the diffraction efficiency over a very broad wavelength range. This could be achieved by using a multi-order diffractive lens (see Chapter 7).

### 16.8.1.2 Origami Objective Lenses

References: Chapters 5 and 10

Origami objective lenses are space-folded lenses that can implement reflective and diffractive structures over doughnut-shaped lens apertures. Off-axis origami lenses can integrate large effective numeric aperture lenses in a very small package. Such an origami objective lens is shown in Figure 16.36.

As there are no chromatic aberrations in a reflective objective, a reflective/diffractive lens element can actually create such longitudinal chromatic aberrations so that the objective can be used with wavefront coding techniques, such as those discussed in Section 16.8.1.4. Such an origami lens can also have focus tuning properties when the spacing between the two surfaces can be changed.

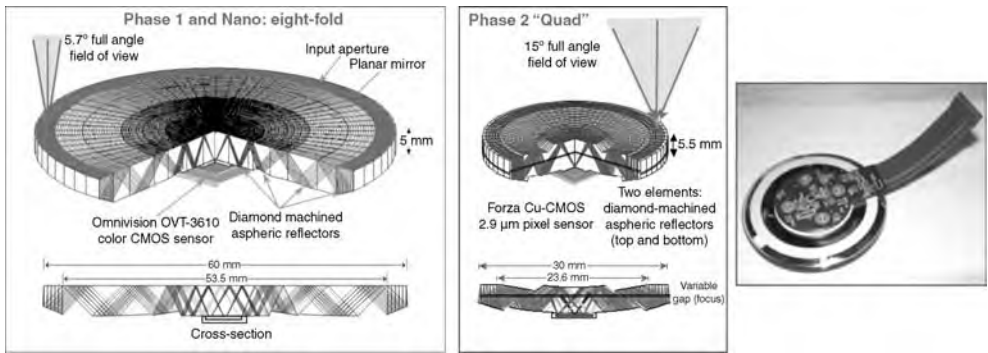


**Figure 16.36** An origami objective lens (courtesy of Professor Joe Ford, of the University of California at San Diego).

**16.8.1.3 Stacked Wafer-scale Digital Camera Objective Lenses**

References: Chapters 4 and 10

Today, miniature digital cameras are ubiquitous, from camera phones to digital SLR cameras, CCTV surveillance cameras, webcams, automotive safety cameras and so on. Such cameras will become smaller in the near future, and will be more and more frequently used as specialized sensors, rather than for taking conventional pictures. These miniature cameras need to work on a good image to start with. Stacking up wafer-level optics and dicing them afterwards is a good way to reduce costs and increase integration for miniature cameras. Figure 16.37 shows the principles of stacked wafer-scale objective lenses.

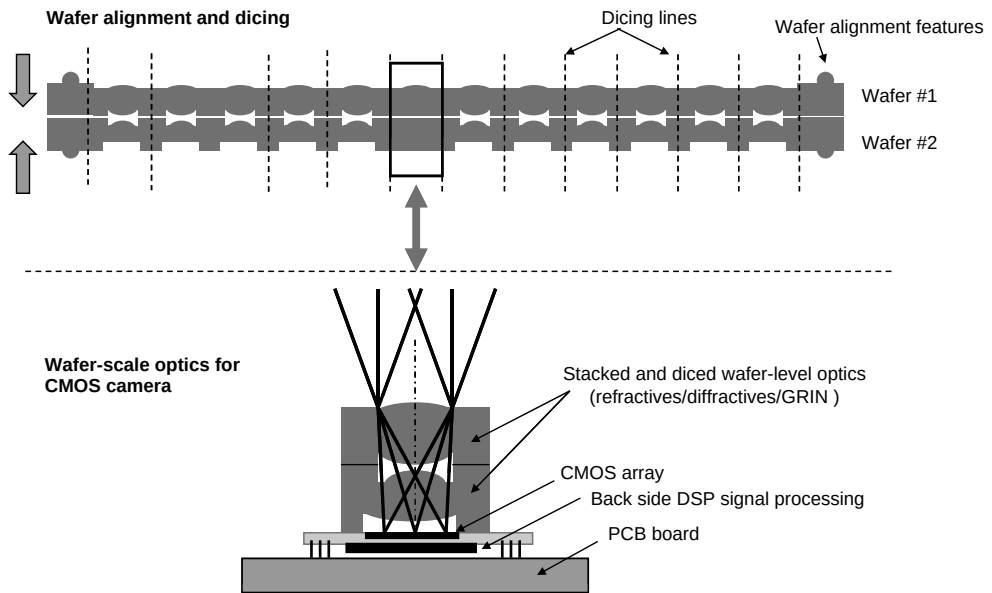


**Figure 16.37** Stacked wafer-scale objective lenses

Figure 16.38 shows the planar wafer stack objective lens architecture that is used by Tessera Inc. in their OptiML™ WLC miniature cameras.

The main advantage of such a technique is the fact that the lenses are pre-aligned onto the wafer prior to final dicing. Thus, the alignment to the CMOS plane is easy and does not need any long and costly adjustments. Also, the process is reflow compatible, so the hybrid opto-electronic integration becomes much simpler, especially in the packaging process.

#### 16.8.1.4 Wavefront-coding Camera Objectives

References: Chapters 4 and 10

Providing focus control in camera objectives is a very desirable feature. Chapter 9 described some of the techniques used in industry to produce dynamic focus by using MEMS techniques, piezo-electric, electro-wetting, LC-based and other exotic micro-technologies. However, none of these techniques make sense for low-cost mass replication. The task on hand is to develop an objective lens that does not require any movement to focus planes at different depths, and that is as efficient and versatile as a dynamic focus lens.

Chapter 9 has described a new lens design technique, the wavefront-coding lens design technique. Such lenses are software enhanced (they need a specific digital algorithm to work properly, and to extend their reach beyond that of traditional imaging lenses).

Such lenses can only be used along with their own specific software correction algorithm, and are therefore referred to as *software-enhanced lenses*. Section 9.5 shows such software-enhanced lenses for a CMOS camera objective lens application.

A well-controlled phase plate (diffractive or refractive) is incorporated in between two standard lenses (or microlenses). The precise prescription of such a wafer-scale phase plate is used in the digital image processing (linear and nonlinear) in order to compute the final aberration-free image.

The signal-processing capability can be integrated in a DSP located on the back side of the CMOS sensor, so that the aberration-free image can be directly output by the sensor chip.

Such software uses the precise prescription of the aberrated lens or phase plate in order to compute a final picture (by deconvolution, a Fourier filtering process or any other digital process). For example, a fixed-focus camera objective that has a considerable longitudinal chromatic aberration (e.g. through the use of diffractive lenses in the planar objective lens stack), can be used to compute digitally well-focused images for various positions of the field (macro, close capture, medium shot and telephoto). This is one application where fixed-focus camera lenses can be used, whereas traditional objectives need to be packaged within complex dynamic focus apparatus.

Lithographic fabrication techniques provide well-controlled refractive and/or diffractive microlens phase profiles (see also Chapter 12). It is thus possible to implement specific aberrations in refractive or diffractive phase plates, in order to integrate complex wavefront-coding techniques, therefore reducing the size and complexity of camera objectives. This produces a new frontier in digital image sensing for various applications not necessarily related to traditional imaging applications.

In future CMOS sensor networks, cameras will no longer be used as traditional photographic devices. Instead, they will often be used as specific sensors, individually or in large distributed network arrays. Such sensors will have their own special tasks and their own onboard signal-processing capabilities (e.g. through a diffractive structured laser illumination projector for 3D sensing). They will work on raw images, process these images digitally on-board, and send back to the network only the desired and necessary information (e.g. a 3D profile or a shape definition).

#### 16.8.1.5 Diffractive Auto-focus Devices

References: Chapters 6 and 14



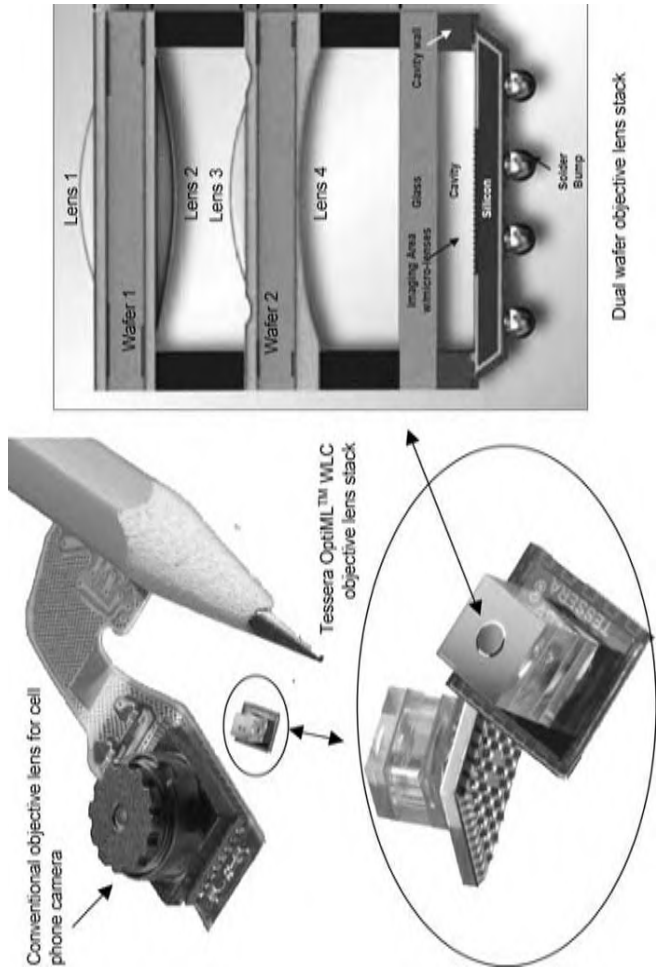
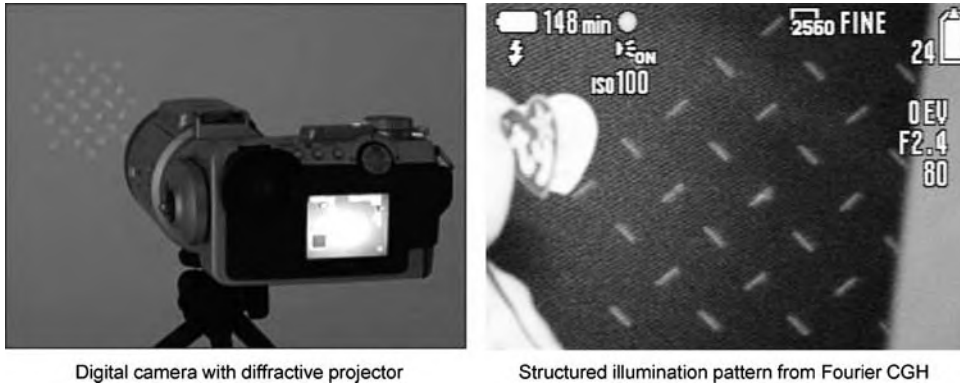


Figure 16.38 A wafer-stacked OptiML™ WLC lens. Reproduced by permission of Tessera Inc.



Digital camera with diffractive projector

Structured illumination pattern from Fourier CGH

**Figure 16.39** A diffractive structured light projector for auto-focus in digital cameras, developed by the Sony Corporation of Japan

Today, diffractive auto-focus devices are integrated in digital cameras as external diffractive far-field structure light projectors, such as the one described in Figure 16.39.

As the pattern projected by the Fourier CGH is a far-field pattern, it is in focus almost everywhere (at least beyond the Rayleigh distance – see Chapter 11), and thus the structured patterns reflected by the object of interest can be used by the objective lens system to tune in the focus.

#### 16.8.1.6 CMOS Lens Arrays

References: Chapters 4, 12 and 14

We have seen in Chapter 4 that microlens arrays can be used to increase the luminosity of CMOS sensors. Such microlens arrays are shown in Figure 16.40.

Such lenses are mostly micro-refractive lenses, with high fill factors. However, they can also be fabricated as highly color-selective diffractive lenses, where fill factors of 100% can be achieved.

#### 16.8.1.7 Digital Holographic Viewfinders

Reference: Chapter 9

Edge-lit holograms can be integrated as edge-lit viewfinders in SLR cameras. Such holograms can be static or even dynamic. Edge-lit holograms (see Chapter 8) can be illuminated by either LEDs or lasers placed on their edges.

### 16.8.2 Optical Data Storage

Optical data storage, and especially Compact Disk (CD) Optical Pick-up Units (OPUs), have integrated holographic and diffractive optics earlier than any other consumer electronic product, starting in the early 1980s.

#### 16.8.2.1 Holographic Grating Lasers

References: Chapters 5, 9 and 14

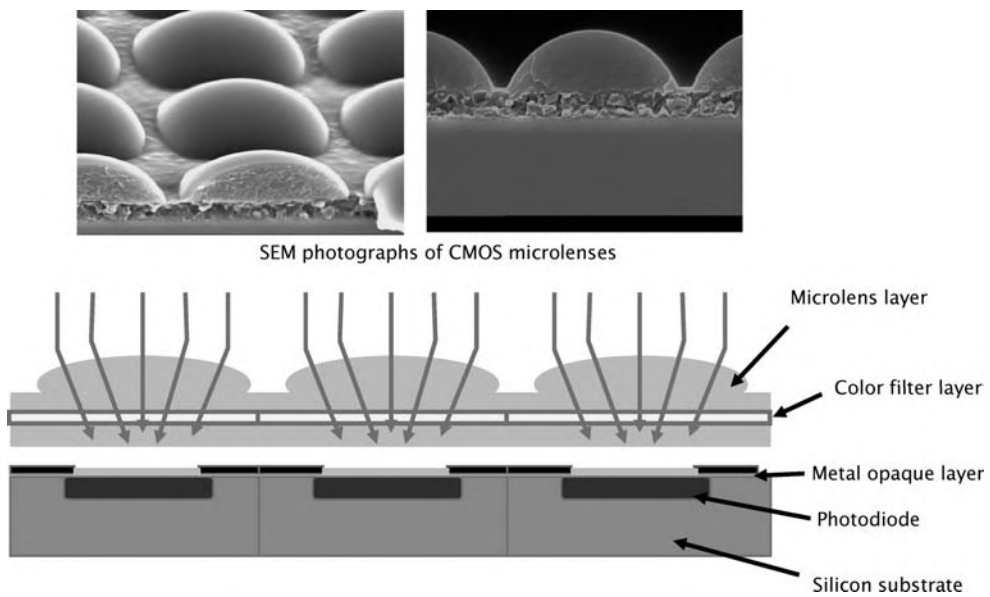


**Figure 16.40** Microlens arrays for CMOS sensors

In an OPU, the laser beam is usually split into three beams, two for tracking and the center one for reading and focus control through quad detectors. The beam splitting is usually done by a holographic grating etched directly onto the laser package (see Figure 16.41).

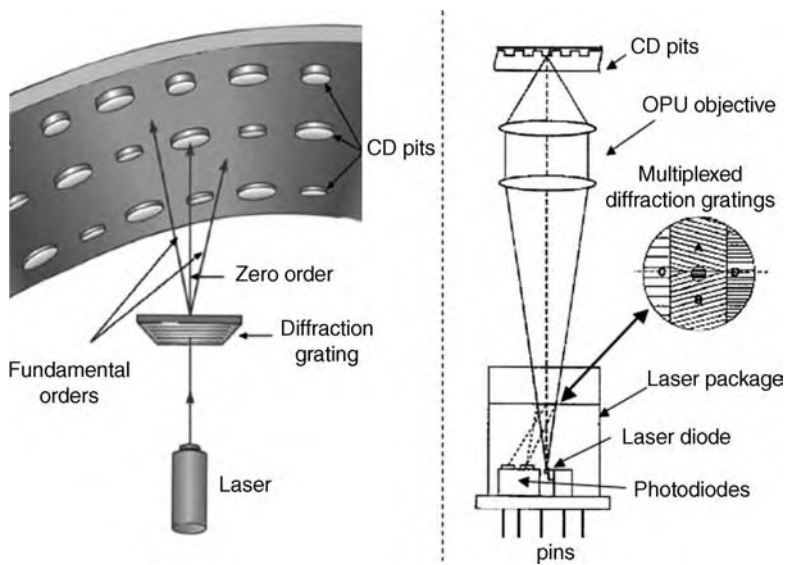


**Figure 16.41** A holographic laser package for beam splitting (tracking)

16.8.2.2 Hybrid Multifocus Lenses for CD/DVD/Blu-ray OPU

References: Chapters 7 and 10

Optical Pick-up Unit (OPU) microlenses are used to read optical storage media such as the CD, DVD or Blu-ray. Such lenses have usually large NAs, and come in different formats. A simple lens for a single wavelength and a single NA is easy to design and fabricate. A lens that has to be capable of reading different media with different NAs, and compensate for different spherical aberrations (due to different disk media thicknesses) over different wavelengths (due to different media dies for RW capabilities) is more difficult to design and fabricate. Such lenses are dual- or triple-focus lenses (see Figure 16.42). See also Chapter 9 for more insight on how these lenses can be fabricated.

Some such lenses are designed and fabricated as hybrid refractive/diffractive lenses (see Chapter 7), while others are fabricated as pure diffractive lenses (see Chapter 5). Figure 16.43 shows such a microlens, which in most cases are produced by lithographic means.

Figure 16.44 shows the hybrid refractive/diffractive dual-focus lens described in Chapter 7 integrated in a CD/DVD OPU. The various quad detectors for focus control and the different laser diodes are also depicted.

Such an integration is a marvel of micro-technology, utilizing every aspect of modern optical technology (multiple-wavelength lasers, integrated Si detectors, dichroic filters, micro-prisms and microlenses, hybrid lenses, holographic gratings, head-gimbal assemblies, fast focus control feedback, fast tracking feedback and so on) all of which are integrated in a translating aluminum assembly selling for less than five dollars.

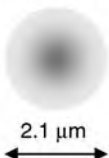
| Drive | CD/CDRW                                                                                                  | DVD/DVDRW                                                                                                | Blu-ray                                                                                                  |
|-------|----------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------|
| OPU   |                       |                       |                       |
| Spot  | <br>2.1 $\mu\text{m}$ | <br>1.3 $\mu\text{m}$ | <br>0.6 $\mu\text{m}$ |
| Specs | $\lambda = 780 \text{ nm}$<br>NA = 0.45<br>1.2 mm disk<br>0.7 GB                                         | $\lambda = 650 \text{ nm}$<br>NA = 0.60<br>0.6 mm disk<br>4.7 GB                                         | $\lambda = 405 \text{ nm}$<br>NA = 0.85<br>0.1 mm disk<br>27 GB                                          |

Figure 16.42 The three lens specifications for CD/DVD/Blu-ray OPU read-out

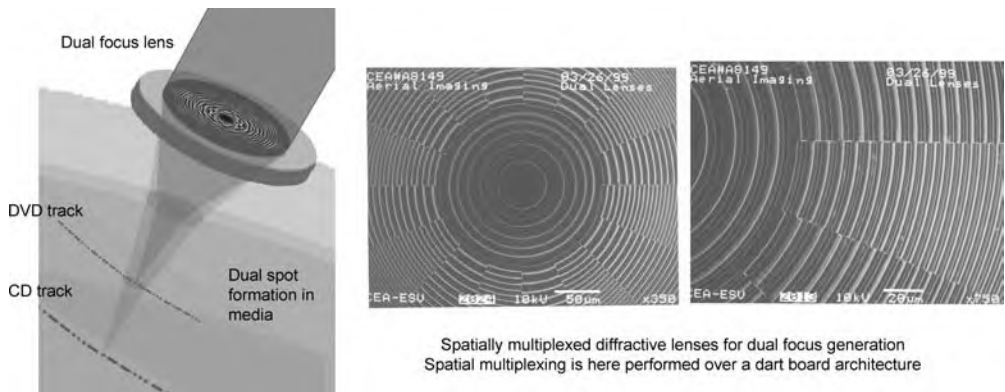


**Figure 16.43** A diffractive microlens for dual-focus generation of CD/DVD OPUs

**16.8.2.3 High-NA Lenses (MO and SIL)**

Reference: Chapter 6

High-NA diffractive lenses have also been developed for magneto-optical drives in Winchester head configurations. Figure 16.45 shows such high-NA diffractive lenses.

Other near-field optical lenses have been developed for high-capacity MO disks. However, this technology seems to be slowing down due to advances in Blu-ray media and holographic media storage technologies.

**16.8.2.4 Holographic Page Data Storage**

Reference: Chapter 8

Holographic page data storage has been discussed in Chapter 8. This application is based on the fact that many holograms can be recorded on the same spot in a holographic medium by angular multiplexing.

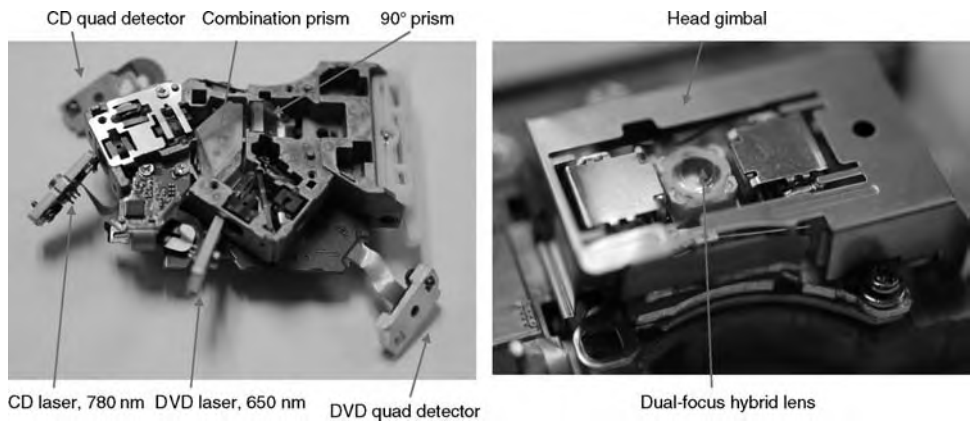


**Figure 16.44** A dual-focus hybrid lens in a CD/DVD OPU unit

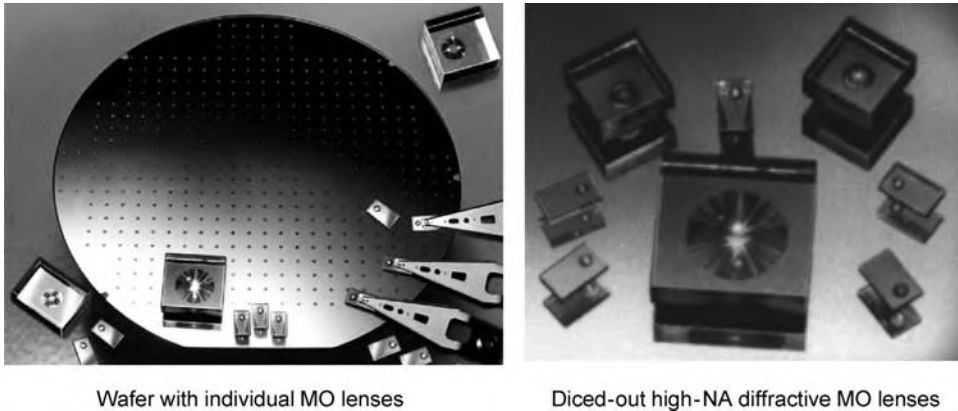


**Figure 16.45** High-NA magneto-optical drive lenses in Winchester flying heads

The effect uses the finite angular bandwidth of the transmission volume hologram. InPhase Inc. of Colorado has been the first company to produce such a holographic page data storage system on the market (see Figure 16.46).

