



Développements et applications de nouvelles techniques de modélisation pour la conception optique contemporaine

Gabor Erdei

► To cite this version:

Gabor Erdei. Développements et applications de nouvelles techniques de modélisation pour la conception optique contemporaine. Optique [physics.optics]. Université Paris Sud - Paris XI, 2002. Français. NNT : . pastel-00716146

HAL Id: pastel-00716146

<https://pastel.hal.science/pastel-00716146>

Submitted on 10 Jul 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ORSAY
N° D'ORDRE:

**UNIVERSITÉ DE PARIS-SUD XI
U.F.R SCIENTIFIQUE D'ORSAY**

South-Paris University, Orsay, France

THÈSE

présentée
pour obtenir

LE GRADE DE DOCTEUR EN SCIENCES
DE L'UNIVERSITÉ DE PARIS XI ORSAY

**BUDAPESTI MŰSZAKI ÉS
GAZDASÁGTUDOMÁNYI EGYETEM
TERMÉSZETTUDOMÁNYI KAR**

Budapest University of Technology
and Economics, Budapest, Hungary

PH.D. DISSZERTÁCIÓ

Ph.D. dissertation by

ERDEI Gábor

Subject

**DEVELOPMENT AND APPLICATION OF NEW MODELLING
TECHNIQUES IN CONTEMPORARY OPTICAL DESIGN**

Supervisor: Dr. Pierre CHAVEL, Directeur de recherche
Consultant: Dr. Hervé SAUER, Maître de conférences

Laboratoire Charles Fabry de l'Institut d'Optique
unité mixte de recherche CNRS/IOTA/UPS UMR8501,
Centre Scientifique, Bât. 503,
91403 Orsay Cedex, France

Supervisor: Pr. RICHTER Péter, Head of department
Consultant: Dr. SZARVAS Gábor, Technical director
of the Optilink Hungary Corp.

Budapest University of Technology and Economics
Department of Atomic Physics
H-1111 Budapest, Budafoki út 8., Hungary

Defended at Budapest on 11th February 2002 in the presence of the Referee Board:

M. JANSZKY József, Dr.	Chairman
M. PRIMOT Jérôme, Dr.	Referee
M. VARGA Péter, Dr.	Referee
M. BAKOS József, Dr.	Secretary
M. CHAVEL Pierre, Dr.	Member
M. FRIGYES István, Dr.	Member
M. SAUER Hervé, Dr.	Member

TABLE OF CONTENTS

TABLE OF CONTENTS	2
PREFACE	3
ACKNOWLEDGEMENTS	4
THESES	5
PUBLICATIONS	9
OTHER, RELATED PUBLICATIONS	9
I GENERAL INTRODUCTION	11
II OVERVIEW OF COMMERCIAL OPTICAL DESIGN PROGRAMMES	13
1 MAIN FEATURES OF OPTICAL DESIGN PROGRAMMES	13
2 EXTENDING THE CAPABILITIES OF COMMERCIAL SOFTWARES	15
III MODELLING AND DESIGN OF HYBRID WAVEGUIDE/BULK SYSTEMS	17
TABLE OF CONTENTS	17
1 INTRODUCTION	18
2 FUNDAMENTALS	19
3 PRELIMINARIES (FORMER CONSIDERATIONS CONCERNING ANISOTROPY AND PRISM-OUTCOUPLED BEAMS)	23
4 RESULTS (EFFICIENT METHOD TO DESIGN WAVEGUIDE LENS SYSTEMS, AND THEIR OPTIMISATION)	24
5 APPLICATION EXAMPLES	29
6 VERIFICATION	34
7 CONCLUSION	35
IV DESIGN OF BEAM TRANSFORMATION SYSTEMS	36
TABLE OF CONTENTS	36
1 INTRODUCTION	37
2 FUNDAMENTALS	39
3 PRELIMINARIES (FORMER METHODS FOR THE DESIGN OF BEAM-SHAPERS)	42
4 RESULTS (NEW METHOD FOR THE DESIGN OF BEAM-SHAPERS)	45
5 APPLICATION EXAMPLES	50
6 VERIFICATION	55
7 CONCLUSION	57
V ACCURATE MODELLING OF HYBRID DIFFRACTIVE/CONVENTIONAL OPTICAL SYSTEMS	58
TABLE OF CONTENTS	58
1 INTRODUCTION	59
2 FUNDAMENTALS	61
3 PRELIMINARIES (STANDARD MODELLING TECHNIQUE, AND THE IMAGING PROPERTIES OF DOEs)	65
4 RESULTS (NEW DOE MODELLING TECHNIQUE AND ITS EFFICIENT IMPLEMENTATION)	72
5 APPLICATION EXAMPLES	75
6 VERIFICATION	82
7 CONCLUSION	86
VI GENERAL CONCLUSIONS	87
REFERENCES	88
APPENDIX A – USERS’ GUIDE TO “RS_DIFFR.EXE”	95
APPENDIX B – EFFECTS OF SAMPLING ON THE ACCURACY OF “RS_DIFFR.EXE”	100
APPENDIX C – PROGRAMME LISTING OF “RS_DIFFR.EXE”	103
APPENDIX D – GIBBS PHENOMENON AT MULTIPLE DIFFRACTION ORDER SUMMATION	107
APPENDIX E – GLOSSARY ON DEFINITIONS AND ABBREVIATIONS	110
APPENDED JOURNAL PAPERS	111
ABSTRACT	150

PREFACE

Recent technologies start to utilise light as a new tool for solving technical problems. The range of optical applications widens rapidly, while the engineering tasks continuously multiply and become more and more complex. This trend, together with the low efficiency of former design methods inspired optical engineers to “upgrade” their design tools. For a few decades superior optical design programs have been available commercially at relatively low prices. Being able to efficiently model not only accustomed optical systems but even exotic ones (such as those containing toroids, aspherics, diffraction gratings, multi-faceted surfaces, inhomogeneous materials, optical fibres etc.), these programmes became the main design tool of optical engineers in a short while. Keeping these programmes up-to-date for the modelling of currently available modern optical devices is an important undertaking of software makers for two reasons. First, the selection of optical devices such a programme is able to handle for the analysis and synthesis of optical systems counts one of its strongest cards on the market of design softwares. Second, exotic optical devices play an increasingly important role in contemporary optical design, thus their models are essential to be included in optical design softwares so that optical engineers are able to utilise them.

My work joins the evolution of commercial optical design programmes in that I made software extensions to them for the modelling of a range of modern optical components/systems that have been absent from their collection (viz. anisotropic waveguide lenses, prism couplers, beam transforming systems and diffractive optics). I realised beam propagation and optical surface modelling in commercial design programmes by simultaneously taking into account several aspects: true physical behaviour, optical design requirements and design software characteristics. Development of the models for the above computer routines and application of these new “engineering tools” constitute my scientific results. I developed new procedures for the modelling of various optical devices to make their design process (analysis and synthesis) more powerful (i.e. accurate, fast, convenient, providing additional degrees of freedom, allowing easy integration into complex systems) in comparison with former methods; I made theoretical considerations for the structural modification of such devices to achieve better performance; using the new modelling procedures I designed optical systems for improved characteristics.

Now briefly about the structure of this Ph.D. dissertation. Since I performed my research activity in three areas of optical design, the best method for the presentation of my results seemed to group them into separate chapters. Each of these chapters are organised as follows: discussion of the results of previous research, summary of the applied models and the assumptions I made, presenting my results together with examples for their application, and finally verification of the results.

ACKNOWLEDGEMENTS

At this point I wish to say thanks to my Hungarian and French supervisors, Dr. RICHTER Péter and Dr. Pierre CHAVEL, who made it possible for me to work together with them in these interesting projects. I would also like to express my gratitude to my consultants, dr. SZARVAS Gábor and dr. Hervé SAUER, who helped me with their excellent scientific knowledge as well as a never-ceasing energy. In addition, several colleagues at both Universities gave me useful advice and a lot of their precious time to teach me for their wisdom. Just a few of them to mention: in Hungary, the Department of Atomic Physics' dr. LŐRINCZ Emőke, dr. KOPPA Pál, Mr. VÁRKONYI Sándor, Mr. NÁDUDVARI Rudolf, the Department of Fabrication Technology's dr. MÉSZÁROS Imre, as well as in France, the Institut d'Optique's dr. Shamlam MALLICK; I apologize for all those not listed here. I am grateful to the Optilink Hungary Ltd., the Hungarian Soros Foundation and the Institut d'Optique for their financial support to my work, as well as SPIE for making the publishment of two articles possible in Optical Engineering. Last, but not at all least, I would like to thank to my parents and to my wife for their sacrificing tolerance and support during all the time I prepared this work.

Budapest, 17th January 2002

ERDEI Gábor

THESES

- 1 *For the design of integrated optical systems I developed a new, efficient ray-tracing method, by which the process of coupling 3D beams into and out of waveguide modes can be simulated, and the optical properties of 2D lens systems implemented in anisotropic planar waveguides can be analysed and optimised [1, 2, 3].*

Current commercial lens design programmes cannot usually evaluate and optimise complex integrated optical systems because of two main problems. First, such systems consist of both bulk (i.e. 3D) and waveguide (i.e. 2D) optical elements, the connection between which is usually performed by coupler prisms. Certain effects of these prisms on image quality cannot be modelled by conventional means, however their extent can be significant (see e.g. an unacceptable amount of image line inclination in transverse direction to the waveguide plane). The second problem is that the waveguide materials are often made of optically anisotropic single crystals ($\Delta n/n \approx 1\text{-}4\%$). If this is not taken into account, large defocusing ($\Delta f/f \approx 8\%$) and ray aberration ($R_{\text{RMS}}/R_{\text{AIRY}} \approx 10$) can occur. Thus, my aim was to produce fast computer routines for the modelling of refracting (and reflecting) surfaces in 2D anisotropic media and of the operation of in- and outcoupler prisms. These routines have been built into a general-purpose commercial optical design programme called OSLO SIX (Sinclair Optics, Inc.), using a specially developed surface model, and OSLO's macro programming capabilities. Optical design software usually offer several methods for their extension, which are rather slow when applied as ray-tracing algorithms. By my surface model I utilize the resources of the programme in a special way, which makes my code run approximately 20-40% faster than a pure macro routine (computation speed depends on the implementation of my routines). The correct operation of the new modelling technique has been checked experimentally on the fabricated prototype of an integrated optical system designed with the help of my routines.

- 2 *I proved by means of geometrical optics that inclination of the image line in transverse direction to the plane of a waveguide lens system is caused by the outcoupler prism; using my new modelling method (described in Thesis 1) I designed an appropriate prism for rectification of transverse image line inclination [2, 3].*

In case of many integrated optical systems, the focus spot of a waveguide lens is formed in air, utilising a frustrated total internal reflection prism to couple the light out of the planar waveguide. Observations showed that, though have not explained why, the image line of such systems shows an inclination in the direction that is perpendicular to the waveguide plane. Rectification of this "curved" image line has not been solved yet, however it is an important problem when the focus spot position is to be analysed by a linear photodetector array (see e.g. integrated acousto-optical spectrum analysers). Using a conventional photodetector bar (e.g. with detector element size of 13

μm), the transverse inclination can be five times larger than the detector element size, causing off-axis spots fall off the detector preventing them to be analysed. By application of pure geometrical optics I have shown that this inclination is in primary connection with the refracting angle of the outcoupler prism. Using my software extensions I designed a new prism for TIPE:LiNbO₃ waveguide, that has been fabricated and tested. According to my calculations, the transverse image line inclination reduced to 0.03% of that introduced by the traditionally used 45° prism. Experimental investigation of the designed system confirmed the rectification of the image line.

- 3 *I developed a new automated optimisation-based ray-tracing method for the design of beam shaping elements used for illumination by spatially coherent light; by application of the new method I designed a number of devices for different structure and beam shaping properties [4,5].*

A usual illumination task is to irradiate a surface homogeneously by a collimated laser beam. The difficulty at solving such problems is caused by that lasers operate usually in the TEM₀₀ resonator mode, i.e. the intensity profile of their output beam is Gaussian. The simplest way for reducing illumination inhomogeneity on the target surface, say to $\pm 5\%$, is to expand the Gaussian beam and cut the unnecessary part by an aperture. With this method about 90% of the input power is lost, implying that high illumination power can only be achieved by special beam transformation systems. For the design of such systems I developed a new approximate method based on geometrical optics. Though the method requires spatially coherent input light (like that of lasers), the intensity shaping is not limited to Gaussian-to-uniform transformation. The computation bases on the principle of automated optimisation, thus can be performed using any commercially available optical design programmes. By the new method I designed several devices of different beam shaping properties (i.e. input beam diameter, output beam diameter etc.), wavelength and structure (using spherical/aspheric surfaces). A design example, demonstrating the operation of the method, transforms the Gaussian beam of a He-Ne laser into a flat-top profile with two aspheric surfaces. The entrance pupil of the device truncates the input beam at a radius of three times the Gaussian spot-size (i.e. the $1/e^2$ intensity radius) to avoid diffraction effects. Ray-tracing predicts that the system has a wave aberration of about $0.07 \cdot \lambda$ RMS OPD, and 86% power transmission at $\pm 5\%$ inhomogeneity. The correct operation of the geometrical optically designed device has been verified by accurate diffraction calculations.

- 4 *I proved by means of geometrical optics that a sequence of optical devices, each performing Gaussian-to-uniform beam transformation, makes up a system that also shapes a Gaussian beam into a homogeneous one with good*

approximation; I used this theory for the design of an all-spherical beam shaping system of improved characteristics.
[6]

Transforming a collimated Gaussian beam into a homogeneous plane wave is a well-known problem, to the solution of which several devices have been developed. Even though such systems can be designed for perfect beam shaping by established methods, the fabrication technology of the necessary optical elements (usually high-precision aspheric or diffractive surfaces) is not available for all optical engineers (either because of cost or inadequate accuracy). To substitute the above special optical elements, one possibility is to make a system of high-precision spherical components exclusively. The design of such a system is not straightforward since no analytical solution exists, but some kind of (local or global) optimum can be sought for numerically. Though my method presented in Thesis 3 can be used for the design task, the optimum is not easy to locate in case of systems comprising more than four lenses due to the large number of degrees of freedom. In addition, the achievable beam shaping power (the ratio of entrance pupil radius to the radius of the Gaussian input beam) of “designable” systems is around unity, which is not high enough in most cases, since diffraction of the input Gaussian beam on the entrance pupil spoils the irradiance profile of the output beam. I pointed out theoretically that by placing low beam shaping power systems one after the other, the beam shaping power can be increased to eliminate diffraction problems, thus maintaining output beam homogeneity. For demonstrative purposes I designed such a system exclusively of spherical surfaces to transform the Gaussian beam of a frequency-doubled solid-state laser into a collimated beam of flat-top profile. The entrance pupil of the device truncates the input beam at a radius of 2.2 times the Gaussian spot-size (i.e. the $1/e^2$ intensity radius) to avoid diffraction effects. The designed system has a wave aberration of 0.029λ RMS OPD and almost 70% overall power transmission at $\pm 5\%$ intensity inhomogeneity.

- 5 *I developed a new procedure for the modelling of surface-relief phase-modulation diffractive optical elements (DOEs), which is fast, allows increased degrees of freedom and provides more accurate results for the analysis of hybrid (diffractive/conventional) systems than the standard DOE modelling method [7].*

Phase-modulation diffractive devices constitute a large subgroup of diffractive optical elements (DOEs). Owing to their special dispersive and structural properties, as well as their high diffraction efficiency, such DOEs are frequently combined with conventional (i.e. refractive and reflective) components to reduce chromatic aberrations and thermal sensitivity of imaging systems. These so-called hybrid systems can be analysed and designed most efficiently by special ray-tracing algorithms. Since the built-in DOE model of commercial lens design programmes has certain deficiencies, the precise evaluation of hybrid systems is possible only by performing supplementary calculations. My work was aimed at calculating the optical characteristics of DOEs as precisely as

possible using a commercial optical design programme, without the need for less convenient external computations. I created a new surface model for surface-relief DOEs (the most widespread realisation form of phase-modulation DOEs), and a number of software extensions by which the analysis and design of hybrid systems become possible using solely a commercial ray-tracing programme. The theoretical background of the operation of my new DOE modelling technique is provided by “zone-decomposition”, a new, high-accuracy DOE modelling method developed in the scope of this work, which is also presented in the dissertation. By the new model I investigated several hybrid systems, some designed in our laboratory for research purposes as well as some found in the literature, and I compared the resulting optical behaviour to that predicted by the standard DOE model of commercial design programmes. Correct operation of the zone decomposition model has been verified by precise diffraction calculations.

PUBLICATIONS

- 1 G. Erdei, G. Szarvas and M. Barabás, “Anisotropic waveguide lens design with a commercial optical design program”, *Proceedings of the 12th International Congress LASER 95*, pp. 906-909, 1996.
- 2 G. Erdei, G. Szarvas and M. Barabás, “Extensions to a lens design program for the ray-optical design of waveguide lenses and prism couplers”, *Journal of Modern Optics*, Vol. 44, No. 2, pp. 415-430, 1997.
- 3 G. Erdei, M. Barabás and G. Szarvas, “2-D anisotropic waveguide lenses”, *Contributed routines to optical design program OSLO SIX Sinclair Optics Inc.*, 1996. (Now distributed with the main software.)
- 4 G. Erdei, G. Szarvas, E. Lőrincz and S. Várkonyi, “Single-element refractive optical device for laser beam profiling”, *Proceedings of the 13th International Congress LASER 97 (Sensors, Sensor Systems and Sensor Data Processing)*, pp. 400-411, 1997.
- 5 G. Erdei, G. Szarvas, E. Lőrincz and P. Richter, “Optimisation method for the design of beam shaping systems”, to be published in *Optical Engineering*, February 2002.
- 6 G. Erdei, G. Szarvas, E. Lőrincz and P. Richter, “Cascading low-quality beam-shapers to improve overall performance”, to be published in *Optical Engineering*, February 2002.
- 7 H. Sauer, P. Chavel and G. Erdei, “Diffractive optical elements in hybrid lenses: modeling and design by zone decomposition”, *Applied Optics*, Vol. 38, No. 31, pp. 6482-6486, 1999.

OTHER, RELATED PUBLICATIONS

- 8 G. Erdei, H. Sauer and P. Chavel, “Accurate modelling of diffractive optical elements by commercial optical design programmes”, to be sent for publishing to *Applied Optics*, 2001.
- 9 P. Kalló and G. Erdei, “Applications of afocal optical systems in metrology, instrumentation techniques and information technology”, *Gépészet '98 (Proceedings of First Conference on Mechanical Engineering)*, Vol. 2, pp. 835-842, 1998.
- 10 G. Erdei, G. Szarvas, P. Kalló and E. Lőrincz, “Telecentric / inverse telecentric objective for optical data storage purposes”, *Optika '98 (Proceedings of 5th Congress on Modern Optics)*, Vol. 3573, pp. 380-383, 1998.

- 11 G. Erdei, J. Fodor, P. Kalló, G. Szarvas and F. Ujhelyi, “Design of high numerical aperture Fourier objectives for holographic memory card writing/reading equipment”, *Proc. of SPIE*, Vol. 4093, No. 67, pp. 464-473, SPIE’s International Symposium on Optical Science and Technology 30 July - 4 August 2000, San Diego, CA.
- 12 G. Erdei, G. Szarvas, E. Lőrincz, J. Fodor, F. Ujhelyi, P. Koppa, P. Várhegyi and P. Richter, “Optical system of holographic memory card writing/reading equipment”, *Proc. of SPIE*, Vol. 4092, No. 14, pp. 109-118, SPIE’s International Symposium on Optical Science and Technology 30 July - 4 August 2000, San Diego, CA.
- 13 G. Erdei, E. Lőrincz, P. Richter, G. Szarvas, S. Várkonyi, Hungarian national patent, Patent Application Number P9700257, “Single element optical device for transformation of the intensity distribution of light beams”, Priority 28/01/1997. (Title freely translated from Hungarian.)
- 14 Hungarian national patent, Patent Application Number P9701456, “Focus-servo system for optical memory card reading head”, Priority 01/09/1997. (Title freely translated from Hungarian.)
- 15 Patent WIPO No. WO09957719A1, “System and method for recording of information on a holographic recording medium, preferably an optical card” , Issued 11/11/1999.

I GENERAL INTRODUCTION

“Harnessing light” – this concept appears to dominate science and industry beyond 2000, the same way as it happened with electricity in the 20th century. Several scientific surveys concluded this currently, experiencing that light is gradually used as a “tool” to solve a widening range of technical problems [32]. The roots of modern optical devices are in the ancient desires to extend the natural limitations of the human visual detection system (in form of mirrors, eyeglasses, magnifiers, telescopes, microscopes, periscopes etc.) and to enlarge the capacity of our visual memory (see photographic still and cinematographic cameras). Except for some rare cases (e.g. the application of focusing mirrors and lenses to make fire), it happened only during the past 40 years that further benefits of optical solutions (non-contact, pollution-free, high-precision, high-resolution, high-energy, high-speed etc.) became well-known, inducing industrial and scientific application of optics in areas other than those connected to human vision: such as micro-lithography, telecommunications, data storage, metrology, production testing, printing, material processing, medical surgery, cosmetics, biosensing etc. This was made possible by the simultaneous development of material sciences (making high-purity glasses, non-linear and anisotropic crystals, liquid-crystals, light-weight but rugged plastics, inhomogeneous materials, special adhesives, photo-sensitive materials, light-emitting polymers etc.), and fabrication technology (injection-moulding, ultra-precision single-point diamond-turning, vacuum-evaporation, micro-machining, asphere generators, replication techniques etc.), due to which new, high-precision optical elements became available at relatively low prices (fibre optics, waveguides, graded-index lenses, diffractive optics, aspherics, lens arrays, microlenses etc.).

The flourishing of optics in science and industry – after the already existing lens designer and optical engineer – gave birth to the profession of optical designers. Their tasks cover those of the former two: they create complex optical systems from light source to detector, including the design of imaging/illumination/interferometric systems, lasers, thin-film coatings, waveguides etc. Optical modelling has ever required heavy computations (since the early beginnings of lens design), so it is not surprising that today’s optical designers need rather efficient tools if they want to achieve results relatively rapidly. Calculation of the optical properties of lens systems was one of the first applications of digital computers (from the 1960’s), but optical design softwares running on widely-available personal computers (PCs) started to commercialise only in the 1980’s. Nowadays, these complex programmes offer the most efficient tools for optical designers. Though there exist several types of optical design programmes (specialised to thin-film calculations, diffraction modelling of beam propagation, field calculation in waveguides etc.), in my dissertation I will deal only with those developed especially for solving problems in the field of optical imaging and illumination systems.

Quite often, the modelling abilities of optical designers are limited when they need to use certain modern optical devices in their systems. This means that either there is no commercial design method for a particular device at all, or the existing method is complicated, imprecise, inefficient or non-user-friendly. These reasons gave me inspiration to get acquainted with the fundamental tool of optical designers, the commercial optical design programme, in order that I can adapt it to the latest achievements of optics research and development (R&D), especially in the field of the design of optical imaging and illumination systems. During my work, I developed new procedures for the geometrical optical description of beam propagation through special optical surfaces and media. I realised these “engineering tools” in the form of efficient software extensions built into commercial optical design programmes. These computer routines can be easily used for both the high-precision analysis and the design of optical systems. Since commercial design programmes use mainly ray-tracing for beam propagation modelling, my new procedures have also been realised within the approximation of geometrical optics. The correct operation of my ray-tracing methods, I confirmed either experimentally or by precise diffraction calculations.

The following part of the dissertation is divided into four main chapters. Chapter II discusses the most important features of commercial optical design softwares, which I think useful for the easy understanding of the subsequent chapters. Chapter III, IV and V present my results in the following sequence: efficient modelling and design of hybrid waveguide/bulk systems, design of beam transformation systems, and accurate modelling of hybrid diffractive/conventional systems. After the main part, five appendices follow giving some more detailed explanation on specific topics not elucidated in the main text. I highlight here Appendix E, which gives a glossary on the most frequently used abbreviations, definitions and nomenclature applied throughout the dissertation.

II OVERVIEW OF COMMERCIAL OPTICAL DESIGN PROGRAMMES

1 Main features of optical design programmes

Optical design programmes have five main functions: acquisition of optical systems (lens entry or description), calculating optical properties of the system (optical evaluation or analysis), automatically modifying the system parameters to improve optical performance (optimisation), determining the effects of fabrication errors on optical characteristics (tolerancing) and making documentations of the design for fabrication and presentation purposes. Below, the most important items of these tasks are discussed briefly, further details can be found in [95], [79] and [92].

Such programmes consider optical systems to be constituted of *surfaces*, and store all optical and geometrical properties (refractive indices of materials, lens thickness, radii of curvature, aspheric coefficients etc.) assigned to these surfaces. The choice of surface types that optical design programmes can model is quite large: it extends from simple spherical refractive/reflective surfaces through Fresnel lenses and aspheric surfaces to diffractive optics and gradient index lenses. From these surfaces, (pratically) arbitrarily complex, three-dimensional (3D) optical systems can be built.

The design of optical systems requires efficient modelling of light propagation through the different optical components, otherwise the design procedure would take an unacceptably long time interval. For this reason, commercial lens design programmes do most part of the modelling job on geometrical optical basis, i.e. by application of ray-tracing methods. In commercial desing programmes the main tool for all optical investigations is *exact* ray-tracing, where the rays proceed starting from a point source selected in object space. By the help of precise geometrical optical calculations, the different optical characteristics of optical systems could be determined in image space (such as focus spot size, spot diagram etc.). The simpler paraxial approximation is used to calculate the effective focal length, paraxial magnification etc. of optical systems. Since the shape of the wavefronts (i.e. equi-phase surfaces) of the light beam exiting from the optical system can be determined by geometrical optics, optical design programmes are also able to calculate the diffraction pattern of the focus spot. To do so, the programmes calculate the Optical Path Lengths (OPLs), measured from a reference surface (usually a plane, or a sphere centered on the current object point and intersecting the *entrance* pupil at the optical axis) along several rays to the *exit* pupil. This way, a phase-map of the exit pupil is created, which accurately describes the output wavefront. Taking account of aberrations at imaging the entrance pupil onto the exit pupil (the so-called pupil aberrations), the amplitude distribution of the input field, as well as the phase-map in the exit pupil, a complex ampitude distribution is determined over the exit pupil. From this complex ampitude-map, the image distance, and the shape of the exit pupil, the Point Spread

Function (PSF) can be calculated in the image plane by directly performing diffraction integrals (see e.g. the Huygens-Fresnel theory), or using the technique of Fast Fourier Transformation (FFT). An equally often used indicator of system performance is the Modulation Transfer Function (MTF), which can also be directly derived from the wavefront given in the exit pupil. Commercial optical design softwares can do sequential and non-sequential ray-tracing. Sequential ray-tracing is usually used for those optical systems, inside which all the rays traverse each surface according to a given, well-known sequence (see e.g. a photographic camera). Non-sequential ray-tracing is applied in cases when the sequence of traversing must be determined for all rays separately (see e.g. a corner cube reflector, where there are three surfaces a ray can hit). Nevertheless, this method is much slower than the previous one. In both cases, rays have one starting, and one endpoint inside the optical system, splitting (on e.g. a semi-reflecting mirror) is not possible.

Maybe the most important functions of commercial design programmes are the built-in powerful optimisation algorithms. The essence of these methods is that the programme automatically improves specific properties (called operands) of the current optical system by changing its construction parameters (called variables) given by the designer. There exist several kinds of optimisation algorithms, the basic types of which are automatic and automated optimisation. In case of both methods, the designer creates an error function (also called a merit function) from the sum of the squared and weighted values of the specified system characteristics minus their nominal values, which the program attempts to minimize. In case of automatic optimisation, the programme changes the variables quasi-stochastically (see e.g. Simulated Annealing or Genetic Algorithm) within a given range, to minimise the error function. This method is also called global optimisation, since it seeks for the absolute best solution for the prescribed characteristics. In case of automated optimisation, the programme attempts to minimize the error function by deterministic algorithms, such as the Damped Least Squares method, Powell's method or the Downhill Simplex method. These methods are also known as local optimisation methods, referring to their local minimum seeking nature. In special cases, an exact solution can also be found for particular properties of the optical system (a necessary condition for this is that there must be at least as many or more variables than constraints to be met). In these cases e.g. the method of Lagrangian multipliers is used (this is often mixed with the use of an error function). To choose the proper optimisation method, to provide a sufficient number of adequate degrees of freedom (variables), and to determine the operands and their weights usually requires an highly experienced optical designer.

Most optical design softwares provide the user some kind of macro programming language. This way, the main programme can be altered by customising the user interface, by adding new functions to the programme for optical analysis and optimisation, and also by creating user defined surfaces (see Section II.2) for modelling new optical elements. In macro routines, all the built-in commands of the

main programme can be used together with the ability of making loops, conditional branching, function calls, I/O and file procedures etc. Different programmes offer different levels of macro programming in the form of interpreted, compiled and executable codes, providing (significantly) different efficiency.

Finally, I present a “map” of currently (i.e. in the year 2000) available commercial design programmes in Tab. II.1. The list may not be complete, but I believe that the most significant and most wide-spread softwares are all indicated in it. Many of these programmes provide tools for data export/import to/from each other and mechanical CAD softwares. Their productivity is improved in a large extent being compatible with Production Test Interferometers (such as a Zygo interferometer, USA) and light-source characterisation systems (e.g. that of the Radiant Imaging company, USA).

Manufacturer	Optical design	Illumination analysis	Other
Optical Research Associates (USA)	CODE V	LIGHT TOOLS	-
Sinclair Optics, Inc. (USA)	OSLO	-	-
Focus Software (USA)	ZEMAX	OPTICAD	LENSVIEW
Radiant Imaging (USA)	-	ZELUM (with ZEMAX)	-
Optis (France)	SOLSTIS	SPEOS	OPTICALC
O++ (France)	-	APILUX	-
Wolfram Research, Inc. (USA)	OPTICA (with MATHEMATICA)	-	-
Lambda Research Corporation (USA)	-	TRACEPRO	-
Breault Research Organization (USA)	SYNOPSYS	ASAP	-
Linco (Spindler & Hoyer) (Germany)	-	-	GLASS MANAGER

Tab. II.1. Summary of commercial optical design softwares on market in the year 2000.

2 Extending the capabilities of commercial softwares

The most flexible optical design softwares allow the user to make extensions to the main programme, by which its surface modelling capabilities can be tailored according to the user’s needs. Basically, there are two such possibilities: user defined surfaces and user defined ray-tracing (different programmes use slightly different names for them). In case of the first one, the surface shape could be given by the user, providing a tool for creating arbitrary refracting/reflecting surfaces not contained among the programme’s range of surface models. When user defined ray-tracing is chosen, not only the surface shape can be prescribed by the user, but even the “law of refraction”. This way, the deflection of a ray can be controlled at any arbitrary point in space corresponding to the user’s needs.

The realisation of both previous cases is carried out through the macro programming abilities of the main programme: the user writes a code containing his/her requirements in the form of a macro routine that is called each time a ray hits the surface during all optical evaluation procedures. At these calls, the main programme passes several parameters to the user-written routine, e.g. the coordinates of the currently traced ray on the previous surface, coefficients that contain information given by the user

describing that surface etc. When the routine finished its operation, it returns to the main programme passing back the ray coordinates after traversing the surface. For user defined surfaces (used in Chapter V), below I briefly summarize the operation of the algorithm performed by the main programme to determine the intersection point of a ray with the surface. (In case of user defined ray-tracing, used in Chapter III, this computation is customized, i.e. it is also made by the user routine.)

To use a user defined surface in ray-tracing, one has to define the required surface shape, from which the programme determines the intersection points of rays with the surface and calculates the direction of refracted rays automatically. The profile ‘F’ of any arbitrary (even discontinuous) user defined surface must be given to the programmes in the following form [97]:

$$F(x, y, z) = 0, \quad (\text{II.1})$$

where (x, y, z) denote a point with its coordinates in a Cartesian system. In order to find the intersection point of a ray with such a surface, we should take notice of that any point P along that (straight) ray can be described by a location vector $\mathbf{r}(x, y, z)$:

$$\mathbf{r} = \mathbf{r}_0 + \mathbf{k} \cdot s, \quad (\text{II.2})$$

where $\mathbf{r}_0$ is the location vector of a reference point Q positioned anywhere on the ray, $\mathbf{k}$ is the wave vector that shows the ray direction, and s is a scalar parameter proportional to the distance between P and Q. Writing this relation into (II.1), we obtain a scalar equation $F(s) = 0$, which has to be solved by the programme. So that the refracted ray direction can be found, the programme also needs to know the first partial derivatives of $F(x, y, z)$. It is useful, if possible, to give the analytic form of the derivative functions to enable fast computation of the surface normal vector; otherwise the program uses the less precise and slower method of finite differences to calculate these derivatives.