There are several challenges to page data storage, such as the crosstalk that takes place between angular multiplexed pages, and the angular and lateral beam steering speeds that are required for data read-out requirements in today's computing applications.

### 16.8.3 Consumer Electronic Displays

Consumer electronics displays include backlit and edge-lit LCD screens used anywhere from large panel TVs down to laptop screens and even cell phones. Projection displays are a fast-moving market and are becoming an important user of laser sources and digital diffractive optics for conventional projection or diffractive projection, especially in mini- or 'pico-' projectors. Finally, 3D displays and other virtual displays are much smaller markets, which include numerous digital optical elements, as we will see in the following sections.

#### 16.8.3.1 The Virtual Keyboard

References: Chapters 6 and 9

The diffractive virtual keyboard was introduced by Canesta early in the present decade, and is now licensed to many contract manufacturers around the world. It is a very impressive device, which projects a diffractive keyboard (a Fourier CGH element) coupled to an IR finger detection system that produces an IR light sheet propagating across the table on which the visible keyboard is projected. The IR light sheet can be produced either by a beam-shaping micro-refractive or by a diffractive element (see Figures 16.47 and 16.48). Figure 16.47 shows a Celluon device.

Figure 16.48 depicts the operating principle of the virtual keyboard with both lasers (the red laser producing the diffractive keyboard and the IR laser producing the light sheet). The IR light is reflected by the fingers and detected by Si photo detectors on the projector base, which enables the detection of the angular direction of the reflected beams by triangulation.

However, this virtual keyboard has never been a real commercial success, mainly because there is no return action from a key stroke (no haptics) and because a perfectly planar surface is needed for it to work

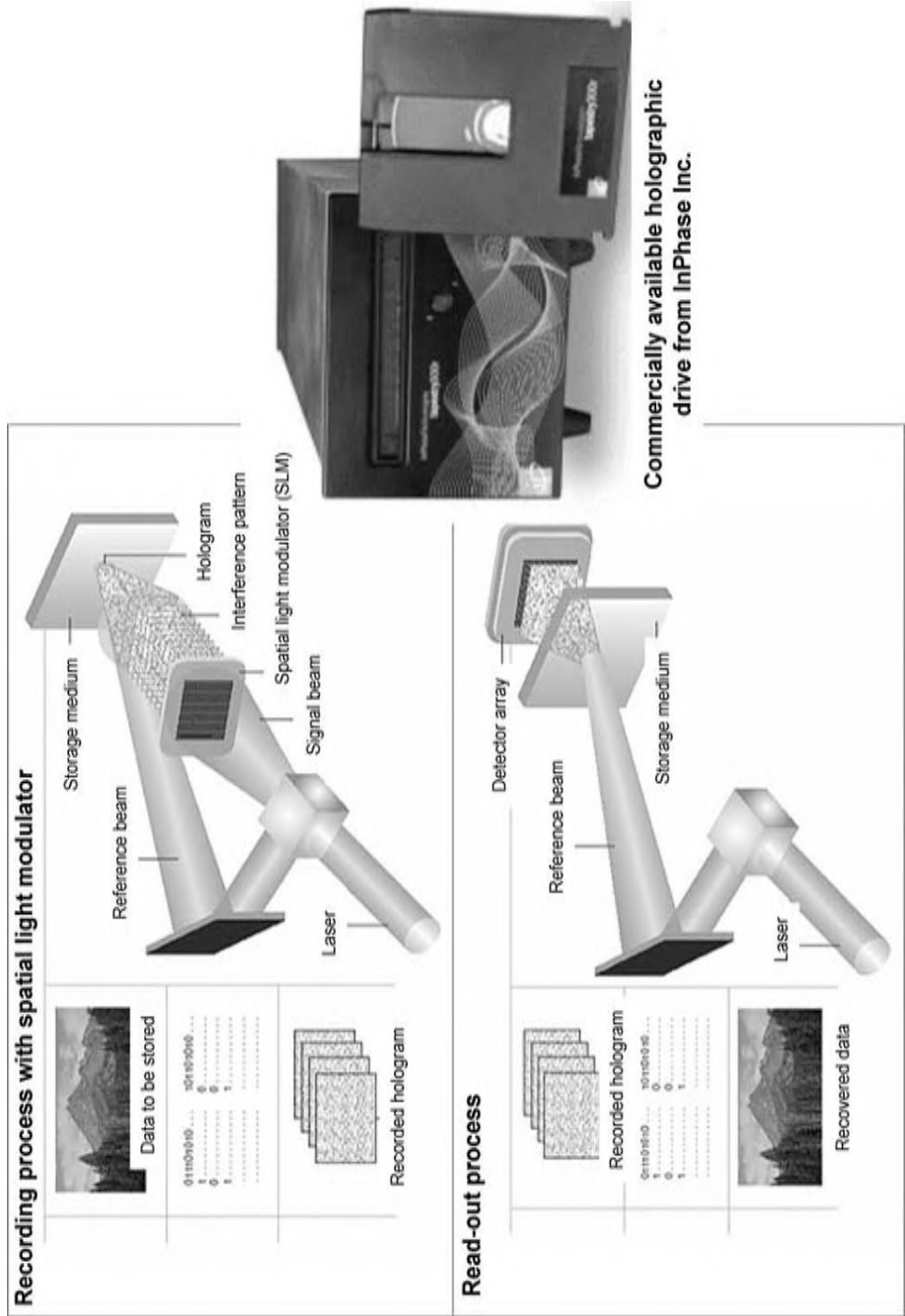


Figure 16.46 The commercially available InPhase Inc. holographic page data storage system

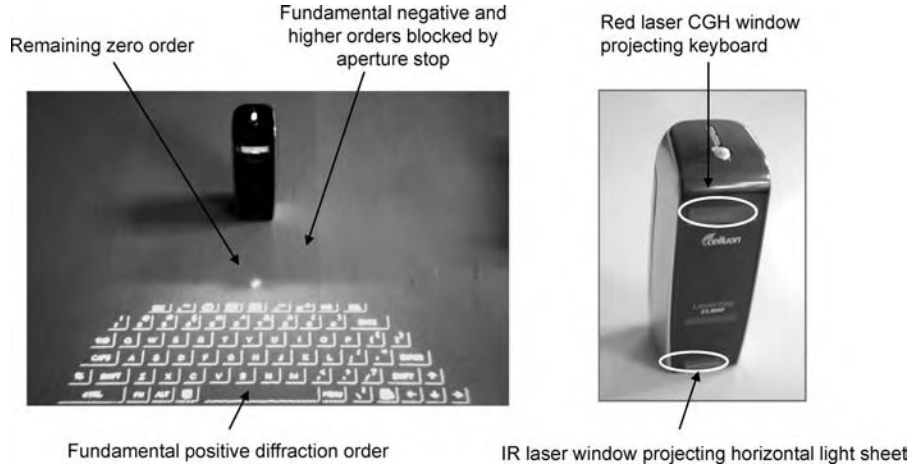


**Figure 16.47** A diffractive virtual keyboard

properly. It remains a high-priced device that does not produce any added value when it is used to replace a foldable computer keyboard. However, it can produce added value when it is used as a switchable interface; for example, in entertainment or transportation (in a car dashboard, an airplane cockpit etc.), where physical interfaces are too bulky to be used. Figure 16.49 shows the projection of a virtual iPod command pad and an audio/video control interface for automotive application.

The virtual console can also be used in complex machinery that requires a time-multiplexed display of various interfaces and commands. Finally, in automotive applications, the detection of the finger position has to be done on uneven surfaces or on curved surfaces (i.e. a dashboard), which is not possible using the above-mentioned technique. Many finger detection schemes are possible, including fiber arrays used in VR, arrays of diffracted beams, the use of structured illumination and so on.

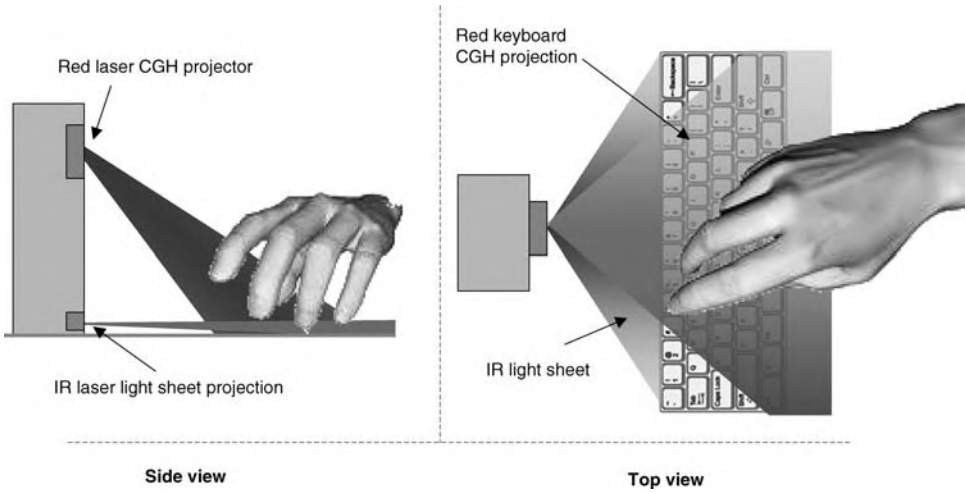


**Figure 16.48** The operational principle of the diffractive keyboard



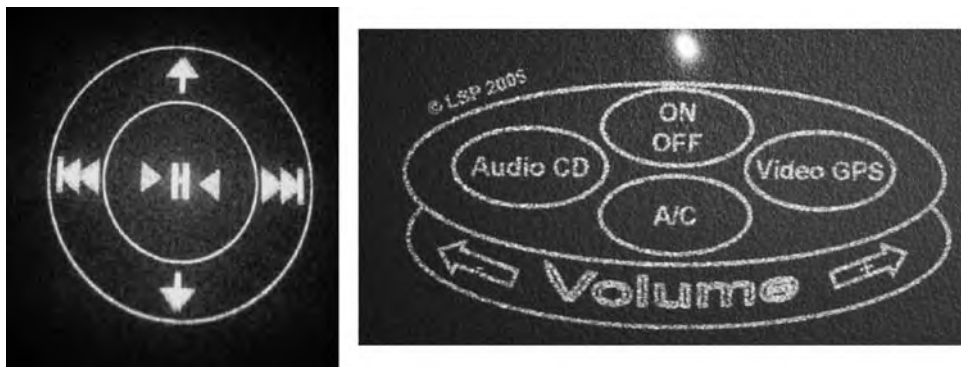


**Figure 16.49** Other types of virtual consoles

### 16.8.3.2 Three-dimensional Displays

References: Chapters 4 and 8

Three-dimensional displays are a large user of micro-optics and other diffractive optics, not to mention 3D display holography.

#### *Lenticular Stereo Displays*

References: Chapters 4 and 5

Stereo displays are implemented via cylindrical or spherical lenticular arrays placed on a sliced photograph set (see Chapter 4) or directly on an LCD computer screen. Such arrays are usually micro-refractive, due to the broadband operation required. However, for laser-based displays and LED backlit displays, diffractive lenticular arrays can also be used, and are easier to manufacture, with tighter tolerances.

#### *Holographic Video*

References: Chapters 8 and 9

Holographic video can be produced by various means. The first attempt was made by the use of crossed acousto-optical modulators at MIT, by Steven Benton and Pierre St Hilaire. Other attempts use reconfigurable digital diffractives, as discussed in Chapter 9. However, there are many 3D dynamic display systems that have been implemented in industry that do not rely on digital optics or micro-optics.

### 16.8.3.3 LCD Displays

Manufacturers of LCD displays for computer screens, flat panel TVs, cell phones and other handheld devices are coming under pressure to increase their efficiency, since battery power still remains a critical issue in portable electronics, especially when a display is used.

Edge-lit displays using LEDs are slowly becoming the standard illumination scheme today for most LCD display applications. However, the light extraction technologies used are not optimal and waste most of the light coupled by the LED on the side of the display. Digital optics, and especially sub-wavelength slanted grating structures, can provide better light extraction efficiencies. Current efforts are being applied to these technologies (see Chapters 10 and 14).

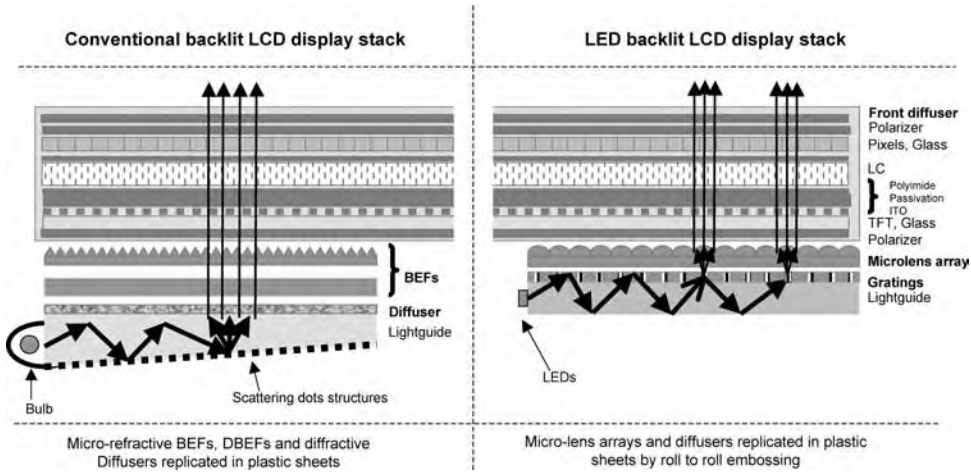


**Figure 16.50** Edge illumination of an LCD screen and light extraction through BEF sheets and holographic diffusers

### Brightness Enhancing Films (BEF)

References: Chapters 5, 9, 10 and 14

Brightness Enhancing Films (BEF) are also referred to as prism sheets or lens films. They are plastic replicated micro-prism films placed under the edge-lit LCD module that enhance the luminance of the display (see Figure 16.50).

Dual Brightness Enhancement Film (DBEF) is a 3M proprietary technology which recycles the light that is normally lost in the rear polarizer. DBEF can be laminated at the back of the LCD window. Replacements for DBEF include holographic diffusers (see Figure 16.50).

### Diffuser Films

References: Chapters 6, 8 and 14

Diffuser films are one of the basic elements of the backlight LCD unit (see Figure 16.50). Normally, backlight units require at least two diffuser films: a top film and a bottom film. The top diffuser film is placed on top of the BEF to protect the BEF. The bottom diffuser is placed between the BEF and a light guide plate for light diffusion. Both diffuser films are PET (polyester) or PC (polycarbonate) based. The top diffuser has higher technical requirements and is priced higher than the bottom diffuser film. In LCD monitors, many panel makers and backlight makers are using a three-diffuser film structure to replace the traditional 'diffuser + BEF + diffuser' structure. In LCD TVs, due to the emerging acceptance of lower brightness, many panel makers are considering the use of a bottom diffuser film to replace the DBEF.

Reflector films are another basic element of the backlight LCD unit. A backlight unit typically requires at least one reflector. The function of the reflector film is to reflect and recycle the light from the side of the light guide plate of the backlight. The raw material of the reflector film is PET (polyester). There are two kinds of reflector films: a white reflector used in all applications and a silver reflector used for small/medium and notebook PC applications.

## 16.8.4 Projection Displays

There are high expectations for LED- and laser-based projection displays, both in front (conventional projectors and pocket projectors) and in rear projection (RPTV) architectures. Current technologies such as

plasma screens and high-pressure arc lamp projection engines are becoming obsolete (they are expensive; they break easily, produce lots of heat, are not environmentally friendly, are bulky, have low color depth and so on). Besides, as display screens get larger, projection engines get smaller and prices are falling.

The first step in the direction toward pico-projectors was to replace the dangerous, bulky and fragile high-pressure mercury lamps by LEDs in the light engine. The next step in digital projectors is the introduction of RGB laser sources to replace LEDs. Lasers produce better color saturation, and are also more efficient than LEDs, thus providing longer battery life, which is critical in portable devices. The introduction of lasers also opens the door to the introduction of digital optics, and especially digital diffractive optics. Since there are only three very thin spectral lines, digital optics and diffractive optics can produce efficient optical functionalities for full color projection, in a small footprint, not only in the light engine but also on the imaging side. However, as for any laser display application, one of the first tasks on hand is to reduce the speckle created by such coherent sources.

#### 16.8.4.1 Speckle Reduction

References: Chapters 6 and 9

The traditional way to reduce speckle (or rather to average the speckle patterns within the integration time of the eye) is to vibrate the screen or use a rotating diffuser (see Chapter 9) within the illumination section, which can be described as phase diversity. Other speckle reduction methods include angular diversity, polarization diversity, amplitude diversity, laser source diversity, wavelength diversity and so on. In source diversity, the speckle level reduces as the square root of the number of sources. However, even though one has taken on the task of reducing the speckle of the projection device (which in all cases is the objective speckle), one has to further take care of the subjective speckle that is created by the viewing device (the human eye). Subjective speckle is more complicated to tackle than objective speckle.

#### 16.8.4.2 Pico-projectors

Pico-projectors are new breeds of digital projectors using either LED or laser light engines and standard microdisplays (DLP, High Temperature Poly Silicon LCD or LCoS). Pico-projectors can be implemented in various ways:

- conventional imaging projectors;
- laser scanning projectors;
- diffractive projectors; or
- edge-lit holographic projectors.

##### *Conventional Imaging Pico-projectors*

LED pico-projectors based on DLP and LCoS are now being integrated in cell phones with up to 20 or 30 lumens, which is enough to project a 2 foot image at a distance of 3 feet, in a dark environment.

Figure 16.51 shows some of the DLP and LCoS conventional imaging pico-projectors available on the market today. Their light engines are all LED based.

Many companies are proposing their projectors as development kits that can be integrated in specific applications such as structure light projection for 3D sensing, or HUD applications (see Figure 16.51, on the lower right).

##### *Laser Scanner Pico-projectors*

Laser scanner pico-projectors are also being brought to the market today. Figure 16.52 shows the Microvision pico-projector, which is based on a single MEMS micromirror and colors time sequencing over three lasers. This projector is also proposed as a development kit to be integrated in specific applications, not necessarily for conventional display applications.



**Figure 16.51** Conventional imaging pico-projectors

Laser scanners have the great advantage of being capable of producing a far-field image, similar to that produced by diffractive pico-projectors, with large angles. Besides, unlike diffractive projectors, laser scanner projectors do not need to compute the Fourier transform of the image.

***Diffractive Pico-projectors***

Diffractive pico-projectors, as their name implies, diffract the image rather than producing an image of a microdisplay. Here, the microdisplay is used to produce the desired phase map, which will diffract the

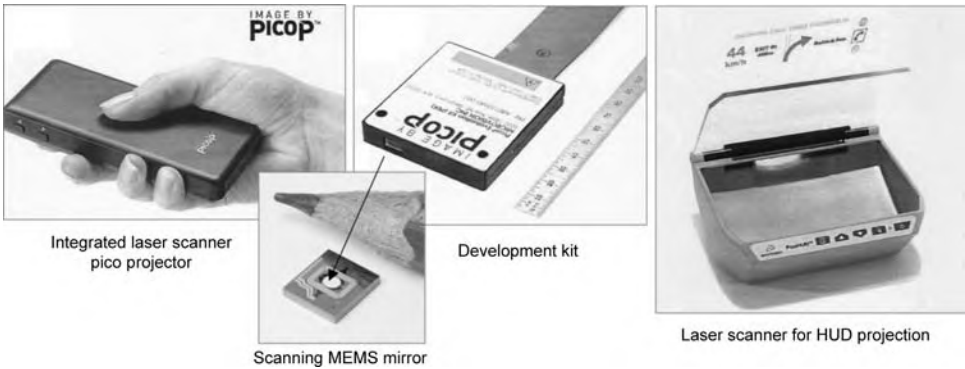


**Figure 16.52** Laser scanner based pico-projectors

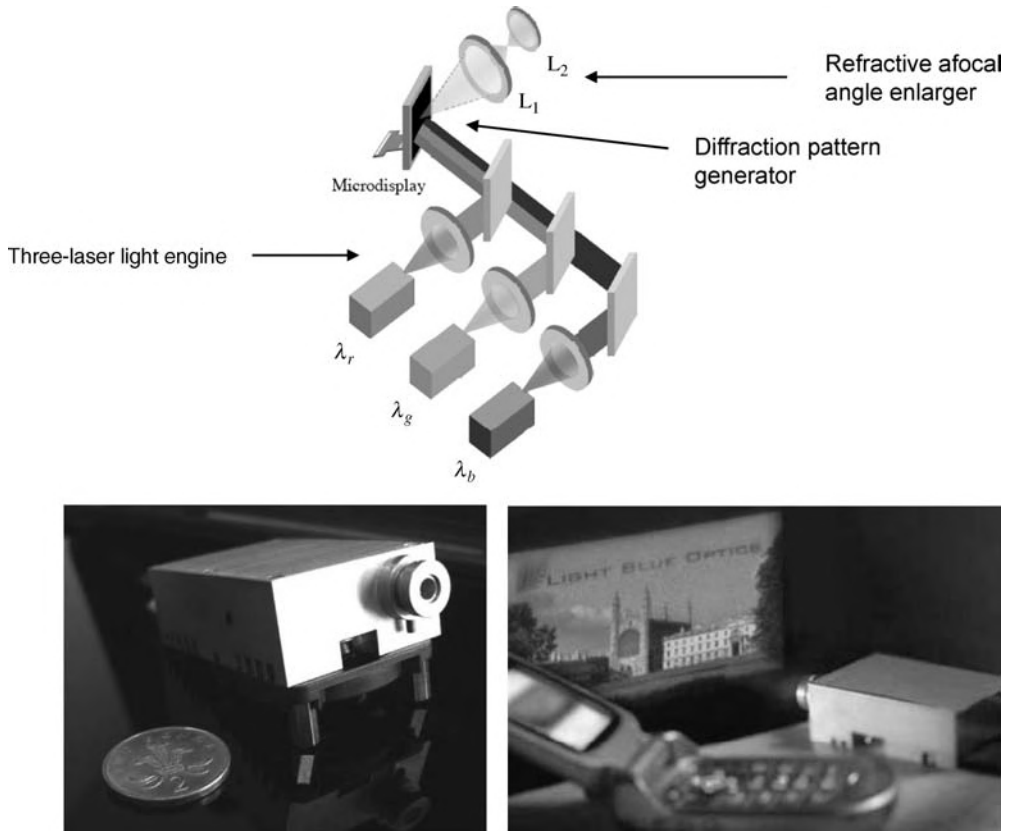


**Figure 16.53** Diffractive pico-projectors

target far-field pattern (Fourier CGH). In a first approximation, the image on the microdisplay is the phase part of the Fourier transform of the real image to be projected. Such a microdisplay can be a reflective LCoS device, which can produce a phase shift of up to  $2\pi$  over the entire visible spectrum. Figure 16.53 shows such an LCoS diffractive pico-projector and its internal configuration, in a front view (left) and in operational mode (right).