III MODELLING AND DESIGN OF HYBRID WAVEGUIDE/BULK SYSTEMS

TABLE OF CONTENTS

TABLE OF CONTENTS.....	17
1 INTRODUCTION	18
2 FUNDAMENTALS.....	19
2.1 PRESUMPTIONS.....	19
2.2 MODEL OF RAY-TRACING IN PLANAR WAVEGUIDES.....	19
2.3 MODEL OF ANISOTROPY	19
2.4 MODEL OF IN- AND OUTCOUPLING	20
2.4.1 Incoupling of an infinite plane wave.....	21
2.4.2 Incoupling of a “synchronised” beam of finite cross section	21
2.4.3 In- and outcoupling of a finite beam with a prism	22
2.5 DEFINITION OF “TRANSVERSE INCLINATION”	22
3 PRELIMINARIES (FORMER CONSIDERATIONS CONCERNING ANISOTROPY AND PRISM-OUTCOUPLED BEAMS)....	23
3.1 INTRODUCTION TO ANISOTROPIC ABERRATIONS	23
3.2 OUTCOUPLING PHENOMENA USING COUPLER PRISMS	23
4 RESULTS (EFFICIENT METHOD TO DESIGN WAVEGUIDE LENS SYSTEMS, AND THEIR OPTIMISATION)	24
4.1 BUILDING THE MODEL OF ANISOTROPY INTO OSLO SIX	24
4.2 MODELLING METHOD OF COUPLER PRISMS FOR OPTICAL DESIGN.....	25
4.2.1 Model of incoupling.....	25
4.2.2 Model of outcoupling	26
4.2.3 Building the models into OSLO SIX	26
4.3 SYSTEM DESIGN FOR REDUCED TRANSVERSE IMAGE LINE INCLINATION	27
4.3.1 Geometrical optical interpretation – first approximation.....	27
4.3.2 Geometrical optical interpretation – second approximation	28
4.3.3 Minimising image line inclination.....	29
5 APPLICATION EXAMPLES	29
5.1 LENS 1.1: PLANAR DOUBLET IN ISOTROPIC MEDIA	29
5.2 LENS 1.2: PLANAR DOUBLET IN ANISOTROPIC MEDIA	30
5.3 LENS 1.3: PLANAR DOUBLET IN ANISOTROPIC MEDIA WITH OUTCOUPLER PRISM.....	30
5.4 LENS 2.1: PLANAR QUADRUPLLET IN ANISOTROPIC MEDIA	31
5.5 LENS 2.2: PLANAR QUADRUPLLET IN ANISOTROPIC MEDIA WITH OUTCOUPLER PRISM.....	31
5.6 LENS 3.1: PLANAR SINGLET IN ANISOTROPIC MEDIA WITH 45° OUTCOUPLER PRISM.....	31
5.7 LENS 3.2: PLANAR SINGLET IN ANISOTROPIC MEDIA WITH SEMI-OPTIMISED OUTCOUPLER PRISM.....	33
5.8 LENS 3.3: PLANAR SINGLET IN ANISOTROPIC MEDIA WITH OPTIMISED OUTCOUPLER PRISM.....	33
6 VERIFICATION.....	34
7 CONCLUSION	35

1 Introduction

Integrated optical devices always require accurate waveguide lenses. The special structural features of such systems make them inconvenient to be analysed by conventional optical design softwares. I developed an appropriate design environment in the popular optical CAD programme OSLO SIX [92] to facilitate the evaluation and optimisation of hybrid integrated optical lens systems (those containing both 2D, i.e. planar waveguide and 3D, i.e. bulk optical elements). The motivation for me to develop a new tool for the design of hybrid waveguide lens systems was that all the existing usual methods for the design of integrated optical systems have some deficiencies, furthermore, none of them can treat in- and outcoupling (i.e. the connection of 2D and 3D systems) [30]. To calculate 2D mirrors embedded in waveguides [74] one also needs a model of 2D anisotropic reflection which is missing from lens design codes. In summary, an appropriate software for the design of integrated optical waveguide lens systems must be able to handle:

- a) 2D lens systems with extended (i.e. far-from-zero) field angle and numerical aperture (*),
- b) 2D lens systems comprising a large number of surfaces with general profiles (*),
- c) 2D refracting/reflecting systems in anisotropic media,
- d) 2D lens systems coupled to 3D systems by frustrated total internal reflection prisms.
- e) Besides, it must be user friendly and fast enough to allow automated optimisation (*).

My objective was to fulfil the above requirements by a design environment in which all the phases of lens design (lens entry, evaluation, optimisation and tolerancing) are supported. I decided to use OSLO SIX because it inherently satisfies the requirements signed with a (*) symbol above, and it accepts user-written extensions (at so-called “User Defined Surfaces”) to handle the unconventional ray-tracing laws needed for the modelling of anisotropic refraction/reflection as well as in- and outcoupling. For this purpose, I prepared a set of routines (software extensions), which form a kind of tool-kit with the following “devices” in it:

- a) Ray-tracing routines for refraction/reflection on 2D surfaces of arbitrary shape in anisotropic waveguides, and for coupling 3D laser beams into/out of waveguide modes by coupler prisms.
- b) Routines that automatically set up the necessary “User Defined Surfaces” and store those special lens data (e.g. the ordinary effective refractive index) the previous routines require.
- c) Routines facilitating 2D lens evaluation (with 2D spot diagrams, ray distribution curves).
- d) Additional routines for the entry of 2D lens systems and the operands for their optimisation.

The above extensions were written in the C-like compiled macro language (CCL) of OSLO SIX [92]. The ray-tracing routines have also been transformed into executable format (viz. DLL) by Sinclair Optics, Inc. to enhance efficiency.

In the following I present the geometrical optical model of waveguide lens systems, as well as the method I used to build it into OSLO SIX. Through design examples and measurement results I demonstrate the usage and benefits of my integrated optical design environment. I also give the geometrical interpretation of transverse image line inclination (an effect that spoil image quality when using a prism to couple light out of the waveguide), and a method for its correction.

2 Fundamentals

2.1 Presumptions

At the Department of Atomic Physics of the Budapest University of Technology and Economics we deal with planar waveguide lenses [64, 105, 77, 74, 30, 113, 111, 65, 84, 56, 103] made on LiNbO_3 , which is optically uniaxially anisotropic, and use frustrated total internal reflection prisms for coupling a laser into and out of the waveguide. My aim was to shape the design environment so that it will be able to handle especially this kind of structure. To investigate anisotropic crystals from optical point-of-view, the orientation of the anisotropic axes should also be given. According to the most usual crystal orientation, the axis of anisotropy of LiNbO_3 lies in the plane of the waveguide (in an otherwise arbitrary position), which I also assume in my model.

2.2 Model of ray-tracing in planar waveguides

Mode propagation in planar waveguides can be described by 2D ray-tracing [108]. Rays in the waveguide represent the direction of energy flow in the mode. I will call these rays “surface rays”, since they propagate in 2D, parallel to the substrate surface of the waveguide. In order to calculate refracted ray directions and optical paths, one needs the refractive indices in the different waveguide regions – for this purpose I use the *effective* refractive indices of a given waveguide mode [66, 108].

2.3 Model of anisotropy

In this subsection, a summary of the model I applied in my computer routines is given; theoretical background of anisotropy can be found in [11], [103], [54], [55], [115] and [17]. Since the substrate material used in our laboratory is LiNbO_3 , I apply the model of uniaxial anisotropy corresponding to the optical properties of this crystal. I approximate the inverse velocity curve (i.e. the 2D “surface” of wave-vectors) with an ellipse [56] as described in [1]. In case of waveguides, anisotropy means that the effective index is direction dependent, which affects ray tracing in two respects. First, to calculate the

angle of refraction of the wavefront normal (wave vector), the direction dependent refractive indices must be substituted into formally the same expression as that of Snell's law. The second effect is the deviation of the direction of the wave vector from that of the ray vector (which is parallel to the Poynting vector). In geometrical optical approximation the intensity distribution of a beam is determined by the Poynting vector (see spot diagrams), therefore one must use the “non-Snellian” refraction law of the *rays* instead of the one related to *wavefront normals*. There is a third phenomenon too: wavefronts emitted by a point source in a uniaxial material are of elliptical shape. Consequently, the algorithm (based on isotropic Fourier optics), by which OSLO computes the diffraction pattern of the focal spot, should also be altered. However, as the change of the intensity distribution due to wavefront ellipticity – and to uniaxial anisotropy in general – was shown [17] to be negligible in most cases of practical interest, the original algorithm can be used.

In case the optical axis of a uniaxially anisotropic substrate material lies in the plane of the waveguide (which is exclusively subject to my consideration here), the curve of (effective) refractive indices can be best approximated by an ellipse [56]. Fig. III.1 shows the connection between the directions of the wave vector and the ray vector in such an anisotropic medium. Fig. III.2 depicts the condition of phase matching for wave vectors at the border line (corresponding to a conventional refractive “surface”) of two adjacent anisotropic regions having different refractive indices.

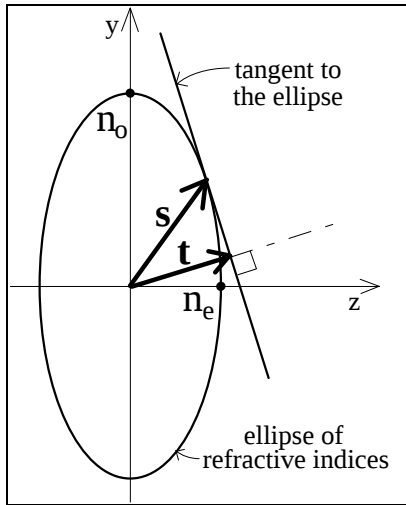


Fig. III.1. Connection between the wave vector $\mathbf{s}$ and the ray vector $\mathbf{t}$ in a two-dimensional medium of uniaxial anisotropy.

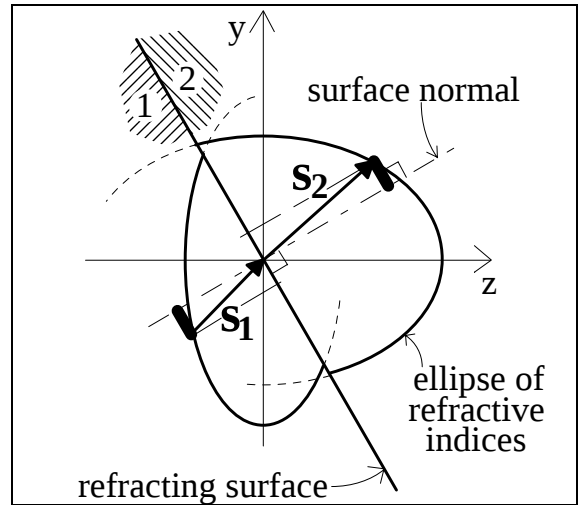


Fig. III.2. The condition of phase matching at a line separating uniaxial regions of different refractive indices ($\mathbf{s}_1$ =incident wave vector, $\mathbf{s}_2$ =refracted wave vector).

2.4 Model of in- and outcoupling

In this subsection, I overview the phenomena that served as a basis to my simple model developed for the purpose of optical design (see Section III.4). Detailed discussion about the theoretical background of frustrated total internal reflection prisms can be found in [109], [106] and [66].

2.4.1 Incoupling of an infinite plane wave

The incident angle of an incoming plane wave must exceed the critical angle of total reflection at the waveguide/substrate interface, otherwise the wave refracts into the substrate. The mode-intensity vs. incident angle function has a local maximum in the waveguide, if the incident wave falls onto the waveguide surface in the so-called “synchronous direction” of a waveguide mode. In this case the condition of phase matching between the desired mode and the incident beam is fulfilled:

$$|\mathbf{k}| \cdot \sin \Theta = \beta, \quad (\text{III.1})$$

where $\mathbf{k}$ is the wave vector of the incident ray and β is the modified propagation constant of the waveguide mode. The propagation constant modification is a consequence of frustrated total reflection in the waveguide at its interface with the coupler prism. The narrower is the air gap (see Fig. III.3), the higher is the change in the propagation constant.

Due to the presence of the low refractive index region of the air-gap between the high-refractive index waveguide and the even higher index coupler prism, the different waveguide modes can be excited separately, with high intensity. The wider is the gap size, the smaller is the angle interval in which incoupling is possible (around synchronous directions). Since the air gap size can never decrease to zero, a small phase-shift also occurs when light traverses from the prism to the waveguide.

2.4.2 Incoupling of a “synchronised” beam of finite cross section

A collimated incident beam of finite cross section falling in according to a synchronised direction excites an exponentially increasing and then decreasing field in the waveguide (Fig. III.3). The limiting

value of mode intensity corresponds to the intensity excited by an infinite plane wave, see previous subsection. The field profile of the mode in the “Z” direction can be characterized by the coupling length (see Fig. III.3). The coupling length is by definition the distance, at which mode intensity increases to $(1-1/e^2)$ fraction of its maximum value at the beginning of the incident beam, and decreases to its $1/e^2$ fraction at the end of it. The coupling length increases exponentially (i.e. coupling becomes weaker) with gap size.

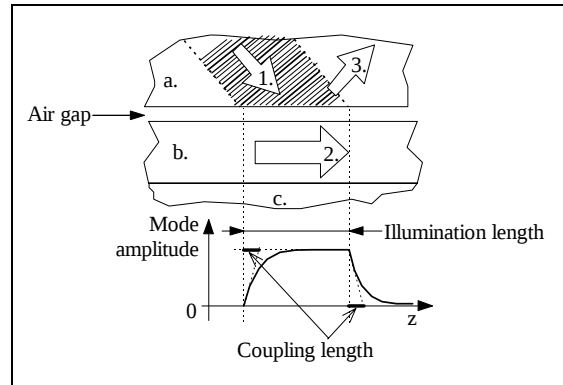


Fig. III.3. Field profile in the waveguide due to a collimated, finite incident beam. White arrows show energy propagation in the incident (1.), guided (2.) and reflected (3.) waves. (a. prism; b. waveguide; c. substrate.)

2.4.3 In- and outcoupling of a finite beam with a prism

If waveguide leakage is stopped at the end of the incident beam, the mode will propagate further in the waveguide without attenuation (neglecting losses caused by scattering on inhomogeneities etc.). This is satisfied if the incoming beam is positioned at the edge of the incoupler prism (see Fig. III.16).

Outcoupling takes place the opposite way (see Fig. III.4). As the mode reaches the edge of the outcoupler prism, its intensity gradually decreases to zero because of mode leakage. The mode profile in the “Z” direction will be the same as the decreasing part of an incoupled finite beam. The outcoupled beam is phase-matched to the waveguide mode. The small phase-shift through the air-gap and the propagation-constant modification here also occur.

2.5 Definition of “transverse inclination”

The image points formed by a waveguide lens are always situated on a single line in the case of a collimated input beam (see Fig. III.16). This is obvious, since the input beam direction can only be varied in 2D, i.e. in the plane of the waveguide (see integrated optical spectrum analyzers, beam deflectors [66]). According to former observations [106] and my calculations presented in Section III.4, the image line is not straight but curved if observed in air through an outcoupler prism. I call this phenomenon “transverse image line inclination”, which nomenclature intentionally implies that this effect is physically independent from the lens aberration known as curvature of field.

Phenomena behind transverse image line inclination can be traced back to the transverse inclination of single rays. To alleviate further explanations, I define “meridional planes” for the outcoupler prism (see Fig. III.4): a beam of parallel surface rays coupled out perpendicularly to the prism edge forms a plane in the prism and a plane in the air. I term these planes “meridional planes” because of their similarity to meridional planes used in classical geometrical optics. Surface rays that are not perpendicular to the prism edge cannot lie in the meridional plane after outcoupling (see Fig. III.4), i.e. they incline from it “transverse” direction. In case of single surface rays, I characterise transverse inclination by the distance between the outcoupled ray and the meridional plane, measured perpendicularly to the waveguide surface (see Fig. III.4). In case of beams, the displacement of the focus spot from the meridional plane in the air is of main interest. I describe the spot position by the transverse inclination of the ray that passes through its center (Fig. III.14 in Section III.5 was drawn in correspondence with this definition).

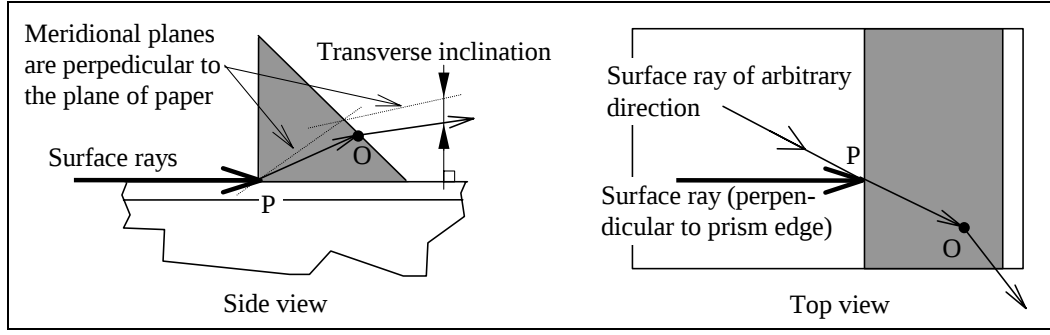


Fig. III.4. Explanation of meridional planes and definition of transverse image line inclination. Point "P" denotes intersection with prism edge, point "O" represents intersection with prism surface. (See also Fig. III.9 and Fig. III.10.)

3 Preliminaries (Former considerations concerning anisotropy and prism-outcoupled beams)

3.1 Introduction to anisotropic aberrations

Effects of optical anisotropy on the imaging properties of planar waveguide lenses have been investigated by several authors. These papers reveal large defocusing and focal spot deterioration (similar to spherical aberration) in optical systems designed with the usual approximation that characterise anisotropic lens materials with isotropic refractive indices. The main types of methods (all of which having some deficiencies) used for correction of anisotropic aberrations are listed below.

- The simplest way is to calculate analytically the necessary lens profile. Due to the approximations incorporated, this is sufficient only for small field-angle 2D isotropic singlets [113, 111].
- 3D design programmes (without modification) can simply treat 2D isotropic systems. In this way wide field-angle (aplanatic, etc.) lens systems can be designed [65, 84].
- If anisotropy has to be taken into account, the designer must develop his own design software [56, 103]. The length of time required to write such a program is usually not acceptable.
- Non-ray-tracing methods. Beam propagation methods [52, 79, 18] can analyze only low numerical aperture lens systems with a narrow field-of-view, and they are also too slow for optimisation.

3.2 Outcoupling phenomena using coupler prisms

Certain papers investigating planar waveguides report that scattered light (always present due to waveguide inhomogeneities as well as surface defects) forms curved lines (each corresponding to one waveguide mode) after coupling a laser beam out of a waveguide using a coupler prism. These lines are known as m-lines, the curvature of which has not yet been explained. Experiments and calculations I made also supports the previous observation. In applications where a prism is used to couple out the image spots of a planar waveguide lens, the image line inclines just as m-lines of scattered light. As usually a CCD bar is used for the analysis of the outcoupled image, it is necessary to maintain straightness to avoid off-axis image spots wander off the CCD detector pixels.

4 Results (Efficient method to design waveguide lens systems, and their optimisation)

Since OSLO was written with three-dimensional optical elements in mind, it is not quite obvious how planar lenses should be modelled by the programme. For purely ray optical simulations one must trace rays only in one plane, however for diffraction analysis a planar lens must be replaced by a cylindrical one (because the wavefront leaving the exit pupil is always considered three-dimensional by commercial optical design programmes). The other difficulty arises from the fact that the built-in ray tracing routine of OSLO can use only one single refractive index per medium, so we must also adapt it for handling anisotropic media that have direction-dependent refractive index.

4.1 *Building the model of anisotropy into OSLO SIX*

My aim was to solve anisotropic refraction/reflection at 2D surfaces of arbitrary shape, as well as to make as fast routines as possible. Therefore, I applied *two* coinciding surfaces instead of *one* User Defined Surface at each anisotropic lens surface (see Fig. III.5). These two surfaces consist of an ordinary lens surface (i.e. a conventional refractive surface) having the prescribed shape, and a User Defined Surface with the same profile. The distance between these surfaces are set to zero.

Ray-tracing is carried out in the following manner, regarding the path of a single ray started from the object point (Fig. III.5). The original OSLO algorithm first traces the ray to the first lens surface, which is an ordinary surface even in the case of an anisotropic medium. Hence, the intersection point coordinates are determined at the surface automatically by OSLO. These coordinates are then passed to my CCL routine at the User Defined Surface, which calculates the refracted ray directions together with the correct optical path and passes these back data to OSLO. After finishing it, OSLO continues to trace the ray towards the next lens surface. By application of two surfaces at the place of one anisotropic-to-anisotropic boundary I could economise on computation time by letting the fast built-in OSLO routine determine the intersection point of the ray and the surface, as well as the surface normal at this position. This way, my algorithm (run at the User Defined Surface) has to correct only the optical path, measured from the previous lens surface, and determine the direction of the refracted ray.

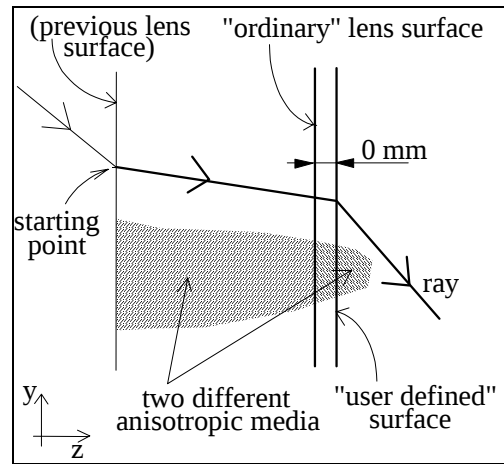


Fig. III.5. Schematic view of a user defined surface and a ray.

The entry of an anisotropic waveguide lens begins with setting up the isotropic version of the lens system (using the extraordinary indices as refractive indices, as usual). To add the above-described User Defined Surfaces and to read-in the data that describe the anisotropic properties (e.g. the ordinary indices) of the lens, I wrote an auxiliary routine that performs these operations automatically, alleviating the designer's task.

4.2 *Modelling method of coupler prisms for optical design*

4.2.1 *Model of incoupling*

Because of phase-matching, only those rays are allowed to couple into a waveguide, whose angle of incidence falls into a given interval around the synchronous direction of the selected mode. After a ray has been incoupled, it will propagate further as a surface ray in the plane of incidence (Fig. III.6). The criteria for coupling a ray into the waveguide are shown in Fig. III.8. Only those rays that hit the waveguide surface close to the prism edge are allowed to incouple. The distance limit is the coupling length. In the programme I realized this by placing a narrow aperture (i.e. a slit of size of the coupling length in "z" direction, and size of the input beam in "y" direction, see Fig. III.6) on the waveguide surface at the prism edge, parallel to the waveguide. Theoretically both the above mentioned angle interval and the coupling length could be calculated from the air-gap size (see Fig. III.3) and the refractive indices. However, as it is much more difficult to measure directly the refractive index and the air-gap size than the angle interval and the coupling length, these latter must be given to the programme by the user. Rays under the critical angle of total reflection are refracted into the substrate. The rest of the rays, those neither incoupled nor refracted into the substrate, are totally reflected on the prism base. The small phase shift between the incident beam and the excited mode (caused by the air gap) is neglected, since it is constant for each ray.

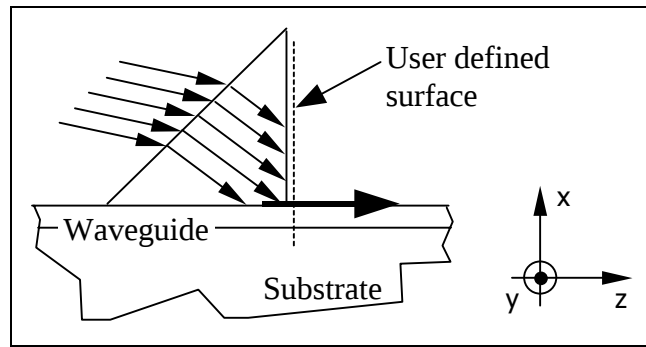


Fig. III.6. Coupling an incident beam into surface rays with a coupler prism.

4.2.2 Model of outcoupling

In my model all rays couple out at the line of the prism edge (see Fig. III.7) as if the coupling length were zero. The divergence angle of the outcoupled beam (in transverse direction to the waveguide plane) can be determined from diffraction on a coupling-length sized aperture (Fig. III.7). Outcoupled rays in the prism are phase-matched to the corresponding surface rays. The phase shift in the air gap is neglected as well as in the case of incoupling.

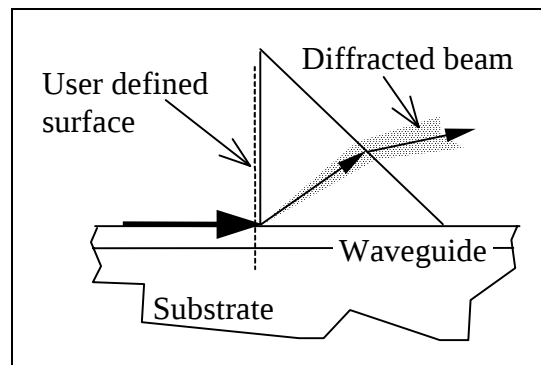


Fig. III.7. Coupling out surface rays using a coupler prism.

4.2.3 Building the models into OSLO SIX

To calculate the intersection point of the incident ray at the waveguide surface, as well as the direction of the incoupled surface ray I place a User Defined Surface at the edge of the incoupler prism (Fig. III.6). The CCL routine at this surface follows the block diagram shown in Fig. III.8 to decide what to do with a particular ray (totally reflect, refract etc.). Another User Defined Surface is placed at the edge of the outcoupler prism (Fig. III.7) to determine the outcoupled ray direction.

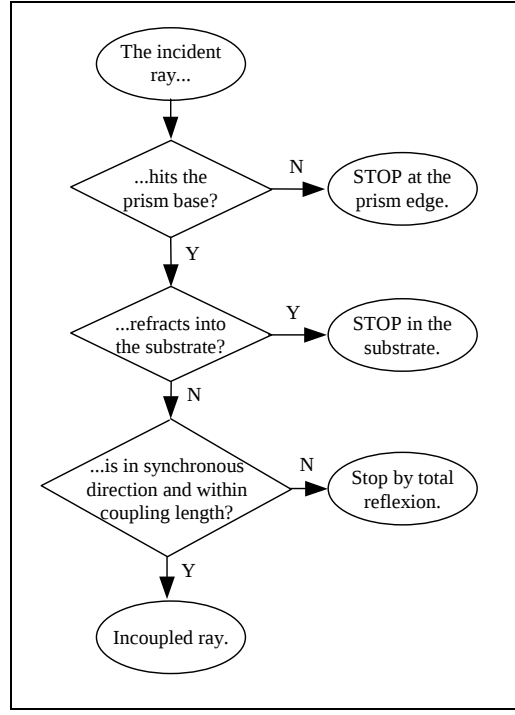


Fig. III.8. Block diagram of the criterium system for incoupling a ray.

4.3 System design for reduced transverse image line inclination

“Transverse image line inclination” is a well-known effect, which has already been presented in [106], but it has not been explained yet. In this subsection I give the geometrical optical interpretation of this effect, and also show a method for its minimisation.

4.3.1 Geometrical optical interpretation – first approximation

Below I prove indirectly that outcoupled rays of different direction cannot lie in the meridional plane in air, assuming that all the rays couple out of the waveguide into the prism in the meridional plane (see plane “T” in Fig. III.9). Let us suppose that the outcoupled rays pass through point “O” on the prism surface (Fig. III.9). From the law of refraction it directly follows that rays in the air must be phase-matched to the corresponding rays in the prism. In addition, the endpoints of all possible wave vectors $\mathbf{k}_{\text{prism}}$ form a circle “C” in the prism. The orthogonal projection of this circle is an ellipse “L” on the prism surface. All the wave vectors in the air ($\mathbf{k}_{\text{air}}$) that are phase-matched to rays in the prism must be on a cylinder “E”, the cross section of which is ellipse “L”. On the other hand, the tips of the possible $\mathbf{k}_{\text{air}}$ vectors form a sphere, the radius of which equals the length of the wave vector in air. Consequently, the refracted rays must be on the intersection curve of these two surfaces. If these rays were in the same plane, this curve would be a circle, but it is impossible, since cylinder “E” has only one circular cross section “C”, the radius of which equals the wave vector length in the prism. It is possible only in the case of normal incidence that all $\mathbf{k}_{\text{air}}$ vectors lie in the meridional plane (i.e. when ellipse “L” reduces to a straight line).

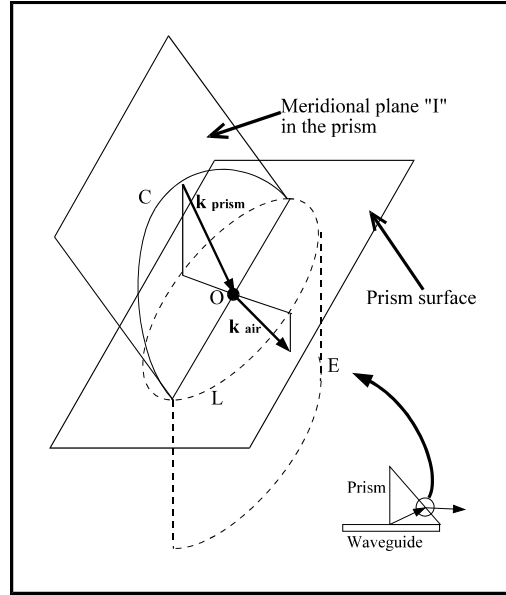


Fig. III.9. Explanation of image line inclination (I). Outcoupled rays traverse from prism ($\mathbf{k}_{\text{prism}}$) to air ($\mathbf{k}_{\text{air}}$).

4.3.2 Geometrical optical interpretation – second approximation

In the previous subsection I assumed that all the outcoupled rays are in the meridional plane in the prism; now I examine the position of these rays in the prism, after outcoupling. For simplicity, let us suppose that all the surface rays pass through point “P”, situated on the prism edge “L” (see Fig. III.10). Since all these rays couple out under the same angle with respect to the waveguide plane (corresponding to the synchronous direction), the surface that these $\mathbf{k}_{\text{prism}}$ wave vectors form in the prism is a cone, denoted by “K” in Fig. III.10. The assumption in the previous subsection that this surface can be approximated by a plane is thus strictly valid only for situations when the surface rays are nearly perpendicular to the prism edge. Although this is often the case, in reality the two effects shown above superimpose to determine the direction of outcoupled rays after the prism, i.e. in air.

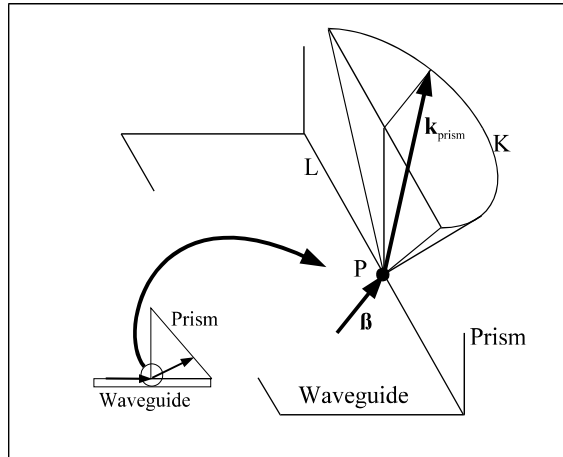


Fig. III.10. Explanation of image line inclination (II). Surface rays (β) couple out into the prism ($\mathbf{k}_{\text{prism}}$).

4.3.3 Minimising image line inclination

From Subsection III.4.3.1 it follows that the appropriate element for outcoupling would be a prism having a surface, which is perpendicular to the meridional plane in the prism (instead of the usually used 45° prism). However, according to Subsection III.4.3.2, outcoupling from the waveguide slightly changes the previously specified prism geometry. Since the prism surface can produce image line inclination in both up and down directions, by proper selection of the prism angle, the curve of the image line can be stretched into almost a straight line. The best prism shape can be determined by automated optimisation (see Lens 3.3 in Subsection III.5.8). It must be noted here, that the optimal angle depends on, among other things, the waveguide effective index and the distance between the image and the outcoupler prism!