Most of these pico-projectors are being sold today as development kits, rather than being implemented in final end-user applications such as cell phones or personal video projectors. It is not quite clear if the real market for such pico-projectors is video projection, as it was for conventional projector technologies, or if the market is rather leaning toward more specific niche applications, such as automotive HUD displays or structured illumination projectors for 3D sensors.

Since a diffractive projector generates a diffracted image in the far field (see Chapter 9), there is theoretically no need for an objective lens and/or a zoom lens, and thus the size of the projector can remain very small. However, due to the small diffraction angles generated by dynamical diffractives, an additional optical compound element is usually necessary to increase these angles, and can take on the size of a small objective lens. Such afocal angle enlargers can be simply inverted telescopes. Diffractive pico-projectors are implemented today with reflective LCoS microdisplays (see Figure 16.53), H-PDLC displays or grating light valve (GLV) MEMS elements.

**Table 16.1** A comparison of various laser-based pico-projector architectures

| Projector technology                                | Technology                   | Source           | Projection | Advantages                            | Disadvantages                           | Digital optics                    |
|-----------------------------------------------------|------------------------------|------------------|------------|---------------------------------------|-----------------------------------------|-----------------------------------|
| Pico-projector, imaging<br>Scanning laser projector | DLP or HTPS microdisplay     | LED/Laser        | Near field | Conventional technology               | Expensive (DLP), large                  | Used in light engine              |
|                                                     | MEMs or acousto-optic        | Laser RGB        | Far field  | No need for objective or microdisplay | Limited to wire framed shapes, speckle  | Can be used in $x$ - $y$ scanning |
| Pico-projector, diffracting,                        | Reflective LCoS microdisplay | Laser RGB        | Far field  | Far field, no need for objective      | Computing power, small angles, speckle, | Used to form the image            |
| Pico-projector, imaging                             | Edge-lit H-PDLC              | Laser or LED RGB | Near field | No microdisplay, see through          | Small projection size                   | Used in light engine              |
| Pico-projector, diffractive                         | Sub-wavelength H-PDLC array  | Laser RGB        | Far field  | No microdisplay, no objective lens    | Non conventional technology             | Used to form the image            |

### 16.8.4.3 Challenges for Diffractive Pico-projectors

References: Chapters 6, 8 and 9

The main challenges for the implementation of diffractive pico-projectors in consumer electronics devices such as cell phones, portable media players or laptops are summarized below:

- heavy computing power required for real-time 2D Fourier transformation of a continuous video stream;
- remaining zero order due to phase inaccuracies in the phase microdisplay;
- multiple orders diffracted (higher and conjugate orders);
- small diffraction angles due to (relatively) large pixels;
- nonuniform illumination compensation from frame to frame (depending on the image content density); and
- parasitic speckle generations, both objective (projector related) and subjective (viewer related).

Table 16.1 compares the various LED and laser-based digital pico-projector architectures that have been developed today.

Today, the only pico-projectors available on the market are based on conventional imaging technologies (LCoS and Ti DLP) and laser scanners (Microvision). Diffractive pico-projectors still have to be transferred from research to industry.

In this chapter we have reviewed several technologies based on binary optics which can be applied to computing devices, such as:

- Diffractive virtual keyboards and other input interfaces;
- Diffractive laser projection displays;
- Holographic page data storage;
- Optical interconnections inter- or intra-chip;
- Optical Ethernet lines.

As the current trend in the computing industry is towards fully portable miniature computing devices that do not sacrifice keyboard or display sizes, there is only one computing device (see Figure 16.54) which incorporates a diffractive optical keyboard and a diffractive laser projection display.

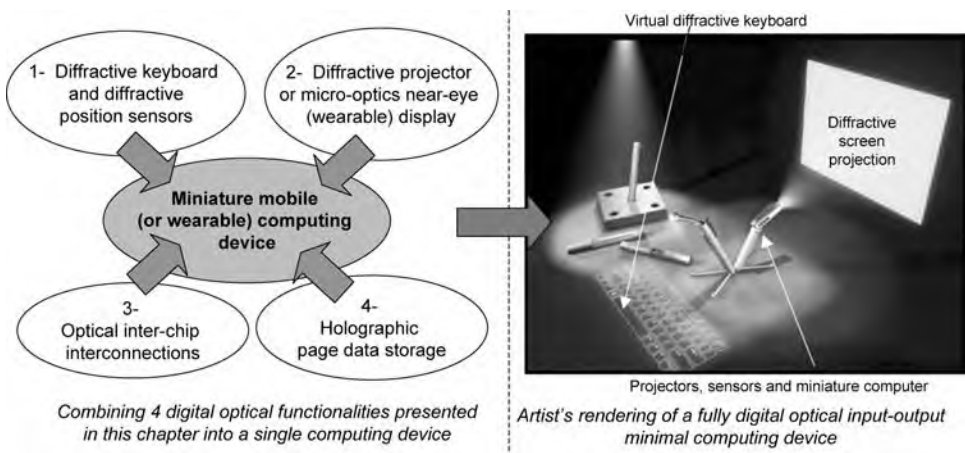


**Figure 16.54** The fully optical input-output portable computing experience.

Although such a device is not yet proposed as a product in industry and is only an artist's rendering, it might only be a few years before it could become available to the consumer electronics market. As we have seen in previous sections this technology is constantly progressing.

## 16.9 Summary

Table 16.2 summarizes the various applications and how to find the respective adequate design, modeling and fabrication techniques in this book.

## 16.10 The Future of Digital Optics

Technological waves and investment hypes have historically generated interest in digital optics technology during the past three decades. Such waves can either be based on healthy markets fuelled by real industrial and consumer needs or on inflated needs generated by venture capital investment.

One has to remember that too much hype for a new technology almost always spells doom for the technology, as has happened early in the present decade with the optical telecom bubble, where investors, entrepreneurs and integrators alike were all singing the praises of the technology before the deadly fall that took place in 2003 (see Figure 16.55). The figure also shows the historical stock chart value for JDU Uniphase (JDSU), which illustrates the phenomenon quite well.

At this point, let us quote *Forbes Magazine* in August 2000, three months after the Internet bubble had crashed and investors were savvy about another bubble:

*After more than 150 years of tireless service in telecommunications, the electron is getting the boot. Suddenly everyone wants this resilient little particle out of their lives. These days, light is king, and the new kingdom—still very much in the making—is the all-optical network, a land of waving fields of bandwidth and bountiful riches.*

*Forbes Magazine, August 2000.*

Figure 16.56 depicts the various technologies that have been triggering successive waves of interest in digital optics, together with descriptions of these various waves:

- *Optical security devices* and *optical data storage* were the first waves to generate interest in diffractives (other than spectroscopy applications). Optical data storage is actually generating sustained interest in diffractives, from the CD signal tracking of the 1980s, to the dual CD and DVD read-out heads of the 1990s, to today's Blu-ray technology and holographic page data storage.
- Optical computing generated a lot of interest in diffractives in the early 1990s, but has never been able to bring diffractives into the mainstream market (see Section 16.2.4). However, Intel has recently announced that optical interconnections are one of the ten key solutions for the development of tomorrow's computing industry.
- The investment hype of the late 1990s in *optical telecoms* – and especially in DWDM Mux/Demux devices – has ill-served interest in digital optics, with the consequence of a disastrous bubble, which burst in 2003. A recent revival of interest in the 10 Gb/s optical Ethernet is correcting this market trajectory.
- *Industrial optical sensors* have shown the most steady growth during past decades, and are still one of the most desirable and least risky markets for the implementation of diffractive optical technology. For example, the optical encoder (motion sensor) market was valued at \$3.5 billion in 2008.



**Table 16.2** A summary of the various applications reviewed in this chapter

| Sector            | Market                 | Product                                | Status         | Reference Chapter |
|-------------------|------------------------|----------------------------------------|----------------|-------------------|
| INDUSTRIAL        | Industrial metrology   | Holographic non destructive testing    | Available      | 8                 |
|                   |                        | Structured illumination for 3D sensing | In development | 6                 |
|                   |                        | Digital holography for 3D sensing      | Available      | 8                 |
|                   | Laser material process | Beam sampling                          | Available      | 5                 |
|                   |                        | Beam marking/engraving                 | Available      | 6                 |
|                   |                        | Welding, cutting                       | Available      | 6                 |
|                   |                        | Surface treatment                      | Available      | 6                 |
|                   |                        | Chip on board multispot welding        | In development | 6                 |
|                   | Industrial packaging   | Diffraction wrapping paper             | Commodity      | 5                 |
|                   |                        | Holographic tags                       | In development | 6                 |
|                   | IC manufacturing       | Phase shift masks                      | Available      | 13                |
|                   |                        | Off axis illumination                  | Available      | 13                |
|                   |                        | Diffraction PCMs                       | In development | 6                 |
|                   |                        | Maskless projection                    | In development | 13                |
|                   |                        | Direct laser write beam shaping        | In development | 6                 |
|                   | LED lighting           | PC LED                                 | Available      | 10                |
|                   |                        | LED beam extraction                    | Available      | 10                |
|                   |                        | LED beam collimation/shaping           | Commodity      | 6                 |
|                   | Automotive             | HUD systems                            | Available      | 8                 |
|                   |                        | Virtual command pads                   | In development | 6                 |
|                   |                        | Contact-less switches                  | In development | 5                 |
|                   |                        | 3D sensing                             | Available      | 6                 |
|                   |                        | LED lighting                           | Commodity      | 6                 |
| DEFENSE/<br>SPACE | Defense/Weaponry       | HMD                                    | Available      | 10                |
|                   |                        | Holographic gun sights                 | Available      | 6                 |
|                   |                        | IR missile optics                      | Available      | 7                 |
|                   | Astronomy              | Laser weapons                          | In development | 5                 |
|                   |                        | Optical cloak                          | Research       | 10                |
|                   |                        | Wavefront sensors                      | Available      | 4                 |
|                   |                        | Null CGHs                              | Available      | 5                 |
|                   |                        | Large-optics alignment optics          | Available      | 5                 |
|                   | Homeland security      | 3D sensing for biometrics              | In development | 6                 |
|                   |                        | Gas sensors                            | Available      | 10                |

(Continued)

**Table 16.2** (Continued)

| Sector                    | Market                  | Product                                    | Status         | Reference Chapter |
|---------------------------|-------------------------|--------------------------------------------|----------------|-------------------|
| CLEAN<br>TECHN-<br>OLOGY  | Optical<br>computing    | Chemical sensors                           | Available      | 10                |
|                           |                         | Camera networks                            | In development | 4                 |
|                           |                         | Optical security devices                   | Commodity      | 5                 |
|                           |                         | Optical clock broadcasting                 | Available      | 10                |
|                           |                         | Optical interconnections                   | In development | 6                 |
|                           |                         | Optical image processing                   | Research       | 6                 |
|                           | Solar PV cells          | AR structures                              | Available      | 10                |
|                           |                         | Solar concentrators                        | Available      | 5                 |
|                           |                         | Solar tracking                             | Available      | 5                 |
|                           | Wireless laser<br>power | Photon trapping                            | In development | 10                |
|                           |                         | Beam steering                              | In development | 9                 |
|                           | NIF                     | Laser damage gratings                      | Available      | 5                 |
| FACTORY<br>AUTO<br>MATION | Position sensors        | Large aperture digital<br>lenses           | Available      | 5                 |
|                           |                         | Diffractive linear encoders                | Available      | 6                 |
|                           |                         | Diffractive rotational<br>encoders         | Available      | 6                 |
|                           |                         | Diffractive multidimen-<br>sional encoders | In development | 6                 |
|                           |                         | Origami objective lenses                   | In development | 7                 |
|                           |                         | Stacked wafer level<br>camera objectives   | In development | 4                 |
|                           |                         | Wavefront coding camera<br>objectives      | Available      | 9                 |
|                           |                         | Diffractive auto-focus<br>device           | Commodity      | 6                 |
|                           |                         | CMOS lens arrays                           | Commodity      | 4                 |
|                           |                         | Diffractive SLR camera<br>view finders     | Available      | 9                 |
|                           | Optical data<br>storage | Holographic grating lasers                 | Commodity      | 5                 |
|                           |                         | Hybrid dual focus<br>OPU lens              | Commodity      | 7                 |
|                           |                         | High NA MO lenses                          | Gone           | 5                 |
|                           |                         | Holographic page<br>data storage           | Available      | 8                 |
|                           | Projection<br>displays  | Virtual keyboard                           | Available      | 6                 |
|                           |                         | 3D displays                                | Available      | 4                 |
|                           |                         | LCD displays                               | Available      | 10                |
|                           |                         | Pico projectors                            | In development | 6                 |
| OPTICAL<br>TELECOMS       | Optical<br>telecoms     | DWDM Mux/Demux                             | Available      | 3                 |
|                           |                         | EDFA gain equalizers                       | Available      | 3                 |

(Continued)

Table 16.2 (Continued)

| Sector               | Market           | Product                            | Status          | Reference Chapter |
|----------------------|------------------|------------------------------------|-----------------|-------------------|
| BIOTECHNOLOGY        |                  | CATV/DWDM splitters                | Available       | 5                 |
|                      |                  | CATV broadcasting                  | Available       | 6                 |
|                      |                  | Polarization combiners             | Available       | 10                |
|                      |                  | Optical transceiver blocks         | In development  | 5                 |
|                      |                  | 10Gb Ethernet                      |                 |                   |
|                      | Fiberless fiber  | Vortex lens fiber couplers         | Available       | 5                 |
|                      |                  | Beam collimation and detection     | Available       | 5                 |
|                      | Biomedical       | Intra-ocular hybrid lens           | Commodity       | 7                 |
|                      |                  | Endoscopic imaging                 | In development  | 5                 |
|                      |                  | Flow cytometry                     | Available       | 6                 |
|                      |                  | Diffraction coherence tomography   | Available       | 8                 |
|                      | Laser treatment  | Laser skin treatment               | In development  | 6                 |
|                      |                  | Chirurgical laser                  | Available       | 5                 |
|                      | Biotech research | Integrated optical sensors         | Available       | 10                |
|                      |                  | Immunoassay sensors                | Available       | 6                 |
|                      |                  | Optical tweezers                   | Available       | 9                 |
| ENTERTAINMENT        | Entertainment    | Laser pointer pattern generators   | Ultra-commodity | 6                 |
|                      |                  | Diffraction food                   | Ultra-commodity | 5                 |
|                      | Packaging        | Optical Variable Devices           | Available       | 5                 |
|                      |                  | Holographic projector screens      | Available       | 10                |
| CONSUMER ELECTRONICS | Digital cameras  | Hybrid SLR camera objective lenses | Available       | 7                 |

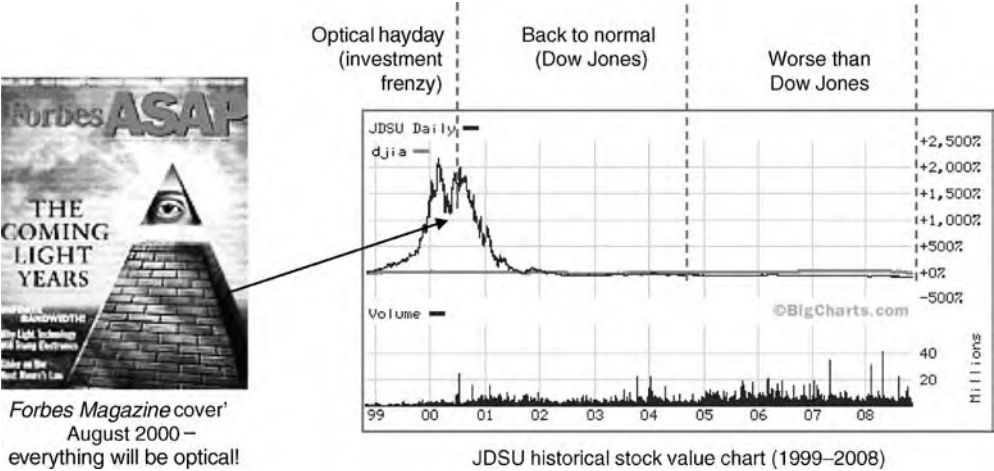


Figure 16.55 The heyday of optical technology, as per *Forbes Magazine* in 2000

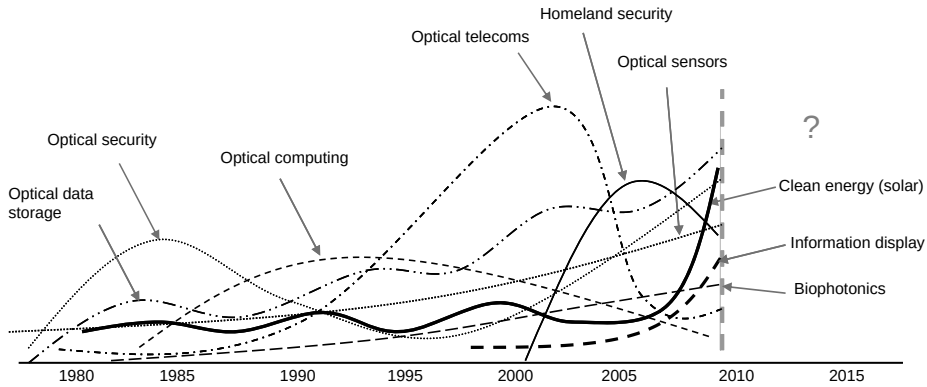


**Figure 16.56** Optical technological waves that have fuelled interest in digital optics over the past 30 years

- The *laser material processing* market (cutting, welding, marking, engraving, heat treatment etc.) crossed the \$6 billion mark in 2008, and is expected to grow steadily throughout 2012. Diffractive optics have an enormous potential in this sector of industry, which is not yet fully recognized at the present time. Diffractive optics have an enormous potential in this sector of industry, which is not yet fully recognized today.
- *Biophotonics* has shown considerable interest in diffractives very recently, and this interest is growing fast, especially in sensor applications.
- *Homeland security* applications were triggered in a special way after 9/11. Emphasis was put on the use of diffractives to implement gas sensors and other chemical sensors. Today's homeland security applications also include HMDs and HUDs, and other near-eye displays for tomorrow's land warrior.
- The current interest in *information display*, and especially in laser-based projection displays (rear and front projection engines), have the potential to build a large market and a stable application pool for the use of diffractive optics for the next decade. Laser- and LED-based RPTV, front projectors, pocket projectors and pico-projectors are expected to yield a \$6 billion market before the end of the decade. There will be lots of niche markets for pico-projectors, such as 3D sensors through structured illumination and HUD displays.
- Finally, *clean energy* – and especially solar energy – is currently the key focus of new venture capital firms for optical technologies such as diffractives and holographics. Any technology that can increase the efficiency of solar cells by a tenth of a percent is worth funding. The main functionalities that digital optics can implement are anti-reflection surfaces, solar concentrators and passive tracking or photon trapping.

The future of digital optics lies in the hands of the global market, which produces demands for better, cheaper and smaller devices, and in the hands of the technology enablers, the entrepreneurs and foundries, which are providing the adequate design and fabrication tools for new and ever-improving digital optical elements, especially in their dynamic form for consumer electronic products. Finally, the fate of digital optics also lies in the hands of high-tech investors and market analysts, who can either curb or artificially increase the real market demand, and thus produce potentially dangerous technological bubbles.

## References

- [1] B. Kress, 'Diffractive Optics Technology for Product Development in Transportation, Display, Security, Telecom, Laser Machining and Biomedical Markets', Short Course, SPIE SC787, 2008.
- [2] L.R. Lindvold, 'Commercial Aspects of Diffractive Optics', Short Course, OVC ApS, CAT Science Park, Frederiksborgvej 399, P.O. Box 30, DK 4000 Roskilde, Denmark.
- [3] J. Turunen and F. Wyrowsky, '*Diffractive Optics for Industrial and Commercial Applications*', Akademie Verlag, Berlin, 1997.
- [4] B.P. Thomas and S. Annamala Pillai, 'A portable digital holographic system for non-destructive testing of aerospace structures', Proceedings of the International Conference on Aerospace Science and Technology, June 26–28, 2008, Bangalore, India.
- [5] A. Haglund, A. Larsson, P. Jedrasik and J. Gustavsson, 'Sub-wavelength surface grating application for high-power fundamental-mode and polarization stabilization of the VCSELs', in '*10th IEEE Conference on Emerging Technologies and Factory Automation, 2005*', ETFA 2005, Vol. 1, 1049.
- [6] J. Pfund, N. Lindlein, J. Schwider *et al.*, 'Absolute sphericity measurement: a comparative study on the use of interferometry and a Shack–Hartmann sensor', *Optics Letters*, **23**, 1998, 742–744.
- [7] J. Pfund, N. Lindlein and J. Schwider, 'Dynamic range expansion of a Shack–Hartmann sensor by using a modified unwrapping algorithm', *Optics Letters*, **23**, 1998, 995–997.
- [8] J. Pfund, 'Wellenfront-Detektion mit Shack–Hartmann-Sensoren', Ph.D. thesis, University of Erlangen, 2001.
- [9] J.W. Goodman, 'Optical interconnections for VLSI systems', *IEEE Proceedings*, **72**, 1984, 850–866.
- [10] J.W. Goodman, F.J. Leonberger, S.-Y. Kung and R.A. Athale, 'Optical interconnections for VLSI systems', *Proceedings of the IEEE*, **72**, 1984, 850–866.
- [11] M.R. Wang, G.J. Sonek, R.T. Chen and T. Jannson, 'Large fanout optical interconnects using thick holographic gratings and substrate wave propagation', *Applied Optics*, **31**, 1992, 236–249.
- [12] H. Zarschizky, H. Karstensen, A. Staudt and E. Klement, 'Clock distribution using a synthetic HOE with multiple fan-out at IR wavelength', in '*Optical Interconnections and Networks*', H. Bartelt (ed.), Proc. SPIE Vol. 1281, 1990, 103–112.
- [13] R. Khalil, L.R. McAdams and J.W. Goodman, 'Optical clock distribution for high speed computers', *Proceedings of SPIE*, **991**, 1988, 32–41.
- [14] A.V. Mule, E.N. Glytsis, T.K. Gaylord and J.D. Meindl, 'Electrical and optical clock distribution networks for gigascale microprocessors' *IEEE Very Large Scale Integration (VLSI) Systems*, **10**(5), 582–594.
- [15] H.P. Herzig, M.T. Gale, H.W. Lehmann and R. Morf, 'Diffractive components: computer-generated elements', in '*Perspectives for Parallel Optical Interconnects*', Ph. Lalanne and P. Chavel (eds), ESPRIT Basic Research Series, Springer-Verlag, Berlin, 1991, 71–108.
- [16] J.P.G. Bristow, Y. Liu, T. Marta *et al.*, 'Cost-effective optoelectronic packaging for multichip modules and backplane level optical interconnects', in '*Optical Interconnect III*', Proc. SPIE Vol. 2400, 1995, 61.
- [17] D. Zaleta, M. Larsson, W. Daschner and S.H. Lee, 'Design methods for space-variant optical interconnections to achieve optimum power throughput', *Applied Optics*, **34**(14), 1995, 2436–2447.
- [18] A.G. Kirk and H. Thienpont, 'Optoelectronic programmable logic array which employs diffractive interconnections', in '*Diffractive and Holographic Optics Technology II*', I. Cindrich and S.H. Lee (eds), SPIE Press, Bellingham, WA, 1995, 235–242.
- [19] T.J. Cloonan, G.W. Richards, A.L. Lentine, F.B. McCormick and J.R. Erickson, 'Free-space photonic switching architectures based on extended generalized shuffle networks', *Applied Optics*, **31**, 1992, 7471–7492.
- [20] D. Zaleta, S. Patra, V. Ozguz, J. Ma and S.H. Lee, 'Tolerancing of board-level free-space optical interconnects', *Applied Optics*, **35**(8), 1996, 1317–1327.
- [21] M. Charrier, M. Goodwin, R. Holzner *et al.*, 'Review of the HOLICS Optical Interconnects Programme (ESPRIT III Project 6276)', in '*IEEE Conference on Electronics, Circuits, and Systems, 1996*', ICECS 96, Vol. 1, 1996, 436–439.
- [22] R.K. Kostuk, 'Simulation of board-level free-space optical interconnects for electronic processing', *Applied Optics*, **31**(4), 1992, 2438–2445.
- [23] J.P. Pratt and V.P. Heuring, 'Designing digital optical computing systems: power distribution and cross talk', *Applied Optics*, **31**(23), 1992, 4657–4662.
- [24] J.L. Lowell, 'VCSEL-based optical interconnections at interbox distances and shorter', in '*Optoelectronic Packaging*', SPIE Vol. CR62, 1996, 229–243.
- [25] B.R. Brown and A.W. Lohmann, 'Complex spatial filtering with binary masks', *Applied Optics*, **5**, 1966, 967–969.



# Conclusion

This book reviews the various aspects of Digital Optics technology as it is known today, from high level design issues down to fabrication and mass replication issues, with special emphasis on the related industrial market segments.

The numerous optical elements constituting the realm of digital optics are defined and presented one by one, as well as their respective design and modeling tools. These include planar waveguides (PLCs), micro-refractives, digital diffractives, digital hybrid optics, digital holographics, digital dynamic optics and digital nano-optics.

The various fabrication and replication tools used today in research and industry are reviewed. Numerous examples of elements fabricated with such techniques are presented.

Typical manufacturing instructions for Digital Optics as well as specific techniques to analyze the fabricated elements, are described.

Finally, an exhaustive list of products incorporating digital optics available on the market today is presented. The application of these products ranges from academia and research to military, heavy industry and, finally, consumer electronics markets.





# Appendix A

## Rigorous Theory of Diffraction

### A.1 Maxwell's Equations

Light is identified as an electric field  $\vec{E}$  and a magnetic field  $\vec{H}$ , linked by Maxwell's equations [1]:

$$\left\{ \begin{array}{l} \text{curl}(\vec{E}) = -\mu \frac{\partial \vec{H}}{\partial t} \\ \text{curl}(\vec{H}) = \varepsilon \frac{\partial \vec{E}}{\partial t} \\ \text{div}(\varepsilon \vec{E}) = 0 \\ \text{div}(\mu \vec{H}) = 0 \end{array} \right. \quad (\text{A.1})$$

where  $\varepsilon$  is the permittivity tensor and  $\mu$  is the permeability tensor related to the medium property in which the wave propagates.

If we restrict our analysis to a linear, isotropic but nonhomogeneous media,  $\varepsilon = \varepsilon(x, y, z)$  and  $\mu = \mu(x, y, z)$  are scalar functions depending on position only (time dependence is not considered here). It is then possible to derive equations in which either the  $\vec{E}$  or  $\vec{H}$  fields appear separately:

$$\left\{ \begin{array}{l} \nabla^2 \vec{E} - \varepsilon \mu \frac{\partial^2 \vec{E}}{\partial t^2} + \text{grad}(\ln \mu) \times \text{rot}(\vec{E}) + \text{grad}(\vec{E} \cdot \text{grad}(\ln \varepsilon)) = 0 \\ \nabla^2 \vec{H} - \varepsilon \mu \frac{\partial^2 \vec{H}}{\partial t^2} + \text{grad}(\ln \varepsilon) \times \text{rot}(\vec{H}) + \text{grad}(\vec{H} \cdot \text{grad}(\ln \mu)) = 0 \end{array} \right. \quad (\text{A.2})$$

### A.2 Wave Propagation and the Wave Equation

If we further restrict our analysis to linear, isotropic and homogeneous media, permittivity and permeability are then scalar constants and all gradient functions are null functions. The previous equations then become the vector wave equation:

$$\nabla^2 \vec{U} - \frac{1}{v^2} \frac{\partial^2 \vec{U}}{\partial t^2} = 0 \quad (\text{A.3})$$

where  $v$  is defined as the propagation velocity in the medium,

$$v = \frac{1}{\sqrt{\epsilon\mu}} \quad (\text{A.4})$$

and  $\vec{U}$  represents either the  $\vec{E}$  or the  $\vec{H}$  field. Rewriting the vector wave equation (Equation (A.3)) for each vector component on a rectangular basis gives rise to

$$\begin{cases} \nabla^2 U_x - \frac{1}{v^2} \frac{\partial^2 U_x}{\partial t^2} = 0 \\ \nabla^2 U_y - \frac{1}{v^2} \frac{\partial^2 U_y}{\partial t^2} = 0 \\ \nabla^2 U_z - \frac{1}{v^2} \frac{\partial^2 U_z}{\partial t^2} = 0 \end{cases} \quad (\text{A.5})$$

As  $U$  represents either  $\vec{E}$  or  $\vec{H}$ , all components of the field have to satisfy the same equation. We can therefore introduce the scalar wave equation as follows [2]:

$$\nabla^2 U - \frac{1}{v^2} \frac{\partial^2 U}{\partial t^2} = 0 \quad (\text{A.6})$$

In the specific case of monochromatic plane waves, the amplitude  $A$  and phase  $\varphi$  of any of the field components can be represented by a complex function of position and time:

$$U(x, y, z; t) = A(x, y, z) \cdot e^{-j\varphi(x, y, z)} \cdot e^{j\omega t} \quad (\text{A.7})$$

where  $\omega = 2\pi v$  and  $\varphi$  is the phase of the wave. The scalar Equation (A.6) can then be rewritten as follows:

$$(\nabla^2 + k^2)U = 0 \quad (\text{A.8})$$

where the wavenumber  $\vec{k}$  is defined as  $k = 2\pi/\lambda$ , with  $\lambda$  the wavelength within the dielectric medium, and  $\vec{r} = x\vec{u}_1 + y\vec{u}_2 + z\vec{u}_3$  is the vector position.

Any space-dependent part of a propagating monochromatic scalar wave obeys the time-independent Helmholtz equation.

### A.3 Towards a Scalar Field Representation

The only approximation that we made to derive the scalar wave equation (Equation (A.6)) concerns the medium in which the wave propagates. For free-space propagation (i.e. no boundary conditions), Equation (A.6) is not an approximation but, rather, the accurate expression for the wave propagation.

In the case in which light propagates through a step index, as in an air/glass interface ( $n = 1$  to  $n > 1$ ; see Figure A.1), the assumption of a homogeneous and isotropic medium is no longer valid. Some deviation arises between the scalar theory and the real diffracted fields.

In the case of a linear, nonisotropic and nonhomogeneous medium,  $\epsilon$  and  $\mu$  are tensors, and Maxwell's vector equation  $\text{curl}(\vec{H}) = \epsilon \cdot (\partial \vec{E} / \partial t)$  can be rewritten as follows:

$$\text{curl}(\vec{E}) = - \begin{pmatrix} \epsilon_{00} & \epsilon_{01} & \epsilon_{02} \\ \epsilon_{10} & \epsilon_{11} & \epsilon_{12} \\ \epsilon_{20} & \epsilon_{21} & \epsilon_{22} \end{pmatrix} \begin{bmatrix} \frac{\partial H_x}{\partial t} \\ \frac{\partial H_y}{\partial t} \\ \frac{\partial H_z}{\partial t} \end{bmatrix} \quad (\text{A.9})$$

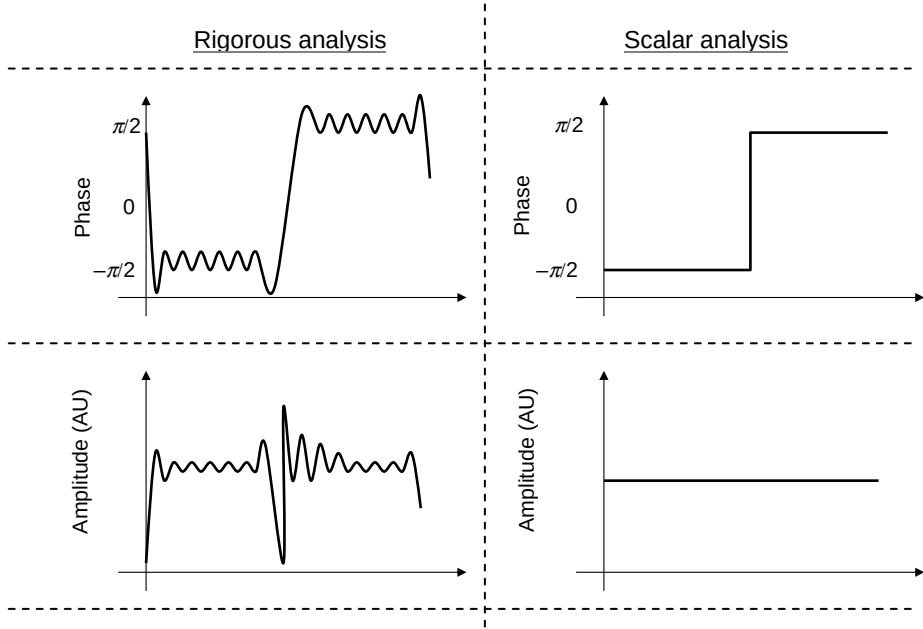


**Figure A.1** Scalar and rigorous expressions of the field at a step-index interface

When Equation (A.9) is decomposed along the first unit vector of a rectangular basis, it becomes

$$\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = \varepsilon_{00} \frac{\partial H_x}{\partial t} + \varepsilon_{01} \frac{\partial H_y}{\partial t} + \varepsilon_{02} \frac{\partial H_z}{\partial t} \quad (\text{A.10})$$

At the boundary, the  $\vec{E}$  and  $\vec{H}$  vector components are not independent. It appears clearly that the  $y$  and  $z$  components of the  $\vec{E}$  field are not only coupled one to each other, but are also coupled to the  $\vec{H}$  field components. As a result, the diffracted expressions for the amplitude and phase differ depending on whether they are evaluated with scalar or rigorous theory (see Figure A.1). Scalar theory expects no amplitude variations and a perfect phase step, whereas rigorous theory shows no sharp discontinuity but ripple oscillations at the interface. It is worth stressing that the difference between scalar and rigorous theory is only noticeable in the immediate vicinity of the interface or at the edges of the limiting aperture. Therefore, as soon as we are a few wavelengths away, the scalar and rigorous theories predict very similar behaviors, and thus the coupling effects at the interface can be ignored.

When it comes to modeling diffractive optics elements and other microstructured optical elements, it is necessary to use rigorous diffraction theory if the smallest feature sizes comprising these elements are less than three to four times the wavelength of light considered [2]. In some cases, scalar theory still gives good predictions of the diffraction efficiency when the smallest features (i.e. half the smallest period for a grating) are about twice the size of the wavelength. Note that divergences between the rigorous and scalar theories mainly affect the diffraction efficiency calculations, not those for the diffracted angles or the overall reconstruction geometry [3, 4].

Scalar theory is used in the vast majority of diffractive optics design today [5, 6]. Rigorous modeling [7] is usually performed on very simple elements such as linear gratings, 2D or 3D photonic crystal structures or simple waveguide structures.

Appendix B introduces the full scalar theory of diffraction.

## References

- [1] M. Born and E. Wolf, 'Rigorous diffraction theory', in *Principles of Optics*, 6th edn, 556–591, Pergamon Press, Oxford, 1993.
- [2] P.G. Rudolf, J.J. Tollet and R.R. MacGowan, 'Computer modeling wave propagation with a variation of the Helmholtz–Kirchhoff relation', *Applied Optics*, **29**(7), 1990, 998–1003.
- [3] J.C. Hurtley, 'Scalar Rayleigh–Sommerfeld and Kirchhoff diffraction integrals: a comparison of exact evaluations for axial points', *Journal of the Optical Society of America*, **63**, 1973, 1003.
- [4] E.W. Marchand and E. Wolf, 'Comparison of the Kirchhoff and Rayleigh–Sommerfeld theories of diffraction at an aperture', *Journal of the Optical Society of America*, **54**, 1964, 587.
- [5] M. Toltzeck, 'Validity of the scalar Kirchhoff and Rayleigh–Sommerfeld diffraction theories in the near field of small phase objects', *Journal of the Optical Society of America*, **8**(1), 1991, 21–27.
- [6] J.A. Hudson, 'Fresnel Kirchhoff diffraction in optical systems: an approximate computational algorithm', *Applied Optics*, **23**(14), 1984, 2292–2295.
- [7] E. Noponen, J. Turunen and A. Vasara, 'Electromagnetic theory and design of diffractive lens arrays', *Journal of the Optical Society of America*, **10**, 1993, 434–455.

# Appendix B

## The Scalar Theory of Diffraction

### B.1 Full Scalar Theory

In order to introduce the concept of the scalar theory of diffraction, let us consider for a start an arbitrary wavefront  $U(x, y; z)$  propagating from  $z = 0$  along the positive  $z$ -axis of a Cartesian referential [1]. The Fourier transform of such a wavefront is given by

$$U(u, v; z) = \iint_{-\infty}^{\infty} U(x, y; z).e^{-2\pi j(ux + vy)} dx dy \quad (\text{B.1})$$

The same arbitrary wavefront  $U(x, y; z)$  is thus the inverse Fourier transform of  $U(u, v; z)$ , defined as

$$U(x, y; z) = \iint_{-\infty}^{\infty} U(u, v; z).e^{-2\pi j(ux + vy)} du dv \quad (\text{B.2})$$

Therefore,  $U(x, y; z)$  can be described as an infinite composition of the set of functions  $e^{-2\pi j(ux + vy)}$  weighted by the coefficients of  $U(u, v; z)$ . Bearing in mind the complex representation of plane waves (Equation (A.7)), it is obvious that these functions can be understood as sets of plane waves propagating in the  $z$  direction with the cosine direction

$$\left( \lambda u; \lambda v; \sqrt{1 - \lambda^2 u^2 - \lambda^2 v^2} \right) \quad (\text{B.3})$$

This means that  $U(x, y; z)$  can be decomposed into an angular spectrum of plane waves  $U(u, v; z)$ . This angular spectrum of plane waves is also the Fourier transform of  $U(u, x; z)$ , or the far-field representation of that same function.

Now let us examine how the angular spectrum propagates from a plane at  $z = 0$  to a plane at  $z = z_0$ . We have therefore to find a relation between  $U(u, v; 0)$  and  $U(u, v; z_0)$  with  $z_0 > 0$ . According to scalar diffraction theory (see Appendix A), the space-dependent part of any propagating field  $U(x, y; z)$  has to obey the Helmholtz equation:

$$(\nabla^2 + k^2).U(x, y; z) = 0 \quad (\text{B.4})$$

Since  $U(x, y; z)$  can be represented by Equation (B.2), the previous equation becomes:

$$\nabla^2 \left( \iint_{-\infty}^{+\infty} U(u, v; z).e^{2\pi j(ux + vy)}.du dv \right) + k^2 \left( \iint_{-\infty}^{+\infty} U(u, v; z).e^{2\pi j(ux + vy)}.du dv \right) = 0 \quad (\text{B.5})$$

We finally get the differential form of the Helmholtz equation:

$$\frac{\partial^2 U}{\partial z^2}(u, v; z) + k^2(1 - \lambda^2 u^2 - \lambda^2 v^2).U(u, v; z) = 0 \quad (\text{B.6})$$

for which an obvious solution is as follows:

$$U(u, v; z) = U(u, v; 0).e^{ikz\sqrt{1 - \lambda^2 u^2 - \lambda^2 v^2}} \quad (\text{B.7})$$

Therefore, the effect of propagation along the  $z$ -axis is only described by the phase factor  $e^{ikz\sqrt{1 - \lambda^2 u^2 - \lambda^2 v^2}}$ .

We have thus demonstrated that the optical disturbance can be decomposed into an infinite sum of plane waves, each traveling in a direction given by the following components:

$$\left\{ \begin{array}{c} u^2 \\ v^2 \\ \sqrt{1 - u^2 - v^2} \end{array} \right. \quad (\text{B.8})$$

where  $u$  and  $v$  have been chosen to be the cosine directors of  $\vec{k}$ . This ensures that we have  $1 - u^2 - v^2 > 0$ . But if we were<sup>1</sup> to have  $1 - u^2 - v^2 < 0$ , we would get the following equation:

$$\left\{ \begin{array}{c} u^2 \\ v^2 \\ j.\sqrt{u^2 + v^2 - 1} \end{array} \right. \quad (\text{B.9})$$

which represents a wave that vanishes rapidly due to the positive real term in the exponential. Such waves are called evanescent waves.

We have demonstrated that if we know the field at a point  $z_0$ , it is possible to evaluate the propagated field at a point  $z$  ( $z > z_0$ ) with very few approximations. We simply have to take into consideration the angular spectrum of the field, multiply each term of the angular spectrum by a  $z$ -linear phase factor, and transform it back using the inverse angular spectrum relation.

The diffracted field can be evaluated by the propagation into the angular spectrum of the plane waves constituting that field. This process is valid if the diffracting medium is linear, isotropic and homogeneous, and if the apertures (or interfaces) dimensions are large in respect to the wavelength. That process, however, is probably not the most convenient tool to solve general-purpose problems. Angular spectrum propagations require that both the diffraction and the calculated surfaces are plane and parallel, which is seldom the case in real applications, where surfaces can be curved rather than be planar. For that reason, other diffracted field propagation methods such as Fresnel–Kirchhoff, although less accurate, have proven their usefulness.

### B.1.1 The Helmholtz–Kirchhoff Diffraction Integral

The Helmholtz–Kirchhoff diffraction integral is the foundation of any scalar diffraction calculations. All propagators that are presented in this book are based on the Helmholtz–Kirchhoff diffraction integral. This relation is presented here to give an overview of approximations that are made when deriving diffraction equations.

Diffraction formulations cannot be considered as ‘recipes’ applied to diffractive optics design. Therefore, we will thoroughly examine the Helmholtz–Kirchhoff integral to get a better insight into diffraction integrals.

The origin of the Fresnel–Kirchhoff and Rayleigh–Sommerfeld diffraction theories lies in Green’s theorem, which expresses the optical disturbance  $U$  at a point  $P_0$  in terms of its values on a surface  $S$ .

<sup>1</sup> Since  $(u, v)$  are cosine directors, we theoretically cannot have  $1 - u^2 - v^2 < 0$ . However, the scalar theory neglects ripples oscillations of the field at the aperture boundary, so the case  $1 - u^2 - v^2 < 0$  might happen and corresponds to real phenomena.

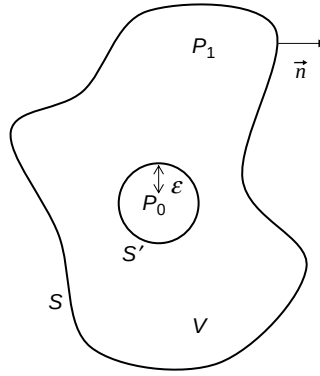


**Figure B.1** The surfaces of integration for Green's theorem  $S = S' + S_\epsilon$

### B.1.2 Green's Theorem

If  $U$  and  $G$  are continuous functions and if their first and second derivatives are single-valued over the surface  $S$  bounding the volume  $V$  [2],

$$\iiint_V (U \nabla^2 G - G \nabla^2 U) dv = \iint_S \left( U \frac{\partial G}{\partial n} - G \frac{\partial U}{\partial n} \right) ds \quad (\text{B.10})$$

where  $n$  is the outward normal of surface  $S$ .