5 Application examples

In this section I present eight different waveguide lenses designed to work in different circumstances. Through these examples one can compare the changes in the operation of integrated optical lens systems if they are not used in that environment where they were designed to work (see also [103] and [1]). The main imaging properties of the lenses presented below are summarized in (Tab. III.1). Since these lenses were designed to work on a LiNbO_3 substrate, the anisotropy is uniaxial and the optical axis lies in the plane of the waveguide (see Fig. III.16). The anisotropic effective refractive indices of the low-index and high-index waveguide regions are also listed in Tab. III.1 for the fundamental TE mode of the waveguide. All these lenses work at the vacuum wavelength of 632.8 nm and have a field angle of 2.5° . In each case when a coupler prism is used, only the prism angle α (see Fig. III.16) is changed; the prism base is 5 mm in length and the prism material is $\text{Bi}_{12}\text{GeO}_{20}$ with a refractive index of 2.55, in each arrangement.

	Effective focal length [mm]	f-number ($f_\#$)	Image NA	Low index n_e / n_o		High index n_e / n_o	
Lens 1.1	23.00	3.3	0.34	2.213	2.213	2.320	2.320
Lens 1.2	"	"	"	2.213	2.293	2.320	2.268
Lens 2.1	22.55	3.2	0.34	2.208	2.293	2.297	2.268
Lens 3.1	16.22	6.2	0.08	2.208	2.293	2.297	2.268

Tab. III.1. Main imaging properties of the discussed lenses and the effective refractive indices of the waveguide layers.

5.1 Lens 1.1: planar doublet in isotropic media

This doublet (Fig. III.11) was designed to work in an isotropic waveguide, without an outcoupler prism. The isotropic waveguide in this case is an approximation of the anisotropic LiNbO_3 waveguide: the effective refractive index equals $n_{\text{extraordinary}}$ (for definition see Fig. III.16 and [54, 55, 115, 17]) in

every waveguide region, since the ordinary axis is perpendicular to the optical axis of the lens system. The lens was corrected to be diffraction limited in full image field by the usage of 10th order aspherical surfaces [92, 103]. The focal properties can be read in Tab. III.2.

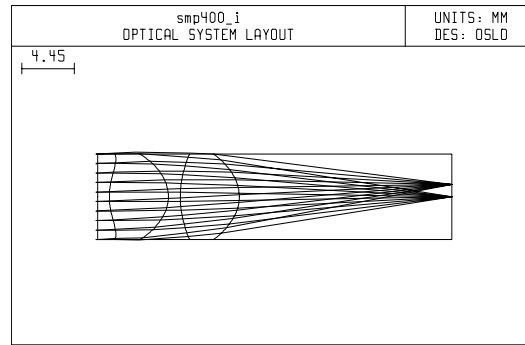


Fig. III.11. Layout of Lens 1.1.

5.2 Lens 1.2: planar doublet in anisotropic media

This lens contains the same surfaces as Lens 1.2 (i.e. it was not re-optimized), but the waveguide is now anisotropic. The effects of anisotropy can well be seen in the second line of Tab. III.2. It shows that anisotropy causes a huge focus shift (see “image distance”), and spoils the focus spot (see “RMS spot size” values and Fig. III.12).

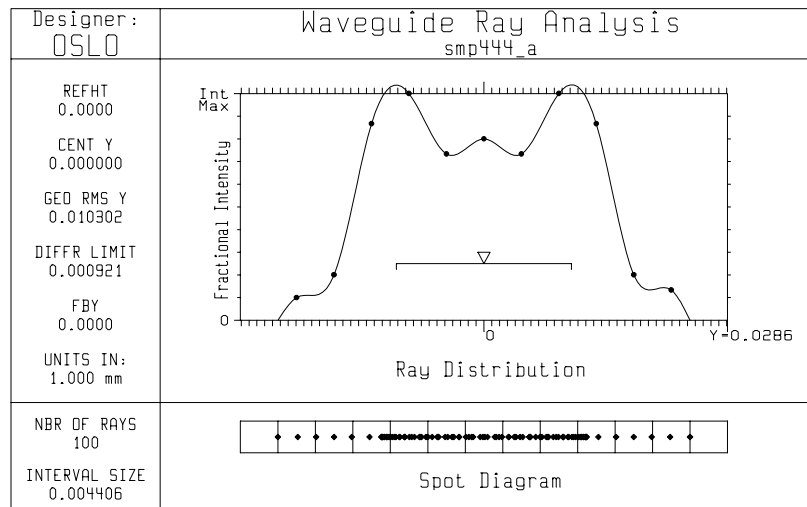


Fig. III.12. On-axis spot diagram and ray distribution curve of Lens 1.2. (The symbol in the middle represents the centroid position and RMS size of the focus spot.)

5.3 Lens 1.3: planar doublet in anisotropic media with outcoupler prism

In this arrangement the outcoupler prism is also placed onto the waveguide surface of Lens 1.2; the prism angle is 45°. The result can be seen in the RMS spot size cells of Tab. III.2. The outcoupler prism acts like a plane parallel glass plate, since the outcoupled rays are almost in the same plane (see Subsection III.4.3). Consequently, the focus spots now contain a large spherical aberration.

	Image distance [mm]	On-axis RMS spot size [μm]	Off-axis RMS spot size [μm]	Airy radius [μm]	Max. image line inclination [μm]
Lens 1.1	18.00	0.1	0.2	0.9	—
Lens 1.2	19.51	10.3	14.7	0.9	—
Lens 1.3	(10.93)	42.1	99.1	0.9	7

Tab. III.2. Main focal properties of the doublet.

5.4 Lens 2.1: planar quadruplet in anisotropic media

This lens is a diffraction limited quadruplet, designed to work in an anisotropic waveguide (see Fig. III.13). The lens was corrected in full field using acircular surfaces of the 10th order. The focal properties are shown in Tab. III.3.

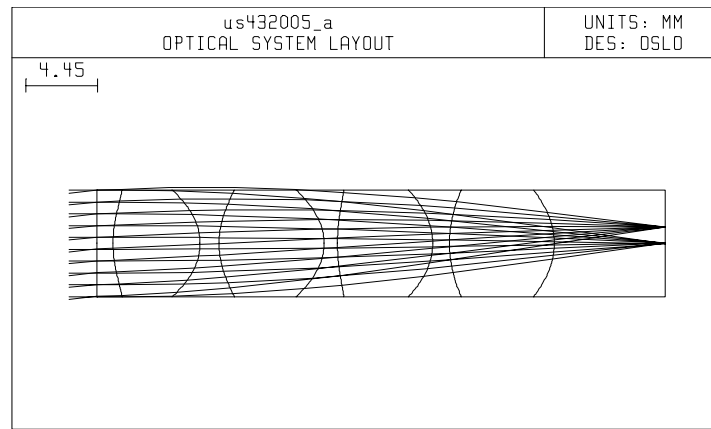


Fig. III.13. Layout of Lens 2.1.

5.5 Lens 2.2: planar quadruplet in anisotropic media with outcoupler prism

This is the same lens as 2.1 with the only difference that a 45° outcoupler prism is placed onto the waveguide surface. The large spherical aberration can well be seen again from the increment of the focus spot sizes in Tab. III.3.

	On-axis RMS spot size [μm]	Off-axis RMS spot size [μm]	Airy radius [μm]	Max. image line inclination [μm]
Lens 2.1	0.2	0.7	0.9	—
Lens 2.2	3.3	4.0	0.9	0.1

Tab. III.3. Main focal properties of the quadruplet.

5.6 Lens 3.1: planar singlet in anisotropic media with 45° outcoupler prism

This singlet (see lens in Fig. III.16) was designed to work in anisotropic waveguide in the presence of the outcoupler prism. Although it is a biconcave type lens, it focuses the light, since the low refractive index region is situated *between* the lens surfaces.



Fig. III.14. Transverse image line inclination of Lens 3.1.

The lens is not diffraction limited, since it was enough to constrain the spot size below $13\text{ }\mu\text{m}$.^{*} The lens surfaces are now conic sections [92, 103], the prism angle is 45° again. The focal properties can be seen in Tab. III.4. It is noteworthy that the maximum image line inclination (for definition of “transverse image line inclination” see Subsection III.2.5) is around $60\text{ }\mu\text{m}$.[†] The precise curve of the half image line (i.e. where the different focus spots, see Fig. III.16, are situated) can be seen in Fig. III.14. The plot shows the focus spot displacement in the X direction, as a function of the angle of the incident beam. The ray distribution curve of the on-axis focus spot can be seen in Fig. III.15.

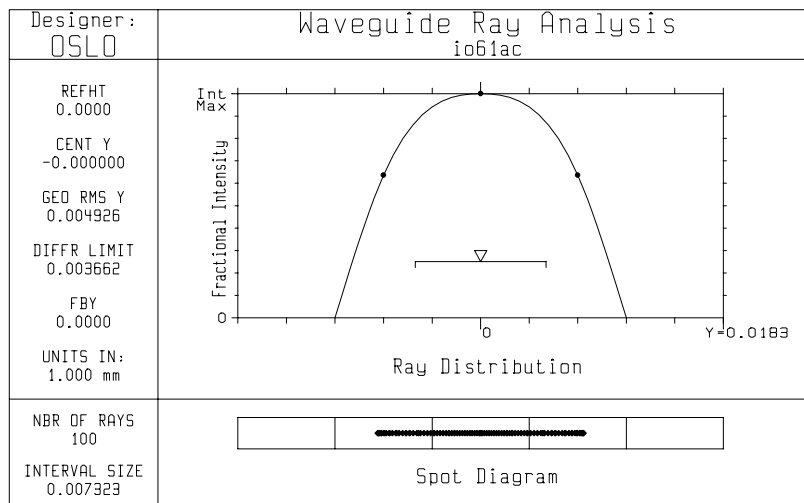


Fig. III.15. Ray distribution curve and spot diagram of the on-axis focus spot of Lens 3.1 (measured along the Y lens axis).

^{*} This lens was designed for an acousto-optical spectrum analyzer that used a CCD bar with pixel size of $13\text{ }\mu\text{m}$.

[†] It is about five times larger than the CCD pixel size used in the spectrum analyser.

5.7 Lens 3.2: planar singlet in anisotropic media with semi-optimised outcoupler prism

This is the same arrangement as the previous one, except for the prism angle was changed to 64.3° . In this special case the outcoupled rays are about perpendicular to the prism surface (for explanation see Subsection III.4.3.1). The maximum image line inclination is reduced by approximately 75 % (see Tab. III.4).

	On-axis RMS spot size [μm]	Off-axis RMS spot size [μm]	Airy radius [μm]	Max. image line inclination [μm]
Lens 3.1	5.0	9.9	3.7	60
Lens 3.2	4.5	9.2	3.5	17
Lens 3.3	4.2	8.7	3.7	0.02

Tab. III.4. Main focal properties of the singlet.

5.8 Lens 3.3: planar singlet with optimised outcoupler prism

This is again the same lens system as before, but the prism angle was optimized to minimize image line inclination (for explanation see Subsection III.4.3.2). The prism angle is now 57.9° , which results in an image line inclination of less than 20 nm. In Fig. III.16 the whole device is shown together with the incoupler prism. In this case the incoupler does not affect the focal properties since a collimated input beam is used.

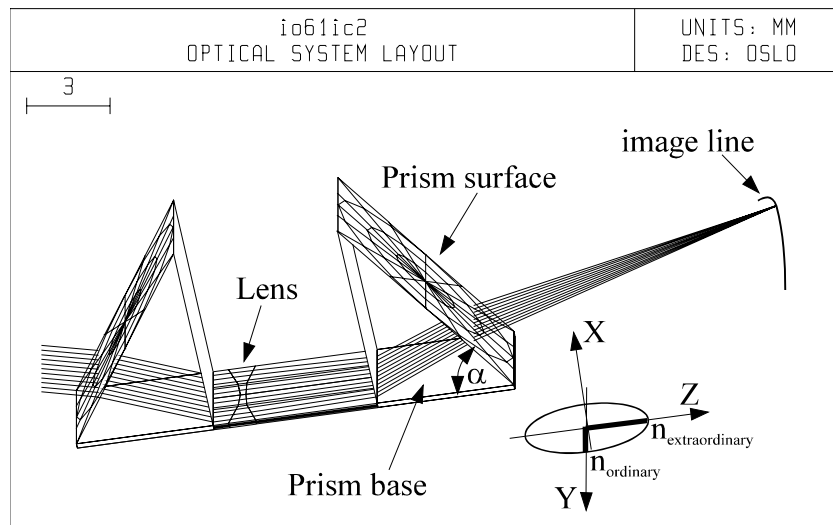


Fig. III.16. The whole optical system of Lens 3.3, containing the in- and outcoupler prisms and the waveguide with the lens in it. The incoupled beam is deflected in the waveguide before the lens by an acousto-optical transducer.

6 Verification

My intention with this measurement was to check whether the real focal properties of the singlet (Lens 3.1) corresponded to the designed values or not. For this purpose I observed the intensity distribution of the on-axis spot and determined its FWHM in Y direction, which I then compared to the calculated value. The measurement was carried out by a CCD detector array, the pixel size of which was $8.5\text{ }\mu\text{m}$. I used 45° coupler prisms and a TE polarized laser beam of 632.8 nm wavelength. The beam diameter was approximately 1 mm . The calculated geometrical RMS spot size of the lens is about $2\text{ }\mu\text{m}$ in air, after outcoupling, while the Airy-radius is about $10\text{ }\mu\text{m}$, assuming an incident beam diameter of 1 mm . The ratio of the geometric and diffraction based spot sizes shows that the measurement results must be compared to the Airy-radius, since the lens is diffraction limited in this case. The measured intensity pattern can be seen in Fig. III.17. The diagram shows one line of the CCD array, the one in which intensity is maximal (CCD lines are parallel to the Y direction). The FWHM of the focal spot is less than $20\text{ }\mu\text{m}$, which means that the lens works approximately as I designed, since the calculated Airy-diameter was $20\text{ }\mu\text{m}$.

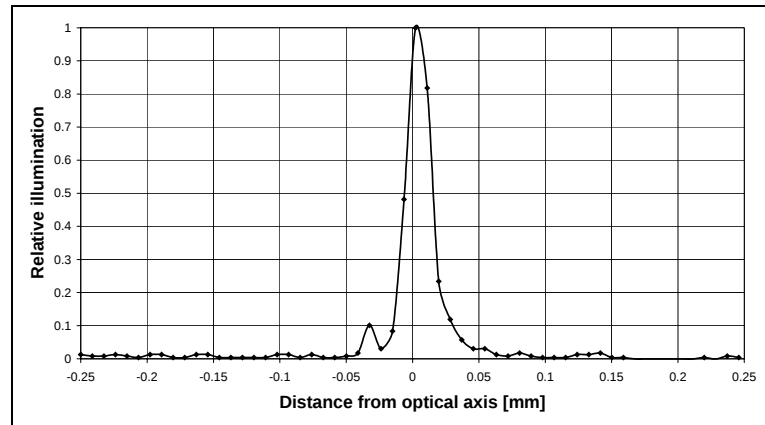


Fig. III.17. Measured intensity distribution of the on-axis focus spot of Lens 3.1. (Asymmetry of the focus spot results from small defects in the waveguide and the prism edge that affect the outcoupled beam.)

I also prepared a measurement setup to determine if my calculations were adequate to rectify the image line of Lens 3.1. The experiment should have been made by deflecting the input collimated beam in the plane of the waveguide using an ultrasonic transducer, thus scanning the whole image line of the planar lens (the lens was intended to work in an integrated, acousto-optical spectrum analyser). Unfortunately the transducer, which was designed for this device, could not reach an acceptable diffraction efficiency. In order to obtain some results for verification, instead of scanning the image spot, I examined the m-lines of the waveguide (without the lens). The m-lines are formed of scattered light that occurs between the waveguide and the outcoupler prism surface (probably due to surface roughness) when a narrow collimated laser beam couples out of the waveguide. Fig. III.18 shows the m-line of the fundamental mode of the waveguide in case of using a 45° prism; here the curvature is

obvious. Fig. III.19 presents the case, when an appropriately designed prism (see prism in Lens 3.1) is used to couple out light from the same waveguide. The m-line is now conspicuously straight.

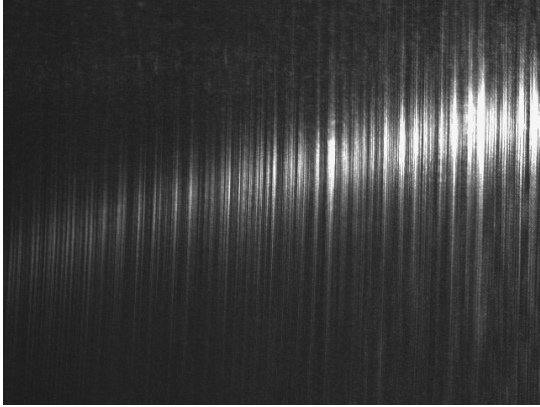


Fig. III.18. Left half of the m-line of a TIPE:LiNbO₃ waveguide with a 45° outcoupler prism. The outcoupled beam itself is situated on the right of the figure.

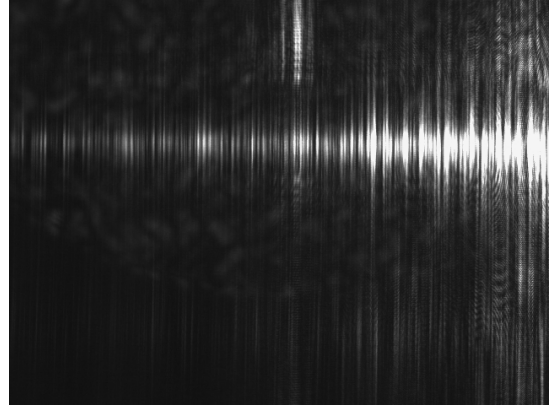


Fig. III.19. Left half of the m-line of a TIPE:LiNbO₃ waveguide with an outcoupler prism of optimised angle. The outcoupled beam itself is situated on the right.

7 Conclusion

I developed a design environment in OSLO SIX, a 3D optical CAD program, to facilitate the entry, evaluation and optimisation of integrated optical lens systems. I demonstrated the usage of this environment through lens design examples. As a confirmation of the accuracy of my models I compared measured and calculated focal characteristics of one of the presented lenses. I also showed, and experimentally confirmed, how the so-called transverse image line inclination can be minimised by modifying the angle of the outcoupler prism.

IV DESIGN OF BEAM TRANSFORMATION SYSTEMS

TABLE OF CONTENTS

TABLE OF CONTENTS.....	36
1 INTRODUCTION	37
2 FUNDAMENTALS.....	39
2.1 PRESUMPTIONS.....	39
2.2 OPTICAL CHARACTERISATION OF BEAM-SHAPERS	39
2.3 DERIVATION OF THE MAP FUNCTION FOR GAUSSIAN-TO-UNIFORM TRANSFORMATION	40
3 PRELIMINARIES (FORMER METHODS FOR THE DESIGN OF BEAM-SHAPERS).....	42
3.1 DIRECT METHODS FOR THE DESIGN OF BEAM-SHAPERS	43
3.2 INDIRECT METHODS FOR THE DESIGN OF BEAM-SHAPERS	44
4 RESULTS (NEW METHOD FOR THE DESIGN OF BEAM-SHAPERS).....	45
4.1 OPTIMISATION METHOD FOR THE DESIGN OF BEAM SHAPING SYSTEMS.....	45
4.2 DESIGN OF GAUSSIAN-TO-UNIFORM TRANSFORMATION SYSTEMS FOR HIGH BEAM SHAPING POWER	46
5 APPLICATION EXAMPLES	50
5.1 LENS 1: SINGLE-ELEMENT ASPHERIC BEAM-SHAPER	50
5.2 LENS 2: ALL-SPHERICAL MULTI-ELEMENT BEAM-SHAPER	52
6 VERIFICATION.....	55
7 CONCLUSION	57

1 Introduction

It is a usual illumination task to irradiate a surface homogeneously using a laser source. The solution poses a major problem, since laser resonators operate usually in the TEM_{00} mode, thus the intensity profile of their output beam is Gaussian. The simplest way for reducing illumination inhomogeneity on a target surface, say to $\pm 5\%$, is to expand the Gaussian beam to the desired extent and cut out the unnecessary part by an aperture. With this method about 90% of the input power is lost, which implies that achieving high illumination power on a target surface requires a more special beam transformation system. There exist several types of beam shaping devices representing different categories in size, complexity and cost; commonly used alternatives to beam expanders include spatial filters, diffusers, multi-faceted beam redistributors, aspheric [43, 83, 59, 53, 57, 40] / spherical systems [89], diffractive (holographic) [67, 110, 61, 99, 29, 24, 25, 26, 82, 27, 112, 34, 75, 69, 47] devices, acousto-optical systems [78, 72] etc., most of which can work in either refraction or reflection. Operation of beam shaping instruments can be described by both geometrical and physical optical models, from which various design methods (analytic, numerical, iterative etc.) have been developed [71, 70].

For the design of beam shaping systems I developed a new approximate method based on geometrical optics. The method presumes spatially coherent input and output light (e.g. plane parallel or spherical beams), but it does not limit intensity shaping only to Gaussian-to-uniform transformation. The computation is based on the principle of automated optimisation, thus it can be applied even for the design of those devices where the system geometry necessary to achieve a prescribed intensity transformation cannot be determined analytically. Another advantage of my method is that it can be conveniently performed using any kind of commercially available optical design programmes (I used [92]). By the new method I designed several devices of different beam shaping properties (i.e. input beam diameter, output beam diameter etc.), wavelength and structures (using spherical/aspheric surfaces). A design example, demonstrating the operation of the method, transforms the Gaussian beam of a He-Ne laser into a flat-top profile with two aspheric surfaces. The entrance pupil of the device truncates the input beam at a radius of three times the Gaussian spot-size (i.e. the $1/e^2$ intensity radius) to avoid diffraction effects. Ray-tracing predicts that the system has a wave aberration of about 0.07λ RMS OPD, and 86% power transmission at $\pm 5\%$ inhomogeneity.

The previously discussed device works according to the most wide-spread working principle of beam shapers: two properly chosen optical elements are placed one after the other, the first realising intensity transformation, and the second collimating the output beam. Even though most systems of this kind can be designed for (theoretically) perfect beam shaping by established methods, the fabrication technology of the necessary optical elements (usually aspheric or diffractive surfaces) is not

available for all optical engineers (e.g. because of its high cost), or the available method may be of inadequate accuracy. For the substitution of such systems one possibility is to make the whole system of high-precision spherical components exclusively, at the expense of increased surface number, weight and overall length. The design of such a system is not straightforward either, since no analytical solution exists to obtain the necessary structure, only some kind of (local or global) optimum can be sought for, usually by numerical (i.e. optimisation) methods [40, 95]. Due to the structural complexity of all-spherical beam shaping systems, their design is not easy even by optimisation methods. The main limiting factors are the large number of degrees of freedom and the complicated relation between construction parameters and optical performance. My design experience shows that in case of quadruplets or simpler systems, in which case the application of optimisation methods still appears feasible, the beam shaping power (i.e. entrance pupil radius to Gaussian spot-size ratio) is hard to increase above unity. This value is too low in most practical cases since it gives rise to significant Fresnel diffraction of the input Gaussian beam on the entrance pupil of the beam-shaper, which spoils the intensity profile of the output beam.

In my dissertation I point out theoretically that by placing a number of low beam shaping power systems (which are easy to design individually) one after the other into a cascade, the beam shaping power can be increased above 2 in order to maintain output beam homogeneity at an acceptable level. For demonstrative purposes I designed such a system exclusively of spherical surfaces to transform the Gaussian beam of a frequency-doubled solid-state laser into a collimated beam of flat-top profile. The entrance pupil of the device truncates the input beam at a radius of 2.2 times the Gaussian spot-size (i.e. the $1/e^2$ intensity radius) to avoid diffraction effects. The designed system has a wave aberration of 0.029λ RMS OPD and almost 70% overall power transmission at $\pm 5\%$ intensity inhomogeneity. For comparison I would like to add that the only optical system (I know about) [89], comprising exclusively spherical components, was designed to truncate the input Gaussian beam at its $1/e$ intensity radius, i.e. the beam shaping power was 0.71. This value is about one third of that of the device discussed here.

In this section of the dissertation, first I briefly overview the geometrical optical model I used as a basis to my calculations, discuss the most important existing design methods and show the working principle of my solution. I also develop the theory of beam shaping device cascades and analytically derive formulae that describe their most important optical characteristics. As application to my design method for the optimisation of beam shaping systems, I show two examples prepared using the commercial optical design software [92]. Calculated geometrical and optical properties of the designed instruments are discussed, diffraction properties of one of the systems are checked by numerical evaluation of the Rayleigh-Sommerfeld diffraction integral (see Appendix A).

2 Fundamentals

The basic idea of beam shaping relies on the intensity law of geometrical optics, according to which light energy cannot propagate out of an area bordered by adjacent rays (since rays show the direction of energy propagation, i.e. that of the Poynting-vector). Correspondingly, for divergent beams the intensity drops in the direction of propagation and vice versa. Hence, ray paths of a beam and its intensity profile on a given surface determine the intensity profile on any arbitrary surface (in the geometrical optical approximation).

2.1 Presumptions

Beam shaping devices influence the processed beam in various ways, such as beam homogeneity, wavefront shape, cross-section shape (circular or rectangular), power transmission and spatial coherence (consider e.g. a plane wave falling onto a diffuser). The required intensity distribution can be achieved on a specific target surface (located in the near- or far-field excluding/including Fraunhofer diffraction in the modelling), or the output beam can be collimated. Due to their large importance in optical design, from now on I will regard only those beam-shapers that produce a homogeneous, circular, spatially coherent (i.e. collimated or spherical) beam in case of a circular Gaussian input beam.

2.2 Optical characterisation of beam-shapers

Three basic optical characteristics of beam shaping devices that perform perfect Gaussian-to-uniform transformation are the *beam shaping power* ‘ η ’, the *magnification ratio* ‘ m ’ and *map function* ‘ P ’ (also known as the “transformation equation”).

The beam shaping power is defined as

$$\eta \equiv \frac{r_{tr}}{\varrho}, \quad (\text{IV.1})$$

where ϱ refers to the $1/e^2$ intensity radius, i.e. the spot size, of the input Gaussian beam [91], and r_{tr} denotes the radius at which it is truncated by the entrance pupil of the beam shaping system. The value of η influences directly the power transmission of the device, and through Fresnel diffraction, the effective homogeneity of the output beam.

The definition of the magnification ratio is

$$m \equiv \frac{R_{max}}{\varrho}, \quad (\text{IV.2})$$

where R_{max} is the theoretical radius of the output beam inside which the *total power* of the input Gaussian beam is distributed homogeneously, see (IV.7). The magnification ratio affects the output beam size and its intensity level.

The map function describes the intensity transformation feature of a given beam shaping device by defining ray paths through the system. Its general formula can be written in the following form:

$$\left. \begin{aligned} X &= f_x(\xi, \gamma) \\ Y &= f_y(\xi, \gamma) \end{aligned} \right\} \quad (\text{IV. 3})$$

where $(\xi; \gamma)$ denote ray coordinates on a surface given in the input space of the beam shaper, and $(X; Y)$ denote those of the same ray on another surface in the output space. In the next subsection I will derive the particular expression of the map function for Gaussian-to-uniform intensity transformation for the case of collimated input and output beams; generalisation to other types of output beams (e.g. spherical) is straightforward. I must emphasize at this point that throughout this chapter, where appropriate, “intensity” is used for the local radiometric description of spatially coherent light beams instead of “irradiance”, the former being a pure characteristic quantity of the beam itself, not depending on the shape of any detecting surface or screen.

To characterise the imperfections of beam-shapers two further quantities are needed, by which the output beam inhomogeneities and power losses can be described: ‘ R_{eff} ’ is the *effective output beam radius*, inside which the intensity inhomogeneity of the output beam is equal or less than a prescribed level, say $\pm 5\%$; and ‘ T ’ is the *power transmission* that refers to the fraction of the whole power of the input Gaussian beam that pass through the effective area (i.e. R_{eff}) of the output beam.

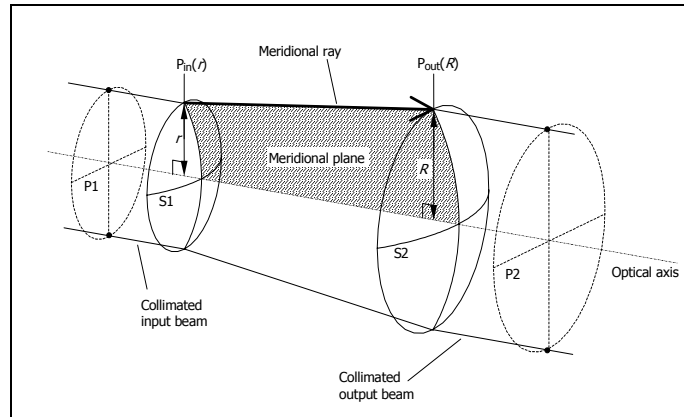


Fig. IV.1. Application of the intensity law of geometrical optics for the case of rotationally symmetric beam-shapers. Surface S1 and S2 denote the input and output surface of an arbitrary beam shaping system; ‘ r ’ and ‘ R ’ are ray coordinates on S1 and S2, respectively, while $P_{\text{in}}(r)$ and $P_{\text{out}}(R)$ represent encircled energy. See details in the text.

2.3 Derivation of the map function for Gaussian-to-uniform transformation

The intensity ‘ I ’ of any circular beam can be given on a rotationally symmetric surface as a function of the ray height ‘ r ’, i.e. the distance measured perpendicularly to the optical axis. The ‘encircled power’ $P(r)$ shows the total power flowing through a given surface of aperture radius ‘ r ’. Its definition for rotationally symmetric beams is obtained by integrating $I(r)$ on the surface:

$$P(r) = \int_0^r 2\pi \cdot r' \cdot I(r') \cdot \frac{\cos(\alpha(r'))}{\cos(\beta(r'))} \cdot dr', \quad (\text{IV.4})$$

where $\alpha(r')$ is the direction of the Poynting-vector relative to the surface normal at r' distance from the axis, and $\beta(r')$ denotes the angle between the surface normal and the optical axis. (It must be mentioned here that $I(r') \cdot \cos(\alpha(r'))$ in the above expression is known as “irradiance”, which quantity has a large technical significance in illumination engineering.) Let the encircled power functions be denoted by $P_{\text{in}}(r)$ and $P_{\text{out}}(R)$ on surfaces S1 and S2 of a beam shaping device, as depicted in Fig. IV.1. These surfaces represent e.g. the plane substrate surfaces of the CGHs in a diffractive instrument, or the aspheric input and output surfaces of a refractive beam-shaper etc. Since all the rays of a beam having rotational symmetry are meridional rays, the map function – that describes the connection between ray coordinates ‘ r ’ on S1, and ‘ R ’ on S2 – can be written in the form $R = f(r)$. The map function can be determined for a given intensity transformation by applying the intensity law of geometrical optics in the case of circular beams (see also Fig. IV.1):

$$P_{\text{out}}(R) = P_{\text{in}}(r), \quad (\text{IV.5})$$

which generally results in an integral equation. Let us assume now that $P_{\text{in}}(r)$ and $P_{\text{out}}(R)$ are to be determined on S1 and S2, i.e. the input and output surfaces of a beam shaping device. In case of collimated input and output beams $\cos(\alpha(r)) \equiv \cos(\beta(r))$, independently from the shape of the surface profiles of S1 and S2. Hence, the integrations can be performed analytically:

$$\begin{aligned} P_{\text{in}}(r) &= P_{\text{Gauss}}(r) = P_{\text{total}} \cdot \left(1 - e^{-2\left(\frac{r}{\ell}\right)^2} \right) \\ P_{\text{out}}(R) &= P_{\text{uniform}}(R) = P_{\text{total}} \cdot \left(\frac{R}{R_{\text{max}}} \right)^2, \end{aligned} \quad (\text{IV.6})$$

where:

$$P_{\text{total}} \equiv \lim_{r \rightarrow \infty} (P_{\text{Gauss}}(r)) = P_{\text{uniform}}(R_{\text{max}}). \quad (\text{IV.7})$$

Substituting (IV.6) into (IV.5), we get the final form of the map function:

$$R = R_{\text{max}} \cdot \sqrt{1 - e^{-2\left(\frac{r}{\ell}\right)^2}}. \quad (\text{IV.8})$$

The above formula can be interpreted as follows: beam-shapers work practically like special afocal systems, in which rays entering the input surface at ‘ r ’ distance from the optical axis leave the output surface at a distance ‘ R ’, so that the R/r ratio is not constant as a function of ‘ r ’. It should be noted that map functions may only be used to characterise those regions of beams, in which rays do not intersect

each other, more precisely if the map function is *strictly monotonic* (if there is a caustic [22] in the beam or its different regions overlap, then diffraction effects make the geometrical model no more valid). It is also noteworthy that the above definition of the map function may be used even if there are additional optical elements (e.g. lenses) between S1 and S2, provided that rays still do not “mix up”. This geometrical optical approximation can be used for the modelling of Gaussian-to-uniform beam transformations with high precision. The validity of the model is limited only by Fresnel diffraction of the input beam on the entrance pupil of the beam shaping system. According to different publications [104, 36, 91] the entrance pupil radius must be 2 to 3 times larger than that of the spot-size [91] of the input Gaussian beam to eliminate diffraction effects. My wave-optical calculations comparing diffraction on different pupil radii are shown in Section IV.6.

3 Preliminaries (Former methods for the design of beam-shapers)

In this chapter I briefly overview the methods previously developed by different authors for the design of beam shaping devices that produce a spatially coherent output beam of uniform intensity with spherical/plane wavefronts [43, 83, 59, 53, 89, 45, 60, 86]. Most of the existing design methods rely on a specific device structure (already mentioned in the introduction), according to which such systems consist of two sequentially placed optical elements (sub-systems): optical element #1 performing intensity shaping, and optical element #2 collimating the output beam. A simple demonstrative example for such a beam shaping system is depicted in Fig. IV.2, which device has been first described by S. R. Jahan and M. A. Karim [59, 53]. The instrument is a refractive device with two aspheric surfaces. The input surface corresponds to optical element #1, while the output surface corresponds to optical element #2. In other realisations of beam shaping systems, e.g. diffractive elements or spherical subsystems can also be used instead of the two aspheric surfaces.

Physical (i.e. diffractive) and geometrical optical models are equally used for the design of beam shapers, applying both analytic and numerical methods (for this latter see further details in the next subsection). The system specification is generally expressed by two requirements: the intensity and the phase condition. The former refers to intensity reshaping (described by the map function), while the latter refers to the shape of the input and output wavefronts (e.g. collimated or spherical beam).

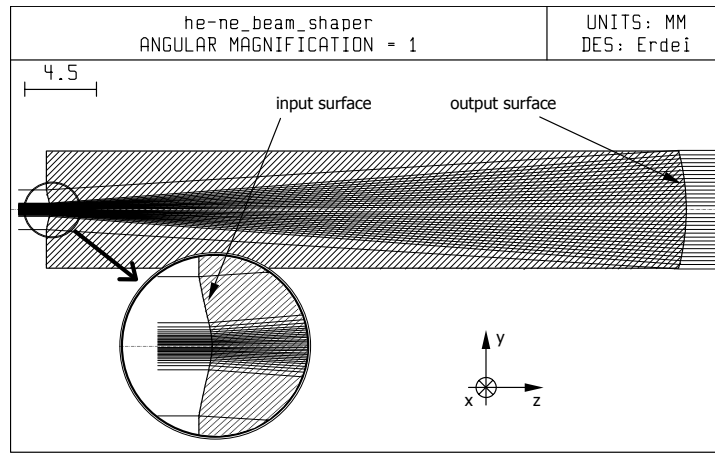


Fig. IV.2. Single-element aspheric refractive beam-shaper. Ray density represents intensity distribution.

3.1 Direct methods for the design of beam-shapers

In case of direct design methods, the construction parameters of a beam shaping device are determined by directly solving equations set up corresponding to the phase and intensity conditions. Such methods can be applied especially for the design of diffractive and aspheric device structures.

In diffractive beam shaping instruments optical element #1 and #2 are realised as Computer Generated Holograms (CGHs) or kinoforms. The structure of optical element #1 and #2 can be usually determined independently of each other (if the CGHs can be considered as thin). The initial step of the design of CGHs is the geometrical optical calculation of the map function [45, 39, 86, 16], which is derived by solving the integral equation obtained from the intensity condition (see IV.4 in Subsection 2.3). Analytical solutions of this equation exist only for simple cases [45], otherwise numerical methods must be applied (e.g. for non-circular beams) [39, 86, 16]. (For cases when the precision of this approximation is inadequate, there are other known methods, such as the iterative non-geometrical procedures using Fourier-optics described in [39]) The phase-contribution of the first CGH is usually computed using the geometrical optical model, while for the design of the second hologram several other methods exist as well. It is usual to record the second hologram experimentally using the output beam of the first CGH as a reference wave. This is made to avoid misalignment problems, since diffractive beam-shapers are rather sensitive to the relative position of the constituting holograms [39, 16]. To calculate the second CGH, the wavefront shape (i.e. the phase of the field) must be determined across its entire surface, for which I found two different methods. In case of the first, the map function and the phase condition are used to set up a differential equation, by the numerical solution of which the required phase can be calculated [45]. The other approximation considers each point of the surface of the first CGH as a local diffraction grating, and applies ray-tracing (extended by optical path length calculation) to determine the phase at the second CGH [86].

The geometrical optical model is also a usual procedure for the direct design of beam-shapers constructed from refractive (or reflective) optical elements [43, 83, 59, 53, 89]. In such devices the functions of optical element #1 and #2 are carried out by continuous aspheric or spherical surfaces or subsystems, which usually cannot be determined separately, making the design process somewhat more complicated. The only all-spherical beam-shaper I found has been constructed from two doublets with a large air-space between them [89]. The system was designed by balancing spherical aberrations using third-order calculations, and was able to reach only $\eta \approx 0.71$ (i.e. the beam is truncated at its $1/e$ intensity radius). In most beam shaping devices both optical element #1 and #2 are realised by aspheric surfaces [43, 83, 59, 53]. The common basis of the design of such systems is also the above introduced map function. Their surface shapes can be determined directly from a differential equation that is derived from the map function and the phase condition. For the solution of this differential equation several numerical methods have been developed. Most of them calculate discrete points of the aspheric surfaces, that are approximated by (e.g. spline) interpolation between the representation points [43, 83, 59, 53].

3.2 Indirect methods for the design of beam-shapers

Optimisation-based design techniques, such as global and local optimisation, are indirect methods, i.e. they determine the construction parameters of the device without the analytic knowledge of the connection between its structure and optical behaviour. Application of such methods is reasonable in cases when the system layout cannot be calculated analytically from the specification (see e.g. all-spherical beam-shapers [89]). Global optimisation is a fully automatic method, which finds theoretically the best values for system parameters to achieve given optical characteristics. Only the type of the parameters (e.g. radius of curvature) and the range of their possible values must be provided. Such methods (see e.g. Adaptive Simulated Annealing, Genetic Algorithm etc.) are either extremely computation demanding or can find solutions only in case of relatively low number of parameters (i.e. degrees of freedom). Local optimisation techniques [92, 95] are basically automated methods, they require the optical engineer to take an active part in the design by supplying a satisfactory initial system for the automated procedure and by continuously controlling the design process to find the best achievable “local” optimum. On the other hand, such methods can be used for more complicated structures. To my knowledge, application of local (automated) optimisation techniques for the design of beam shaping devices has not been reported yet.

4 Results (New method for the design of beam-shapers)

In this section I present a local optimisation-based general method for the design of beam shaping systems. As a second topic I propose a procedure that alleviates the design of high beam shaping power beam transformation systems from simple, conventional (e.g. spherical) components.

4.1 Optimisation method for the design of beam shaping systems

As I concluded in Subsection IV.3.2, to my knowledge no local optimisation methods have been applied for the design of beam-shapers until now. In this subsection I present a method that fills this gap by providing a new design tool for optical engineers. The main advantage of the method is that it does not require special softwares, since it is integrated into the convenient, accustomed environment of commercial design programmes, offering an aid at the design of complex optical systems containing a wide range of beam shaping elements.

It has been presented in Subsection IV.2.3 that intensity and phase variations in a circular beam of light can be described (with given limitations) by geometrical optical means using meridional ray-tracing. In my method I use the map function to specify the required intensity transformation for the optical design programme. First, I define a collimated input beam by the ray heights $r(i)$ of a set of ‘n’ meridional parallel rays on the input surface of the beam shaping system ($i = 1, 2, \dots, n$). The corresponding coordinates on the output surface will serve as the operands for the optimisation, i.e. the individual terms in the error function:

$$\left. \begin{aligned} o_{\text{int}}(i) &= R(i) - R_0(i) \\ o_{\text{wvf}}(i) &= L(i) - L_0(i) \end{aligned} \right\}, \quad (\text{IV.9})$$

where the ‘ o_{int} ’ operands specify the required intensity distribution on the output surface, and the ‘ o_{wvf} ’ operands determine the wavefront profile of the output beam. In (IV.9) ‘ R ’ denote ray height measured from the optical axis on the output surface of the beam-shaper, while ‘ L ’ refers to the direction cosine of the output rays, defined as

$$L \equiv \mathbf{t} \cdot \mathbf{u}, \quad (\text{IV.10})$$

where $\mathbf{t}$ and $\mathbf{u}$ denote unity vectors, $\mathbf{t}$ pointing into the direction of the ray, and $\mathbf{u}$ lying perpendicularly to the optical axis in the meridional plane of the rays. This expression of the wavefront error makes it unnecessary to trace a 3D bundle of rays through the system, as it is usually needed at wavefront calculations, in order to result in a fast optimisation algorithm. In (IV.9) the operands o_{int} and o_{wvf} are given by the difference of the *real* and *prescribed* ray coordinates (these latter are denoted by the subscript

'0'), so that the operand values tend to zero as the system performance improves with optimisation. In (IV.9) $R_0(z)$ are calculated by substituting the ray coordinates $r(z)$, given on the input surface, into the map function (IV.8), as well as $L_0(z) = 0$ corresponding to the plane wavefronts of the collimated output beam. The error function 'ERR' is constructed from the operands such as:

$$\text{ERR} \equiv \sqrt{\frac{\sum_{i=1}^n (w_{\text{int},i} \cdot |o_{\text{int}}(i)|^2 + w_{\text{wvf},i} \cdot |o_{\text{wvf}}(i)|^2)}{\sum_{i=1}^n (w_{\text{int},i} + w_{\text{wvf},i})}}, \quad (\text{IV.11})$$

where $w_{\text{int},i}$ and $w_{\text{wvf},i}$ are weighting factors.

After specifying the error function and the initial system (see Section IV.5), the automated optimisation proceeds as follows: in each iteration cycle the optical design programme determines the coordinates of the sample rays on the output surface by tracing them through the optical system, evaluates ERR, and change the given system parameters (given by the user) in order to reduce its value. When ERR reaches its minimum, the output beam characteristics of the provisional system will correspond most closely to the requirements (i.e. ray heights calculated on the output surface will approximate those specified by the map function, and all the rays of the output beam will travel parallel to the optical axis). Further details on automated optimisation can be found in [95] and [92].

The above method has several advantages and also some disadvantages, of course. Since it is integrated into a commercial optical design programme, my method may be used for the design of several kinds of beam shaping optical systems (e.g. spherical or aspheric lens configurations). The method is simple, which means that the initial optical system as well as the necessary operands and variables can be defined very easily. Although the method is approximate – since the incident beam is modelled by discrete rays, between which the system performance is not evaluated –, an arbitrarily high precision can be achieved by increasing the number of the sample rays. In its present form the only disadvantage of my method is the common drawback of all automated design procedures, viz. the possibly slow convergence. The convergence speed depends on several parameters, a few of which are: type of the equation used to describe the surface profiles, type of the optimisation algorithm, and choice of the initial system. A concrete design example will be given in Subsection IV.5.1.

4.2 Design of Gaussian-to-uniform transformation systems for high beam shaping power

In case of certain beam-shaper configurations (such as an all-spherical system of high η) it is impractical to apply a direct (i.e. analytic) method to find an acceptable optical system. On the other hand, in case of all-spherical realisations, indirect (i.e. optimisation) methods can only be used for the design of low beam shaping power systems, which are sufficiently simple in structure. To eliminate the

above problems I propose to combine simple beam shaping devices of low η into a sequential setup (i.e. cascade), as shown in Fig. IV.7. The elementary beam-shapers are relatively easy to design individually by optimisation, while their cascading increases the value of η of the resultant system. Below, I introduce a terminology for discussion of such beam-shaper cascades, as well as derive simple formulae to give an analytical proof of their operation and reveal the expectable performance.

In order to generalise the discussion, first I introduce the ‘ x ’ and ‘ y ’ normalised coordinates, measured on the input and output surface of a beam-shaper, respectively:

$$x \equiv r/\varrho \quad ; \quad y \equiv R/R_{\max} . \quad (\text{IV.12})$$

(For definitions of ϱ and $R_{\max}$, see explanations at (IV.1) and (IV.2) respectively.) Using these coordinates, the map function can be written in the following general form:

$$y = F(x) = \sqrt{1 - e^{-2x^2}} . \quad (\text{IV.13})$$

The formula in (IV.13) I consequently call the *generalised map* function (denoted by ‘ F ’) of perfect Gaussian-to-uniform beam shaping.

In case of two beam-shapers put one after the other (with parameters ϱ_1 , η_1 , m_1 and ϱ_2 , η_2 , m_2 respectively), the resultant map function looks like:

$$y = F_2(x) = F(F(x) \cdot \eta_{12}) , \quad (\text{IV.14})$$

where:

$$\eta_{12} \equiv \frac{R_{\max,1}}{\varrho_2} . \quad (\text{IV.15})$$

This parameter I named the *cross-beam shaping power*, for it matches the normalised coordinates of the two subsequent beam-shapers together, as well as its expression is similar to the definition of the beam shaping power. Since equation (IV.14) formally does not correspond to the general formula (IV.13), I introduce two coefficients ‘ c ’ and ‘ d ’ in order to fit F_2 to F (see also Fig. IV.3):

$$y = F_2(x) \equiv \frac{F(F(x) \cdot d(\eta_{12}) \cdot c(\eta_{12})) \cdot \eta_{12}}{c(\eta_{12})} . \quad (\text{IV.16})$$

By requiring the following conditions:

$$\lim_{x \rightarrow \infty} [F_2(x)] := \lim_{x \rightarrow \infty} [F(x)] = 1 \quad ; \quad \left. \frac{dF_2(x)}{dx} \right|_{x=0} := \left. \frac{dF(x)}{dx} \right|_{x=0} = \sqrt{2} , \quad (\text{IV.17})$$

the coefficients will be

$$c(\eta) = \sqrt{1 - e^{-2\eta^2}} \quad ; \quad d(\eta) = \frac{1}{\eta \cdot \sqrt{2}}. \quad (\text{IV.18})$$

This value of ‘c’ provides that the resultant generalised map function F_2 will always be normalised to unity as $x \rightarrow \infty$, while the present choice of ‘d’ ensures that the slope of the two curves will have the same value at $x = 0$. It is important to emphasize, that (IV.16) is still not exactly the same formula as (IV.13), only a good approximation of it, where the goodness of the approximation (i.e. the quality of the resultant beam shaping) depends on the value of η_{12} . By examining the maximum relative fitting error between $F(x)$ and $F_2(x)$ one can find that a worst case occurs when $\eta_{12} \approx 1$. Fig. IV.3 presents the generalised map function $F(x)$ and the fitted $F_2(x)$ curve in case $\eta_{12} = 1$. In cases when the simple fitting procedure presented above results in an unacceptably large error, the output beam homogeneity can be improved by slightly modifying the value of ‘d’ (i.e. the slope of $F_2(x)$ at $x = 0$) relative to that calculated by (IV.18) to achieve a better fit averaged to the whole curves.

As a consequence of cascading two beam-shapers, the resultant system will have different ϱ , η and m parameters than those of the constituent subsystems. The above method presented for fitting the resultant map function to the general formula of ideal beam shaping predicts the following resultant ϱ and $R_{\max}$ for the $F_2(x)$ function:

$$\begin{aligned} \rho_{\text{res}} &= \rho_1 \cdot d(\eta_{12}) \cdot c(\eta_{12}) \\ R_{\max, \text{res}} &= R_{\max, 2} \cdot c(\eta_{12}), \end{aligned} \quad (\text{IV.19})$$

while the beam shaping power growth (expressed relative to η_1) is:

$$\frac{\eta_{\text{res}}}{\eta_1} = \frac{\varrho_1}{\varrho_{\text{res}}} = \frac{1}{d(\eta_{12}) \cdot c(\eta_{12})}. \quad (\text{IV.20})$$

Equation (IV.19) shows that the cascade of two beam-shapers will transform a Gaussian input beam of spot-size ϱ_{res} into a (more-or-less) homogeneous output beam of radius $R_{\max, \text{res}}$. The output beam homogeneity depends on the goodness of fitting $F_2(x)$ to $F(x)$.

The above formulae of $F_2(x)$, ϱ_{res} and $R_{\max, \text{res}}$ can be easily generalised for any number of sequentially placed (i.e. cascaded) elementary beam-shapers. (IV.21) and (IV.22) show their values for an example of five devices placed one after the other:

$$\begin{aligned} \varrho_{\text{res}} &= \varrho_1 \cdot d(\eta_{12}) \cdot d(\eta_{23}) \cdot d(\eta_{34}) \cdot d(\eta_{45}) \cdot c(c(c(c(\eta_{12}) \cdot \eta_{23}) \cdot \eta_{34}) \cdot \eta_{45}) \\ R_{\max, \text{res}} &= R_{\max, 5} \cdot c(c(c(c(\eta_{12}) \cdot \eta_{23}) \cdot \eta_{34}) \cdot \eta_{45}) \end{aligned} \quad (\text{IV.21})$$

$F_5(x) =$

$$\frac{F(F(F(F(x \cdot d(\eta_{12}) \cdot d(\eta_{23}) \cdot d(\eta_{34}) \cdot d(\eta_{45}) \cdot c(c(c(c(\eta_{12}) \cdot \eta_{23}) \cdot \eta_{34}) \cdot \eta_{45})) \cdot \eta_{12}) \cdot \eta_{23}) \cdot \eta_{34}) \cdot \eta_{45})}{c(c(c(c(\eta_{12}) \cdot \eta_{23}) \cdot \eta_{34}) \cdot \eta_{45})}$$

(IV.22)

For engineering purposes the output beam intensity distribution $I_{\text{out}}(y)$ is of main interest. Its analytical form, normalised to the intensity of a perfectly homogeneous output beam of identical total power, can be written as follows in the case of *two* devices cascaded:

$$I_{\text{out}}(y) = I_{\text{out}}(F_2(x)) = I_{\text{in}}(x \cdot \rho_{\text{res}}) \cdot \frac{2}{I_0} \cdot \frac{x}{F_2(x) \cdot F_2'(x)}. \quad (\text{IV.23})$$

In the above formula, $I_{\text{in}}(x \cdot \rho_{\text{res}})$ represents the input intensity distribution, which takes the following form for a Gaussian input beam:

$$I_{\text{in}}(r) = I_0 \cdot e^{-2 \cdot \left(\frac{r}{\varrho}\right)^2}, \quad (\text{IV.24})$$

where I_0 denotes the peak intensity. The intensity of the perfect output beam is:

$$I_{\text{out,perf}} = \frac{I_0}{2} \cdot \left(\frac{\varrho_{\text{res}}}{R_{\text{max,res}}} \right)^2, \quad (\text{IV.25})$$

where “perfect” refers to a beam, in which the total power of the input Gaussian beam of $\varrho_{\text{res}} 1/e^2$ intensity radius is homogeneously redistributed in an area of $R_{\text{max,res}}$ radius. Output beam intensity distributions are plotted in Fig. IV.4 at three different η_{12} values: 0.4, 1 and 6. The corresponding effective radii R_{eff} are: 0.94, 0.56 and 1.0 times $R_{\text{max,res}}$ (at $\pm 5\%$ inhomogeneity allowed), while T (i.e. the power transmission) is 94%, 31% and 100%, respectively. From these figures it can be seen that the curves are much more flat-top than a Gaussian beam, which has 10% of the total power at $\pm 5\%$ inhomogeneity as depicted in Fig. IV.4, curve ‘d’. It is also obvious that the near-perfect homogeneity can be achieved at low or high cross-beam shaping power values.

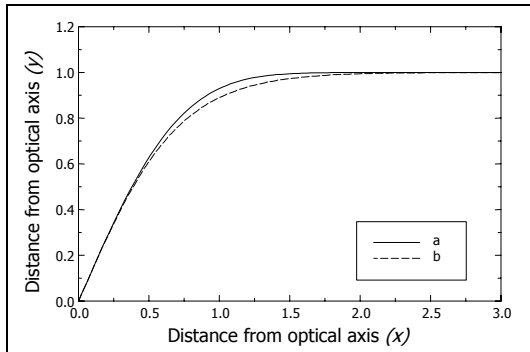


Fig. IV.3. Connection between normalised coordinate ‘ x ’ of a ray on the input surface, and that of the same ray ‘ y ’ on the output surface of a beam-shaper. (a) represents $F(x)$ of perfect Gaussian-to-uniform intensity transformation, while (b) represents $F_2(x)$ of two optical devices placed one after the other, each performing perfect beam transformation in case of $\eta_{12} = 1$. See details in the text.

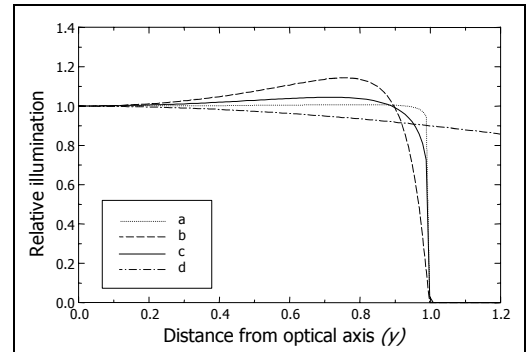


Fig. IV.4. Analytically calculated output beam intensity distribution (as a function of ‘ y ’) of two optical devices placed one after the other, each performing perfect Gaussian-to-uniform transformation. a.) $\eta_{12} = 0.4$, b.) $\eta_{12} = 1$, c.) $\eta_{12} = 6$, d.) Gaussian beam (of $\pm 5\%$ inhomogeneity measured within a circle of radius $y = 1$). See details in the text.

At this point it might be apparent to the reader that the choice of the map function of the elementary beam-shapers could be a serious question. In the above calculations I used Gaussian-to-uniform transforming devices, since their cascades provide acceptable resultant performance in most cases as well as since their analytic discussion is simple. However, it must be noted that with this solution a theoretically perfect output beam homogeneity is not achievable. A different choice of the map function that provides better beam homogeneity can be subject to further studies.

5 Application examples

5.1 Lens 1: single-element aspheric beam-shaper

This design example is intended to demonstrate the operation of the optimisation method presented in Subsection IV.4.1 by presenting the design process of a single-element aspheric beam-shaper shown in Fig. IV.2. First, I specify the parameters of the input and output beams. Let the $1/e^2$ intensity radius of the Gaussian input beam be $\varrho = 0.415$ mm. The maximum output beam radius is $R_{\max} = 3.7$ mm (i.e. $m = 8.92$), and $\eta = 3$. The beam shaping power has been chosen large enough so that diffraction effects of the input beam on the entrance pupil (it coincides here with the aperture of the first surface of the device) become negligible. The number of sample meridional rays used for optimisation is $n = 50$, equally distributed over the entrance pupil between $r = 0 \dots 3 \cdot \varrho$.

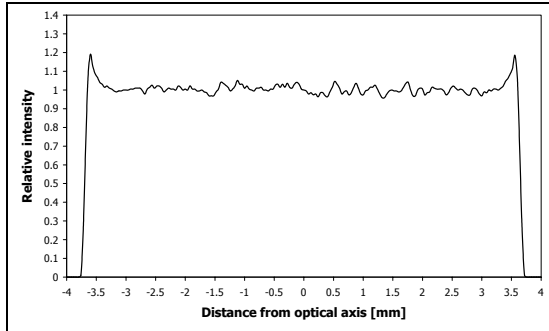


Fig. IV.5. Intensity distribution (i.e. ray number per unity area) of the designed beam-shaper as a function of ray height measured along the output surface. The plot is a cross-section of a 2D image obtained by random ray-tracing the beam-shaper.

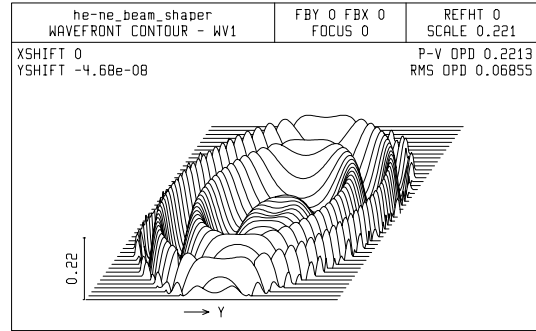


Fig. IV.6. Wavefront plot of the designed beam-shaper at its output surface. The diagram is shown inside a circle of radius $R_{\max}$ (i.e. 3.7 mm). Wavefront aberration is expressed in wavelengths (i.e. 632.8 nm).

Further requirements of the device are: diffraction limited collimation at $\lambda = 632.8$ nm (i.e. the RMS OPD of the output beam be lower than $0.07 \cdot \lambda$), and maximum $\pm 5\%$ intensity inhomogeneity. The length of the instrument in this case is determined by fabrication technology. The shorter is the device, the higher is the curvature of its surface profiles, assuming the same beam shaping requirements. Since short radii of curvature are difficult to fabricate, I chose the length to be 40 mm, which is large enough for fabrication, and sufficiently short for most applications. The device material is PMMA ($n_{632.8 \text{ nm}} =$

1.4884), which is easy to shape by diamond turning. The profile of the input and output surfaces is described by the equation:

$$z = \sum_{i=1}^{10} a_{2i} \cdot r^{2i}, \quad (\text{IV.26})$$

where ' r ' and ' z ' are explained in Fig. IV.2 (their dimension is [mm]), while ' a_2 ', ' a_4 ', ... , ' a_{20} ' are aspheric coefficients (see specification of dimensions in Tab. IV.1).

	Input surface	Output surface
a_2 [mm ⁻¹]	-4.4214577675e-001	-3.5239906294e-002
a_4 [mm ⁻³]	6.4285593631e-001	3.9667527920e-004
a_6 [mm ⁻⁵]	-8.3453440121e-001	-1.9517199170e-004
a_8 [mm ⁻⁷]	7.2845160942e-001	2.1312666900e-005
a_{10} [mm ⁻⁹]	-3.7017841709e-001	6.5867451512e-006
a_{12} [mm ⁻¹¹]	7.0082640464e-002	-2.0138808071e-006
a_{14} [mm ⁻¹³]	2.3748134060e-002	2.0763916086e-007
a_{16} [mm ⁻¹⁵]	-1.2988618600e-002	-8.5071700180e-009
a_{18} [mm ⁻¹⁷]	6.8499283280e-004	2.8320055530e-011
a_{20} [mm ⁻¹⁹]	3.4966402780e-004	4.6798609296e-012

Tab. IV.1. Calculated coefficients of the aspheric equation (IV.26) in case of the single-element twice-aspheric beam-shaper.

The automated design process started from an initial system (by changing a_2 , ... , a_{20} at both surfaces), where only the second-order terms were given at the input/output surfaces so that the device formed a usual afocal beam expander. The initial magnification was set close the paraxial magnification of the final system (m_{paraxial}), calculated from the map function:

$$m_{\text{paraxial}} \equiv \left. \frac{df(r)}{dr} \right|_{r=0} = \sqrt{2} \cdot \frac{R_{\text{max}}}{\rho} \equiv \sqrt{2} \cdot m. \quad (\text{IV.27})$$

To adjust the order of magnitude of intensity and wavefront operands to each other, I applied the following weighing factors: $w_{\text{int},i} = 1$ and $w_{\text{wvf},i} = 1000$. By alternating application of the damped least squares method and Powell's method, the optimisation converged into a device that fulfils all the requirement in about one hour (running on a 366 MHz PC).

The parameters of (IV.26) calculated by automated optimisation are listed in Tab. IV.1, the device itself can be seen in Fig. IV.2. The results of the simulated evaluation of the designed beam-shaper can be seen in Fig. IV.5 and Fig. IV.6. The former shows the intensity profile of the output beam at the

output surface, where the beam inhomogeneity is about $\pm 5\%$ of the average intensity of the whole beam, measured inside a circle of 3.44 mm radius (R_{eff}). The power transmission is $T \approx 86\%$. Fig. IV.6 presents the output wavefront, here the nominal RMS OPD is less than $0.07 \cdot \lambda$, i.e. the design is diffraction limited.

5.2 Lens 2: all-spherical multi-element beam-shaper

The below example is a beam-shaper cascade that works according to the theory presented in Subsection 4.2. The primary design requirement is to increase the beam shaping power above 2, in order to minimise diffraction effects that would otherwise disturb the output beam intensity profile. The concrete value of the parameters Q_{res} , $R_{\text{max,res}}$ is of small importance, since the system can be re-scaled to any size of input beam, and the necessary magnification specific to the actual design task can be implemented by placing an appropriate beam expander behind the device. A secondary design requirement is that the beam-shaper may contain only spherical lenses, the number of which should be as low as possible (this is not straightforward for cases when $\eta > 2$).

The highest reachable η of an all-spherical beam-shaper is in connection with its magnification. It is well-known that the aberrations of an optical system increase as it departs from the field-of-validity of the paraxial approximation. At all-spherical beam-shapers this is the case when the expansion ratio increases and the overall system length remains constant. Since beam shaping is performed by balancing spherical aberration of the lenses, a larger difference of ‘m’ from unity can result in a higher η . Preliminary study of all-spherical beam shaping systems showed that quadruplets are the best choice to alleviate automated optimisation and to maximise beam shaping power at the same time. I designed two such elementary beam-shapers: one of magnification well above unity (‘a’) and one below it (‘b’), so that $f_a(r_{\text{tr},a}) = r_{\text{tr},b}$ and $r_{\text{tr},a} = f_b(r_{\text{tr},b})$. In the cascade of two such beam shapers all energy emerging from the first device is fed into the second one. Since the output beam size of the second device is also the same as that of the input beam of the first, any number of these two-fold cascades can be put one after the other without subsequent loss of energy (neglecting Fresnel losses at refracting surfaces). From subsystems ‘a’ and ‘b’ at least 5 must be cascaded to reach $\eta = 2$, resulting in 20 lenses.

	m	η	ρ [mm]
a	6.11	0.600	0.80
b	0.127	0.778	4.5

Tab. IV.2. Design parameters of the elementary beam-shapers ‘a’ and ‘b’ depicted in Fig. IV.7.

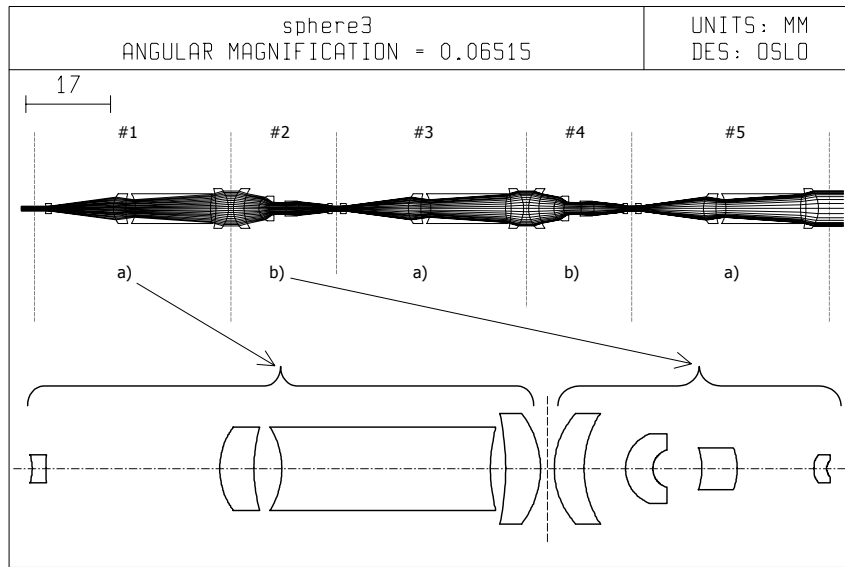


Fig. IV.7. Layout of the all-spherical beam-shaper consisting of 5 subsystems in a cascade.