According to Huygen's principle, the chosen auxiliary Green's function<sup>2</sup> is

$$G(P_1) = \frac{1}{r_{01}} e^{-jk\vec{r}_{01}} \quad (\text{B.11})$$

where  $\vec{r}_{01} = \vec{P}_0 \vec{P}_1$ . The auxiliary function  $G$  represents a spherical wave expanding from  $P_0$ . Since all functions mentioned in Green's theorem are supposed to be continuous, the  $P_0$  singularity (i.e.  $r_{01} = 0$ ) of  $G$  has to be removed from the integration domain. The new surface  $S'$  is the same as  $S$  (see Figure B.1).

Keeping in mind that both  $U$  and  $G$  have to obey the Helmholtz equation:

$$\begin{cases} (\nabla^2 + k^2)U = 0 \\ (\nabla^2 + k^2)G = 0 \end{cases} \quad \begin{cases} \nabla^2 U = -k^2 U \\ \nabla^2 G = -k^2 G \end{cases} \quad (\text{B.12})$$

the first integral term of Green's theorem can be rewritten as

$$\iiint_{V'} (U(-k^2 G) - G(-k^2 U)) dv = \iiint_{V'} (UG - GU) dv = 0 \quad (\text{B.13})$$

which means that the second term of Green's theorem also has to be zero:

$$\begin{aligned} & \iint_{S'} \left( U \frac{\partial G}{\partial n} - G \frac{\partial U}{\partial n} \right) ds \\ &= \iint_S \left( U \frac{\partial G}{\partial n} - G \frac{\partial U}{\partial n} \right) ds + \iint_{S_\epsilon} \left( U \frac{\partial G}{\partial n} - G \frac{\partial U}{\partial n} \right) ds \quad (\text{i}) \\ &= A + B \quad (\text{ii}) \\ &= 0 \quad (\text{iii}) \end{aligned} \quad (\text{B.14})$$

Now let us evaluate the second term in the first line of the previous equation.

<sup>2</sup> It can be shown (with considerable effort) that the final result is independent of the choice of the Green's function  $G$ .

Since the surface  $S_\varepsilon$  is a sphere, it is then obvious (see Figure B.1) that vectors  $\vec{n}$  and  $\vec{r}_{01}$  are collinear. Moreover,  $G$  represents a spherical expanding wave, and the wave vector  $\vec{k}$  is collinear to the normal vector  $\vec{n}$  of the sphere, so that we have

$$\begin{aligned}
 \frac{\partial G}{\partial n} &= \frac{\partial G}{\partial r_{01}} \cdot \frac{\partial r_{01}}{\partial n} \\
 &= \frac{\partial}{\partial r_{01}} \left( \frac{1}{r_{01}} e^{jk r_{01}} \right) \cdot \frac{\partial r_{01}}{\partial n} \\
 &= \left[ \frac{1}{r_{01}} \cdot \frac{\partial}{\partial r_{01}} (e^{jk r_{01}}) + e^{jk r_{01}} \frac{\partial}{\partial r_{01}} \left( \frac{1}{r_{01}} \right) \right] \cdot \frac{\partial r_{01}}{\partial n} \\
 &= \left[ \frac{1}{r_{01}} \cdot jk \cdot e^{jk r_{01}} + \frac{1}{r_{01}^2} e^{jk r_{01}} \right] \cdot 1 \\
 &= \frac{1}{r_{01}} \cdot e^{jk r_{01}} \left( jk - \frac{1}{r_{01}} \right)
 \end{aligned} \tag{B.15}$$

Replacing  $dG/dn$  by its value in the second term in the first line of Equation (B.14) yields

$$B = \iint_{S_\varepsilon} \left[ U \cdot \frac{1}{r_{01}} \cdot e^{jk r_{01}} \left( jk - \frac{1}{r_{01}} \right) - \frac{1}{r_{01}} e^{jk r_{01}} \cdot \frac{\partial U}{\partial n} \right] ds \tag{B.16}$$

where  $ds$  represents the elementary surface of sphere  $S_\varepsilon$  of radius  $\varepsilon$ . Using the solid angle  $\Omega = S/\varepsilon^2$ , the value of which is by definition  $4\pi$  over the full surface  $S$ , the previous equation can be rewritten as

$$B = \int_0^{4\pi} \left[ U \cdot \frac{1}{r_{01}} \cdot e^{jk r_{01}} \left( jk - \frac{1}{r_{01}} \right) - \frac{1}{r_{01}} e^{jk r_{01}} \cdot \frac{\partial U}{\partial n} \right] \varepsilon^2 d\Omega \tag{B.17}$$

The total radius  $\varepsilon$  of the sphere  $S$  can be considered as infinitely small, so taking the zero limit of the previous relation gives rise to

$$\begin{aligned}
 \lim_{\varepsilon \rightarrow 0} (B) &= \lim_{\varepsilon \rightarrow 0} \int_0^{4\pi} \left[ U \frac{1}{\varepsilon} \cdot e^{jk \varepsilon} \left( jk - \frac{1}{\varepsilon} \right) - \frac{1}{\varepsilon} \cdot e^{jk \varepsilon} \cdot \frac{\partial U}{\partial n} \right] \varepsilon^2 d\Omega \\
 &= \int_0^{4\pi} \left[ \lim_{\varepsilon \rightarrow 0} (jk \varepsilon U e^{jk \varepsilon}) - \lim_{\varepsilon \rightarrow 0} (U e^{jk \varepsilon}) + \lim_{\varepsilon \rightarrow 0} \left( \varepsilon e^{jk \varepsilon} \frac{\partial U}{\partial n} \right) \right] d\Omega \\
 &= \int_0^{4\pi} [0 - U(P_0) + 0] d\Omega \\
 &= 4\pi U(P_0)
 \end{aligned} \tag{B.18}$$

Reporting the result  $B = 4\pi U(P_0)$  in Equation (B.14) gives the final form of Kirchhoff's integral theorem, which expresses the optical disturbance at a point  $P_0$  in terms of its values over the surface  $S$ :

$$\begin{aligned}
 U(P_0) &= \frac{1}{4\pi} \iint_S \left[ U \frac{\partial}{\partial n} \left( \frac{1}{r_{01}} e^{jk r_{01}} \right) - \frac{1}{r_{01}} e^{jk r_{01}} \frac{\partial U}{\partial n} \right] ds \quad (i) \\
 U(P_0) &= \frac{1}{4\pi} \iint_S \frac{1}{r_{01}} e^{jk r_{01}} \left[ U \left( jk - \frac{1}{r_{01}} \right) - \frac{\partial U}{\partial n} \right] ds \quad (ii)
 \end{aligned} \tag{B.19}$$



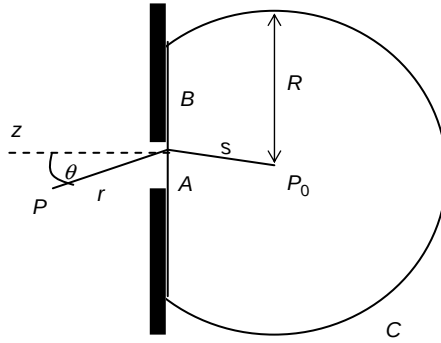


**Figure B.2** The volume and surfaces of integration used to derive the Fresnel–Kirchhoff integral

### B.1.3 The Fresnel–Kirchhoff Diffraction Integral

In order to derive an adequate version of Equation (B.19) for diffractive optics (or refractive optics, for that matter), we will consider the integration surfaces and volumes depicted in Figure B.2.

The Helmholtz–Kirchhoff diffraction integral can be evaluated by using the three surfaces depicted in Figure B.2 (A, B and C):

- surface A is the aperture;
- surface B creates an aperture stop around surface A; and
- surface C is a portion of a sphere of radius  $R$  centered on  $P_0$ .

Thus, the Helmholtz–Kirchhoff diffraction integral [3, 4] can be rewritten as:

$$U(P_0) = \frac{1}{4\pi} \iint_{A,B,C} \frac{1}{s} e^{jks} \left( U \left( jk - \frac{1}{s} \right) - \frac{\partial U}{\partial n} \right) ds \quad (\text{B.20})$$

The Helmholtz–Kirchhoff diffraction integral expresses the disturbance at a given point in terms of its values and the values of its first derivative on the surrounding volumes A, B and C. In Equation (B.20),  $U$  is the disturbance to be determined and  $G = (\exp(jks))/s$  is the auxiliary Green's function.

Note that the Helmholtz–Kirchhoff integral theorem implies that the conditions of validity of the scalar theory are met. That is, we are only considering large diffracting apertures compared to the wavelength of light used.

Let us evaluate the integral in Equation (B.20) over the three surfaces depicted in Figure B.2. It can be shown that the contribution of the spherical cap C to the integral is zero. In order to evaluate the contribution of the two other surfaces, we need to make the following assumptions:

- Assumption #1: on surface B, we have  $U=0$  and  $dU/dn=0$  (i.e. on the region in the geometrical shadow of the screen, the derivative is null).
- Assumption #2: on surface A, we have  $U=U^i$  and  $dU/dn=dU^i/dn$  (i.e. over that surface, the disturbance and its first derivative are exactly what they would be without the opaque screen).

In that case, the integral in Equation (B.20) reduces to the integral over surface  $A$ :

$$U(P_0) = \frac{1}{4\pi} \iint_A \frac{1}{s} e^{jks} \left( U \left( jk - \frac{1}{s} \right) - \frac{\partial U}{\partial n} \right) ds \quad (\text{B.21})$$

If we further consider that the aperture is illuminated by a spherical diverging wave, the disturbance  $U$ , the auxiliary function  $G$  and their first derivatives are as follows:

$$\begin{aligned} U &= \frac{1}{r} e^{jkr}, & \frac{\partial U}{\partial n} &= \left( jk - \frac{1}{r} \right) e^{jkr} \cos(\vec{n}, \vec{r}) \\ G &= \frac{1}{s} e^{jks}, & \frac{\partial G}{\partial n} &= \left( jk - \frac{1}{s} \right) e^{jks} \cos(\vec{n}, \vec{s}) \end{aligned} \quad (\text{B.22})$$

Considering that we only evaluate the optical disturbance at points far away from the aperture, we furthermore have  $k \gg 1/r$ , which simplifies the previous relations down to

$$\begin{aligned} U &= \frac{1}{r} e^{jkr}, & \frac{\partial U}{\partial n} &= jk \cdot e^{jkr} \cos(\vec{n}, \vec{r}) \\ G &= \frac{1}{s} e^{jks}, & \frac{\partial G}{\partial n} &= jk \cdot e^{jks} \cos(\vec{n}, \vec{s}) \end{aligned} \quad (\text{B.23})$$

By integrating these expressions as shown in Equation (B.20), we obtain the Fresnel–Kirchhoff diffraction integral:

$$U(P_0) = \frac{1}{j\lambda} \iint_A \frac{e^{jk(r+s)}}{rs} \left[ \frac{\cos(\vec{n}, \vec{r}) - \cos(\vec{n}, \vec{s})}{2} \right] ds \quad (\text{B.24})$$

Finally, in the very specific case of normally incident plane wave illumination, Equation (B.24) becomes

$$U(P_0) = \frac{1}{j\lambda} \iint_A \frac{e^{jk(r+s)}}{rs} \left[ \frac{1 + \cos(\theta)}{2} \right] ds \quad (\text{B.25})$$

It has been experimentally proven [1, 2] that Fresnel–Kirchhoff diffraction theory predicts the behavior of the diffracted fields with excellent accuracy [3, 4]. This is the reason why the Fresnel–Kirchhoff diffraction integral is widely used today [5].

However, the Fresnel–Kirchhoff theory has some inconsistencies. The assumptions that Kirchhoff made for surface  $B$  require that both the field and the derivative are null. It can be shown that if these requirements are met, the field must be identically null everywhere. Moreover, the approximation  $k \gg 1/r$ , although valid when we are many wavelengths away from the aperture, leads to wrong results as soon as we get closer to the aperture (a diffracting screen). This is why the light distributions at the aperture are different from what Kirchhoff's assumptions involve.

In the following sections, we will see that either one of the two assumptions is enough:  $U=0$  or  $dU/dn=0$ . Depending on the chosen assumption, we get either the first or the second Rayleigh–Sommerfeld diffraction integral for the diffracted field.

### B.1.4 Rayleigh–Sommerfeld Diffraction Theory

Although the Fresnel–Kirchhoff diffraction integral gives rise to excellent results, some inconsistencies in this theory have driven the need for a more mathematically accurate formulation of the diffraction integral. As we have seen in the previous section, Kirchhoff's two assumptions  $U=0$  and  $dU/dn=0$  lead to the mathematical conclusion that the field must be zero everywhere in space. To solve this contradiction, Rayleigh showed that either  $U=0$  or  $dU/dn=0$  is enough to derive another relation: the Rayleigh–Sommerfeld diffraction integral [5, 6].

The Rayleigh–Sommerfeld diffraction integral also relies on the Helmholtz–Kirchhoff integral theorem (see Equation (B.19a)). This implies that we are still dealing with apertures that are large compared to the wavelength of light used:

$$U(P_0) = \frac{1}{4\pi} \iint_A \frac{e^{jks}}{s} \left( U \left( jk - \frac{1}{s} \right) - \frac{\partial U}{\partial n} \right) ds \quad (\text{B.26})$$

We also keep the same Green auxiliary function  $G = (\exp(jks))/s$ , but the volume  $V$  of integration differs. Considering two symmetric points, it is possible to derive both versions of the Rayleigh–Sommerfeld diffraction formula:

$$U(P_0) = \frac{1}{2\pi} \iint_A U(x', y', z') \frac{\partial}{\partial z'} \left( \frac{e^{jks}}{s} \right) ds \quad (\text{B.27})$$

The first solution is obviously based on Kirchhoff's first assumption  $U=0$ , while the second one is based on Kirchhoff's second assumption  $dU/dn=0$ . As we did in the case of Fresnel–Kirchhoff integral, we suppose that the field inside the aperture opening is exactly what it would be without the aperture. If we further consider a diverging spherical wave, we then obtain

$$U = \frac{1}{r} e^{jkr}, \quad \frac{\partial U}{\partial n} = \left( jk - \frac{1}{r} \right) e^{jkr} \cos(\vec{n}, \vec{r}) \quad (\text{B.28})$$

The final Rayleigh–Sommerfeld relations thus become

$$\begin{cases} U_1(x, y, z) = \frac{1}{2\pi} \iint_A \frac{e^{jkr}}{r} \left( jk + \frac{1}{s} \right) \frac{e^{jks}}{s} \cos(\vec{n}, \vec{s}) . ds \\ U_2(x, y, z) = \frac{-1}{2\pi} \iint_A \frac{e^{jkr}}{r} \left( jk - \frac{1}{r} \right) \frac{e^{jks}}{s} \cos(\vec{n}, \vec{r}) . ds \end{cases} \quad (\text{B.29})$$

The Rayleigh–Sommerfeld diffraction integral is believed to be more accurate than the Fresnel–Kirchhoff formulation because of its mathematical consistency, and also because of its ability to reproduce closely the diffracted field right behind the aperture. However, it has been shown experimentally [1, 6] that the Fresnel–Kirchhoff diffraction formulation gives more accurate results than Rayleigh–Sommerfeld theory (assuming that we are many wavelengths away from the diffracting aperture) [4, 5]. Moreover, the Rayleigh–Sommerfeld theory is limited to a plane surface, which is a severe limitation, since we usually deal with curved surfaces in optics. On the contrary, the Fresnel–Kirchhoff relation can handle surfaces of any shape [3].

Finally, if we restrict our study to large distances from the aperture ( $k \gg 1/r$ ) and to normally incident plane wave illumination,  $U_1(x, y, z)$  and  $U_2(x, y, z)$  reduce to

$$\begin{cases} U_1(x, y, z) = \frac{j}{\lambda} \iint_A \frac{e^{jkr}}{r} \frac{e^{jks}}{s} \cos(\vec{n}, \vec{s}) . ds & \text{(i)} \\ U_2(x, y, z) = \frac{-j}{\lambda} \iint_A \frac{e^{jkr}}{r} \frac{e^{jks}}{s} \cos(\vec{n}, \vec{r}) . ds & \text{(ii)} \end{cases} \quad (\text{B.30})$$

It seems obvious that considering small angles, the obliquity factor of the previous relations is unity. In this very specific case, the Fresnel–Kirchhoff and the two Rayleigh–Sommerfeld formulations are equivalent.

## B.2 Scalar Diffraction Models for Digital Optics

Here, we develop the two models that are used most often in diffractive design and modeling: the Fresnel diffraction model and the Fraunhofer diffraction model, respectively, for implementing numerical reconstructions in the near and far fields (the angular spectrum).

### B.2.1 Fresnel Diffraction

We have seen in Equation (B.30b) that the first Rayleigh–Sommerfeld integral can be written as

$$U(P_0) = \frac{1}{j\lambda} \iint_A U(P) \cdot \frac{e^{jk_r}}{r} \cos\theta \cdot ds \quad (\text{B.31})$$

where  $\theta$  is the angle defined by the normal  $\vec{n}$  to the surface  $A$  and vector  $\vec{r}$ . Examining Figure B.2, it is obvious that  $\cos(\theta) = z/r$ . So without any approximations, the previous relation can become

$$U(P_0) = \frac{z}{j\lambda} \iint_A U(P) \cdot \frac{e^{jk_r}}{r^2} du \cdot dv \quad (\text{B.32})$$

The Rayleigh–Sommerfeld diffraction integral (Equation (B.30b)) was derived assuming that the aperture dimensions are much larger than the wavelength of light considered, in addition to the restriction to an observation plane lying far away from the aperture. In that particular case, the  $r$  distance can be approximated as follows:

$$\begin{aligned} r &= \sqrt{(x-u)^2 + (y-v)^2 + z^2} \\ &= z \sqrt{1 + \left(\frac{x-u}{z}\right)^2 + \left(\frac{y-v}{z}\right)^2} \\ &= z \left( 1 + \frac{1}{2} \left(\frac{x-u}{z}\right)^2 + \frac{1}{2} \left(\frac{y-v}{z}\right)^2 \right) \end{aligned} \quad (\text{B.33})$$

Substituting that approximation in the exponential part of Equation (B.32), and taking  $r \approx z$  for the denominator results in

$$U(p_0) = \frac{z}{j\lambda} \iint_A U(P) \cdot \frac{1}{z^2} e^{jkz \left( 1 + \frac{1}{2} \left(\frac{x-u}{z}\right)^2 + \frac{1}{2} \left(\frac{y-v}{z}\right)^2 \right)} du \cdot dv \quad (\text{B.34})$$

which yields the Fresnel diffraction formulation:

$$U(p_0) = \frac{e^{jkz}}{jz\lambda} \iint_A U(P) \cdot e^{\frac{jk}{2z}((x-u)^2 + (y-v)^2)} du \cdot dv \quad (\text{B.35})$$

The previous equation can be developed as follows:

$$U(p_0) = \frac{e^{jkz}}{jz\lambda} \cdot e^{\frac{jk}{2z}(x^2 + y^2)} \cdot \iint_A \left[ U(P) \cdot e^{\frac{jk}{2z}(u^2 + v^2)} \right] \cdot e^{-\frac{jk}{2z}(xu + yv)} du \cdot dv \quad (\text{B.36})$$

Equation (B.36) highlights the fact that, according to the Fresnel diffraction theory, the optical disturbance at  $Z = Z_0$  is basically the Fourier transform of the product of the optical disturbance  $U(P)$  with a quadratic phase factor.

### B.2.2 Fraunhofer Diffraction

We have seen in the previous section that if the observation plane is far from the diffracting aperture, the diffracted field given by the Kirchhoff diffraction integral can be simplified to the Fresnel diffraction formulation (Equation (B.36)).

In the case of a very distant observation plane,  $z \gg (k/2)(u^2 + v^2)$ , the previous relation can be further simplified and leads to the Fraunhofer diffraction integral [7]:

$$U(p_0) = \frac{e^{jkz}}{jz\lambda} \cdot \iint_A U(P) \cdot e^{-\frac{jk}{z}(xu + yv)} du \cdot dv \quad (\text{B.37})$$

Equation (B.37) shows that the optical disturbance  $U(P_0)$  in a plane far from the diffracting aperture can easily be determined by taking the Fourier transform of the complex transmittance of the diffracting aperture. This is the very basis of the entire Fourier optics theory, which we will use repeatedly throughout this book.

## B.3 Extended Scalar Models

In some applications where neither 100% scalar or full vector theory gives rise to acceptable models to predict efficiency for digital diffractive optics, extended scalar models have been developed, which are detailed in Chapter 11.

Such extensions of scalar efficiency predictions include optimal local etch depth modulations and local geometrical shadow.

## References

- [1] A. Sommerfeld, 'Optics', Academic Press, San Diego, 1949.
- [2] H.S. Green and E. Wolf, 'A scalar representation of electromagnetic fields', *Proceedings of the Physical Society A*, **66**, 1953, 1129–1137.
- [3] E.W. Marchand and E. Wolf, 'Consistent formulation of Kirchhoff's diffraction theory', *Journal of the Optical Society of America*, **56**, 1966, 1712–1722.
- [4] E. Wolf and E.W. Marchand, 'Comparison of the Kirchhoff and the Rayleigh–Sommerfeld theories of diffraction at an aperture', *Journal of the Optical Society of America A*, **54**(5), 1964, 587–594.
- [5] G.C. Sherman, 'Application of the convolution theorem to Rayleigh's integral formulas', *Journal of the Optical Society of America*, **57**(4), 1967, 546–547.
- [6] G.C. Sherman and H.J. Bremermann, 'Generalization of the angular spectrum of plane waves and the diffraction transform', *Journal of the Optical Society of America*, **59**(2), 1969, 146–156.



# Appendix C

## FFTs and DFTs in Optics

The first attempts to describe linear transformation systems go back to the Babylonians and the Egyptians, likely via trigonometric sums. In 1669, Sir Isaac Newton referred to the spectrum of light, or light spectra (specter = ghost), but he had not yet derived the wave nature of light, and therefore he stuck to the corpuscular theory of light (for more insight into the duality of light as a corpuscule and a wave, see also Chapter 1).