The layout of the final beam shaping system is depicted in Fig. IV.7. Design parameters of the elementary devices are shown in Tab. IV.2 at $\lambda = 473$ nm. The elementary beam-shapers were designed for these parameters using my optimisation method described in Subsection IV.4.1. Construction parameters are summarised in Tab. IV.3 and IV.4; distance between the subsequent devices is around 1 mm.

Surface number	Radius of curvature [mm]	Thickness [mm]	Aperture radius [mm]	Glass material (Schott)
1	-3.756	1	1	SF11
2	14.172	12.5	1	AIR
3	4.989	2.5	3	SF11
4	11.072	2	3	AIR
5	-5.776	15	3	SF11
6	11.650	1.1	3	AIR
7	-19.123	2.5	4	SF11
8	-6.548	-	4	AIR

Tab. IV.3. Construction parameters of elementary beam-shaper 'a' shown in Fig. IV.7.

Surface number	Radius of curvature [mm]	Thickness [mm]	Aperture radius [mm]	Glass material (Schott)
1	5.952	2.14	4	SF11
2	7.332	2.99	4	AIR
3	2.669	1.96	2.5	SF11
4	1.403	3.52	1.35	AIR
5	-4.930	2.56	1.5	SF11
6	-4.276	5.51	1.5	AIR
7	1.606	0.95	1	SF11
8	0.851	-	0.6	AIR

Tab. IV.4. Construction parameters of elementary beam-shaper 'b' shown in Fig. IV.7.

Prior to presenting the optical performance of the complete system, I perform on it the analytical calculation described in Subsection IV.4.2 to obtain the theoretically available homogeneity of the output beam intensity profile of the five-element cascade. From Tab. IV.2 the values of the cross-beam shaping powers are: $\eta_{12} = \eta_{34} = 1.09$, $\eta_{23} = \eta_{45} = 0.716$, where

$$\eta_{12} \equiv \frac{R_{\max,1}}{Q_2} ; \quad \eta_{23} \equiv \frac{R_{\max,2}}{Q_3} ; \quad \eta_{34} \equiv \frac{R_{\max,3}}{Q_4} ; \quad \eta_{45} \equiv \frac{R_{\max,4}}{Q_5}. \quad (\text{IV.28})$$

The above values result in $Q_{\text{res}} = 0.243$ mm, $R_{\max,\text{res}} = 3.59$ mm according to (IV.21), and an output beam intensity profile shown in Fig. IV.8. To maximise the energy contents of the $\pm 5\%$ inhomogeneity range of the output beam, the size of the $1/e^2$ intensity radius (spot-size) of the input beam must be changed relative to the system (this corresponds to altering ‘d’ in (IV.16) as presented in Subsection IV.4.2). For this purpose the value of Q_{res} has to be scaled down to $Q_{\text{res, scale}} = 0.221$ (or equivalently, the system must be scaled up relative to the original Q_{res}).

The resulting output beam intensity curve after re-scaling is also shown in Fig. IV.8. With this modification the effective radius of the resulting system is $R_{\text{eff}} = 0.77 \cdot R_{\text{res}} = 2.75$ mm, inside which about 60% of the power of the input beam is collected (at inhomogeneity of $\pm 5\%$). After this re-scaling, the resulting magnification and beam shaping power will have the following values: $m_{\text{res, scale}} = 16.2$, $\eta_{\text{res, scale}} = 2.17$, which means about a 2.8 times growth in η .

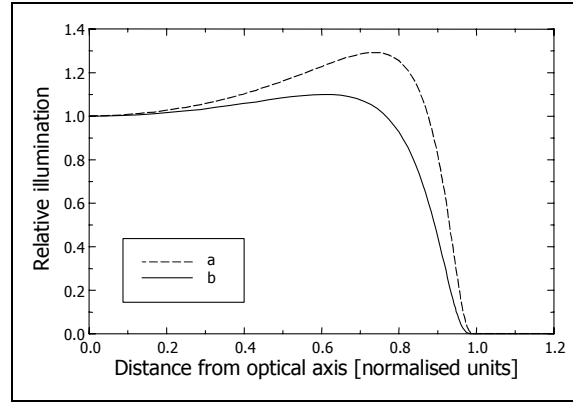


Fig. IV.8. Analytically calculated output beam intensity profile of the beam-shaper shown in Fig. IV.7. a - before re-scaling, b - after re-scaling. See details in the text.

Optical evaluation of the designed beam shaping system by the optical design programme results in a very similar behaviour to that predicted by the above discussed analytic calculation. Fig. IV.9 shows the output beam intensity profile after re-scaling, while the output wavefront is depicted in Fig. IV.10. Since the designed system cannot perfectly realise the ideal map function assumed in the analytic calculation, the resulting parameters will also have slightly different values: $Q_{\text{res, scale}} = 0.210$, $R_{\max,\text{res}} = 3.52$, $m_{\text{res, scale}} = 16.8$ and $\eta_{\text{res, scale}} = 2.29$.

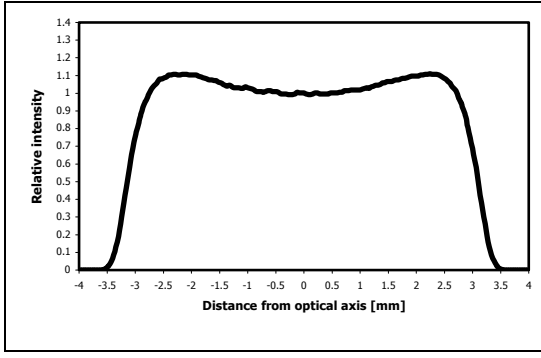


Fig. IV.9. Output beam intensity profile of the beam shaping system shown in Fig. IV.7. The plot is a cross-section of a 2D image obtained by random ray-tracing the beam-shaper.

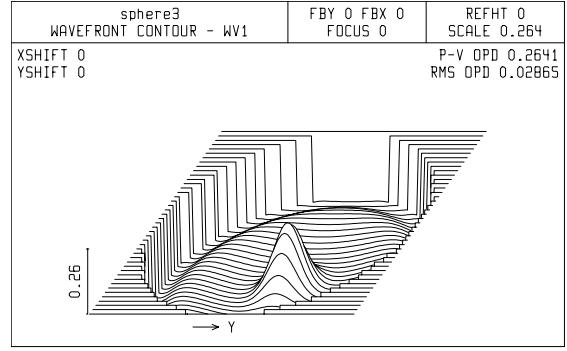


Fig. IV.10. Output wavefront of the beam shaping system shown in Fig. IV.7.

6 Verification

In the literature [91] one can find references to diffraction effects that take place when a Gaussian beam is truncated by a clear aperture (i.e. one that does not modify the phase) and then suffers free-space propagation at a given distance. These calculations predict that the aperture size must be about 3 times larger than the $1/e^2$ intensity part of the Gaussian beam to avoid diffraction fringes. However, it is not obvious that the same rule-of-thumb holds for the case of beam-shapers, where the truncating aperture (usually the input surface of the device) modifies also the phase of the beam. Certain calculations presented in [36] imply that η must be more than 2. Due to the importance of diffraction effects, I performed similar calculations, systematically changing η between 0.71 and 3 (the former value corresponds to truncating the input Gaussian beam at its $1/e$ intensity part).

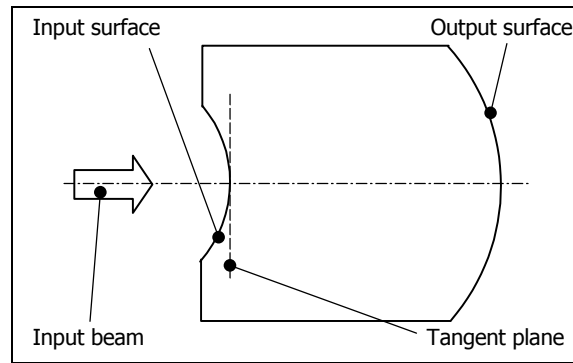


Fig. IV.11. Test setup for calculation of the diffraction pattern on the output surface of a beam-shaper. Rayleigh-Sommerfeld integration is performed over the tangent plane of the input surface, where the complex field distribution is determined by the geometrical approximation.

The test system used in the calculation (see Fig. IV.11) is the aspheric beam-shaper presented in Subsection IV.5.1. The basis of my diffraction calculations is the Rayleigh-Sommerfeld approximation, for the implementation of which I prepared my own routine (see Appendix A, B and C).

The results of the calculation are summarised in Fig. IV.12. The plots reveal that the output beam intensity profile inhomogeneity reduces to $\pm 5\%$ at about $\eta = 2$, and the ripples vanish over $\eta > 2.5$. Hence, for most practical problems it is enough to maintain the beam shaping power between 2 and 2.5, beyond which diffraction effects no longer take place.

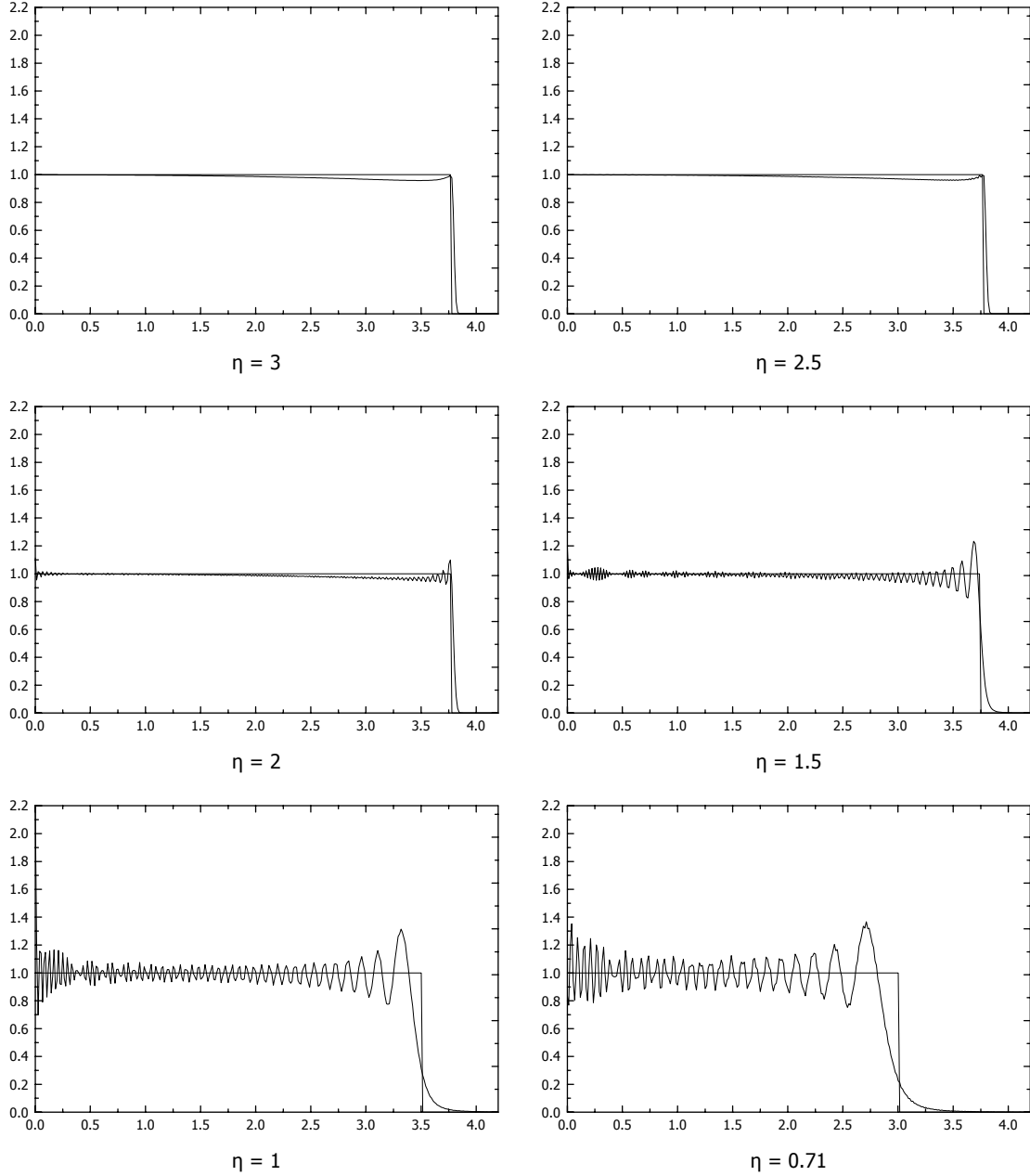


Fig. IV.12. Diffraction intensity distribution on the output surface of the single-element aspheric beam-shaper. Horizontal axes depict distance from optical axis in [mm], vertical axis depict intensity relative to the ideal value calculated from geometrical optics. The diagrams also show the geometrical size of the output beam. The six diffraction patterns represent different cases with respect to the truncation radius of the input Gaussian beam relative to its $1/e^2$ intensity radius (i.e. the beam shaping power).

The demonstrative calculation presented above also implies that the geometrical optical model (viz. ray-tracing) can be applied for the design of beam-shapers with a surprisingly high precision (provided the beam shaping power is sufficiently high). Analytic examination of exploring the validity

limits of ray-tracing in case of beam-shapers, e.g. by using the eikonal function, is an intriguing challenge that is still left open here as it points far beyond the scope of this discussion.

7 Conclusion

I have developed a new method for the design of laser beam-shapers. The method is a special application of automated optimization, in which I used the geometrical optical model for the near-field description of circular beams (i.e. excluding diffraction). Main advantage of the method is its generality: theoretically it is applicable to implement any kind of intensity transformation with arbitrary precision. Its limitations occur when the beam shaping system is of high degrees of freedom: here the optimisation converges slowly, and the optimum is hard to locate. I integrated the method into the advanced environment of OSLO SIX – optical design software of Sinclair Optics, Inc. – in order that its usage is simple and convenient. I demonstrated operation of the method through the optimization of a single-element aspheric beam-shaper that converts a Gaussian input beam into an output beam of uniform intensity profile. Verification of my calculations has been made by diffraction calculations performed on the geometrically designed device. Another result of my work in the field of beam-shaper design is a simple method for increasing the beam shaping power of beam transforming systems. The method is especially useful in case of all-spherical beam shaping systems, where high η is hard to achieve by current design techniques. In the dissertation I presented theoretical background of my method and designed an all-spherical system, comprising 20 lenses, to demonstrate its operation.

V ACCURATE MODELLING OF HYBRID DIFFRACTIVE/ CONVENTIONAL OPTICAL SYSTEMS

TABLE OF CONTENTS

TABLE OF CONTENTS.....	58
1 INTRODUCTION	59
2 FUNDAMENTALS.....	61
2.1 PRESUMPTIONS.....	61
2.2 OPTICAL CHARACTERISATION OF PHASE-MODULATION DOES	62
3 PRELIMINARIES (STANDARD MODELLING TECHNIQUE, AND THE IMAGING PROPERTIES OF DOES)	65
3.1 THE STANDARD METHOD FOR DOE MODELLING: DECOMPOSITION INTO DIFFRACTION ORDERS	65
3.2 EFFICIENT IMPLEMENTATION OF ORDER DECOMPOSITION IN COMMERCIAL DESIGN PROGRAMMES	67
3.3 IMAGING PROPERTIES OF HYBRID LENSES IN TERMS OF THE FIRST DIFFRACTION ORDER.....	67
3.4 IMAGING BY A HYBRID LENS IN THE PRESENCE OF MULTIPLE DIFFRACTION ORDERS	69
4 RESULTS (NEW DOE MODELLING TECHNIQUE AND ITS EFFICIENT IMPLEMENTATION)	72
4.1 IMPROVED ACCURACY MODELLING OF DOES: ZONE DECOMPOSITION.....	72
4.2 EFFICIENT IMPLEMENTATION OF ZONE DECOMPOSITION FOR SURFACE-RELIEF DOE MODELLING	73
5 APPLICATION EXAMPLES	75
5.1 LENS 1: ACHROMATIC IR DOUBLET.....	76
5.2 LENS 2: VISIBLE-RANGE APOCHROMATIC QUADRUPLLET.....	78
5.3 LENS 3: FAST ACHROMATIC IR DOUBLET.....	79
6 VERIFICATION.....	82
7 CONCLUSION	86

1 Introduction

Diffraction optics is a rapidly developing branch of modern optics [20]. Due to their numerous advantages (light weight, small thickness, easy mass manufacturability) diffractive components are more and more widely used in optical imaging systems. Since their realisation requires high-resolution surface-patterning techniques at low working wavelengths, their usage was initially confined to IR applications [42, 21]. However, recent fabrication technologies (ultra-precision single-point diamond turning, laser writing, holography, lithography) make them also available for the visible wavelength range [51, 107]. In the field of optical imaging, Diffractive Optical Elements (DOEs) play the role of focusing/aberration correcting lenses. Because of the special dispersive behaviour of DOEs, they are frequently combined with conventional components (i.e. refractive and reflective surfaces) to reduce chromatic aberrations [99, 29, 24, 25, 26, 82, 27, 112, 34, 75] and/or thermal sensitivity of optical systems [69]. Application of phase-modulation DOEs in these hybrid systems significantly reduces fabrication costs and structural complexity, while maintains high power transmission and yields a good or even enhanced optical quality.

My research work connects to the modelling of such hybrid systems. The analysis and design of these systems can be most efficiently carried out by commercial optical design programmes [79, 92]. However, such softwares use single diffraction order ray-tracing for the modelling of diffractive surfaces [100, 92, 51], which method does not take into account the effects of the non-nominal diffraction orders, though these can be significant in certain cases (usually when the evaluation wavelength of the DOE differs from its design wavelength, see Subsection V.3.4, or in case of binary optics, see [100], [101] and Subsection V.2.1). The straightforward solution, viz. coherent summation of diffraction orders is impractical for most optical design purposes due to its slowness (usually a very high number of orders must be summed up to obtain a satisfactory approximation of the output field of a DOE). There are several investigations that examine the transition between refractive and diffractive properties of DOEs [92, 85, 87], and entirely diffraction calculations [35, 28, 81, 41, 32, 88] both performed to reveal the real optical behaviour of hybrid systems. Presently, these calculations/considerations can only be used as supplements to the built-in modelling tool of commercial optical design programmes, which makes the design process somewhat complicated.

To avoid the above inconveniences and meet optical designers' needs, a new method has been developed for the modelling of DOEs, called "zone decomposition". The surface of most diffractive components contains annular-shaped regions, inside which the optical properties (phase and amplitude transmission) do not change abruptly; within certain limitations these "zones" can be regarded as conventional optical elements, i.e. lenses or mirrors. "Decomposition" means that the output field of

the separately analysed lens elements must be summed up coherently to obtain the resulting output field. It must be noted that in contrast to diffraction order decomposition, here a finite summation is sufficient to obtain the output field accurately. The calculation bases partly on geometrical optics and partly on diffraction: ray-tracing is used to determine the complex field distribution over the exit pupil of the hybrid system (from the optical path length measured along each traced ray, see Section II.1), from which the PSF (i.e. Point Spread Function) of the system is computed by scalar diffraction. It is one of the most important features of zone decomposition that the above described algorithm can be entirely integrated into commercial lens design programmes resulting in an efficient tool for the modelling of hybrid systems.

The first application of zone decomposition for the modelling of hybrid systems has been published in [31, 88]. There the computation was performed in three stages: first, ray-tracing data of the conventional part of the hybrid system was extracted from a lens design programme, then the optical path length (OPL) contribution of the DOE was added, and finally the diffraction intensity distribution of the image spot was determined by a one dimensional Fourier-Bessel transformation, also called zero-order Hankel-transformation [23]. This computation well demonstrated the usefulness, accuracy and rapidness of zone decomposition, though for optical design purposes it showed certain disadvantages. Since a significant part of the calculation was performed outside the lens design program, some of the convenient services of the software, especially that of automated optimisation, became unavailable. In addition, in that particular implementation, off-axis objects (i.e. non-rotationally symmetric cases) could not be analysed either.

To eliminate the above deficiencies, I made an efficient implementation of zone decomposition directly for usage in commercial optical design programmes. The new DOE model is realized by the User Defined Surface (UDS) option of such programmes that allows fast ray-tracing, since summation of the output field of the lens elements (i.e. the zones) is made automatically (see details in Subsection V.4.1). By application of the UDS model the range of analysable hybrid systems broadens with those also working with off-axis object points, as well as their optimisation becomes possible even in the presence of multiple diffraction orders (especially using diffraction PSF, MTF and RMS wavefront optimisation, see [79], [92] and Subsection V.5.3). Using the UDS model I evaluated several hybrid systems to show the differences between the new model and the standard method of commercial optical design programmes, i.e. single diffraction order ray-tracing.

In what follows, first I present the basic optical properties of DOEs and their standard modelling method together with its deficiencies. Then I show the principle of zone decomposition in detail and its efficient implementation in the form of UDS surfaces. After this, several application examples follow, and a section that investigates the correct operation of the UDS model.

2 Fundamentals

2.1 Presumptions

In general, the term “diffractive” indicates an optical surface that modifies the properties of the incident light in such a way that the propagation of the output field can no longer be described by purely geometrical optical means. In the present chapter I deal with one type of diffractive surfaces called “Diffractive Optical Element” (DOE), which is distinguished from other diffracting structures (such as a small single hole, slit etc.) by that its output field can be decomposed into *discrete* diffraction orders. A few examples of DOEs are thin and thick holograms, diffraction gratings, Fresnel zone-plates, kinoforms etc., all of which can work either in transmission or reflection.

DOEs diffract the traversing/reflecting light by modulating its amplitude and/or its phase. Since phase-modulation DOEs have (theoretically) 100% transmission/reflection, they are more frequently used in practice than amplitude DOEs that obscure (i.e. waste) a significant part of the input light. Phase modulation DOEs can be classified as thick and thin from operational point-of-view. Thick DOEs, such as volume holograms, have a thickness of many times the wavelength and utilize the Bragg phenomenon. On the contrary, the thickness of “thin” phase-modulation DOEs lies in the order of the wavelength, providing an operation within the Raman-Nath regime [22, 94]. Thin elements can be recorded holographically or can be generated by a computer. An attractive feature of computer generated thin DOEs (also known as kinoforms [44]) is that they can be designed at given wavelengths to transform all the transmitted light into one single diffraction order [76, 68, 63].

In this chapter of my dissertation I deal with the modelling and design of computer generated, thin phase-modulation DOEs. Such elements perform diffractive behaviour by a surface structure that introduces equal phase steps into the spatial phase-distribution of the input field at given positions of the DOE surface.* The annular-shaped continuous-phase regions between the phase-steps of the DOE surface are called zones. Most current manufacturing techniques (direct diamond turning, photolithography, laser writing, replication etc.) realise the above phase-modulation by varying the material thickness of a transparent/reflective substrate material, resulting in a relief-like surface profile

* For the sake of completeness, I must mention here a modification of diffractive elements, the so-called variable-phase step DOEs. Such elements work according to the same principle as “conventional” DOEs, except for their output field cannot be decomposed into discrete diffraction orders, due to phase-step variation over the surface. In Subsection V.4.1, one of their representatives, the so-called superzone kinoforms will be briefly described.

(see f_{doe} in Fig. V.1). Typical realisation forms are blazed and binary DOE profiles (see Fig. V.2), the substrate of the former often having an aspheric shape (f_{sub} in Fig. V.1). In case of both previous structures, the relief edges (at phase-steps) are usually parallel to the optical axis.

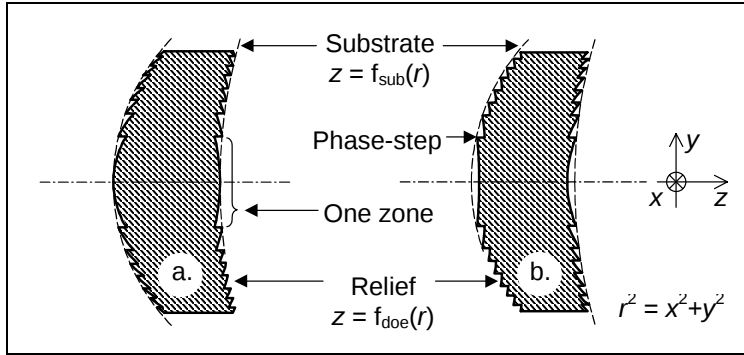


Fig. V.1. Demonstrative examples of thin phase-modulation (surface-relief) DOEs. The profiles are depicted much thicker than reality for better visibility (a. convergent-, b. divergent-output field relief profiles).

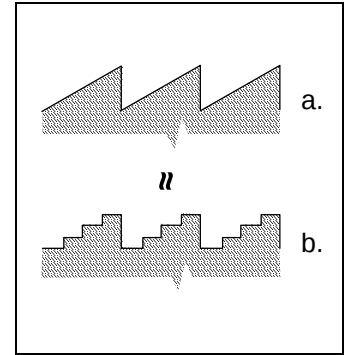


Fig. V.2. Blazed (a.) and binary (b.) DOE relief profiles. b. is used in photolithographic techniques as an approximation of a.

Throughout this chapter I assume scalar approximation to be valid for diffraction modelling of light propagation. This method is the fastest available diffraction algorithm implemented in commercial optical design programmes, providing accurate results within a large field of validity. However, it excludes from discussion those DOEs, the lateral (i.e. X-Y) zone dimension of which is smaller than a few times the wavelength. In addition, since the focal-plane intensity pattern of highly convergent beams cannot be calculated with pure scalar theory, it also excludes from discussion all imaging systems that have too high numerical aperture (viz. $\text{NA} > \approx 0.5$).

2.2 Optical characterisation of phase-modulation DOEs

Modelling of phase-modulation DOEs is usually divided into two parts: transmission through the element and beam propagation behind it; this subsection deals with the first part. Light propagation through a DOE can be precisely described in terms of rays, since diffraction effects over the short propagation distances inside the element are negligible. Different approaches developed for the modelling of beam propagation behind DOEs will be discussed in detail in Sections V.3 and V.4.

In this subsection, I confine attention to ideal DOEs, which means that fabrication issues (i.e. optical properties resulting from e.g. the binary approximation of a blazed relief profile, manufacturing errors etc.) will not be discussed here. The simplest model of such a DOE is an optical element of extended surface and zero thickness. At one given working wavelength λ (measured in vacuum), the transmittance of such an infinitely thin element can be described by the complex amplitude transmission 't', which is function of only the lateral coordinates 'x' and 'y':

$$u_{\text{out}}(x, y) = t(x, y) \cdot u_{\text{in}}(x, y), \quad (\text{V.1})$$

where $u_{\text{in}}(x, y)$ and $u_{\text{out}}(x, y)$ denote the input and output field complex amplitudes along the surface of the DOE, and the x, y arguments are interpreted over the whole interval of $[-\infty; +\infty]$. In this approximation, computer generated thin phase-modulation DOEs are traditionally described by a spatially continuous function, called the characteristic function $H(x, y)$. The Optical Path Length (OPL) contribution $H'(x, y)$ that the DOE adds to the incident field can be expressed with $H(x, y)$ as follows:

$$H'(x, y) \equiv \text{mod}'(H(x, y)), \quad (\text{V.2})$$

where mod' is a modified form of the modulo function, defined as

$$\text{mod}'(z) \equiv l \cdot \lambda_0 \cdot \text{mod}(z - \text{OPL}_0, l \cdot \lambda_0) + \text{OPL}_0 \quad \text{and} \quad \text{mod}(a, b) \equiv \frac{a}{b} - \text{floor}\left(\frac{a}{b}\right). \quad (\text{V.3})$$

The connection between $H'(x, y)$ and $H(x, y)$ is demonstrated graphically in Fig. V.3: the modulo function “cuts” the characteristic function into zones, each of which correspond to one DOE zone. In (V.3) λ_0 is the nominal or design wavelength, at integer-times multiples of which the diffraction efficiency-wavelength function has (theoretically) 100% value. The positive integer l determines the OPL level at which $H(x, y)$ is to be cut: $l\lambda_0$ equals to the height of OPL “steps” in $H'(x, y)$. l is usually set to 1, other values ($l \in \mathbb{N}^+$) result in the so-called higher-order or “harmonic” DOEs [42, 102]. l affects the basic diffractive properties of the DOE, such as wavelength dispersion, spatial separation of the orders etc. The parameter OPL_0 specifies the lower limit of $H'(x, y)$, its value usually falls between $-l\lambda_0$ and 0. (Effects of OPL_0 on the optical behaviour of DOEs is discussed in Subsection V.3.1.)*

The phase of the output field Φ_{out} on the surface of a DOE can be expressed with $H'(x, y)$ as:

$$\Phi_{\text{out}}(x, y) = \Phi_{\text{in}}(x, y) + \Phi'(x, y), \quad \Phi'(x, y) \equiv \frac{2\pi}{\lambda} \cdot H'(x, y), \quad (\text{V.4})$$

where Φ_{in} represents the phase of the input field on the DOE surface and Φ' denotes the phase contribution of the DOE to the input field. Since the amplitude transmittance of phase-modulation DOEs is unity (neglecting losses caused by e.g. Fresnel reflection), their transmission function is:

* For variable phase-step DOEs, mentioned earlier, formula (V.2) does not provide an adequate surface description, since (for certain reasons) in such elements the value of ‘ l ’ is changed from zone to zone, either discretely or continuously. The surface description formula (V.21) to be presented in Subsection V.5.2 offers a proper solution for a wide range of such DOEs.

$$t(x, y) = a(x, y) \cdot e^{i\Phi(x, y)}, \quad (\text{V.5})$$

where the aperture function $a(x, y)$, usually a circular or rectangular window, has been incorporated in $t(x, y)$ in order to include the finite lateral dimensions of real DOE surfaces. It must be emphasized here that in (V.5), as well as all subsequent formulae, the phase (and accordingly the optical path length) measured along a ray is considered as increasing in the direction of light propagation.

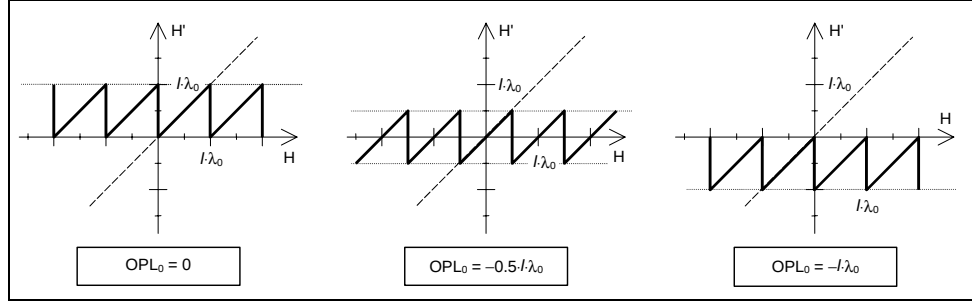


Fig. V.3. Connection between H' and the characteristic function H for different OPL_0 parameters. (The dashed line represents $H' \mapsto H$.)

In case rays propagate through the DOE parallel to its optical axis, the OPL contribution $H'(x, y)$ can be expressed by the surface profile $f_{\text{doe}}(x, y)$ as follows:

$$H'(x, y) = -(n_2(\lambda) - n_1(\lambda)) \cdot f_{\text{doe}}(x, y), \quad (\text{V.6})$$

where n_1 and n_2 are refractive indices as depicted in Fig. V.4. (For the sign definition of f_{doe} , see also Fig. V.1.) Effects of material dispersion on the imaging properties of DOEs are usually neglected relative to the effects of dispersion resulting from the diffractive behaviour of the surface (dispersion characteristics of diffractive and refractive surfaces are discussed in Subsection V.3.3). Correspondingly, $n_1 := n_1(\lambda_0) \approx n_1(\lambda)$ and $n_2 := n_2(\lambda_0) \approx n_2(\lambda)$.

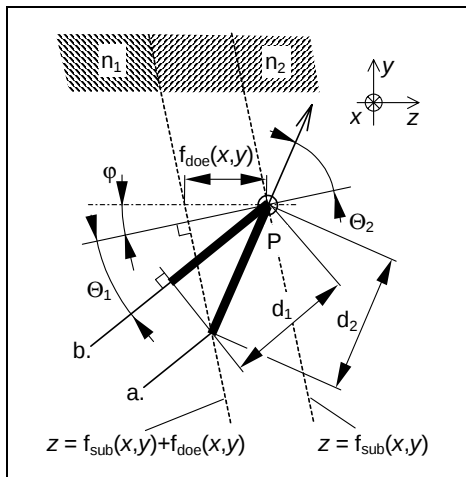


Fig. V.4. Calculating the OPL contribution of a DOE at point P along ray a . in its plane of incidence. OPL along ray b . serves as a reference. (Note that f_{doe} is negative in the figure.)