During the 18th century, two outstanding problems would arise:

1. The orbits of the planets in the solar system: Joseph Lagrange, Leonhard Euler and Alexis Clairaut approximated observation data with a linear combination of periodic functions. Clairaut actually derived the first Discrete Fourier Transform (DFT) formula in 1754!
2. Vibrating strings: Euler described the motion of a vibrating string by sinusoids (the wave equation). But the consensus of his peers was that the sum of sinusoids only represented smooth curves.

Eventually, in 1807, Joseph Fourier presented his work on heat conduction, which introduced the concept of a linear transform. Fourier presented the diffusion equation as a series of (infinite) sines and cosines. Strong criticism at the time actually blocked publication: his work was finally published in 1822, in *Théorie analytique de la chaleur* (*Analytic Theory of Heat*).

### C.1 The Fourier Transform in Optics Today

The Fourier Transform is a fast and efficient insight into any signal's building blocks. This signal can be of an optical nature, an electronic nature, an acoustic nature and so on [1]. Figure C.1 shows the analytic process enabled by the Fourier transformation as a general problem-solving tool.

The Fourier transform is a linear transform that has a very broad range of uses in science, industry and everyday life today. The applications of the Fourier transform in industry today include the following:

- telecoms (cell phones, . . .);
- electronics (digital and analog electronics, DSP, etc.);
- multimedia (audio, video, MP1, MP2, MP3, MP4 players, . . .);
- imaging, image processing, wavefront coding, and so on;
- research (X-ray spectrometry, FT spectrometry, radar design, etc.);

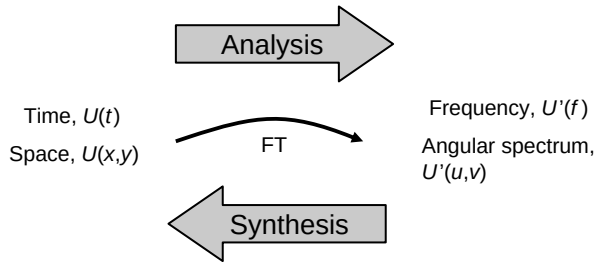


**Figure C.1** Analysis and synthesis in the Fourier domain

- medical (PET scanner, CAT scans & MRI diagnosis, etc.);
- speech analysis (voice-activated “devices”, biometry, etc.);
- and of course optics (Fourier optics, etc.)!

During the 19th and 20th centuries, two paths were followed to implement Fourier transforms: the continuous path and the discrete path. Chapter 11 has reviewed diffraction modeling techniques that are used to model the diffraction effects of light through digital diffractives within the scalar regime of diffraction.

Fourier extended the analysis to arbitrary functions. Johann Dirichlet, Siméon Poisson, Bernhard Riemann and Henri Lebesgue addressed the convergence of the Fourier series. Other Fourier transform variants were derived for various needs. The first notable applications of the Fourier Transform (FT) were to solve complex analytic problems; for example, solving Partial Differential Equations (PDEs). The FT is a powerful tool that is complementary to time domain analysis techniques. Some of the transforms available for the designer’s toolbox include the Laplace transform, the Z transform, and other fractional and wavelet transforms, as described in Section C.6 (wavelets, ridgelets, curvelets etc.).

In 1805, Carl Gauss first described the use of a Fast Fourier Transform (FFT) algorithm (the manuscript, in Latin, went unnoticed and was only published in 1866!). IBM’s J.W. Cooley and John Tukey ‘rediscovered’ the FFT algorithm in 1966, and published it as ‘An algorithm for the machine calculation of complex Fourier series’. Other Discrete FT (DFT) variants have been proposed for various applications (e.g. warped DFT for filter design and signal compression). Today, DFTs and FFTs are well refined and optimized for specific computing requirements. For example, throughout this book we make use of a complex 2D FFT routine published by ‘Numerical Recipes in C’. Figure C.3 summarizes the continuous and discrete FTs for periodic and aperiodic functions. In Figure C.2, the acronym ‘FS’ means Fourier Series, the acronym ‘FT’ denotes a Fourier Transform, ‘DFS’ denotes a Discrete Fourier Series, ‘DTFT’ denotes a Discrete Transform Fourier Transform and ‘DFT’ denotes a Discrete Fourier Transform.

The property of linearity in a transformation system allows the decomposition of a complex signal into a sum of elementary signals. In Fourier analysis, these decompositions are performed on sinusoidal functions (sines and cosines), also called *basis functions*.

Let us consider a 1D FT of a signal  $U(x)$ , where  $U(x)$  is defined over the entire 1D space, as follows:

$$\left\{ \begin{array}{l} -\infty \leq x \leq +\infty \\ U(u) = \int_{-\infty}^{+\infty} U(x) \cdot e^{-j2\pi xu} \cdot dx \end{array} \right. \quad (C.1)$$



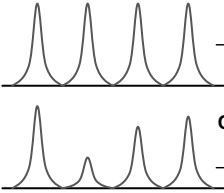
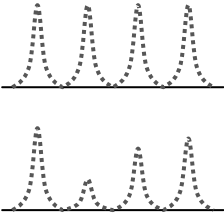
| Input signal                                                                      |                        |             | Angular spectrum |                                                                           |
|-----------------------------------------------------------------------------------|------------------------|-------------|------------------|---------------------------------------------------------------------------|
|  | Periodic (period $T$ ) | <b>FS</b>   | Discrete         | $c_m = \frac{1}{T} \int_0^T U(x).e^{-j\frac{2m\pi}{T}x} dx$               |
|                                                                                   | Aperiodic              | <b>FT</b>   | Continuous       | $U(v) = \int_{-\infty}^{+\infty} U(x).e^{j2\pi vx} dx$                    |
|  | Periodic (period $T$ ) | <b>DFS</b>  | Discrete         | $\tilde{c}_m = \frac{1}{N} \sum_{n=0}^{N-1} U(n).e^{-\frac{j2\pi mn}{N}}$ |
|                                                                                   | Aperiodic              | <b>DTFT</b> | Continuous       | $U(v) = \sum_{n=-\infty}^{+\infty} U(n).e^{-j2\pi vn}$                    |
|                                                                                   |                        | <b>DFT</b>  | Discrete         | $\tilde{c}_m = \frac{1}{N} \sum_{n=0}^{N-1} U(n).e^{-\frac{j2\pi mn}{N}}$ |

**Figure C.2** Continuous and discrete Fourier Transform (FT) and Fourier Series (FS) for period and aperiodic functions, where  $n$  is the harmonic number

where  $U(u)$  is the Fourier transform of  $U(x)$ . The inverse 1D FT is the representation of  $U(x)$  in terms of the basis exponential functions:

$$U(x) = \int_{-\infty}^{+\infty} U(u).e^{j2\pi xu}.du$$

(C.2)

Equation C.2 is also referred to as the *analysis equation*. Equation C.1 is the corresponding *synthesis equation* (see also Figure C.1).

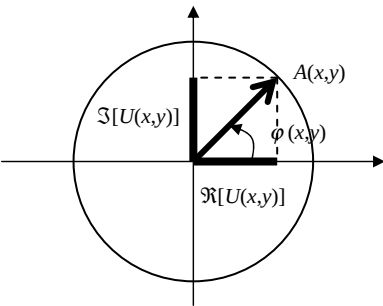


**Figure C.3** The complex plane representation of the phase and complex amplitude of the wavefront  $U(x,y)$

The multidimensional (2D, 3D or beyond) FT belongs to the set of separable unitary functions, as the transformation kernel is separable along each spatial direction. For example, the 2D transform kernel  $U(x_1, y_1, x_2, y_2) = U(x_1, y_1) * U(x_2, y_2)$ .

The 2D FT  $U(u, v)$  of a signal  $U(x, y)$  defined over a 2D area of infinite dimensions is as follows:

$$\begin{cases} -\infty \leq x, y \leq +\infty \\ U(u, v) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} U(x, y) \cdot e^{-j2\pi(xu + yv)} \cdot dx \cdot dy \end{cases} \quad (\text{C.3})$$

where the  $(u, v)$  are the spatial frequencies corresponding to the  $-x$  and  $-y$  directions, respectively. The inverse 2D FT is thus as follows:

$$U(x) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} U(u, v) \cdot e^{j2\pi(xu + yv)} \cdot du \cdot dv \quad (\text{C.4})$$

## C.2 Conditions for the Existence of the Fourier Transform

The three sufficient conditions (Dirichlet's conditions) of existence of the continuous FT are:

- A. The signal must be absolutely integrable over the infinite space.
- B. The signal must have only a finite number of discontinuities and a finite number of maxima and minima in any finite subspace.
- C. The signal must have no infinite discontinuities.

However, one can curb these conditions. A notable example is the Dirac function  $\delta(x, y)$ , which is only defined over a single point:

$$\delta(x, y) = \lim_{N \rightarrow \infty} (N^2 e^{-N^2 \pi(x^2 + y^2)}) \quad (\text{C.5})$$

The Dirac function obviously does not satisfy the FT condition (C), but can nevertheless have a transformation pair in the Fourier space. The FT  $\Delta(u, v)$  of the Dirac function  $\delta(x, y)$  is as follows:

$$\Delta(u, v) = \lim_{N \rightarrow \infty} \left( e^{-\pi \frac{(u^2 + v^2)}{N^2}} \right) = 1 \quad (\text{C.6})$$

Two other functions that do not satisfy condition A are as follows:

$$\begin{cases} U(x, y) = 1 \\ U(x, y) = \cos(2\pi ux) \end{cases} \quad (\text{C.7})$$

With such functions, the FT is still defined by incorporating generalized functions such as the above delta function and defining the FT in the limit. The resulting transform is often called the *generalized Fourier transform*.

## C.3 The Complex Fourier Transform

The functions  $U(x, y)$  that we use throughout this book are complex functions representing the amplitude and phase of the propagating wavefront [1]. It is therefore desirable to derive a complex FT (CFT) that can handle such complex functions.  $U(x, y)$  can be written as follows:

$$U(x, y) = A(x, y) \cdot e^{-i\varphi(x, y)} \quad (\text{C.8})$$

where the complex amplitude  $A(x, y)$  and the phase  $\varphi(x, y)$  are defined as follows:

$$\begin{cases} A(x, y) = \sqrt{(\Im[U(x, y)])^2 + (\Re[U(x, y)])^2} \\ \varphi(x, y) = -\arctan\left(\frac{\Im[U(x, y)]}{\Re[U(x, y)]}\right) \end{cases} \quad (\text{C.9})$$

Figure C.3 shows the complex plane representation of the phase and complex amplitude of the wavefront  $U(x, y)$ .

Chapter 5 uses the complex plane representation of the various Computer-Generated Hologram (CGH) cells in order to assess the quality of the diffractive element. Sometimes the complex plane gives a much better assessment of the physical phenomenon than the amplitude and phase planes.

## C.4 The Discrete Fourier Transform

### C.4.1 The Discrete Fourier Series (DFS)

A periodic function  $U(x, y)$  satisfying the Dirichlet condition can be expressed as a Fourier series, with the following harmonically related sine/cosine terms:

$$U(u, v) = a_0 + \sum_{m=1}^{+\infty} \left[ a_m \cdot \cos\left(\frac{2\pi m x}{T}\right) - b_m \cdot \sin\left(\frac{2\pi m x}{T}\right) \right] \quad (\text{synthesis}) \quad (\text{C.10})$$

$$\begin{cases} a_0 = \frac{1}{T} \int_0^T U(x) \cdot dx \\ a_m = \frac{2}{T} \int_0^T U(x) \cdot \cos\left(\frac{2\pi m x}{T}\right) \cdot dx \\ -b_m = \frac{2}{T} \int_0^T U(x) \cdot \sin\left(\frac{2\pi m x}{T}\right) \cdot dx \end{cases} \quad (\text{analysis}) \quad (\text{C.11})$$

where  $a_m$  and  $b_m$  are the Fourier coefficients and  $m$  is the harmonic number.

Figure C.4 shows the Fourier decomposition of a 1D square function and also the Gibbs phenomenon at the discontinuities of the original function.

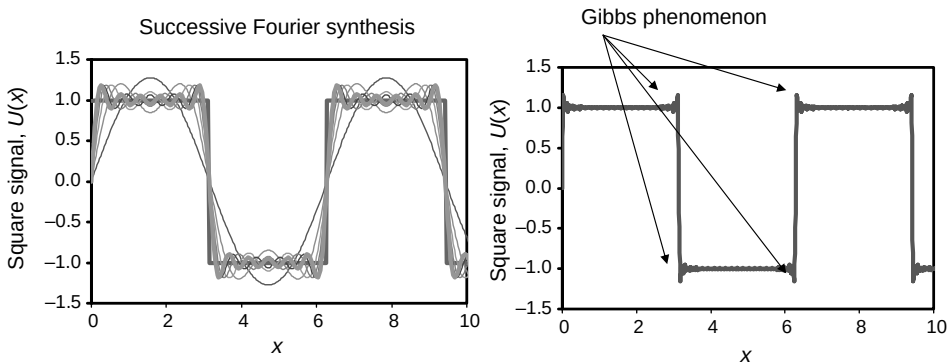


**Figure C.4** The Fourier decomposition of a square function and the Gibbs phenomenon

### C.4.2 The Discrete Fourier Transform (DFS)

The DFT  $U'$  can be written simply in the complex form as the sum of the various magnitude and phase terms sampled in the signal window  $U$  according to the Nyquist rate (see the next section):

$$\begin{cases} U'_{m,n} = \sum_{k=1}^K \sum_{l=1}^L U_{m,n} \cdot e^{-i\pi \left( \frac{km}{M} + \frac{ln}{N} \right)} \\ A'_{n,m} = \sqrt{U'_{m,n} \cdot U'^{*}_{m,n}} \\ \varphi_{n,m} = \arctan \left( \frac{\Im(U'_{m,n})}{\Re(U'_{m,n})} \right) \end{cases} \quad (\text{C.12})$$

Equation C.12 shows the complex amplitude and phase of a single pixel in the reconstruction field. In order to reconstruct a small window in that field (the far field or the near field, by using DFT over a Fresnel integral – see Chapter 11), one has to compute all  $M \times N$  complex pixels within that window.

### C.4.3 Sampling Theorem and Aliasing

Aliasing can arise when the signal  $U(x, y)$  is sampled at a rate that is not sufficient to capture the amount of detail in that signal [2–5]. For a diffractive element, the sampling must be high enough to capture the highest frequency in the element, which is often the smallest local grating period included in the diffractive element (see Chapters 5 and 6).

In order to avoid aliasing, the sampling theorem (Nyquist or Shannon criteria) stipulates that the sampling rate should be at least twice the maximum frequency in the signal  $U(x, y)$ . This means that for a diffractive element in which the smallest local grating has a period of  $P$ , the sampling rate should be at least  $P/2$  – which is, by the way, the smallest feature to fabricate in a binary local grating, also known as the Critical Dimension (or CD), and which is a very useful parameter when it comes to choosing an adequate fabrication technology for that digital diffractive element (see Chapter 12).

Figure C.5 shows an undersampled signal that would create aliasing (on the left) and a signal that has been properly sampled (on the right). The minimum sampling rate (twice the highest frequency) is called the Nyquist rate.

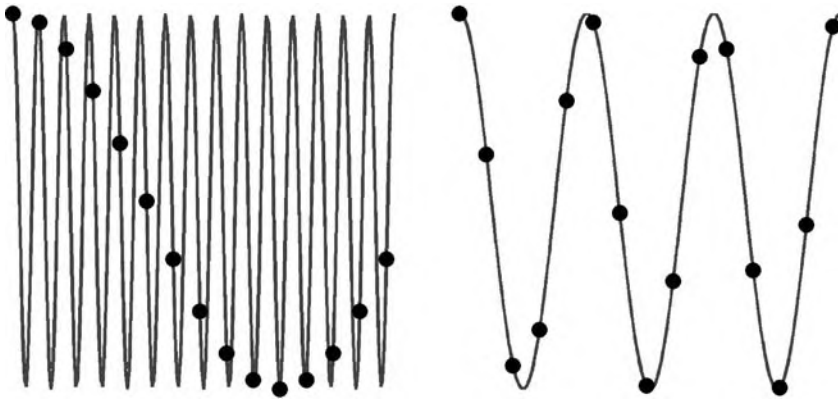


**Figure C.5** Undersampling (left) and proper sampling (right) of a unidimensional signal (sinusoidal grating)

### C.4.4 The Fast Fourier Transform (FFT)

The Fast Fourier Transform (FFT) is a very efficient algorithm for performing a discrete Fourier Transform (DFT). Although the first principle of FFT was used by Gauss in the early 19th century, it was Cooley and Tukey in 1965 who first published the algorithm.

The complexity of a simple DFT, as seen in the previous sections, can be expressed as  $O(N^2)$ . The aim is to reduce that complexity by using an adapted algorithm: the FFT algorithm. The principle of the FFT algorithm is to divide and conquer, or to separate the DFT into two DFTs that will be added together after parallel processing:

$$\text{DFT}(U, N)'' = {}''\text{DFT}\left(V, \frac{N}{2}\right)'' + {}''\text{DFT}\left(W, \frac{N}{2}\right)'' \quad (\text{C.13})$$

In other words, computing a DFT of  $N$  coefficients is equivalent to computing two DFTs with  $N/2$  coefficients. If one applies this idea recursively, one gets the FFT algorithm.

The even and odd Fourier coefficients over half the samples are as follows:

$$\begin{cases} a_{2k} = \frac{1}{N} \sum_{n=0}^{N/2-1} (U_n + U_{n+N/2}) H_{N/2}^{-nk} = \text{DFT}(V, N/2) \\ a_{2k+1} = \frac{1}{N} \sum_{n=0}^{N/2-1} (U_n - U_{n+N/2}) H_{N/2}^{-nk} = \text{DFT}(W, N/2) \end{cases} \quad (\text{C.14})$$

where  $H_N = e^{\frac{2j\pi}{N}}$

Thus, the complexity of computing the FT by using the FFT algorithm rather than a simple DFT reduces from  $N^2$  to  $N \log_2 N$ .

Figure C.6 shows the complete binary tree of the FFT algorithm, with  $\log(m)$  levels. Practical FFT algorithms are iterative, going across each level in the tree, starting at the bottom.

Due to the highly interconnected nature of the FFT calculation, several interconnection schemes have been proposed to ‘wire’ an FFT. For example, one popular architecture is the Butterfly FFT algorithm (with and without transpose). Hard wiring of FFT algorithms has therefore been implemented in ASICs and DSPs. Such a hard-wired FFT is much faster than any software-implemented

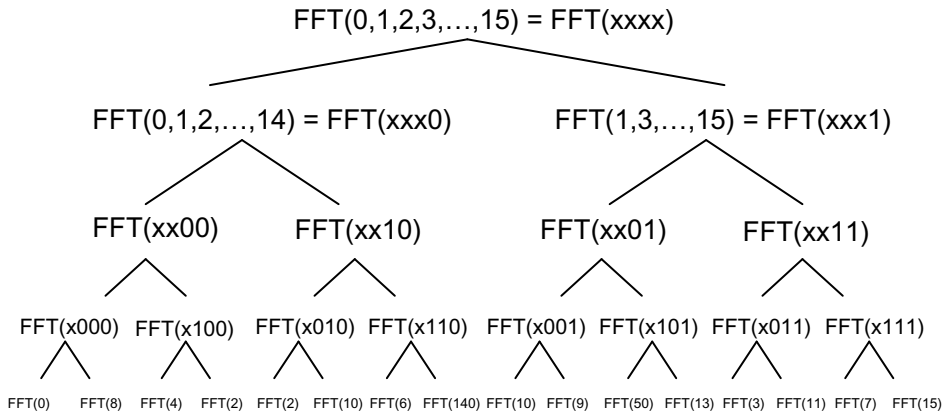


**Figure C.6** The FFT algorithm tree

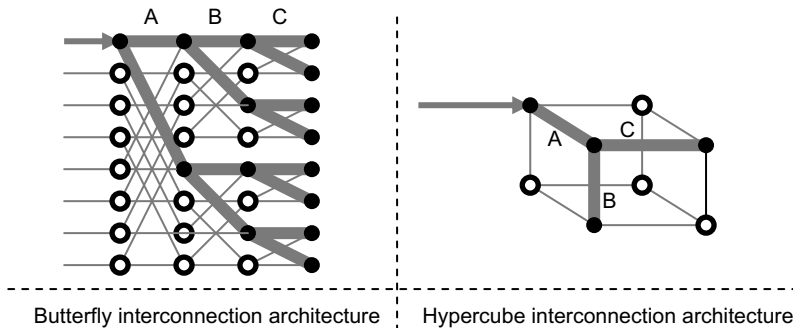


**Figure C.7** Butterfly and hypercube interconnection architectures, which can implement hard-wired ‘optical’ or ‘electronic’ 1D and 2D FFTs

FFT, but is dedicated to a specific task. For example, hard-wired FFTs in butterfly architectures are found today in diffractive projectors, where the image on the microdisplay is the Fourier transform of the pattern projected in the far field by that projector (for details of that technology, see Chapter 16). Figure C.7 shows the butterfly and hypercube interconnection architectures used in order to compute FFTs by hard-wiring or by software.

Digital diffractive optics (as in arrays of fan-out gratings) are actually very good candidates for the physical implementation of hard-wired butterfly or hypercube interconnection architectures in Opto-Electronic (OE) modules (as seen in Figure C.7). These elements and architectures are also used in parallel opto-electronic computing and Multi-Chip Modules (MCMs – see also Chapter 16).

Imagine that each node in Figure C.7 in the butterfly or hypercube interconnection scheme is actually a fan-out grating that splits the incoming beam into  $n$  beams that carry on to the next stage of interconnection, which includes a detector and another laser (as in VCSEL-based smart pixel arrays). It is quite ironic that such OE modules can actually implement hard-wired FFTs (and other functions such as Poisson’s equation etc.), not electronically, as is done in hard-wired ASICs, but optically in free space, and therefore can help to compute digitally a Fraunhofer diffraction pattern, by using Fraunhofer diffractive elements.

## C.5 The Properties of the Fourier Transform and Examples in Optics

Table C.1 summarizes the properties of 2D Fourier transforms, which are a generalization of 1D Fourier transforms. Here, we describe some analogies between mathematical operators and Fourier optics using diffractive elements:

- *Optical linearity*: Assume that we have two diffractive elements, one of which is a 1 to 3 fan-out grating while the other is a 1 to 4 fan-out grating. When illuminated by a single Gaussian laser beam, the sandwiched pair (equivalent to the addition of the two complex amplitudes) will yield an array of  $12 \times 12$  spots. This is an optical example of the linearity property of the FT.
- *Optical convolution*: A simple optical convolution example is the far-field pattern of a CGH illuminated by an LED (a partially coherent light source, with no spatial coherence). The far-field reconstruction is (approximately) the Point Spread Function (PSF) of the CGH (as illuminated, for example, by a laser centered on the LED spectrum), which is the FT of the CGH, convoluted by the Fourier transform of the LED aperture. The other effect not analyzed here is the partial temporal coherence. This result is also described in Chapter 11, and is an example of optical convolution.

**Table C.1** Properties of Fourier transforms

|                                     | $f(x, y)$                                                                                                                                                                           | $F(u, v) = TF(f(x, y))$                                                               |
|-------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|
| Linearity                           | $a.g(x, y) + b.h(x, y)$                                                                                                                                                             | $a.G(u, v) + b.H(u, v)$                                                               |
| Convolution                         | $g(x, y) * h(x, y)$                                                                                                                                                                 | $G(u, v).H(u, v)$                                                                     |
| Correlation                         | $g(x, y) \otimes h(x, y)$                                                                                                                                                           | $G(u, v).H^*(u, v)$                                                                   |
| Modulation                          | $g(x).h(y)$                                                                                                                                                                         | $G(u).H(v)$                                                                           |
| Separable function                  | $g(x).h(y)$                                                                                                                                                                         | $G(u).H(v)$                                                                           |
| Space invariance                    | $g(x - x_0, y - y_0)$                                                                                                                                                               | $G(u, v)$                                                                             |
| Frequency shift                     | $g(x, y).e^{2j\pi(u_0x + v_0y)}$                                                                                                                                                    | $G(u - u_0, v - v_0)$                                                                 |
| Differentiation in space domain     | $\frac{\partial^k}{\partial x^k} \frac{\partial^l}{\partial y^l} g(x, y)$                                                                                                           | $(2\pi j u)^k.(2\pi j v)^l.G(u, v)$                                                   |
| Differentiation in frequency domain | $(2\pi j x)^k.(2\pi j y)^l.g(x, y)$                                                                                                                                                 | $\frac{\partial^k}{\partial u^k} \frac{\partial^l}{\partial v^l} G(u, v)$             |
| Laplacian in space domain           | $\left(\frac{\partial^2}{\partial x^2} \frac{\partial^2}{\partial y^2}\right)g(x, y)$                                                                                               | $-4\pi^2.(u^2 + v^2).G(u, v)$                                                         |
| Laplacian in frequency domain       | $-4\pi^2.(x^2 + y^2).g(x, y)$                                                                                                                                                       | $\left(\frac{\partial^2}{\partial u^2} \frac{\partial^2}{\partial v^2}\right)G(u, v)$ |
| Signal squared                      | $ g(x, y) ^2$                                                                                                                                                                       | $G(u, v) * G^*(u, v)$                                                                 |
| Spectrum squared                    | $g(x, y) * g^*(x, y)$                                                                                                                                                               | $ G(u, v) ^2$                                                                         |
| Axis rotation                       | $g(\pm x, \pm y)$                                                                                                                                                                   | $G(\pm u, \pm v)$                                                                     |
| Parseval's theorem                  | $\int\limits_{-\infty}^{+\infty} \int\limits_{-\infty}^{+\infty} g(x, y).h^*(x, y).dx.dy = \int\limits_{-\infty}^{+\infty} \int\limits_{-\infty}^{+\infty} G(u, v).H^*(u, v).du.dv$ |                                                                                       |
| Real function $g(x, y)$             | $G(u, v) = G^*(-u, -v)$                                                                                                                                                             |                                                                                       |
| Real and even function $g(x, y)$    | $G(u, v)$ is real and even                                                                                                                                                          |                                                                                       |
| Real and odd function $g(x, y)$     | $G(u, v)$ is real and odd                                                                                                                                                           |                                                                                       |

- *Optical modulation:* Let us take the simple example of a transparent CGH and an opaque aperture. Setting the aperture on top of the CGH is like multiplying the amplitude distribution of the aperture stop by the phase distribution of the CGH. The result is therefore the FT of the CGH (the far-field pattern) multiplied by the FT of the aperture. If this aperture is a square, the far-field reconstruction will be modulated by a sinc envelope. If the aperture is a circle, the reconstruction will be modulated by a Bessel function. If there is no aperture, but the laser has a Gaussian profile, the reconstruction is modulated by a Gaussian envelope. This is an example of optical modulation.
- *Optical space invariance:* The Fraunhofer far-field pattern of a Fourier CGH is simply the complex 2D FT of the element. When one moves the CGH around in the  $x$  and  $y$  directions, the far-field reconstruction does not move. This is why one can replicate Fourier CGHs in  $x$  and  $y$ . The same thing happens for Fresnel elements. An array of microlenses produces the same far-field pattern when

the lenslet array moves laterally. However, this is not true in the near field, but only in the far field (FT). This is an example of optical space invariance.

- *Optical frequency shift:* Take a linear grating that reconstructs a spot in the far field at a given diffraction angle (given by the grating equation). A modulation of the grating function by a complex exponential (e.g. a sine function) will shift the diffraction angle (the spot in the far field) by the amount of the sine modulation frequency in the grating plane. This can be similar to a grating carrier modulation. This is an example of the optical frequency shift property of the FT.

## C.6 Other Transforms

We have seen that the FT is a good representation of a shift-invariant linear transform. However, the Fourier transform provides a poor representation of nonstationary signals, and provides a poor representation of discontinuous objects (the Gibbs effect). Multiscale transforms have therefore been introduced in the form of cosine transforms [6], fractional Fourier transforms [7], wavelet, ridgelet and curvelet transforms, Wigner transforms (or Wigner distributions) and so on.

We will briefly review below the fractional Fourier transform, the wavelet transform and the Wigner transform.

### C.6.1 The Fractional Fourier Transform

The Fractional Fourier Transform (FRFT) is a linear transformation that generalizes the Fourier transform. It can be thought of as the Fourier transform to the  $n$ th power, where  $n$  need not be an integer:

$$\text{FRFT}^n[U(u_n)] = U_n(u_n) = \int_{-\infty}^{+\infty} K_n(u_n, x) \cdot U(x) dx \quad (\text{C.15})$$

where the FRFT kernel  $K_n$  is defined as follows:

$$K_n(x_n, x) = K_\phi \cdot e^{i\pi(u_n^2 \cot\phi - 2u_n x \csc\phi + x^2 \cot\phi)} \quad (\text{C.16})$$

where

$$\phi = \frac{n\pi}{2} \quad \text{and} \quad K_\phi = \frac{e^{-\frac{i}{2}\left(\frac{\pi \cdot \text{sgn}(\phi)}{2} - \phi\right)}}{\sin^{1/2}(\phi)}$$

The Fourier transform can thus be applied in fractional increments, and can transform a complex amplitude signal into an *intermediate* domain:

$$\begin{cases} \text{FRFT}^1[U(x)] = U'(u) \\ \text{FRFT}^2[U(x)] = U(-x) \\ \text{FRFT}^3[U(x)] = U'(-u) \\ \text{FRFT}^4[U(x)] = U(x) \end{cases} \quad (\text{C.17})$$

The FRFT can be generally understood to correspond to a rotation in space/frequency phase space; for example, a domain that could exist between the hologram plane and the angular spectrum (not 'near' as in space, but 'near' as in a description). Efforts to develop a Discrete Fractional Fourier Transform (DFRFT) require a orthogonal set of DFT eigenvectors closely resembling Gauss–Hermite functions.

Fractional Fourier transforms have been applied to optical filter design, optical signal analysis and optical pattern recognition, as well as phase retrieval (see also the Gerchberg–Saxton algorithm in Chapter 6). The FRFT can be used to define fractional convolution, correlation and other operations, and can also be further generalized into the Linear Canonical Transformation (LCT). An early definition of the FRFT was given by V. Namias [8], but it was not widely recognized until it was independently reinvented around 1993 by several groups of researchers [9].



### C.6.2 Wavelet Transforms

Wavelet transform coefficients are partially localized in both space and frequency, and form a multiscale representation of a complex amplitude map with a constant scale factor, leading to localized angular spectra bands or sub-bands with equal widths on a logarithmic scale. The Continuous Wavelet Transform (or CWT) is defined as follows:

$$CWT_{a,b}[U(x)] = \frac{1}{\sqrt{a}} \int U(x) \cdot \Psi^* \left( \frac{x-b}{a} \right) dx \quad (C.18)$$

where  $\Psi$  denotes the mother wavelet. The parameter  $a$  represents the scale index, which is the reciprocal of the frequency  $u$ . The parameter  $b$  indicates the space shifting (or space translation). Wavelet analysis represents a windowing technique with variable-sized regions, and does not use a space-frequency region but, rather, a space-scale region.

Wavelet transforms are used extensively in image compression algorithms by exploiting the fact that real-world images tend to have internal morphological consistency, local luminance values, oriented edge continuations and higher-order correlations, such as textures and so on. Holographic and diffractive elements tend to have the same characteristics, to an even higher degree.

### C.6.3 The Wigner Transform

Another transform, close to the Fourier transform, which has direct usefulness in optics and especially holography, is the Wigner transform, which was originally published by E. Wigner in 1932, as a phase-space representation of quantum-mechanical systems. Such a phase-space representation can be applied to diffractive optics and holography since the reconstruction and holographic plane is either a Fourier or a Fresnel transform (to a scalar approximation).

We have seen previously that the Fourier transform  $U'(u)$  of a complex wavefront  $U(x)$  is defined as follows:

$$U'(u) = \int_{-\infty}^{+\infty} U(x) \cdot e^{-i2\pi \cdot ux} dx \quad (C.19)$$

The Wigner transform is defined as a bilinear transformation, as follows (here, we are using a 1D representation for the sake of simplicity):

$$W(x, u) = \int_{-\infty}^{+\infty} U \left( x + \frac{x'}{2} \right) \cdot U^* \left( x - \frac{x'}{2} \right) \cdot e^{-i2\pi \cdot ux'} dx' = \int_{-\infty}^{+\infty} U' \left( u + \frac{u'}{2} \right) \cdot U'^* \left( u - \frac{u'}{2} \right) \cdot e^{+i2\pi \cdot u'x} du' \quad (C.20)$$

The fact that the Wigner distribution (or transform) is a function of both the signal itself and its Fourier transform hardly goes unnoticed here. Such representations are highly desirable in holography, where the reconstructed signal is either a Fourier transform (far field) or a Fresnel transform (near-field reconstruction).

One important characteristic of the Wigner distribution is its projection rules (see Figure C.8):

$$\begin{cases} \int_{-\infty}^{+\infty} W(x, u) \cdot du = |U(x)|^2 \\ \int_{-\infty}^{+\infty} W(x, u) \cdot dx = |U'(u)|^2 \end{cases} \quad (C.21)$$

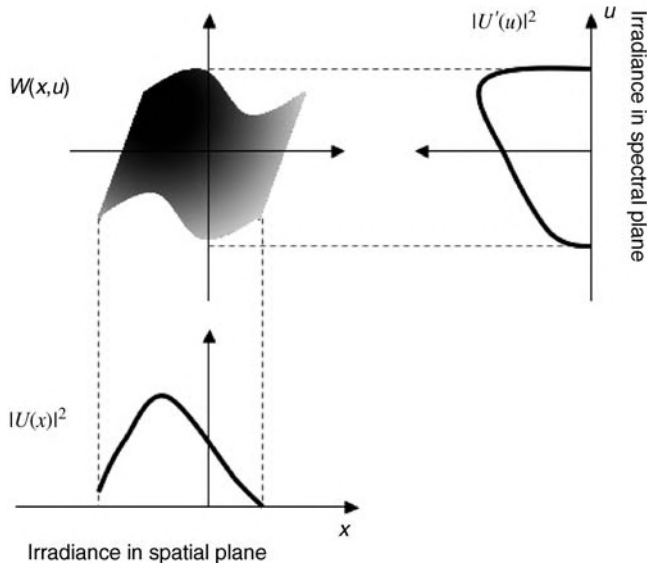


**Figure C.8** The Wigner distribution and irradiance projections on the spatial and frequency axes

Figure C.8 shows the intensity projection of the signal and its Fourier transform in the Wigner plane. For example, a quadratic phase profile (lens) and its Wigner distribution can be expressed as follows:

$$\begin{cases} U_{lens}(x) = e^{-i\frac{\pi}{\lambda f}x^2} \\ W_{lens}(x, u) = \delta\left(u + \frac{x}{\lambda f}\right) \end{cases} \quad (C.22)$$

Propagation in Fourier and Fresnel space can be analyzed in Wigner space. One example is the optimization of the setting of the optimal offset point of a Fourier or Fresnel diffractive element (or hologram), so that the reconstructions (diffraction orders) in either the near field or the far field are spatially separated. A plane wave illumination is represented in Wigner space as a shift, and a  $2f$  optical system as a rotation. The Wigner distribution is therefore an adequate tool to optimize the diffractive design parameters or the holographic recording parameters (or the space bandwidth product).

## References

- [1] E.O. Brigham, *'The Fast Fourier Transform and Its Applications'*, Signal Processing Series, Prentice-Hall, Englewood Cliffs, NJ, 1988.
- [2] E.A. Sziklas and A.E. Siegman, 'Diffraction calculations using the Fast Fourier transform methods', *Proceedings of the IEEE*, **62**, 1974, 410–412.
- [3] H. Hamam and J.L. de Bougrenet de la Tocnaye, 'Efficient Fresnel-transform algorithm based on fractional Fresnel diffraction', *Journal of the Optical Society of America A*, **12**(9), 1995, 1920–1931.
- [4] W.H. Lee, 'Sampled Fourier transform hologram generated by computer', *Applied Optics*, **9**, 1970, 639–645.
- [5] H. Farhoosh, Y. Fainman and S.H. Lee, 'Algorithm for computation of large size fast Fourier transforms in computer generated holograms by interlaced sampling', *Optical Engineering*, **28**(6), 1989, 622–628.

- [6] D. Kerr, G.H. Kaufmann and G.E. Galizzi, 'Unwrapping of interferometric phase-fringe maps by the discrete cosine transform', *Applied Optics*, **35**(5), 1996, 810–816.
- [7] H.M. Ozaktas, M.A. Kutay and Z. Zalevsky, 'Fractional Fourier optics', *Journal of the Optical Society of America A*, **12**(4), 1995, 743–751.
- [8] V. Namias, 'The fractional order Fourier transform and its application to quantum mechanics', *Journal of the Institute of Mathematics and Its Applications*, **25**, 1980, 241–265.
- [9] L.B. Almeida, 'The fractional Fourier transform and time–frequency representations', *IEEE Transactions on Signal Processing*, **42**(11), 1994, 3084–3091.



# Index

- Abbe V number 9, 96, **161–164**, 465
- Aberrations
- Diffractive lens aberrations* **96–99**, 161–162, 241, 250, 515, 558
  - GRIN lens aberrations* **50**, 53
  - Holographic lens aberrations* **163**, 202–203
  - Hybrid lens aberrations* 167, **176–178**, 555–556
  - Refractive lens aberrations* 96, 106, **158–161**, 168, 417
  - Wavefront aberrations* 62, 68, **106–107**
- Achromat singlet 20, 60, **163–165**, 176–178, 556
- Adaptive optics 68, **237**, 533
- Add-Drop module **44–45**
- Aerial Image 328, 373, 397, **414–415**, 431–432, 445–445, 450–451
- Alignment errors 82, 508–510, **514–517**
- Alligator lens **60**
- Anamorphic lens (see Lens)
- Anti-counterfeiting 88, 528, **535**
- Anti Reflection Surface (ARS) 49, 260, 267, **269**, 539
- Artificial materials **8**, 288
- Athermal singlet 96, **165–166**
- Atomic Force Microscope (AFM) 470, **497–498**
- Attenuated Phase Shift Mask **445**
- Arrayed Waveguide Gratings (AWG) (see Gratings)
- Beam
- Beam homogenizer* **66–68**, 315
  - Beam Propagation Method (BPM)* **322–323**
  - Beam sampler* 74, **526**
  - Beam shaper* 58, 67, 72, 73, **104–105**, 152–153, 315, 320–321, 497, 530
  - Beam splitter* 73, 75, **151–152**, 172, 273, 550, 552
  - Beam steering* **64–65**, 285
- Binary
- Binary grating* 5, **79–80**, 82, 94, 210, 227, 256, 270
  - Binary grey scale mask* **396–397**
  - Binary mask* 71, 351, **395**, 397, 434
- Birefringence **273–275**, 461
- Blazed grating 9–10, **83–84**, 91, 226–227, 266, **325**, 342, 400, 423–424
- Blu-ray disk 98, 168, 243, **562**
- Boron Phosphorous Silicate Glass (BPSG) **386–387**
- Brightness Enhancement Film (BEF) **60**, 568
- Bragg
- Bragg grating* 39, 88, 169, 181, **193–199**
  - Bragg condition* **199**
  - Bragg coupler* **169**, 237, 546
- Broadband diffractive lens (see Lens)
- Burch encoding **144**
- Caltech Intermediate Format (CIF) **356**, 358
- Casting (UV) 211, **464**
- CD-ROM 57, 74, 81, 102, 166–167, 243, **562**
- Chemically Assisted Ion Beam Etching (CAIBE) 378, **390**, 438
- Chirped grating 77, **88**, 165, 235, 270–271
- Chromatic Aberration **96–98**, 160–162, 178, 251, 555, 558
- Clean room **487–489**
- CMOS sensor 66, 172, 242, 251, 534, 557, **561**
- Coarse Wavelength Division Multiplexing (CWDM) 21, 33, **168**
- CodeV 106, 173, **327**
- Coma 160, **162–163**
- Computer Aided Design / Manufacturing (CAD/CAM) 94–95, **173–175**, **327–329**
- Computer Generated Hologram (CGH) 18, 73, **111–152**, 181, 204–206, 249, 363, 406, 436–437, 439, 448, 470, 510
- Compound eyes **49**
- Conjugate order **73**, 80, 188, 120, **204**
- Contact lithography **366–367**, 372, **373–374**
- Cost function 113, **124**, 129–130, 136–140
- Cross Connect (Optical Module) **45**

- Curvature 91, 160, **174**  
 Cylindrical lens (see Lens)
- Dammann grating 77, **86–87**  
 Daisy lens (see Lens- Extended depth of focus)  
 Dense Wavelength Division Multiplexing (DWDM) 21, 35, 36, **37**, 40–44, 74, 190, 521, 544, 546  
 Depth Of Focus (DOF) 77, **96**, 417, 549  
 Design For Manufacturing (DFM) 327–328, **413**  
 Design Rule Check (DRC) **147–149**  
 Detuning factor (diffractive) **92**, 127, 317  
 Detuning parameter (holographic) **193–194**, 195  
 Diamond ruling 60, 340, **342**, 345, 455  
 Diamond turning 340, **342**, 344, 345, 455  
 Di-Chromated Gelatin (DCG) 187, 207, **208–209**, 211  
 Diffraction  
*Diffraction Phenomenon* **6–10**  
*Diffraction efficiency* 10, **80–84**, 93–95, 99, 124–125, 142, **192**, 195–197  
*Diffraction orders* **73–76**, 122, 122, 189–190, 205, 254  
 Diffraction models  
*Fourier diffraction (see Fourier)*  
*Fraunhofer diffraction (see Fraunhofer)*  
*Fresnel diffraction (see Fresnel)*  
*Rayleigh-Sommerfeld diffraction (see Rayleigh-Sommerfeld)*  
 Diffractive  
*Diffractive Diffuser (see Diffuser or Holographic Diffuser)*  
*Diffractive grating (see Grating)*  
*Diffractive interferogram lenses* **106–108**  
*Diffractive lenses* 76, **91–104**  
*Moiré DOE (M-DOE)* **240–241**  
*Diffractive Optical Element (DOE)* 17–18, 71, 73, **90–94**  
 Diffuser 72–73, **153–154**, 238–239, 554, 568  
 Digital  
*Digital/Analog optics* 2–3, 4  
*Digital Camera* 57, **250–251**, 330–331, 557, 560, 577  
*Digital Holography* **214–215**, 497, 524–525  
*Digital Light Processor (DLP)* 42, 68, **221**, 222, 245, 449, 569  
*Digital Versatile Disk (DVD)* 57, 102, 167, 243, 332–333, 391, **562**  
 Direct write 347, **388–390**, 409, 418–427, 517  
 Direct Binary Search (DBS) **135–136**, 138  
 Discrete Fourier Transform (DFT) 128, 141, **301–302**, 306–307, 308, 316–317, 318, **597**  
 Dispersion (spectral) 13, 56, 57, 96, 161–165, **190**, 204, 523  
 Distributed Feed-Back Laser (DFB) **38–39**, 169, 283  
 Distributed Bragg Reflector Laser (DBR) **38–39**, 169, 283  
 Double exposure (Holography) **212**, 524  
 Double exposure (Lithography) **450–451**, 514  
 Dry etching (see also RIE, RIBE and CAIBE) **378–379**, 476, 490–491  
 Dual Brightness Enhancement Film (DBEF) **60**, 568  
 Durer, Albrecht **396**  
 Effective index  
*Effective index (waveguide)* **30**, 32, 88, 233  
*Effective Medium (see also EMT)* **267–273**, 539  
*Effective Medium Theory (EMT)* 248, **267–273**, 296, 396–397, 504  
 Electron  
*Electron Beam Pattern Generators* 356–357, **370–371**, 401, 420  
*Electron Proximity Effects (2D and 3D EPE)* **420–423**  
*Electron Proximity Compensation (EPC)* **423–430**  
 Electronic Design Automation (EDA) 327–328, 356, 359, 362, **418–420**, 433–434  
 Encoder (linear, 2D and shaft encoder) 104, 239–240, 470, **541–545**, 545  
 Encoding (phase) **140–148**  
 Embossing 347, 453, 458, **459–464**, 476  
 EMBF format **356–357**, 359, 370  
 Error diffusion (encoding) 142, **154–147**  
 Extended Depth of Focus lens (see Lens)  
 Extended scalar theory 296, 310, **324–326**, 595  
 Etch depth  
*Etch depth* 67, **80–81**, 92, 93, 97, 147, 317  
*Etch depth errors (modeling)* 81, **316–317**, 379–381, 409, 443, 491–492  
*Etch depth errors (analysis)* 511, **513–514**  
*Etch depth measurements (see Profilometry)*  
 Fabless operation **479–480**, 519  
 Fabrication errors  
*Random fabrication errors* 317, 346, **442**  
*Systematic fabrication errors* 295, 316, 351, **379**, **388**  
 Fabry-Perot 38, 49, **269**  
 Fan-out grating 121, 140, 148, **151–152**, 224, 260, 517–518, 527, 550, 551–552  
 Far-field Pattern 76, 86, 121, **151–152**, 186, 204, 310, 319  
 Far-field /Near-field (definition) 152–153, **310**  
 Finite Distance Time Domain (FDTD) 72, **263–267**, 296, 327  
 Flat top generator (see also top Hat) 66–67, **105**, 153, 321  
 Fly's eye arrays **467**