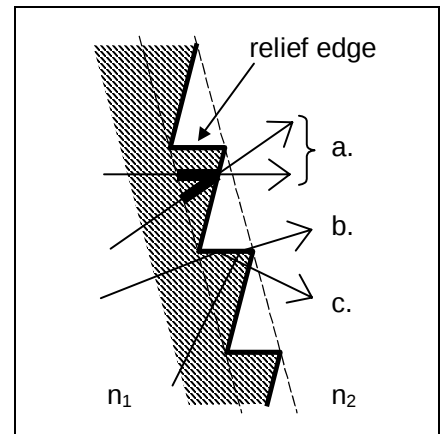


Fig. V.5. Ray-tracing properties of surface-relief DOEs. (a.) – OPL contribution depends on ray direction (see (V.7)); (b.) – ray refracted on relief edge (shadowing); (c.) – totally reflected ray on relief edge (shadowing).

The infinitely thin approximation can be applied for the modelling of computer generated thin phase-modulation DOEs with very high precision. However, formula (V.6) does not provide a sufficient model when rays passing through the diffractive element travel at a large angle relative to the optical axis. The function that describes the OPL contribution of the surface must be multiplied with an inclination factor in order to take account of the angle dependence of the optical path inside the DOE (see Fig. V.4). Assuming that the DOE profile (i.e. $z = f_{\text{sub}}(x, y) + f_{\text{doe}}(x, y)$) and the substrate profile (i.e. $z = f_{\text{sub}}(x, y)$) are locally parallel to each other (approx. I), as well as that the wavefront can be approximated by a plane wave in the vicinity of the ray-DOE intersection point (approx. II), the resulting H' is:

$$H'(x, y) = -(n_2 \cdot \cos(\Theta_2) - n_1 \cdot \cos(\Theta_1)) \cdot \cos \varphi \cdot f_{\text{doe}}(x, y), \quad (\text{V.7})$$

where

$$H'(x, y) = n_2 \cdot d_2 - n_1 \cdot d_1 \quad ; \quad d_2 = -f_{\text{doe}}(x, y) \cdot \frac{\cos \varphi}{\cos \Theta_2} \quad ; \quad d_1 = d_2 \cdot \cos(\Theta_2 - \Theta_1). \quad (\text{V.8})$$

In most practical cases (viz. when the d_1 and d_2 distances lie in the order of a few wavelengths) the above-mentioned (I) and (II) approximations cause only a negligible error in the expression of (V.7).

The amplitude transmission of surface-relief DOEs also differ from the ideal value of unity. Because of the shadowing effect of relief edges (see Fig. V.5), a small fraction of the input light is always obscured, giving rise to a weak, more-or-less uniform stray light that slightly decreases the contrast of the image. In most cases (i.e. excluding lenses of high NA) the effects of shadowing can be neglected.

3 Preliminaries (Standard modelling technique, and the imaging properties of DOEs)

3.1 The standard method for DOE modelling: decomposition into diffraction orders

Basics of this theory can be found in [100], to which I made amendments in order to take account of the effects of l and OPL_0 on the output field of DOEs. Denoting by $\Phi(x, y)$ the phase contribution that corresponds to the OPL contribution $H(x, y)$, it follows from (V.2) and (V.4):

$$\Phi'(x, y) = \frac{2\pi}{\lambda} \cdot \text{mod}'\left(\frac{\lambda}{2\pi} \cdot \Phi(x, y)\right). \quad (\text{V.9})$$

Since $\Phi'(x, y)$ is a periodic function of $\Phi(x, y)$, $e^{i\Phi'(x, y)}$ is also periodic in $\Phi(x, y)$, consequently it can be written in the form of a Fourier series expansion:*

$$e^{i\Phi'(x, y)} = \sum_{m=-\infty}^{\infty} c_m \cdot e^{i \frac{\lambda}{L \cdot \lambda_0} m \cdot \Phi(x, y)} \Rightarrow t'(x, y) = a(x, y) \cdot \sum_{m=-\infty}^{\infty} c_m \cdot e^{i \frac{\lambda}{L \cdot \lambda_0} m \cdot \Phi(x, y)}, \quad (\text{V.10})$$

where the Fourier series has been substituted into (V.5). Resulting from my calculations, the coefficients c_m of the series are

$$c_m = \text{sinc}\left(\frac{L \cdot \lambda_0}{\lambda} - m\right) \cdot e^{i\pi \left(\frac{L \cdot \lambda_0}{\lambda} - m\right) \left(1 + 2 \cdot \frac{\text{OPL}_0}{L \cdot \lambda_0}\right)}. \quad (\text{V.11})$$

Substituting (V.10) into (V.1), and moving all coefficients into the summation, one obtains the expression for the output field, right behind the DOE:

$$u_{\text{out}}(x, y) = \sum_{m=-\infty}^{\infty} a(x, y) \cdot u_{\text{in}}(x, y) \cdot c_m \cdot e^{i \frac{\lambda}{L \cdot \lambda_0} m \cdot \Phi(x, y)} = \sum_{m=-\infty}^{\infty} u_m(x, y), \quad (\text{V.12})$$

where the terms $u_m(x, y)$ are complex amplitudes of the so-called “diffraction orders”. The above equation shows that diffraction orders represent light beams of continuous wavefronts, the propagation of which can be described by geometrical optics (as far as diffraction effects caused by the aperture function $a(x, y)$ are negligible). This property of diffraction order decomposition is used for modelling DOEs in commercial optical design programmes, where the basic modelling tool is ray-tracing.

The absolute square of the coefficients c_m describe the diffraction orders from energetic point-of-view. The quantity that shows the ratio of the power of the m^{th} order (P_m) to the total incident power (P_{in}) is called the diffraction efficiency η_m ; its relation to the coefficients in (V.12) is:

$$\eta_m \equiv \frac{P_m}{P_{\text{in}}} = |c_m|^2. \quad (\text{V.13})$$

Ray-tracing models of DOEs usually assume that only one diffraction order (most often the first one) has a non-zero diffraction efficiency, i.e. $c_1=1$ and $c_{m \neq 1}=0$. It is interesting to note, that the value of OPL_0 does not affect η_m , it only causes a change in the phase offset between the diffraction orders.

* The careful reader can notice a minor departure between (V.10) and Eq. 2.15 of [100]. Since the period of the $|\Phi(x)|_\alpha$ function (defined in [100]) is α instead of unity, as stated in the reference, the right form of Eq. 2.15 looks like:

$$e^{i2\pi\Phi'(x)} = \sum_{m=-\infty}^{\infty} c_m \cdot e^{i2\pi m \cdot \Phi(x) \frac{1}{\alpha}}.$$

3.2 Efficient implementation of order decomposition in commercial design programmes

Most of the optical design programmes use ray-tracing methods for the analysis of optical systems. (Remember that with this method not only geometrical optical evaluation is possible, but diffraction effects of the aperture stop can also be taken into account by calculating diffraction PSF, MTF etc. from the geometrical wavefront, see Section II.1.) Such programmes evaluate and optimise all kind of DOEs by applying the single-order ray-tracing model, which has been presented in Subsection V.3.1. A more detailed discussion of the method to be presented below is given in [92].

Optical design programmes regard the vicinity of each point of the surface of a DOE as a plane linear diffraction grating, and determines the wave-vector $\mathbf{k}_{\text{out}}$ of the output ray at this point by applying the diffraction grating equation to the wave-vector $\mathbf{k}_{\text{in}}$ of the incident ray, see (V.14) below.

$$\left. \begin{aligned} k_{\text{out},x} &= k_{\text{in},x} + m \cdot \frac{\partial \Phi_0(x,y)}{\partial x} \\ k_{\text{out},y} &= k_{\text{in},y} + m \cdot \frac{\partial \Phi_0(x,y)}{\partial y} \end{aligned} \right\} ; \quad |\mathbf{k}_{\text{in}}| = |\mathbf{k}_{\text{out}}| = \frac{2\pi}{\lambda} \quad (\text{V.14})$$

$$\Phi_0(x,y) = \frac{2\pi}{l \cdot \lambda_0} \cdot H(x,y) = \frac{\lambda}{l \cdot \lambda_0} \cdot \Phi(x,y) \quad (\text{V.15})$$

Generally, it is enough to know the periodicity and direction of ruling of the local linear grating to perform the above-presented ray-tracing algorithm. In the case of phase modulation DOEs, these are determined from the gradient of $\Phi_0(x,y)$, as shown in (V.15), where the zero subscript denotes that Φ is calculated at the nominal wavelength. The phase (i.e. the OPL), measured along each ray must also be corrected after passing through the DOE by the contribution $\Phi_m(x,y)$ of the diffractive surface:

$$\Phi_m(x,y) = m \cdot \Phi_0(x,y) . \quad (\text{V.16})$$

The most significant property is at once the largest disadvantage of this method: for the sake of time efficiency, it performs the analysis of the optical system only in one order at a time, which thus excludes the effects of other orders, especially in cases when $\lambda \neq \lambda_0$.

3.3 Imaging properties of hybrid lenses in terms of the first diffraction order

In the field of IR optical imaging (and recently even in case of visible systems) DOEs are widely used as correction elements for chromatic aberration. In the wavelength range of 3-5 μm and 8-13 μm , the choice of materials of different material dispersion is strongly limited (germanium is the most widespread such material), so it is hard to build achromatic doublets. If such a system must be corrected for chromatic aberration, a different method could be extremely helpful. For this purpose DOEs offer an adequate solution due to their special chromatic properties: in contrast to refractive surfaces, DOEs

have a negative dispersion, as illustrated in Fig. V.6, which makes it possible to use a DOE of positive power for the compensation of the chromatic aberrations of an also positive refractive surface [100]. (Eliminating the need for a negative power component from an achromatic system also reduces the resultant monochromatic aberrations [96].) Those optical systems that are built of both conventional (i.e. refractive/reflective) and diffractive surfaces are called “hybrid systems”. I must mention here, that DOEs can also be used for the correction of sphero-chromatism (i.e. undesired spherical aberration variation with wavelength), and for reduction of thermal effects in lens systems (i.e. those caused by thermal dilation and refractive index-temperature variation). Never-theless, the latter two phenomena will not be discussed here.

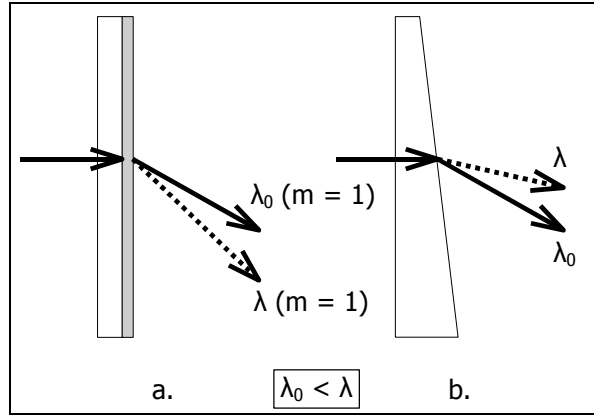


Fig. V.6. Wavelength dependence of deflection angle in case of a DOE a., and a refractive surface b. DOEs deflect rays of long wavelength with a larger angle than rays of short wavelength. It is an opposite behaviour to that of refractive surfaces.

To achieve achromatic focusing in hybrid doublets, a special condition must be satisfied, called the paraxial achromatic equation [100]. Let us consider the case of a thin lens that consists of a diffractive ($m := 1$) and a refractive surface of focal lengths $F_d(\lambda)$ and $F_r(\lambda)$, respectively. The wavelength-dependence of the focal lengths of these surfaces are:

$$F_d(\lambda) = F_d(\lambda_0) \cdot \frac{\lambda_0}{\lambda} \quad \text{and} \quad F_r(\lambda) = F_r(\lambda_0) \cdot \left[1 - \frac{D \cdot (\lambda - \lambda_0)}{n(\lambda_0) - 1} \right]^{-1}, \quad (\text{V.17})$$

where $n(\lambda) \approx n(\lambda_0) - D \cdot (\lambda - \lambda_0)$.

In (V.17) ‘D’ is a constant that characterises dispersion of the lens material. (Note, that in the above expression of $F_d(\lambda)$ it is assumed that the DOE operates in its first diffraction order, and that the nominal wavelength of the DOE is λ_0 .) The case of wavelength-independent $F(\lambda)$ can be described by:

$$\frac{1}{F(\lambda)} \equiv \frac{1}{F(\lambda_0)}, \quad \text{where} \quad \frac{1}{F(\lambda_0)} = \frac{1}{F_r(\lambda_0)} + \frac{1}{F_d(\lambda_0)}, \quad (\text{V.18})$$

as it follows from the formula that gives the resultant focal length of a thin doublet (see [96]). Substituting (V.17) into (V.18), the requested condition can be easily derived:

$$\frac{F_d(\lambda_0)}{F_r(\lambda_0)} = \frac{n(\lambda_0) - 1}{\lambda_0 \cdot D}. \quad (\text{V.19})$$

3.4 Imaging by a hybrid lens in the presence of multiple diffraction orders

In the previous paragraph, the DOE was assumed to work in one single (viz. the first) diffraction order. This requirement can never be satisfied perfectly if the device is operated not at its nominal wavelength (or for binary optics etc.). In such cases, an infinite number of orders are present (the coherent superposition of which forms the image), thus the image quality will decrease in a certain extent. In this subsection, I discuss those basic phenomena that can be observed when non-nominal orders of the DOE are taken into account in the optical analysis of hybrid systems, especially when the working wavelength (λ) of the DOE is different from the nominal one (λ_0). Until more precise design methods have been developed, considerations originated in the phenomena presented below served as a basis to estimate the effects of higher diffraction orders on the image quality of hybrid lenses.

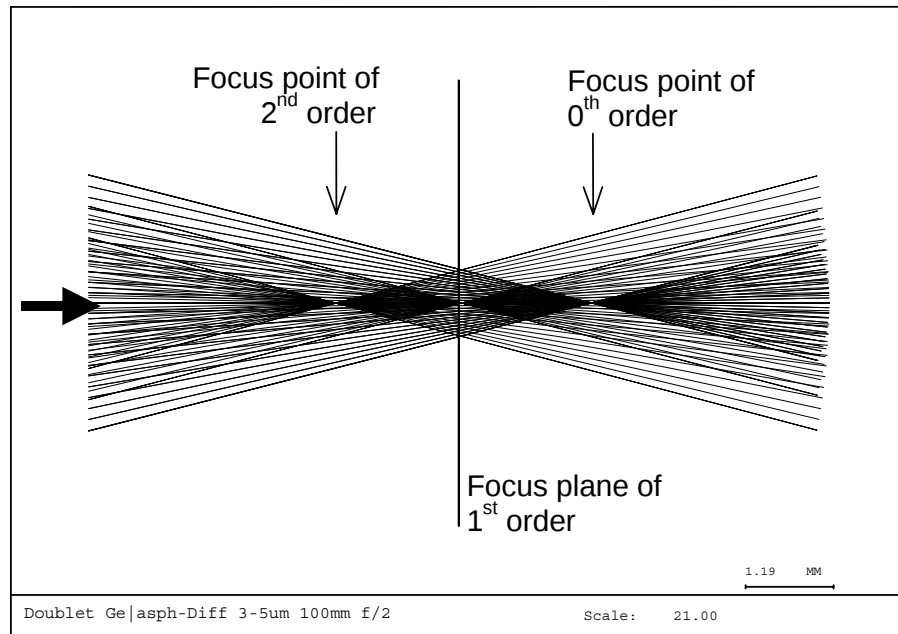


Fig. V.7. Geometrical optical representation of the vicinity of the focal plane of the first diffraction order of a hybrid lens, at $\lambda_0/\lambda = 1.33$. In the figure only the first, second and zeroth orders are depicted. (The figure does not reflect, neither intends to do so, any difference between diffraction efficiencies of the presented orders.)

For demonstration purposes, let us consider a refractive singlet having a DOE on one surface, which together build a hybrid doublet (this lens will be discussed in detail in Subsection V.5.1). I assume here that the DOE has a perfect relief profile, thus fabrication errors can be neglected. The lens performs diffraction limited infinite conjugate imaging in the 3-5 μm wavelength range (the f-number is

$f/2$), its nominal wavelength is $4\text{ }\mu\text{m}$. In order that the effects of multiple orders will be apparent, I evaluate the lens in the focal plane of the first diffraction order at $\lambda = 3\text{ }\mu\text{m}$, i.e. $\lambda_0/\lambda = 1.33$ (see Fig. V.7). Since $\lambda \neq \lambda_0$, an infinite number of orders are present, all of which are defocused relative to the first one (the lower orders are nearly diffraction limited spherical waves, for $m = -1..4$, while the higher ones are aberrated in a greater extent).

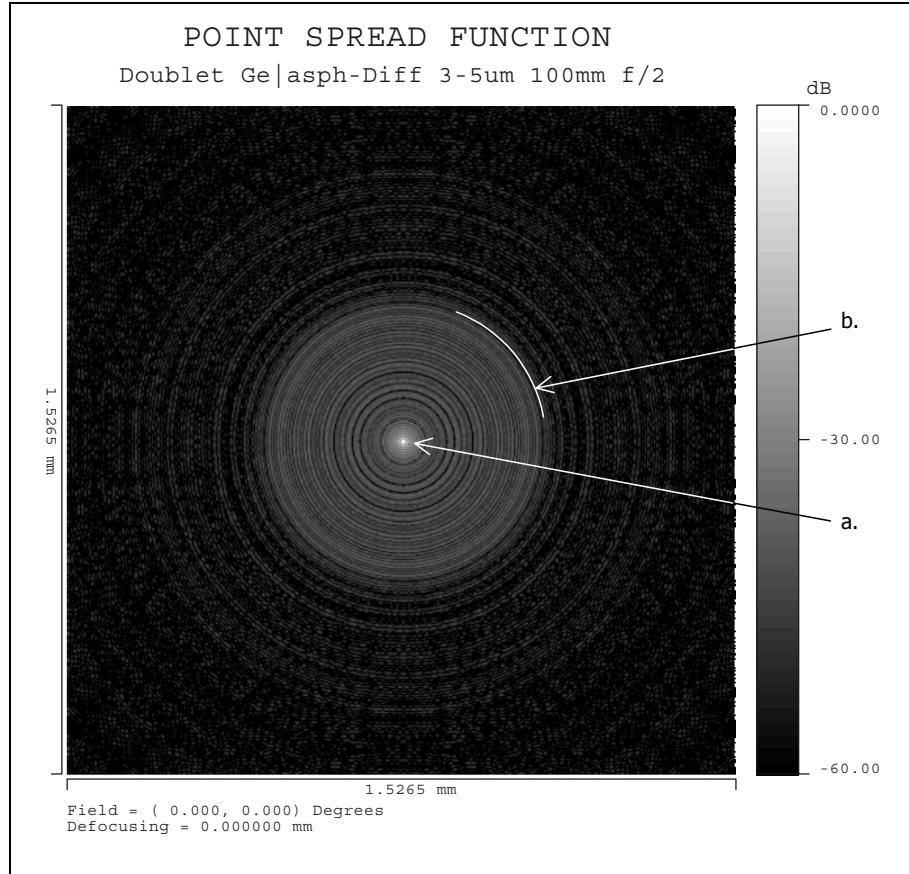


Fig. V.8. Diffraction intensity distribution in the focal plane of the first order of a hybrid lens, at $\lambda_0/\lambda = 1.33$. The Airy-pattern of the first order can be seen in the middle (a.), surrounded by the (identical-sized) light-disks of the second and zeroth orders (b.). Note the Newton-fringes caused by interference of the second and the zeroth orders. ($\lambda_0 = 4\text{ }\mu\text{m}$.) See details in the text.

The beam diameter of orders symmetric to the first one (i.e. 0 and 2, -1 and 3 etc.) are of the same size in the focal plane of the first order (see Fig. V.7). Diffraction intensity distribution of the focus spot is shown in Fig. V.8.* The Airy pattern of the first order can be seen at the center, while the defocused second and zeroth orders surround it in the form of a light-disk (Newton-fringes caused by the interference of the zeroth and second order are quite distinct). It is the diffraction efficiency of the first order ($\eta_1 = 0.684$) that primarily determines the Strehl-ratio of the focus spot.

* The plot has been computed here using the new DOE modelling method to be presented in Subsection V.4.2.

The effects of the higher orders are apparent in the MTF of the lens too, see Fig. V.9.* For interpretation of these curves one must know two things. First, an MTF curve is always normalized to the total power of the incident beam. Consequently, the MTF of the first order is $\eta_1 \approx 0.684$ at zero spatial frequency, and its shape is that of a diffraction limited MTF. Second, the intensity pattern of the higher orders is a much larger disk than that of the first one, thus their intensity level is orders of magnitude less than that of the first one. For this reason, the diffraction *intensity* patterns, hence the MTFs of the first and the higher orders can be added up (with certain limitations, see [28, 88]) to approximate the resultant curve. This can be seen in Fig. V.9, where the difference between the resultant MTF and that of the first order is a sinc-like function (practically a Dirac-delta at zero spatial frequency), dominated by the MTF of the second order, which is a strongly defocused spherical wave [44].

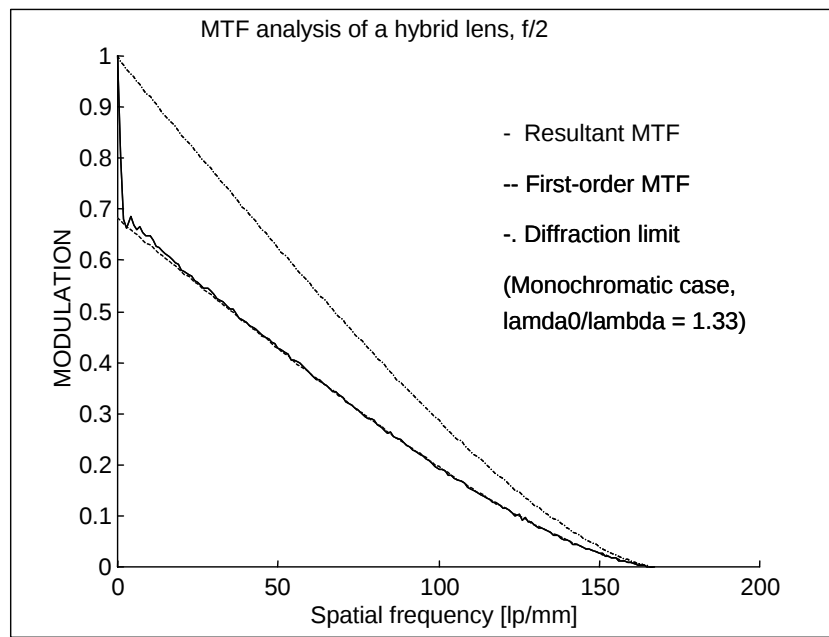


Fig. V.9. Diffraction MTF curve of a hybrid lens in the focal plane of the first order, at $\lambda_0/\lambda = 1.33$. The MTF of the first order shown in the figure is a diffraction limited curve scaled to 0.684 (i.e. η_1) at zero spatial frequency. ($\lambda_0 = 4 \mu\text{m}$.) See details in the text.

Unfortunately, the field of validity of the above simple ideas (also known as PSF and MTF scaling by diffraction efficiency) is quite limited: their accuracy decrease when the higher diffraction orders focus close to the first one, or when they are strongly aberrated. In these cases coherent summation of a large number of orders should be made, which is most often impractical for optical design purposes. In the next section I present a method that offers a solution to the above problems, together with its efficient implementation in commercial optical design programmes.

* The diagram has been computed here using the new DOE modelling method to be presented in Subsection V.4.2.

4 Results (New DOE modelling technique and its efficient implementation)

In this section I present the concept of a new surface-relief DOE model together with its efficient implementation in the form of software extensions, as well as I discuss the advantages and disadvantages of its usage for the analysis and design of hybrid systems compared to those of the standard model of commercial optical design programmes.

4.1 Improved accuracy modelling of DOEs: zone decomposition

In Subsection V.3.1 it was mentioned that in order to correctly describe the output field of a DOE by order decomposition, many diffractive orders have to be summed up coherently. Such a calculation is laborious and slow, moreover, optical design programmes presently do not support it as a built-in feature. The method of “zone-decomposition” has been developed to eliminate the need for coherent summation of diffraction orders when a precise optical evaluation of a DOE is necessary.

As described in Subsection V.2.2, the $H'(x, y)$ function of DOEs consists of separate zones, inside which the OPL contribution of the surface changes continuously, and at the boundaries of which the OPL contribution has discontinuities equalling $l\lambda_0$. From this fact follows the idea of zone decomposition: *inside any one of its zones, a DOE can be represented by a conventional lens, the aperture of which is determined by the contour of the surrounding zone boundaries; the output fields of the lens elements are calculated at an infinitesimal distance after the DOE by geometrical optical means, from which the field distribution in the image plane is calculated by scalar diffraction, and then the results for each zone are summed up coherently* [88]. Demonstration of the correct operation of zone decomposition will be given in Section V.6, where a comparison is made between its results and those of coherent summation of multiple diffraction orders.

From practical computing point-of-view, the obvious advantage of the above method over order-decomposition is that the number of zones is always finite and relatively low, thus no huge summation is required to obtain precise results. Another advantage is that this model still works (with appropriate implementation) in case of binary optics as well as variable phase-step DOEs, i.e. optical surfaces that behave like conventional DOEs, but where diffraction orders cannot be defined. (Such optical elements are subject to contemporary research, see for instance superzone kinoforms* [63, 42].) From optical design point-of-view, the largest advantage is that with proper implementation (e.g. in a commercial optical design software) the zone decomposition model can be efficiently applied for the modelling of hybrid systems. The only disadvantage of the method is that the precision of the resulting model may

* For these devices l , cf. (V.3), has different discrete values over larger regions (superzones) that cover several DOE zones.

depend on the position of the DOE inside the hybrid system, as diffraction theory implies. From this point-of-view, the next two cases (including four subcases) can be distinguished:

- a) the DOE is the *last* optical surface before the image plane and it coincides with the exit pupil; the DOE is positioned *inside* the hybrid system and it coincides with the aperture stop (i.e. the exit pupil and the DOE are conjugates of each other)
- b) the DOE is the *last* optical surface before the image plane and it is not located at the exit pupil; the DOE is positioned *inside* the hybrid system and it is not the aperture stop (i.e. the exit pupil and the DOE are not conjugates of each other).

As it was described in Section II.1, design softwares determine diffraction patterns in the image plane by performing diffraction computations on the geometrically-calculated complex field distribution in the exit pupil. From this respect, case a) is a direct application of zone decomposition as described in the previous paragraph, but when neither the DOE nor its image does not correspond to the exit pupil, see case b), there is a potential source of error in the model. In case b), Fresnel-diffraction may affect beam propagation behind the DOE in such an extent that the geometrical optical approximation may provide inaccurate results for calculating the field distribution on the exit pupil. Since Fresnel-diffraction becomes more and more significant as zone dimensions decrease, the distance – at which ray-tracing (i.e. the OPL calculation) is precise enough for the characterisation of output field propagation between the DOE and the aperture stop – is determined by the size of DOE zones. It is an assumption that in case b) the effects of Fresnel diffraction on the imaging properties of hybrid systems is negligible, however its correct proof has not yet been given.

4.2 Efficient implementation of zone decomposition for surface-relief DOE modelling

From the point-of-view of lens design programmes, the relief structure is a multi-faceted surface. One of the standard methods for the modelling of such surfaces by commercial ray-tracing programmes is to use non-sequential ray-tracing, where each surface-segment is regarded as a *separate* surface. In case of large number of zones this method promises to be very slow, thus I decided to built the $\mathcal{Z} = f_{\text{sub}}(r) + f_{\text{doc}}(r)$ surface profile (see Fig. V.1 in Subsection V.2.1 for definition), by using another possibility of these programmes, called User Defined Surface in CODE V [79], and User Defined Sag surface in OSLO SIX [92] (in both cases the abbreviation of this surface type is UDS).*

* I must mention here, that CODE V also offers a special surface type called UD2, developed especially for the modelling of Fresnel, i.e. multi-faceted surfaces, which are similar in structure to surface-relief DOEs. Since my implementation promised to result in a simpler code, I have not checked yet the usefulness (regarding efficiency) of the UD2 surface.

An interesting feature of commercial lens design programmes is that they are able to find the intersection point of a ray with a UDS surface even if it has discontinuities in discrete points. I utilised this possibility by defining the DOE surface as one *single* (however discontinuous) function, as described below, resulting in a fast modelling tool for lens design programmes. The efficiency of my method lies in the fact that summation of the output field for the separate zones are made automatically, by simply tracing a bundle of rays (i.e. a light beam) through the hybrid system. Another advantage of describing a DOE surface by a single function, see (V.21) below, is that it allows an implementation in which zone-boundaries are determined automatically. This is a great gain, since the calculation of zone boundaries (required by certain former DOE modelling techniques, see [47, 48, 46, 49, 50, 73]) needs a time-consuming computation. My method has also certain disadvantages of course, e.g. it cannot model reliefs where the edges are not parallel to the Z axis (this is a straightforward consequence of the way I implemented zone decomposition using a UDS surface). Nevertheless, assuming the relief edges to be parallel to the Z axis does not cause a large error in ray-tracing, if the relief-steps are in the order of the wavelength, which is often the case (see the shadowing effect described in Subsection V.2.2).

I describe the substrate profile $\mathcal{Z} = f_{\text{sub}}(r)$ of surface relief DOEs by the standard aspheric equation [79, 92, 96]:

$$f_{\text{sub}}(r) = \frac{C \cdot r^2}{1 + \sqrt{1 - (K + 1) \cdot r^2}} + a \cdot r^4 + b \cdot r^6 + c \cdot r^8 + d \cdot r^{10}, \quad (\text{V.20})$$

where ‘C’ means curvature, ‘K’ is the conic constant and a, b, c, d are aspheric coefficients. For a given characteristic function $H(r)$, to be realised by the DOE, I specify the relief profile $\mathcal{Z} = f_{\text{doe}}(r)$ by (V.21) and (V.22), assuming $l = 1$ and $\text{OPL}_0 = -\lambda_0$, which are the most often used values.

$$f_{\text{doe}}(r) \equiv f(r) \cdot \text{mod}'(H(r)) = f(r) \cdot H'(r) \quad (\text{V.21})$$

$$f(r) \equiv f_0 + f_2 \cdot r^2 + f_4 \cdot r^4 + \dots \quad \text{and} \quad H(r) \equiv h_0 + h_2 \cdot r^2 + h_4 \cdot r^4 + h_6 \cdot r^6 + \dots \quad (\text{V.22})$$

According to (V.22), the characteristic function $H(r)$ is defined by a polynomial, and $f(r)$ determines the geometrical size of the relief-profile steps. By setting the following values:

$$f_0 = \frac{1}{n_1 - n_2} \quad \text{and} \quad f_2 = f_4 = \dots = 0, \quad (\text{V.23})$$

one obtains the same formula as (V.6).*

In Subsection V.2.2 I mentioned that in case rays hit the DOE surface at large angle to the optical axis, (V.6) may provide an inaccurate surface profile (see Fig. V.5, case (a.)), which can spoil the results of the optical analysis of hybrid systems. This problem can be solved by choosing $f_{\text{doe}}(r)$ corresponding to (V.7) (see also [47, 48, 46, 49, 50, 73]). In my solution, the step heights are changed over the DOE surface by the higher-order terms in the polynomial expression of $f(r)$. The values of the parameters f_0 , f_2 , f_4 etc. can be easily found for any specific hybrid system either manually (guess-and-try), analytically (directly fitting $f(r)$ onto the coefficient of $H'(r)$ in the rearranged expression† of (V.7)) or by the more systematic automated optimisation (which is a standard method of commercial design programmes). Examples for these calculations will be shown in Subsection V.5.3.