- f-number **95**, **96**
- Fourier*
- Fourier Transform (see also DFT, FFT and IFTA) **74**, **116**, **122**, **130**, **249**, **300**, **595**, **597–606**
  - Fourier Pattern Generators (see also Laser Pointer Pattern Generator) **151–152**, **315**, **531**, **532**
  - Fourier Transform Lens **75**, **116**, **205**, **206**
  - Fast Fourier Transform (FFT)
  - Fractional Fourier Transform **606–607**
  - Fourier propagator **128**, **300–307**, **318**
  - Fourier CGH **114**
- Focussed Ion Beam (FIB) **409**, **414**
- Form birefringent elements **223–224**, **272–273**, **274**, **275**, **326**
- Fracture process **106**, **351**, **354**, **355**, **358–364**
- Fraunhofer*
- Fraunhofer diffraction regime **185–186**, **189–190**, **204**, **595**
  - Fraunhofer propagator **128**, **300–307**, **323**, **318**
- Free Spectral Range **79**
- Fresnel*
- Fresnel Transform **128**, **299**, **442**, **591**, **594**
  - Fresnel Diffraction **128**, **591**, **594**
  - Fresnel CGH **114**, **121**
  - Fresnel Focussators **152–153**, **202**, **527**, **550**
  - Fresnel Propagator **128**, **300–307**, **323**, **318**
  - Fresnel Zone Plate (FZP) **11**, **12**, **18**, **77**, **91–92**, **184**
- Gabor, Denis (see Hologram – Gabor)
- Gaussian to flat top beam shaper (see Flat Top and Beam Shaper)
- GDSII format **346**, **351**, **353**, **356**, **358–359**
- Genetic Algorithm **129**, **139**
- Geometrical optics **5**, **9**, **298**
- Gerchberg-Saxton algorithm **130–132**
- Gibbs phenomenon **5**, **601**
- Grating*
- Grating equation **10–11**, **76–78**, **297**
  - Grating efficiency (see Diffraction Efficiency)
  - Reflection Grating **38**, **44**, **77–78**, **190**, **274**, **523**
  - Transmission grating **74**, **197**, **234**, **523**
  - Binary Grating **5**, **79–80**, **82**, **94**, **210**, **227**, **256**, **270**
  - Multilevel Grating **80–83**
  - Blazed Grating **9–10**, **83–84**, **91**, **226–227**, **266**, **325**, **342**, **400**, **423–424**
  - Sawtooth Grating **61**, **77**, **83–84**, **206**, **267**, **340**
  - Sinusoidal Grating **85**
  - Resonant Grating **275–277**
  - Echelette Grating **43–44**, **84**, **454–455**, **546**
  - Fan-out grating **121**, **140**, **148**, **151–152**, **224**, **260**, **517–518**, **527**, **550**, **551**
  - Grating Light Valve (GLV), **245–246**
  - VIPA grating **77**, **87**
  - Grating strength **193–194**, **197**
  - Holographic Grating **17**, **77**, **79**, **84–85**, **170**, **186–187**, **536**, **560**, **576**
  - Arrayed waveguide grating **20**, **29**, **42–46**
  - Zero order grating **72**, **254**, **263**, **275–276**, **326**
- Gray tone*
- Binary gray tone mask **146**, **346**, **396–397**, **481**
  - Gray scale mask **398–399**
  - Gray tone lithography **47**, **72**, **340**, **346**, **395**, **398–406**
- Grey code **545**
- Green's function **6**, **298–299**, **588–589**
- GRIN lens **49–54**, **546–547**
- GRIN lens array **58–59**
- Hardbake (see also softbake) **376**, **487**, **489**
- Half tone mask **146**, **346**, **396–397**, **481**
- Head Mounted Display (HMD) **247**, **249**, **532**
- Head Up Display (HUD) **211**, **247**, **249**, **530–531**
- Helmholtz (Hermann Ludwig Ferdinand von) formulation **7**, **256**, **322**, **587–588**
- Hexamethyldisilazane (HMDS) **487**, **488**
- High Energy Beam Sensitive (HEBS) mask **400–402**, **403**, **406**, **481**
- High Temperature Poly-Silicon (HTPS) **221**, **247**
- Hologram **18**, **71**, **72**, **111**, **163**, **181**, **232**, **279**, **280**
- Gabor Hologram **181–182**, **185**, **199–201**
  - Leith and Upatnieks Hologram **181–182**, **340**
  - Kogelnik Hologram **192–193**
  - Computer generated Hologram (CGH) **18**, **73**, **111–152**, **181**, **204–206**, **249**, **363**, **406**, **436–437**, **439**, **448**, **470**, **510**
  - Thin/Thick Hologram **186/187**
- Holographic*
- Holographic Optical Element (HOE) **18**, **72**, **163**, **199–206**, **211**
  - Holographic origination **203–204**, **340**
  - Holographic recording **182–183**, **210**, **232**, **340**
  - Holographic angular multiplexing **213–214**, **565**
  - Holographic materials **207–210**, **211**
  - Holographic diffuser **72**, **153–154**, **568**
  - Holographic tag **528**
  - Holographic keyboard **564–567**
  - Holographic scanner **536–537**
  - Holographic wrapping paper **17**, **528**
  - Holographic bar code **535–536**
  - Holographic non destructive testing **524**, **575**
  - Holographic-Polymer Dispersed Liquid Crystal (H-PDLC) **72**, **170**, **209**, **211**, **232–235**, **572**
- Huygens (Christian) **6**, **7**, **318**, **589**

- Hybrid  
*Hybrid optics* 19–20, **157**  
*Hybrid achromat* **163–165**  
*Hybrid athermal* **165–166**
- IC (Integrated Circuit) 2, 253
- Imprint (Nano-) 211, 340, **472–475**, 476
- Immersion lithography 8, 419, **447–448**
- Index (Refractive) **8**
- Injection Molding 453, 458, **464–471**
- Ion Exchange process, **386**
- Insertion Loss (IL) **32–33**
- Interconnections (optical) 127, 151, 223, **537–539**
- Interferometric microscope **498–502**
- Intra-ocular lens **549**
- Iterative Fourier Transform Algorithm (IFTA) 129, **131–133**, 249, 553
- Indium Tin Oxide (ITO) electrode **219**, 388
- Job Deck File **485–486**
- Kinoform **18**, 142, 144
- Kirchhoff (Gustav) integral 6, 324, **588–592**
- Kogelnik (Herwig) theory **192–199**, 255, 259
- Laser  
*Laser material processing* **525–527**, 576  
*Laser pointer* 17, 511, 532, **553**  
*Laser scanner projector* 569, **570**, 572  
*Laser scanner* **536–537**  
*Laser skin treatment* **550**  
*Laser Beam Writer (LBW)* **369–370**, 529  
*Laser beam Ablation (LBA)* **430–431**
- Layout 329, 351, 353, **355–356**
- LCoS (Liquid Crystal on Silicon) 245, **246**, 249–250, 569, 571
- Left Handed Materials **288–289**
- Leith (Emmet) and Upatnieks (Juris) **181–182**, 340
- Lens  
*Refractive lens* 10, 12, **13**, 96, 158–159  
*GRIN lens* **49–54**, 546–547  
*Hybrid lens* 161–168, 330–332, 527, 563  
*Diffraction lens* 76, **91–104**  
*Holographic lens* 18, 72, 163, **199–206**, **211**  
*Micro lens* 57, 167, 558, 562  
*Micro lens array* **61–68**, 72, 114, 386–387, 406–408, 449, 533, 560–561  
*Superzone lens* **97**  
*Broadband or multiorder diffractive lens* **97–99**  
*Extended depth of focus lens* **102–103**, 240, 442, 549  
*Toroidal lens* **100**, 107–108  
*Cylindrical lens* 60, 64, **100**  
*Vortex lens* 26, 35, 53, 77, **103–104**, 202, 336–337, 502, 547–548
- Anamorphic lens **100–101**  
*Perfect lens* 127, **290**
- Lenticular array **64**
- Light Emitting Diode (LED)  
*LED source* **319–320**  
*LED modeling* **321**  
*LED light extraction* 530  
*LED secondary optics* 530
- Linear optical encoders **542**
- Lippmann Photography **188**
- Liquid Crystal (LC) 170, **220**, 230, 232
- Liquid Crystal Display (LCD) **219–220**, 229–231
- Liquid lens 8, **244**, 251
- Lithography  
*Lithography (optical)* 205, 273, 340, 350–353, **371–375**  
*Step and repeat lithography* 366–367, 368, 372, **373–375**, 413–414  
*Microlithography* 221, **347**  
*Nanolithography* **393–394**  
*Lithography and GAlvanoforming (LIGA)* **453–455**  
*Immersion Lithography* 8, 419, **447–448**  
*Deep UV (DUV) Lithography* 372, 414, **446**  
*Deep proton irradiation lithography* 450  
*Soft Lithography* **474–475**  
*X-ray Lithography* 366, 372, **453–455**  
*Maskless Lithography* 245, **448–449**
- Lohmann (Adolf) encoding 18, **142–144**, 340
- Magneto-Optical Disk (MO) **563**, 564
- Mask (see also Photomask and Reticle)  
*Masking layer* **366–367**  
*Mask aligner (see also Stepper)* **373**  
*Maskless lithography* 245, **448–449**  
*Masking misregistrations* 81, 391, 492, 496, **514–518**  
*Binary mask* 71, **351–352**, 421, 434  
*X-ray mask* **453–454**  
*Gray scale mask* **398–399**  
*Gray tone mask* 146, 346, **396–397**, 481
- Maxwell's (James Clerk) equations 255, 263, 280, **583**
- MCM (Multi Chip Modules) 39, **537–538**
- M-DOE (Moire-Diffractive Optical Elements) **240–241**
- MEBES format **356–357**, 359, 370
- Metamaterial 522, 1, 8, **288–292**, 539
- Micro  
*Micro Electro Mechanical Systems (MEMS)* 42, 72, 217, **220–221**, 222, 223–225  
*Micro Opto Electro Mechanical Systems (MOEMS)* **223–225**  
*Microlithography (see Optical Lithography)*  
*Micro-display* 219–220, 221, **244–249**, 569–571  
*Micro lens* 57, 167, 558, 562  
*Microprism (see prism)*



- Micro lens array* **61–68**, 72, 114, 386–387, 406–408, 449, 533, 560–561
- Micromirror array* 42, 68, **221–222**, 224–226, 245, 569
- Mode
- Multimode fiber* **24–25**
  - Single mode fiber* 21, **24–25**, 30
  - Mode matching* **30–31**, 33–35
  - Mode propagation* **26**, 28, 526
  - Cladding mode* **29**
  - Mode selection* 277, **526**
  - Multi-mode interference filter (MMI)* **169**
- Modeling (numerical) 1, 5, 192, 255, **295**
- Molding **455–460**
- Multilevel element **80–82**, 271, 340, 350–352
- Multifocus lens **113–114**, 420, 537–538, 550, 562
- Nanoimprint 211, 340, **472–475**, 476
- Nano-optics 3, 4, 19, **253**, 254
- Nano-lithography (see Lithography)
- Near field
- Near field propagator (see Fresnel Propagator)*
  - Near-field Pattern* 76, 121, **152–153**, 186, 204, 310
  - Near-field / Far-field* 152–153, **310**
- Nickel shim (see Shim)
- Numerical Aperture (NA)
- NA of waveguide* **24**, 26
  - NA of lens* **95**, 413, 556
- Null lens 77, **105–106**, 534
- Numerical modeling **299**
- Nyquist criteria **117–118**, 300–303, 311, 315, 602
- Ophthalmology 102, **548–549**
- Optical
- Optical lithography* 205, 273, 340, 350–353, **371–375**
  - Optical Proximity Compensation (OPC)* 358, **431–440**
  - Optical Coherence Tomography (OCT)* **488**, 525, 550
  - Opto-Electronics (OE)* 39, **537–539**
  - Optical Pick Up Unit (OPU)* 57, 72, **166–168**, 560–563
  - Optical anticounterfeiting* 211, 528, **535**, 555
  - Optical cloaking* **290–291**, 532
  - Optical computing* 87, 127, 223, **536–539**, 574, 578
  - Optical mouse* **485–486**
  - Optical data storage* 127, 123–214, **560–565**, 574, 578
  - Optical tweezers* 548, **551–553**
  - Optical Variable Device (OVD/OVID)* **88–89**, 554
- Optimization algorithms 72, 127–128, **129–140**
- Origami lens **172–173**, 556–557
- Orthogonal Cylindrical Diffractive Lens (OCDL) **100**
- Oversampling 122, **311–312**
- Polydimethylsiloxane (PDMS) Stamp **475**
- Pico projectors 218, 246, 249, **569–572**
- Phase Shift Mask (PSM) 328, **443–445**, 528
- Phase
- Phase grating* 74, 197, 234, 523
  - Phase quantization* 74, 146, 351–352, **355**, 362
  - Phase multiplexing* **149–150**
  - Phase encoding* **140–148**
  - Virtual Phase Array (VIPA)* 77, **87**
  - Phasar* 20, 29, **42–46**
- Photolithography (See Lithography)
- Photomask (see Mask)
- Photonic Crystal 4, 8, 17, 72, 254, **279–288**, 289
- Photo-reduction **346–347**, 353
- Photorefractive **210**
- Photoresist 145, 203, **210**, 211, 274, 341–344, 375–376, 386–387
- Physical optics **1–3**, 298–299, 327
- Planar Lightwave Circuits (PLC) 4, 17, 21, **28–29**, 30–36, 521–522, 546
- Planar optics 3–4, **171–172**
- Plasma etching **378–380**, 390–391, 529
- Plasmon (surface) 8, **277–280**, 289, 290, 522, 534, 551
- Plastic injection molding (see Molding)
- Point Spread Function (PSF) 319–320, 419–420, 424–426, 431
- Polarization
- Polarization effects* 7, 19, 23, 44, 88, 96, 148, **197–198**, 220, 322
  - Polarization combining* 151, 214, **273–275**, 546–547
  - Polarization mode selection* **526**
  - Polarization selective elements* 223–224, **254–255**
  - Polarization recovery* 230–231
  - Polarization maintaining fiber* **25–26**
  - Polarization dependent loss (PDL)* **33**, 44
- Poly Methyl MetAcrylate (PMMA) 453, 465, 470
- Pre-bake, post-bake 376, 487, **489**
- Prism
- Prisms* **9–11**
  - Micro-prism* **55–57**
  - Micro-prism array* **59–61**
  - Prism coupling* **31–32**
  - Prism dispersion* 57
  - Switchable prism* **230–231**
  - Super-Prism effect* **285**
- Profilometry (surface) **511**, **513–514**
- Process Control Monitors (PCM) 368, 406–407, **484–485**

- Process Definition-Manufacturing Instructions (PDMI) **487**, 492–493
- Propagators (numerical)  
*Fourier propagators* 128, **300–307**, 318  
*Fresnel propagators* 128, **300–307**, 323, 318
- Quality factor (hologram) **186/187**
- Quantisation noise 67, 73, **126**
- Quantisation (phase) **150**
- Quarter wave plate **275**
- Quarternary lens **382**
- Raman-Nath diffraction regime 211, 255, **259**
- Rayleigh  
*Rayleigh-Sommerfeld diffraction theory* 6, 7, 255, 256, 309, **592–593**  
*Rayleigh distance* **310–311**, 560  
*Rayleigh criterion* **190–191**, 417–418  
*Rayleigh scattering* **22**, 33
- Reflective grating (see Grating)
- Resonant gratings (see Grating)
- Reflection hologram (see Hologram)
- Rigorous Couple Wave Analysis (RCWA) 72, 192, **255**, **259–263**
- Reactive Ion Beam Etching (RIE-RIBE) **378–380**, 390–391
- Reticle (see Mask)
- RET (Reticle Enhancement Techniques) 418–419, **433–434**, 444–445
- Replication 1, **453**
- Rotational encoders (optical) 104, 239–240, 470, **541–545**, 545
- Rigorous Diffraction Models 72, 192, 255, 296, 326–327, **583**
- Rytov model **267**
- Reflow (resist) **386–387**
- Refractive  
*Refractive index* **8**  
*Refraction angle* **10**  
*Refractive/diffractive (hybrid) elements* 19–20, **157**  
*Refractive micro-optics* 9, **47**, 333, 386–387  
*Refractive lens* 10, 12, **13**, 96, 158–159  
*Refractive microlens* **57**, 167, 558, 562  
*Refractive microlens array* **61–68**, 72, 114, 386–387, 406–408, 449, 533, 560–561  
*Effective refractive index* **267–273**, 539  
*Equivalent refractive lens (ERL)* **296–297**
- Reconfigurable optics 19, 72, 211, 217, **244–249**, 567
- Roll to roll embossing 460, 464, **473–474**
- Sag equation **91**
- Sampling (see Nyquist)
- Sawtooth grating (see Grating)
- SBWP (Space Band Width Product) **113**, 123, 125, 135, 139, 223
- Scalar diffraction model (see Diffraction Models)
- Scratch-o-grams **16**
- Shack-Hartmann wavefront sensor 68, 236–237, 249, **533**
- Shadow (geometrical) 83, 310, **325–326**
- Shim **455–458**, 476
- Simulated annealing **135–137**, 140
- Slab waveguide (see Waveguide)
- SLM (Spatial light Modulator) 220, 225–226, **245**, 249
- Snell's law 7, **9–10**, 159–160, 195, 288, 296
- Solar  
*Solar concentrator* 211, 393, 522, **539–540**  
*Solar tracking* **539–540**  
*Solar trapping* **541**
- Softbake (see also Hardbake)
- Soft lithography 340, **474–475**
- Sol-Gel materials 453, **471–472**, 476
- Spectral dispersion (see Dispersion)
- Spectroscopic gratings 17, 166, 190, 207, **522–523**
- Speckle  
*Speckle modeling* 313–315  
*Speckle reduction* **238–239**, 249–250
- Spherical aberration (see Aberration)
- Spherical GRIN lens **53**
- Spin coating 375, **488–489**, 517
- Spot array generator 72, 95, 121, 140, 148, **151–152**, 224, 260, 517–518, 527, 550, 551–552
- Strehl ratio 97, **127**, 299, 333
- Stepper 366–367, 368, 372, **373–375**, 413–414
- Stochastic algorithms 72, 127–128, **129–140**
- Subwavelength gratings (see Gratings)
- Superzone lens (see Lens)
- Super prism effect **285**
- Super lens effect 127, **290**
- Surface Plasmon Polariton (SPP - see Plasmon)
- Surface profilometry (see Profilometry)
- Sweatt model 161–162, 173, **296–297**
- Switchable optics 1, 19, 72, 211, **223–235**
- Systematic fabrication errors 295, 316, 351, **379**, **388**
- Talbot  
*Talbot illuminators* **89**  
*Talbot self imaging* **89**, 406–408  
*Talbot distance* **90**
- Thin film coating 38, 42, 49, 58, 168–169, **269**
- Thin/Thick hologram **186/187**
- Top hat (Gaussian to top hat) 66–67, **105**, 153, 321
- Total Internal Reflection (TIR) **23**
- Transmission hologram (see Hologram)
- Tunable optics 19, 72, 217, **235–244**, 553
- Unidirectional algorithms **135**
- UV casting 211, 458, **464**, 476

- V number (see Abbe V-Number)
- VCSEL (Vertical Cavity Surface Emitting Laser) 21, 34, 38, **39–40**, 64, 319–320, 526, 537–539
- Vector diffraction theory (see Diffraction Models)
- Virtual Phase Array (VIPA) (see Grating – VIPA)
- Virtual keyboard **564–567**
- VLSI (Very Large Scale Integration) 1, **414**, 494
- VOA (Variable Optical Attenuator) **42**, 169, 237, 546
- Vortex lens 26, 35, 53, 77, **103–104**, 202, 336–337, 502, 547–548
- Wafer
- Wafer scale optics* 55, 57–58
  - Wafer material* **347–349**
  - Wafer sizes* **367**
  - Wafer quality* **350**
  - Wafer processing* **375–385**
  - Wafer dicing* **383–385**
  - Wafer Spin Rinse Dry* **487**
  - Wafer back/front alignment* **406–408**
- Wave
- Wave optics* **1–3**, 298–299, 327
  - Wave equation* **483–484**
  - Wave plate* 275
- Wavefront
- Wavefront analysis* 77, **105–106**, 534
  - Wavefront coding* **250–251**, 532, 558–559
  - Wavefront coupling* **260**
  - Wavefront sensor* **68**, 236–237, 365, 533
- Waveguide
- Fiber waveguide* 21, **24–25**
  - Channel waveguide* **28–29**
  - Planar Lightwave Circuit (PLC)* 4, 17, 21, **28–29**, 30–36, 521–522, 546
  - Slab waveguide* 17, **27–28**
  - Waveguide mode matching* **30–31**, 33–34
  - Waveguide grating* **169**
  - Arrayed Waveguide* 20, 29, **42–46**
  - Waveguide Grating Routers* **43**
  - Waveguide propagation losses* **22**
  - Waveguide modes* **26–27**
  - Waveguide parameters* **26**
- Wavelength Division Multiplexing (WDM, *see* DWDM and CWDM)
- Wavelet Transform **607**
- Wet bench **375**
- Wet etching **376–377**
- Wigner Transform **607–608**
- X-ray lithography 366, 372, **453–455**
- X-ray mask **454**
- Yang-Gu algorithm **132–133**
- Young (Thomas) double slit experiment **5–7**
- Zemax 106, 173–174, 327, **329**
- Zones (Fresnel) 11–12, 18, **91–92**