5 Application examples

In this subsection I demonstrate the usage of the new DOE model by presenting examples for the analysis of hybrid systems. Though I made the software extensions to both CODE V and OSLO SIX, all the demonstrations to be presented below have been made using only one of these programmes (OSLO SIX), for consistency. In what follows I will compare the most interesting characteristics of hybrid systems obtained by UDS modelling, with those obtained by single-order ray-tracing. The examined characteristics include: wavefront-shape in the exit pupil (OPD curve), intensity distribution in the close vicinity of the focus spot (diffraction Point Spread Function, PSF), together with its Fourier-transform (diffraction Modulation Transfer Function, MTF), and radial energy distribution in the image plane (diffraction Encircled Energy) that reveals the characteristics of the focal intensity pattern far from the optical axis. I will also calculate the normalised maximum value of the PSF plot (called Strehl-ratio), which gives the peak of the diffraction intensity pattern relative to that of the reference spherical wave having the same NA and total power than those of the real output beam. In the following I will present both monochromatic and polychromatic calculations, at which it is important to note that the latter always refers to an analysis made at discrete (usually 3 selected) wavelengths, and not over a continuous range. (For precise definition of optical design concepts and image-evaluation diagrams see [96].) In addition to the new DOE surface-model, I will also use in this

* In the present implementation of zone decomposition, the values of l and OPL_0 cannot be modified directly. However, their change is possible by properly setting the parameters f_0 , f_2 , f_4 and h_0 , h_2 , h_4 , h_6 .

† The rearranged form of (V.7) is: $f_{\text{doe}}(x, y) = [- (n_2 \cdot \cos(\Theta_2) - n_1 \cdot \cos(\Theta_1)) \cdot \cos(\varphi)]^{-1} \cdot H'(x, y)$. Note that intersecting the surface at an (x, y) point with rays of different angles can result in different $f_{\text{doe}}(x, y)$ profiles for the same $H'(x, y)$ function.

subsection two other software extensions, those I wrote for plotting the relief-profile of DOEs and the OPD curves of hybrid systems.

5.1 Lens 1: achromatic IR doublet

This germanium doublet (shown in Fig. V.10) has been designed as a demonstrative example. The system performs diffraction-limited, infinite-conjugate on-axis imaging in the 3-5 μm wavelength range; the design wavelength of the diffractive component is 4 μm . The effective focal length of the system is 100 mm, the relative aperture is $f/2.0$ (NA 0.25). The substrate $z = f_{\text{sub}}(r)$ of the DOE is a spherical surface, plus 4th- and 6th-order aspheric terms. The surface-relief profile $z = f_{\text{doe}}(r)$ is depicted in Fig. V.11, it is described by a 4th-order polynomial. The step-heights of relief edges are uniformly 1.322 μm . For the calculation of $f_{\text{doe}}(r)$ I used (V.6) without any corrections, i.e. using only f_0 in $f(r)$.

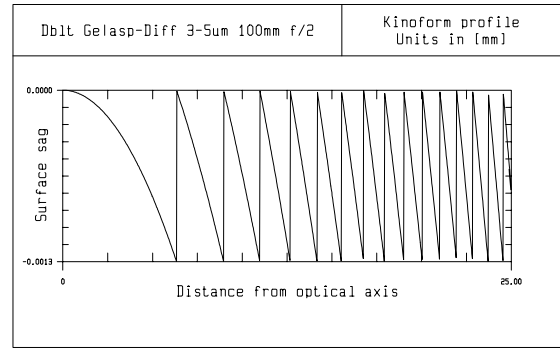
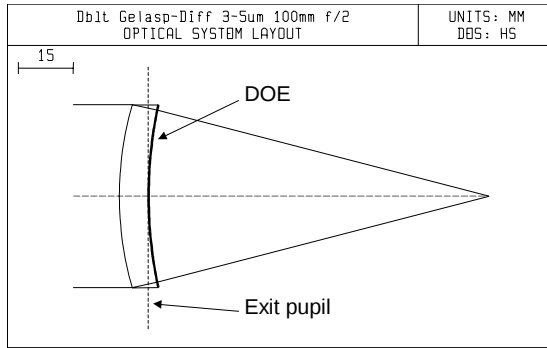


Fig. V.10. System layout of the germanium hybrid doublet. **Fig. V.11.** Surface relief profile of the germanium doublet.

Fig. V.12 shows the wavefront-shape in the exit-pupil at the design-wavelength. The saw-tooth structure of the OPD curve results from the fact that higher terms have not been added to $f(r)$ (see also Subsection V.4.2). The larger steps in Fig. V.13 are caused by that this curve has been calculated at a wavelength different from the nominal one. Dash-dot curves show in these figures the OPD diagram of the same hybrid system using the 1st diffraction order ray-tracing model of the DOE (i.e. the standard DOE model of commercial lens design programmes).

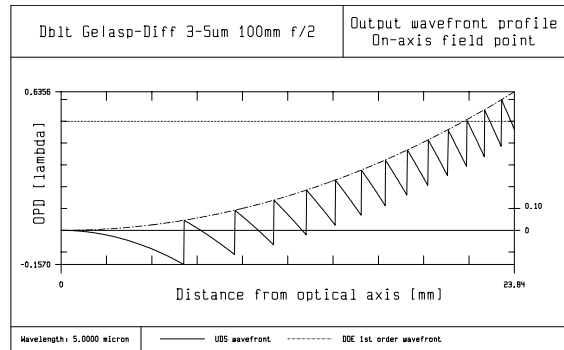
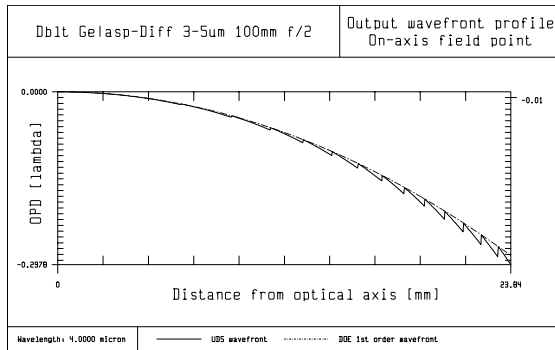


Fig. V.12. Output wavefront-profile of the germanium hybrid doublet. Solid line shows the OPD curve calculated by the UDS model of DOEs at the design wavelength, and **Fig. V.13.** Output wavefront-profile of the germanium hybrid doublet. Solid line shows the OPD curve calculated by the UDS model of DOEs at $\lambda = 3 \mu\text{m}$, and dash-dot line

dash-dot line shows the same in case of 1st diffraction order ray-tracing. ($\lambda = 4 \mu\text{m}$.)

Fig. V.14 shows the polychromatic PSF curves of the hybrid system calculated by the UDS model and 1st-order ray-tracing. The Strehl-ratio is 0.670 in the case of the UDS model, and 0.699 in the case of 1st-order ray-tracing; the difference is more than 4%. Fig. V.15 depicts the Encircled Energy plots for both DOE models; the large difference implies the presence of higher, strongly defocused diffraction orders that transport a significant portion of the total energy of the output beam.

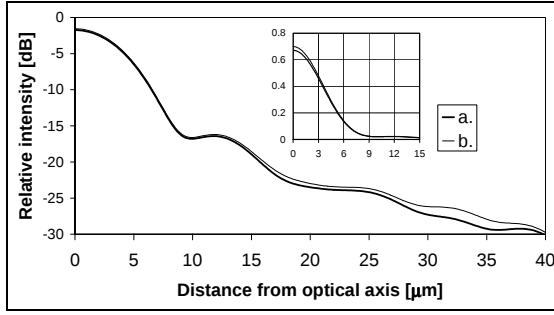


Fig. V.14. Intensity distribution in the focal plane of the germanium hybrid doublet. Thick line (a.) shows the polychromatic PSF curve calculated by the UDS model of DOEs, and thin line (b.) shows the same in the case of 1st diffraction order ray-tracing. (Inset shows the plot with linear intensity scale.)

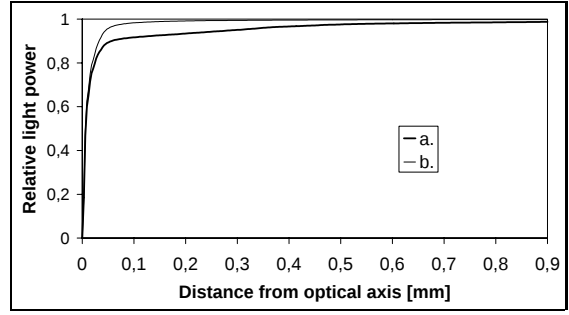


Fig. V.15. Radial energy distribution in the focal plane of the germanium hybrid doublet. Thick line (a.) shows the polychromatic Encircled Energy curve calculated by the UDS model of DOEs, and thin line (b.) shows the same in the case of 1st diffraction order ray-tracing.

Fig. V.16 shows polychromatic MTF diagrams, in which the zero-spatial frequency peak of the curve calculated by the UDS model reveals the presence of defocused higher diffraction orders (especially the 0th and the 2nd ones). The shape of the curve is in good agreement with the predictions of MTF scaling with diffraction efficiency presented in [28], [41] and [88].

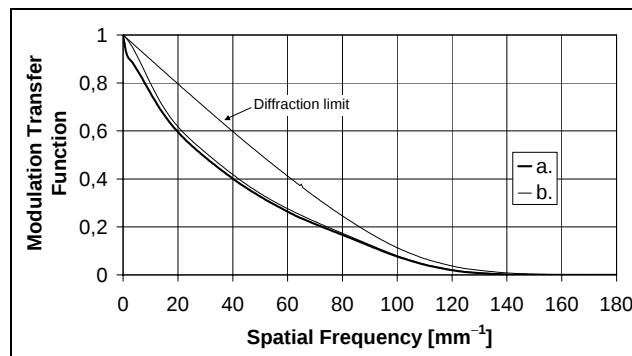


Fig. V.16. Modulation Transfer Function of the germanium hybrid doublet. Thick line (a.) shows the polychromatic MTF curve calculated by the UDS model of DOEs, and thin line (b.) shows the same in the case of 1st diffraction order ray-tracing.

5.2 Lens 2: visible-range apochromatic quadruplet

The next example shows the analysis of a commercially available hybrid system, viz. an apochromatic quadruplet.* The system performs diffraction-limited, infinite-conjugate on-axis imaging in the 480-780 nm wavelength range. The effective focal length of the system is 21.1 mm; the relative aperture is $f/1.7$ (NA 0.29). The substrate $\mathcal{Z} = f_{\text{sub}}(r)$ of the DOE is a plane surface; the surface-relief profile $\mathcal{Z} = f_{\text{doe}}(r)$ is described by an 8th-order polynomial. For the calculation of $f_{\text{doe}}(r)$ I used (V.6) without any corrections, i.e. using only f_0 in $f(r)$.

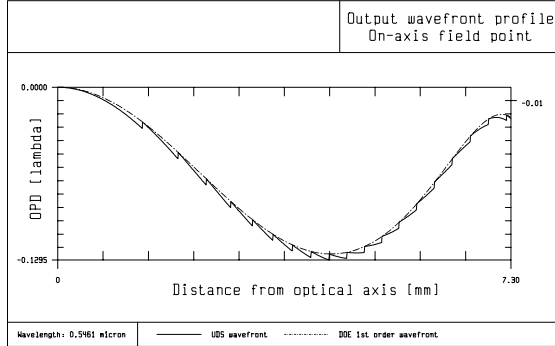


Fig. V.17. Output wavefront-profile of the apochromatic hybrid quadruplet. Solid line shows the OPD curve calculated by the UDS model of DOEs at $\lambda = 546.1$ nm (close to the design wavelength), and dash-dot line shows the same in the case of 1st diffraction order ray-tracing.

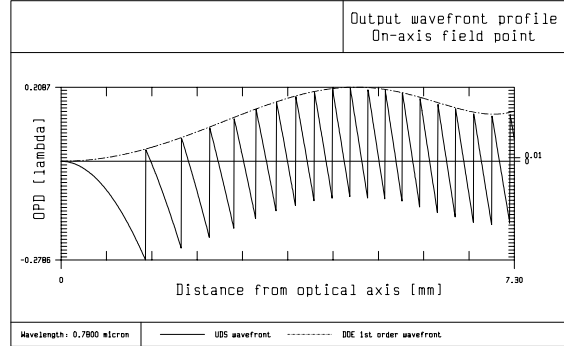


Fig. V.18. Output wavefront-profile of the apochromatic hybrid quadruplet. Solid line shows the OPD curve calculated by the UDS model of DOEs at $\lambda = 780$ nm, and dash-dot line shows the same in the case of 1st diffraction order ray-tracing.

Fig. V.17 shows the wavefront-shape in the exit-pupil at a wavelength slightly different from the nominal one, so that the saw-tooth structure of the curve becomes visible (precisely at the design wavelength the saw-tooths disappear since all rays that traverse the DOE surface are almost parallel to the optical axis). The larger steps in Fig. V.18 are caused by that this curve has been calculated at a wavelength significantly different from the nominal one. Dash-dot curves show the OPD diagram of the same hybrid system using the 1st diffraction order ray-tracing model of the DOE.

* Melles Griot® APO014 Dapromat™ – Melles Griot was so kind as to provide us with the layout of the lens (exclusively for academic research purposes). Due to the confidentiality agreement between the company and IOTA, detailed information about the lens cannot be given here for understandable proprietary reasons.

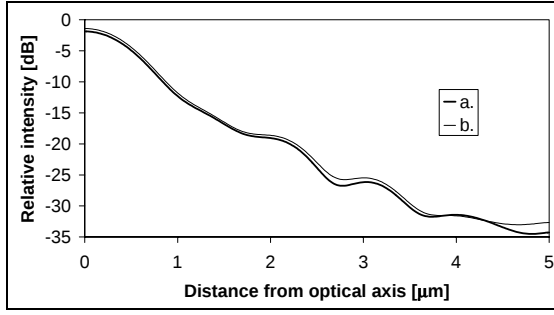


Fig. V.19. Intensity distribution in the focal plane of the apochromatic hybrid quadruplet. Thick line (a.) shows the polychromatic PSF curve calculated by the UDS model of DOEs, and thin line (b.) shows the same in the case of 1st diffraction order ray-tracing.

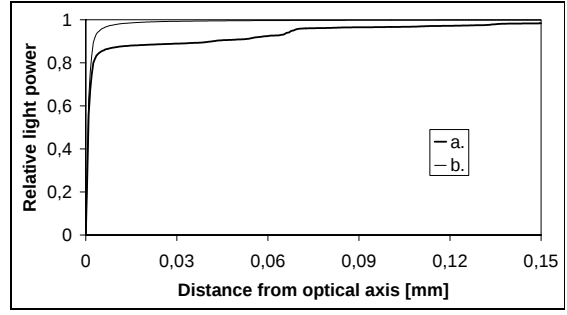


Fig. V.20. Radial energy distribution in the focal plane of the apochromatic hybrid quadruplet. Thick line (a.) shows the polychromatic Encircled Energy curve calculated by the UDS model of DOEs, and thin line (b.) shows the same in the case of 1st diffraction order ray-tracing.

Fig. V.19 shows the polychromatic PSF curves of the hybrid system calculated by the UDS model and 1st diffraction order ray-tracing. The Strehl-ratio is 0.650 in case of the UDS model, and 0.724 in case of 1st-order ray-tracing; the difference is more than 10%. Fig. V.20 depicts the Encircled Energy plots in case of the two different DOE models; the large difference implies the presence of higher, strongly defocused diffraction orders that transport a significant portion of the total energy of the output beam. Fig. V.21 depicts polychromatic MTF diagrams, in which the zero-spatial frequency peak of the curve calculated by the new model shows the presence of defocused higher diffraction orders.

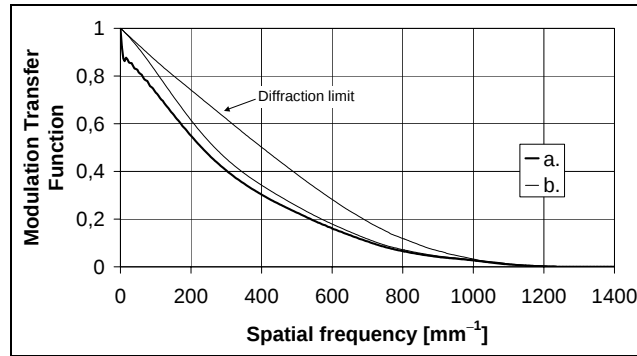


Fig. V.21. Modulation Transfer Function of the apochromatic hybrid quadruplet. Thick line (a.) shows the polychromatic MTF curve calculated by the UDS model of DOEs, and thin line (b.) shows the same in the case of 1st diffraction order ray-tracing.

5.3 Lens 3: fast achromatic IR doublet

A large advantage of the new DOE model is that the additional parameters of the function $f(r)$ provide useful degrees of freedom in the definition of the relief profile. Below I show a simple example for benefitting from the surface parameters $f_0, f_2, f_4 \dots$ in the design of a hybrid system.

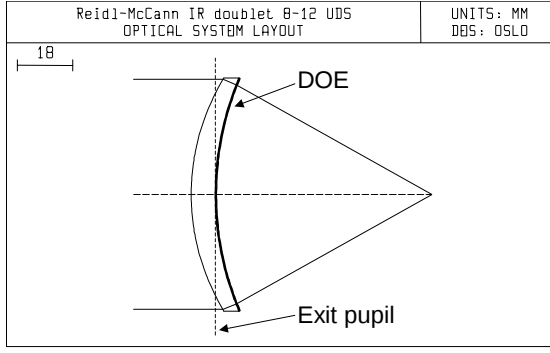


Fig. V.22. System layout of the fast hybrid doublet.

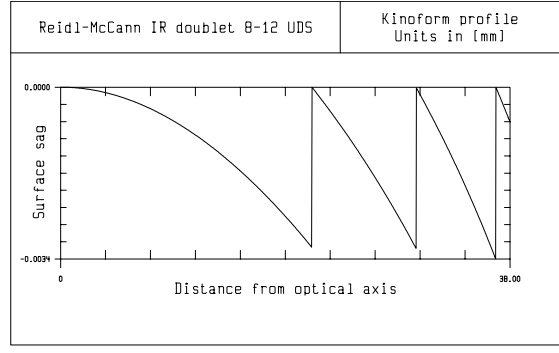


Fig. V.23. Relief-profile of the DOE of the fast hybrid doublet after optimisation (Case III).

The germanium lens shown in Fig. V.22 performs diffraction-limited, infinite-conjugate on-axis imaging in the 8-12 μm wavelength range [82, 92]; the design wavelength (λ_0) of the diffractive component is 10 μm . The effective focal length of the system is 77 mm, the f-number is f/0.9 (NA 0.49). The input surface of the system is a rotation-ellipsoid plus a 6th-order asphere; the substrate $\tilde{z} = f_{\text{sub}}(r)$ of the DOE is a spherical surface; the characteristic function $H(r)$ is given by a 2nd-order polynomial. Below, I compare the optical characteristics (Strehl ratio) of this hybrid system in three cases, in all of which the DOE relief profile was calculated by different methods. The relief profile shown in Fig. V.23 corresponds to Case III (explanation see below).

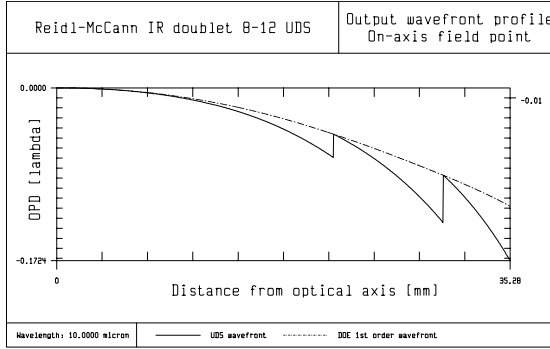


Fig. V.24. Output wavefront-profile of the fast hybrid doublet before optimisation (Case I). Solid line shows the OPD curve calculated by the UDS model of DOEs at the design wavelength, and dash-dot line shows the same in the case of 1st diffraction order ray-tracing.

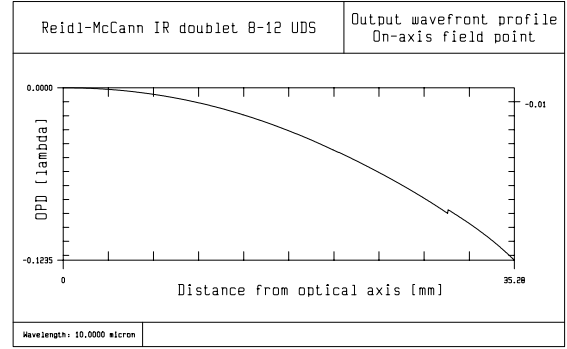


Fig. V.25. Output wavefront-profile of the fast hybrid doublet after optimisation (Case II).

For the determination of the function $f_{\text{doe}}(r)$ first I calculated the parameters of the function $f(r)$ by direct usage of (V.6) (Case I). The resulting output wavefront-profile can be seen in Fig. V.24, calculated at the nominal wavelength. The large steps in the curve show that the surface-relief does not realise $H'(r)$ correctly – this is the case when the higher-order terms of $f(r)$ should be used.

To minimise the OPL steps I slightly changed the f_2 and f_4 parameters in $f(r)$ by calculating their values analytically, using the following approximation of (V.7) (Case II):

$$H'(r) \approx -(n_2 - n_1) \cdot f_{\text{doe}}(r) \cdot \cos \varphi, \quad (\text{V.24})$$

where $r^2 = x^2 + y^2$. The explanation of this approximation can be understood from Fig. V.22: in case of this lens all the rays travel about 90° degrees relative to the DOE surface, thus their angle of incidence and refraction can be approximated by zero (i.e. $\Theta_2 \approx 0$ and $\Theta_1 \approx 0$). From Eq. (V.24) it follows:

$$f_{\text{doe}}(r) \approx \frac{H'(r)}{n_1 - n_2} \cdot \frac{1}{\cos \varphi(r)} = \frac{H'(r)}{n_1 - n_2} \cdot \frac{1}{\sqrt{1 + \left(\frac{r}{R}\right)^2}} \approx \frac{H'(r)}{n_1 - n_2} \cdot \left(1 + \frac{1}{2 \cdot R^2} \cdot r^2 - \frac{1}{2 \cdot R^4} \cdot r^4\right), \quad (\text{V.25})$$

where R is the radius of curvature of the DOE substrate ($R = 97.06$ mm; $n_1 = 4.003227$ at $\lambda = 10$ μm ; $n_2 = 1$). Comparing the above formula with (V.21) the value of the parameters in the $f(r)$ function are:

$$f_0 = \frac{1}{n_1 - n_2} \quad ; \quad f_2 = \frac{1}{(n_1 - n_2) \cdot 2 \cdot R^2} \quad ; \quad f_4 = -\frac{1}{(n_1 - n_2) \cdot 8 \cdot R^4}. \quad (\text{V.26})$$

The concrete values of these parameters are listed in Tab. V.1, together with the monochromatic and polychromatic Strehl ratios, while the output wavefront profile is depicted in Fig. V.25. The validity of the above approximation is confirmed by the evaluation results: the Strehl ratio calculated at the nominal wavelength differs only with 1‰ from the same value of first diffraction order modelling, as well as their output wavefront profiles almost perfectly coincide. The usefulness of this calculation manifests only at the nominal wavelength: though the monochromatic Strehl ratio increased with 3%, the polychromatic Strehl ratio decreased relative to that of Case I.

	UDS, Case I	UDS, Case II	UDS, Case III	First diffraction order
f_0	0.333	0.333	0.333	-
f_2	0	$1.767 \cdot 10^{-5} \text{ mm}^{-2}$	$-6.487 \cdot 10^{-5} \text{ mm}^{-2}$	-
f_4	0	$-4.690 \cdot 10^{-10} \text{ mm}^{-4}$	$5.062 \cdot 10^{-8} \text{ mm}^{-4}$	-
Strehl ratio (at λ_0)	0.921	0.952	0.903	0.953
Strehl ratio (polychr.)	0.889	0.874	0.900	0.979

Tab. V.1. Parameters of the function $f(r)$ before and after optimisation.

To increase the polychromatic Strehl ratio, I used the optimisation capabilities of OSLO SIX to determine the coefficients in $f(r)$ (Case III). The error function was the polychromatic Strehl ratio itself, while the variables were f_2 and f_4 . The output wavefront profile is plot in Fig. V.26. The results are summarised in Tab. V.1, from which it can be seen that the polychromatic Strehl ratio could be increased with 1% above that of Case I. Comparing Case III to first diffraction order PSF calculation reveals an 8% difference between the Strehl ratios, which again clearly shows the benefits of UDS modelling of diffractive surfaces over the conventional model.

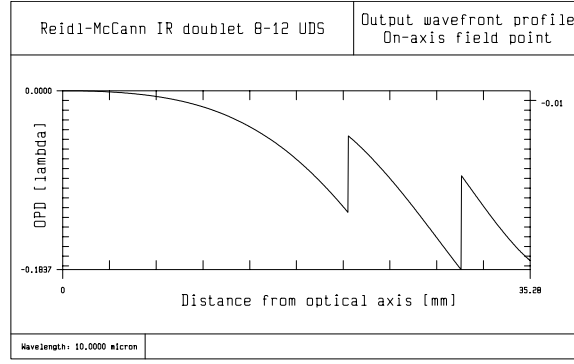


Fig. V.26. Output wavefront-profile of the fast hybrid doublet (at its design wavelength) after optimisation to increase the polychromatic Strehl ratio (Case III).

6 Verification

In this subsection, by demonstrative calculations, I confirm the correct operation of the zone decomposition DOE modelling method implemented in commercial lens design programmes as UDS surfaces. To perform the previous test one needs a reference method that is known to precisely describe the operation of DOEs. In Subsection V.3.1, it was presented that coherent summation of infinite number of diffraction orders theoretically results in exactly the output field of a DOE. In my numerical computations, I chose this method as a reference, in which series truncation means one of the error sources. Since a series given by (V.12) has a very slow convergence [88], a large number of orders must be summed up to obtain precise results. In the below calculations, I use the $m = -49..51$ diffraction orders, which number is sufficiently large to be accurate at verifying the UDS model, while small enough to keep computation time within executable limits. (The -49 and 51 orders have been chosen to be symmetric to the first one in order to have about the same diffraction efficiency.) The other error source in multiple diffraction order summation is the neglect of material dispersion inside the DOE. As the following computation will point out, the dispersive effects may be significant at non-nominal wavelengths in case of surface-relief phase-modulation DOEs.

For the verification of the UDS model, I use a test lens system, in which the DOE is the last surface on a plane substrate that coincides with the aperture stop, thus the exit pupil. In such a configuration, my implementation of zone decomposition theoretically results in the real output field of the system, since no Fresnel-diffraction can occur from the DOE to the exit pupil. In the same configuration, multiple order summation should give the same results with an insignificant error resulting from series truncation. I confirm the correct operation of the UDS model by comparing its results to those of order decomposition. For evaluation of the optical behaviour of the test system, I use the diffraction Encircled Energy (or Radial Energy Distribution) curve. This is more advantageous than direct application of the PSF curve in that this latter function consists of high spatial frequency diffraction and interference fringes, which makes the comparison of PSFs very complicated. By

contrast, the Encircled Energy contains all information about the focal plane intensity distribution and eliminates its sharply-changing nature.

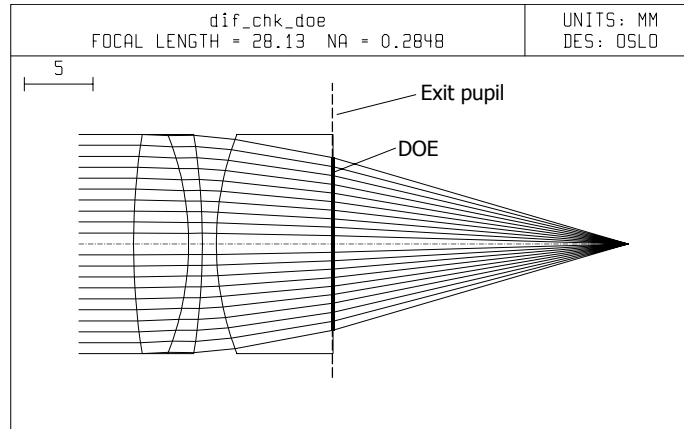


Fig. V.27. System layout of the test lens.

The test system I use is an achromatic hybrid quadruplet (designed for the purpose of the present computation) shown in Fig. V.27 (see also Tab. V.2) that performs diffraction limited infinite conjugate imaging at 540 nm and 780 nm according to first diffraction order ray-tracing. 540 nm is the design wavelength of the DOE, while 780 nm is the evaluation wavelength, which has been chosen intentionally far from the nominal one so that the effects of higher orders on the focus spot become characteristic. The Encircled Energy curves have been calculated at 780 nm in the focal plane of the first diffraction order. So that the results of multiple diffraction order summation and that of the UDS model could be compared to each other, the effects of material dispersion inside the DOE relief must be taken into account. This is essential, since the neglect causes a relative error between the two DOE modelling methods of about 3%, which is one order of magnitude higher than that caused by series truncation (see the end of this subsection). To eliminate this error I used the *evaluation* wavelength refractive index of the substrate material to calculate f_{doe} . In order to take account of the inclination factor, introduced by (V.7), I also added higher order terms to the $f(r)$ polynomial; the value of the coefficients are listed in Tab. V.2.

	Radius of curvature [mm]	Thickness [mm]	Aperture radius [mm]	Glass material (Schott)
SRF 1	52.456822	4	8	BK7
SRF 2	-22.006691	1	8	SF11
SRF 3	-52.456822	1.023244	8	AIR
SRF 4	22.006691	8.5	8	BK7
SRF 5	-	0.015	6.3	PMMA
SRF 6	-	21.514163	6.3	AIR (DOE)

	h_2 [mm ⁻¹]	h_4 [mm ⁻³]	h_6 [mm ⁻⁵]
$H(r)$ coefficient	-2.2394550690e-004	2.5406435099e-006	-1.7103034439e-008

	f_0 [-]	f_2 [mm ⁻²]	f_4 [mm ⁻⁴]
--	-----------	---------------------------	---------------------------

$f(r)$ coefficient	2.0662614852	-1.8841345807e-003	1.5073076646e-005
--------------------	--------------	--------------------	-------------------

Tab. V.2. Lens data of the test system.

To calculate the Encircled Energy curve, first the complex field distribution has to be determined in the exit pupil. For this calculation I used OSLO SIX in both the case of the UDS model and diffraction order decomposition. The second step is to compute the PSF by performing a diffraction integral over the exit pupil. To obtain the PSF curve does not impose any questions in case of the UDS model, while the summing up of diffraction orders offers two ways that are theoretically equivalent, however the realisation of which is rather different:

- a) first the complex field distribution is calculated over the exit pupil for each diffraction orders, second the PSF of the separate orders are calculated, third the curves are summed up coherently;
- b) first the complex field distribution is calculated over the exit pupil for each diffraction orders, second the results are summed up coherently for the separate orders, third the PSF is calculated.

Both methods has certain disadvantages. Method a) requires the calculation of a large number of PSF functions, which is very time consuming. The realisation of the method is also problematic, since the higher diffraction orders are strongly defocused and aberrated, and the PSF of such beams cannot be computed by all commercial lens design programmes. Method b) requires only one PSF calculation, however its direct implementation is presently not possible within commercial lens design programmes. As presented in Section IV.6, I prepared a high-precision diffraction routine for calculation of 2D diffraction patterns of rotationally symmetric beams (see Appendix A, B and C). Since this routine is rather slow at its present stage, I had to choose method b) in my computations for better time efficiency.

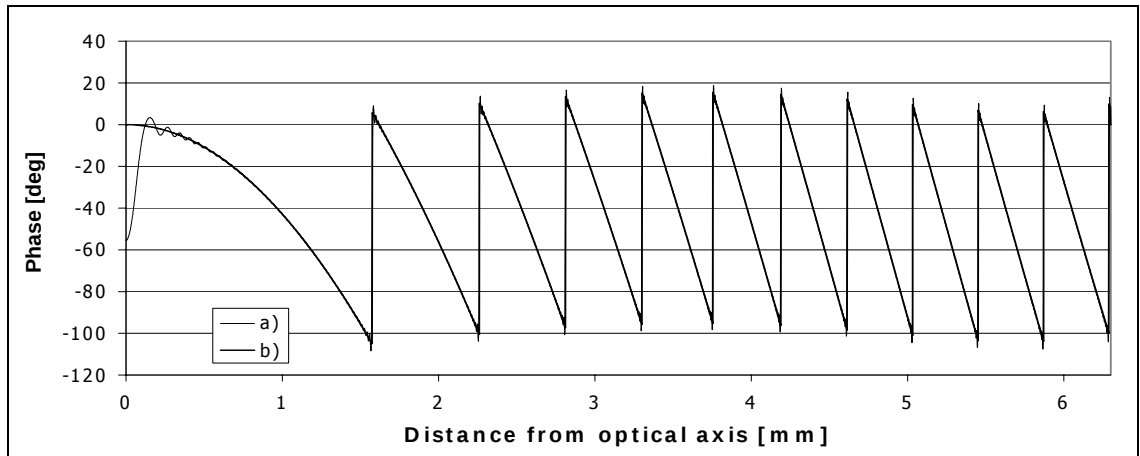


Fig. V.28. Output wavefront profile of the test lens at 780 nm. Thin line shows the phase calculated by summing up the $m = -49..51$ diffraction orders (a), thick line shows that in case of the UDS model (b).

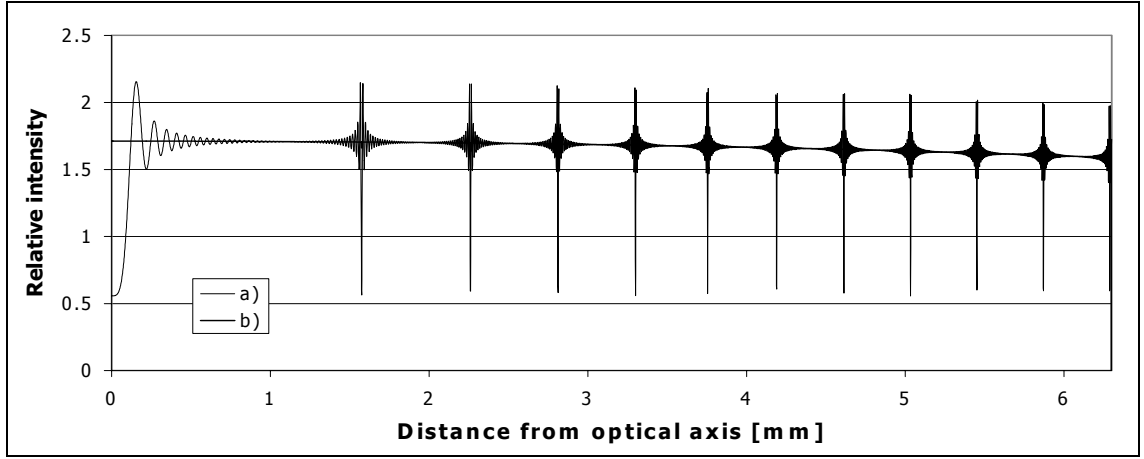


Fig. V.29. Intensity distribution over the exit pupil of the test lens. Thin line shows intensity calculated by summing up the $m = -49..51$ diffraction orders (a), thick line shows that in case of the UDS model (b).

The output wavefront profiles can be seen in Fig. V.28, while the shape of the output intensity distributions are depicted in Fig. V.29. The phase in Fig. V.28 is defined as follows:*

$$\Delta\Phi(r) = \frac{\text{OPL}_{\text{ref}}(r) - \text{OPL}(r)}{\lambda} \cdot 360^\circ, \quad (\text{V.27})$$

where λ is the evaluation wavelength, $\text{OPL}(r)$ is the optical path length measured from the entrance pupil to point r of the exit pupil, and $\text{OPL}_{\text{ref}}(r)$ is the optical path length of the reference spherical wave† measured at the same point. At phase jumps in the UDS profile, $n \cdot 360^\circ$ values (n is an integer) have been removed to constrain the jump between -180° and 180° . At drawing the intensity curves in Fig. V.29 I assumed unity intensity in the entrance pupil (of radius 8.012 mm). For an explanation of the high spatial frequency oscillations in the wavefront and intensity curves in Figs. V.28 and V.29, see Appendix D.

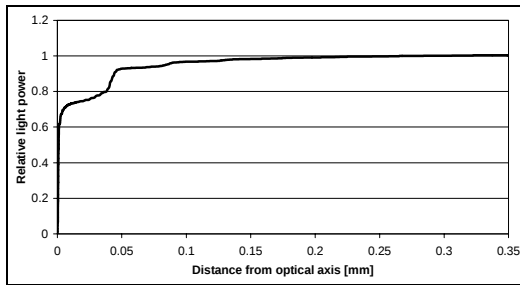


Fig. V.30. Radial energy distribution of the test lens calculated by summing up the $m = -49..51$ diffraction orders.

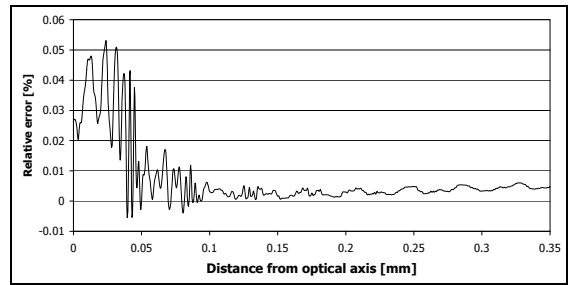


Fig. V.31. Error of the Radial energy distribution curve of the test lens calculated by the UDS model relative to that calculated by summing up the $m = -49..51$ diffraction orders

* This OPD sign convention is used in OSLO SIX.

† The reference is a perfect spherical wave centered on the intersection point of the image plane and the optical axis.

orders.

In my calculations, I compared the radial energy distribution (or Encircled Energy) curves – computed by the UDS model and multi order summation – by determining the relative error of their difference. As an error limit to declare the instance of coincidence between the curves I chose 3%, the error that occurred when material dispersion inside the DOE relief was neglected in my calculations (this is always the case for the standard DOE model of optical design softwares). The Encircled Energy Curve for the UDS model can be seen in Fig. V.30, while the relative error of multi order summation is presented in Fig. V.31. Since the maximum of the error is less than 1%, I can claim that the two curves coincide (it is most probable that truncation of the series expansion accounts for the small remaining error).

7 Conclusion

In this chapter of my dissertation, I propose a method for the efficient implementation of the zone decomposition concept that has been developed for the optical evaluation of diffractive optical elements (DOEs) in order to obtain a more accurate model than those presently available in commercial optical design programmes. My work was particularly aimed at creating a new surface model of DOEs that can be built into commercial design programmes to provide optical engineers with efficient means for the precise analysis and design of hybrid systems. I realised this model for the case of the surface-relief structure of phase-modulation computer-generated DOEs, which is nowadays of most practical importance, and built it into two of the most wide-spread lens design programmes, CODE V and OSLO SIX. The efficiency of my method lies in the fact that it eliminates the need to perform any calculations in addition to the standard computation made by commercial design programmes, which would otherwise be necessary when the optical performance of a hybrid system must be known to high precision. Further benefits of my design method include the following four items. First, it can be applied even for those DOEs when the standard model of commercial design programmes cannot be used at all (e.g. variable phase-step DOEs). Second, it is useful when PSF and MTF scaling provide inaccurate results (e.g. in the case of closely focused diffraction orders of low zone number DOEs). Third, by its help, the effect of scattered light caused by non-nominal diffraction orders can be taken into account in the design process. Fourth, it offers the only means when directly the relief profile has to be optimised in order to improve imaging quality. In my dissertation I demonstrate the usage and benefits of the new surface model by presenting several design examples for its application.

VI GENERAL CONCLUSIONS

For the sake of efficiency, today's optical designers use commercial programmes to analyse and optimize optical systems. These commercial design programmes provide very convenient engineering tools for the design of accustomed optics (such as e.g. spherical refractive/reflective systems), however they show certain deficiencies when used in case of some modern, presently exotic optical systems. During my Ph.D. work I prepared extensions to such programmes in order to make them capable of the modelling of planar anisotropic waveguide lenses, prism couplers to planar waveguides, Gaussian-to-uniform laser beam shapers and hybrid diffractive/conventional systems, as well as I developed new considerations for the design of prism couplers and all-spherical beam shapers. By usage of my new routines and design considerations and I designed optical systems of improved optical quality.

REFERENCES*

- 16 C. C. Aleksoff, K. K. Ellis and B. D. Neagle, "Holographic conversion of a Gaussian beam to a near field uniform beam", *Optical Engineering*, Vol. 30, No. 5, pp. 537-543, 1991.
- 17 M. Barabás and G. Szarvas, "Fourier description of the propagation and focusing of an extraordinary beam in a planar uniaxial medium", *Applied Optics*, Vol. 34, No. 1, pp. 11-21, 1995.
- 18 G. P. Bava, P. Rosina and I. Montrosset, "Numerical analysis of planar Fresnel lenses", *Journal of Modern Optics*, Vol. 35, No. 6, pp. 863-869, 1988.
- 19 J. Bengtsson, "Kinoform-only Gaussian-to-rectangle beam shaper for a semiconductor laser", *Applied Optics*, Vol. 35, No. 20, pp. 3807-3813, 1996.
- 20 C. G. Blough and G. M. Morris, "Hybrid lenses offer high performance at low cost", *Laser Focus World*, Vol. 31, No. 11, pp. 67-74, 1995.
- 21 C. G. Blough, M. Rossi, S. K. Mack and R. L. Michaels, "Single-point diamond turning and replication of visible and near-infrared diffractive optical elements", *Applied Optics*, Vol. 36, No. 20, pp. 4648-4654, 1997.
- 22 M. Born and E. Wolf, *Principles of Optics*, Pergamon Press, 1980.
- 23 J. Brunol and P. Chavel, "Fourier transformation of rotationally invariant two-variable functions: computer implementation of Hankel transformation", *Proc. of IEEE*, July 1977.
- 24 D. A. Buralli and G. M. Morris, "A flat-field diffractive landscape lens: design and performance", *SPIE Vol. 1136 Holographic Optics II: Principles and Applications*, pp. 134-139, 1989.
- 25 D. A. Buralli and G. M. Morris, "Design of a wide field diffractive landscape lens", *Applied Optics*, Vol. 28, No. 18, pp. 3950-3959, 1989.
- 26 D. A. Buralli and G. M. Morris, "Design of diffractive singlets for monochromatic imaging", *Applied Optics*, Vol. 30, No. 16, pp. 2151-2158, 1991.
- 27 D. A. Buralli and G. M. Morris, "Design of two- and three-element diffractive Keplerian telescopes", *Applied Optics*, Vol. 31, No. 1, pp. 38-43, 1992.
- 28 D. A. Buralli and G. M. Morris, "Effects of diffraction efficiency on the modulation transfer function of diffractive lenses", *Applied Optics*, Vol. 31, No. 22, pp. 4389-4396, 1992.

* The author's publications are listed separately on p. 9; references are indexed continuing that list (in alphabetical order).

- 29 D. A. Buralli, G. M. Morris and J. R. Rogers, "Optical performace of holographic kinoforms", *Applied Optics*, Vol. 28, No. 5, pp. 976-983, 1989.
- 30 W. S. C. Chang, S. Forouhar, J.-M. Delavaux and R.-X. Lu, "Fabrication and performance of diffraction lenses", *Proc. of SPIE*, Vol. 321, Integrated Optics II, pp. 9-14, 1982.
- 31 P. Chavel, G. Narcy and H. Sauer, "Should hybrid diffractive optics always be modeled using an expansion into orders?", *Proc. of OSA Annual Meeting*, Rochester, 20-24 October 1996.
- 32 Committe on Optical Science and Engineering, Board on Physics and Astronomy, National Materials Advisory Board, Commission on Physical Sciences Mathematics and Applications, Commission on Engineering and Technical Systems, National Research Council, *Harnessing Light: Optical Science and Engineering for the 21st Century*, National Academy Press, Washington D.C., 1998.
- 33 Y. Danziger, E. Hasman, A. A. Friesem and A. W. Lohmann, "Multilevel diffractive elements for generalized wavefront shaping", *Optical Engineering*, Vol. 35, No. 9, pp. 2556-2565, 1996.
- 34 N. Davidson, A. A. Friesem and E. Hasman, "Analytic design of hybrid diffractive-refractive achromats", *Applied Optics*, Vol. 32, No. 25, pp. 4770-4774, 1993.
- 35 N. Davidson, R. Duer, A. A. Friesem and E. Hasman, "Blazed holographic gratings for polychromatic and multidirectional incidence light", *Journal of Optical Society of America*, Vol. 9, No. 7, pp. 1196-1199, 1992.
- 36 F. M. Dickey and S. C. Holswade, "Gaussian laser beam profile shaping", *Optical Engineering*, Vol. 35, No. 11, pp. 3285-3295, 1996.
- 37 T. Dresel, M. Beyerlein and J. Schwider, "Design of computer-generated beam-shaping holograms by iterative finite-element mesh adoption", *Applied Optics*, Vol. 35, No. 35, pp. 6865-6873, 1996.
- 38 M. Duparré, M. A. Golub, B. Lüdge, V. S. Pavelyev, V. A. Soifer, G. V. Uspleniev and S. G. Volotovskii, "Investigation of computer-generated diffractive beam shapers for flattening of single-modal CO₂ laser beams", *Applied Optics*, Vol. 34, No. 14, pp. 2489-2497, 1995.
- 39 M. T. Eismann, A. M. Tai and J. N. Cederquist, "Iterative design of a holographic beamformer", *Applied Optics*, Vol. 28, No. 13, pp. 2641-2650, 1989.
- 40 N. C. Evans and D. L. Shealy, "Design and optimisation of an irradiance shaping system with a genetic algorithm method", *Applied Optics*, Vol. 37, No. 22, pp. 5216-5221, 1998.
- 41 D. Faklis and G. M. Morris, "Spectral properties of multiorder diffractive lenses", *Applied Optics*, Vol. 34, No. 14, pp. 2462-2468, 1995.

- 42 M. B. Fleming and M. C. Hutley, "Blazed diffractive optics", *Applied Optics*, Vol. 36, No. 20, pp. 4635-4643, 1997.
- 43 B. R. Frieden, "Lossless conversion of a plane laser wave to a plane wave of uniform irradiance", *Applied Optics*, Vol. 4, No. 11, pp. 1400-1403, 1965.
- 44 J. W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill, 1968.
- 45 C. Y. Han, Y. Ishii and K. Murata, "Reshaping collimated laser beams with Gaussian profile to uniform profiles", *Applied Optics*, Vol. 22, No. 22, pp. 3644-3647, 1983.
- 46 Y. Han, L. N. Hazra and C. A. Delisle, "Exact surface-relief profile of a kinoform lens from its phase function", *Journal of Optical Society of America*, Vol. 12, No. 3, pp. 524-529, 1995.
- 47 L. N. Hazra, Y. Han and C. Delisle, "Plane kinoform lenses for axial stigmatism in finite conjugate imaging", *Optics Communications*, Vol. 91, No. 1,2, pp. 1-4, 1992.
- 48 L. N. Hazra, Y. Han and C. Delisle, "Curved kinoform lenses for stigmatic imaging of axial objects", *Applied Optics*, Vol. 32, No. 25, pp. 4775-4784, 1993.
- 49 L. N. Hazra, Y. Han and C. A. Delisle, "Geometric optical limits in the synthesis of planar kinoform lenses for axial stigmatism", *Optical Engineering*, Vol. 34, No. 5, pp. 1296-1302, 1995.
- 50 L. N. Hazra, O. F. Meknassi and C. A. Delisle, "Zone characteristics of planar kinoform lenses of any specific order for stigmatic imaging between extra-axial conjugate points", *Journal of Optical Society of America*, Vol. 12, No. 11, pp. 2463-2470, 1995.
- 51 H. P. Herzig, *Micro-Optics Elements Systems and Applications*, Taylor and Francis, 1997.
- 52 A. Ishikawa, M. Izutsu and T. Sueta, "Beam propagation method analysis of optical waveguide lenses", *Applied Optics*, Vol. 29, No. 34, pp. 5064-5068, 1990.
- 53 S. R. Jahan and M. A. Karim, "Refracting systems for Gaussian-to-uniform beam transformations", *Optics and Laser Technology*, Vol. 21, No. 1, pp. 27-30, 1989.
- 54 W. Jiang and V. M. Ristic, "Study of anisotropy effect in planar lenses for integrated optics", *Journal of Modern Optics*, Volume 35, No. 6, pp. 849 - 862, 1988.
- 55 W. Jiang and V. M. Ristic, "Definition of anisotropic aberrations in planar lenses", *IEEE Journal of Quantum Electronics*, Vol. QE-22, No. 10, pp. 1934-1935, 1986.
- 56 P. Jiang, P. J. R. Laybourn and G. C. Righini, "Homogeneous planar lens on an anisotropic waveguide", *Journal of Modern Optics*, Volume 39, No. 1, pp. 121 - 132, 1992.

- 57 W. Jiang, D. L. Shealy and J. C. Martin, "Design and testing of a refractive reshaping system", *Proc. of SPIE* 2000, pp. 64-75, 1993.
- 58 Z. Jiang, Q. Lu and Z. Liu, "Relative phase-shifts of apertured Gaussian beams and transformation of a Gaussian beam through a phase aperture", *Applied Optics*, Vol. 36, No. 4, pp. 772-778, 1997.
- 59 M. A. Karim, A. K. Cherri, A. A. S. Awwal and A. Basit, "Refracting system for annular laser beam transformation", *Applied Optics*, Vol. 26, No. 12, pp. 2446-2449, 1987.
- 60 M. A. Karim, A. M. Hanafi, F. Hussain, S. Mustafa, Z. Samberid and N. M. Zain, "Realization of a uniform circular source using a two dimensional binary filter", *Optics Letters*, Vol. 10, No. 10, pp. 470-471, 1985.
- 61 N. L. Kazansky, V. V. Kotlyar and V. A. Soifer, "Computer-aided design of diffractive optical elements", *Optical Engineering*, Vol. 33, No. 10, pp. 3156-3166, 1994.
- 62 M. Kuittinen, P. Vahimaa, M. Honkanen and J. Turunen, "Beam shaping in the nonparaxial domain of diffractive optics", *Applied Optics*, Vol. 36, No. 10, pp. 2034-2041, 1997.
- 63 R. E. Kunz and M. Rossi, "Phase-matched Fresnel elements", *Optics Comm.*, Vol. 97, No. 1,2, pp. 6-10, 1993.
- 64 P. J. R. Laybourn and G. C. Righini, "New design of thin-film lens", *Electronics Letters*, Vol. 22, No. 6, pp. 343-345, 1986.
- 65 P. J. R. Laybourn, G. Molesini and G. C. Righini, "Homogeneous refracting lenses for integrated optical circuits", *Journal of Modern Optics*, Volume 35, No. 6, pp. 1029 - 1048, 1988.
- 66 D. Lee, *Electromagnetic Principles of Integrated Optics*, John Wiley & Sons New York, 1986.
- 67 W. H. Lee, "Method for converting a Gaussian laser beam into a uniform beam", *Optics Communications*, Vol. 36, No. 6, pp. 469-471, 1981.
- 68 L. B. Lesem, P. M. Hirsch and J. A. Jordan Jr., "The Kinoform: A New Wavefront Reconstruction Device", *IBM J. Res. Develop.*, pp. 150-155, 1969.
- 69 C. Londoño, W. T. Plummer, P. P. Clark, "Athermalization of a single-component lens with diffractive optics", *Applied Optics*, Vol. 32, No. 13, pp. 2295-2302, 1993.
- 70 P. H. Malyak, "Two-mirror unobscured optical system for reshaping the irradiance distribution of a laser beam", *Applied Optics* 31(22), pp. 4377-4383, 1992. K. Nemoto, T. Nayuki, T. Fujii, N. Goto and Y. Kanai, "Optimum control of the laser beam intensity profile with a deformable mirror", *Applied Optics*, Vol. 36, No. 30, pp. 7689-7695, 1997.

- 71 J. H. McDermit and T. E. Horton, "Reflective optics for obtaining prescribed irradiative distributions from collimated sources", *Applied Optics*, Vol. 13, No. 6, pp. 1444-1450, 1974.
- 72 M. D. McNeill and T.-C. Poon, "Gaussian-beam profile shaping by acousto-optic Bragg diffraction", *Applied Optics*, Vol. 33, No. 20, pp. 4508-4515, 1994.
- 73 O. F. Meknassi, C. A. Delisle and L. N. Hazra, "Extra-axial stigmatic imagery with a thick planar kinoform", *Applied Optics*, Vol. 36, No. 26, pp. 6577-6582, 1997.
- 74 S. Misawa, M. Aoki, S. Fujita, A. Takaura, T. Kihara, K. Yokomori and H. Funato, "Focusing waveguide mirror with a tapered edge", *Applied Optics*, Vol. 33, No. 16, pp. 3365-3370, 1994.
- 75 M. D. Missig and G. M. Morris, "Diffractive optics applied to eyepiece design", *Applied Optics*, Vol. 34, No. 14, pp. 2452-2461, 1995.
- 76 K. Miyamoto, "The Phase Fresnel Lens", *Journal of the Opt. Soc. of America*, Vol. 51, No. 1, pp. 17-20, 1961.
- 77 Z. Nikolov and B. Pantchev, "Single aplanatic homogeneous refracting waveguide lenses without field curvature", *Optics Letters*, Vol. 17, No. 20, pp. 1429-1431, 1992.
- 78 Y. Ohtsuka and A. Tanone, "Acousto-optic intensity modification of a Gaussian laser beam", *Optics Communications*, Vol. 39, No. 12, pp. 70-74, 1981.
- 79 Optical Research Associates, *CODE V Reference Manual*, 1997.
- 80 G. Perrone and I. Montrosset, "A correction to the two-dimensional BPM for the analysis of waveguide lenses", *Journal of Modern Optics*, Vol. 39, No. 1, pp. 35-41, 1992.
- 81 D. A. Pommet, M. G. Moharam and E. B. Grann, "Limits of scalar diffraction theory for diffractive phase elements", *Journal of Optical Society of America*, Vol. 11, No. 6, pp. 1827-1834, 1994.
- 82 M. J. Reidl and J. T. McCann, "Analysis and performance limits of diamond-turned diffractive lenses for the 3-5 and 8-12 micrometer regions", *Proc. SPIE*, Vol. CR38, pp. 153-163, 1991.
- 83 P. W. Rhodes and D. L. Shealy, "Refractive optical systems for irradiance redistribution of collimated radiation: their design and analysis", *Opt. Soc. of America*, Vol. 19, No. 20, pp. 3545-3553, 1980.
- 84 G. C. Righini and G. Molesini, "Design of optical-waveguide homogeneous refracting lenses", *Applied Optics*, Vol. 27, No. 20, pp. 4193-4199, 1988.
- 85 M. Rossi, R. E. Kunz and H. P. Herzig, "Refractive and diffractive properties of planar micro-optical elements", *Applied Optics*, Vol. 34, No. 26, pp. 5996-6007, 1995.

- 86 F. S. Roux, "Intensity distribution transformation for rotationally symmetric beam shaping", *Optical Engineering*, Vol. 30, No. 5, pp. 529-536, 1991.
- 87 T. R. M. Sales and G. M. Morris, "Diffractive-refractive behaviour of kinoform lenses", *Applied Optics*, Vol. 36, No. 1, pp. 253-257, 1997.
- 88 H. Sauer, G. Narcy and P. Chavel, "Kinoform modeling for hybrid optical system design", *Proceedings of Diffractive Optics '97 Savonlinna*, pp. 174-175, 1997.
- 89 D. Shafer, "Gaussian to flat-top intensity distributing lens", *Optics and Laser Technology*, pp. 159-160, June 1982.
- 90 G. Sharp and A. Kathman, "Laser beam shapes for the future", *Industrial Laser Review*, No. 12, 1994.
- 91 A. E. Siegman, *Lasers*, University Science Books, Mill Valley, California, ISBN 0-935702-11-5.
- 92 Sinclair Optics, *OSLO SIX Reference Manual*, 1997.
- 93 S. Sinzinger and M. Testorf, "Transition between diffractive and refractive micro-optical components", *Applied Optics*, Vol. 34, No. 26, pp. 5970-5976, 1995.
- 94 R. S. Sirohi and M. P. Kothiyal, *Optical Components, Systems, and Measurement Techniques*, Marcel Dekker, 1991.
- 95 W. J. Smith, *Modern lens design: a resource manual*, McGraw-Hill Inc., 1992.
- 96 W. J. Smith, *Modern Optical Engineering: The Design of Optical Systems*, McGraw-Hill, 1990.
- 97 G. H. Spencer and M. V. R. K. Murty, "General ray-tracing procedure", *Journal of Optical Society of America*, Vol. 52, pp. 672-678, 1962.
- 98 R. M. Stevenson, M. J. Norman, T. H. Bett, D. A. Pepler, C.N. Danson and I. N. Ross, "Binary-phase zone plate arrays for the generation of uniform focal profiles", *Optics Letters*, Vol. 19, No. 6, pp. 363-365, 1994.
- 99 T. Stone and N. George, "Hybrid diffractive-refractive lenses and achromats", *Applied Optics*, Vol. 27, No. 14, pp. 2960-2971, 1988.
- 100 G. J. Swanson, "Binary Optics Technology: The Theory and Design of Multi-Level Diffractive Optical Elements", *MIT Lincoln Laboratory*, Technical Report 854, 1989.
- 101 G. J. Swanson, "Binary Optics Technology: Theoretical Limits on the Diffraction Efficiency of Multilevel Diffractive Optical Elements", *MIT Lincoln Laboratory*, Technical Report 914, 1991.

- 102 D. W. Sweeney and G. E. Sommargren, "Harmonic diffractive lenses", *Applied Optics*, Vol. 34, No. 14, pp. 2469-2475, 1995.
- 103 G. Szarvas, M. Barabás, P. Richter and L. Jakab, "Design of multielement acircular waveguide lens systems in anisotropic media", *Optical Engineering*, Vol. 32, No. 10, pp. 2510 - 2516, 1993.
- 104 X. Tan, B. Y. Gu, G. Z. Yang and B. Z. Dong, "Diffractive phase elements for beam shaping: a new design method", *Applied Optics*, Vol. 34, No. 8, pp. 1314-1320, 1995.
- 105 K. Tatsumi, T. Nakaguchi and S. Ito, "Wide field angle bi-aspherical waveguide lens in LiNbO₃ fabricated by proton exchange", *Electronics Letters*, Vol. 24, No. 9, pp. 546-548, 1988.
- 106 P. K. Tien and R. Ulrich, "Theory of Prism-Film Coupler and Thin-Film Light Guides", *Journal of the Optical Society of America*, Vol. 60, No. 10, pp. 1325-1337, 1970.
- 107 J. Turunen and F. Wyrowsky, *Diffractive Optics for Industrial and Commercial Applications*, John Wiley & Sons, 1998.
- 108 R. Ulrich and R. J. Martin, "Geometrical optics in thin film light guides", *Applied Optics*, Vol. 10, No. 9, pp. 2077-2085, 1971.
- 109 R. Ulrich, "Theory of the Prism-Film Coupler by Plane-Wave Analysis", *Journal of the Optical Society of America*, Vol. 60, No. 10, pp. 1337-1350, 1970.
- 110 W. B. Veldkamp and C. J. Kastner, "Beam profile shaping for laser radars that use detector arrays", *Applied Optics*, Vol. 21, No. 2, pp. 345-356, 1982.
- 111 J. M. Verdiell, M. A. Newkirk, T. L. Koch, R. P. Gnall, U. Koren, B. I. Miller and L. L. Buhl, "Aspheric waveguide lenses for photonic integrated circuits", *Applied Physics Letters*, Vol. 62, No 8, pp. 808-810, 1993.
- 112 A. P. Wood, "Design of infrared hybrid refractive-diffractive lenses", *Applied Optics*, Vol. 31, No. 13, pp. 2253-2258, 1992.
- 113 Z. D. Yu, "Waveguide optical planar lenses in LiNbO₃ - theory and experiments", *Optics Communications*, Vol. 47, No. 4, pp. 248-250, 1983.
- 114 U. D. Zeitner, H. Aagendal and F. Wyrowski, "Comparison of resonator-originated and external beam shaping", *Applied Optics*, Vol. 38, No. 6, pp. 980-986, 1999.
- 115 W. Zhou and V. M. Ristic, "Anisotropic aberrations in planar waveguide lenses", *IEEE Journal of Quantum Electronics*, Vol. 25, No. 4, pp. 749-754, 1989.

APPENDIX A – Users' guide to “rs_diffraction.exe”

The “rs_diffraction.exe” programme has been prepared in order to calculate the diffraction pattern of focused (and aberrated) light far from focus, and that of collimated (and aberrated) light inside the Fresnel-range or even closer to the diffracting aperture, in the scalar approximation. So as to achieve precise results, the programme directly implements the Rayleigh-Sommerfeld (RS) integration formula [44], without any further approximations. The accuracy of the computation is thus limited (within the field of validity of the RS formula) only by the number of sampling intervals applied in the numeric integration used by the programme (this can be set by the user). Whether the sampling is fine enough can be decided by thorough error analysis, for which an example is presented in Appendix B. (An estimation for the required sampling is given later in this appendix.) Integrating the output irradiance over an area that theoretically incorporates almost all the input power, and comparing this value to that of the known real input power, also indicates undersampling (this is a necessary but not sufficient condition).

The programme performs diffraction calculation on some monochromatic input field (of wavelength “lambda”) to calculate the complex amplitude of the diffracted output field at user-defined image points (of number “data_points”). The input field must be rotationally symmetric to the Z axis, and should be specified to the programme in a plane parallel to the XY directions; the complex amplitude distribution of the input field should be given along the Y axis. The integration is performed over the input plane inside a circular aperture of radius “a”. The phase of the input field must be given in degrees, relative to that of the reference wave (the phase is assumed to increase in the direction of wave propagation); the amplitude may be specified in arbitrary units (denoted by AU).

The reference wave is a perfect spherical wave of aperture radius “rfs_a”, focused at a distance of “rfs_p” from the input plane (called the geometrical focus). If $rfs_p = 0$, then the center of the reference wave is assumed to be at infinity, in which case the phase must be given by raw phase data on the input plane. The intensity (i.e. the absolute value of the Poynting vector on the square) of the perfect wave is constant on a sphere of “rfs_p” radius centered on the geometrical focus point (called the reference sphere), and is calculated so that the total power of the reference wave is the same as that of the input field (“total_power”). The numerical aperture of the reference wave is calculated from “rfs_p” and “rfs_a”, and is used to determine its on-axis intensity. The phase of the perfect wave is taken to be zero at the intersection of the plane and the Z axis.

The image points can be positioned in the ZY plane. The phase of the field in these points are given relative to the geometrical phase measured between the input plane center and the on-axis image point (i.e. $360^\circ \cdot z_0 / \lambda$). In case of perfect focusing, at the geometrical focus point ($z_0 = rfs_p$) the

diffraction phase is -90° relative to the geometrical phase, see the phenomenon of phase anomaly [22]. The calculated intensity is normalized to that of the reference wave in its geometrical focus (except for the case of infinite reference sphere radius, when output intensity is not normalized); this is also known as the Strehl normalization. The on-axis intensity of a perfect spherical wave, such as the reference wave in this case, is analytically calculated from the Rayleigh-Sommerfeld formula, to be presented below in (A.1). The position of image points can be specified two ways: by defining the radius of a circular aperture (“patch”) in an output plane over which the field is to be calculated (in this case data points are uniformly set along the Y axis of the output plane), or by a second input file (in this case data points can be positioned on any curve in the ZY plane). In both cases, the on-axis distance between the input plane and the coordinate system of the output field has to be defined by the user (“z0”). In order to tell the programme that data points are given by a file, the value of “patch” must be set to zero.

Input data must be given in an input text file (see a sample after rs_diff.exe), the name of which is prompted for after the programme has been started (full path definition required). The input file begins with 10 parameters (separated by EOL characters), constituting a kind of header structure:

- [0] number of sample intervals on the input plane “intervals”
- [1] number of data points on the output plane “data_points”
- [2] wavelength given in the medium of the image space [mm] “lambda”
- [3] distance between the input plane and the geometrical focus point [mm] “rfs_p”
- [4] aperture radius of the reference sphere on the input plane [mm] “rfs_a”
- [5] aperture radius on the output plane [mm] “patch”
- [6] on-axis distance between the input plane and the output coordinate system [mm] “z0”
- [7] total power of the input beam [AU²·mm²] “total_power”
- [8] name of the second input file with full path
- [9] name of the output file with full path.

Then the input field data follow: intervals+1 times four numbers, arranged in a 2D array. The items in each row are separated by one “space”, and each row ends with an end-of-line (“EOL”) character. The value of “a” is also calculated from the input file, as shown in Tab. A.1.

	[0] distance from optical axis [mm]	[1] intensity [AU ²]	[2] amplitude [AU]	[3] phase [degree]
[0]	$y_0 (= 0)$	$I_0 (= U_0^2)$	U_0	Φ_0
[1]	$y_1 (= a / \text{intervals})$	$I_1 (= U_1^2)$	U_1	Φ_1
[...]	$y_{...}$	$I_{...}$	$U_{...}$	$\Phi_{...}$
[intervals]	$y_{\text{intervals}} (= a)$	$I_{\text{intervals}} (= U_{\text{intervals}}^2)$	$U_{\text{intervals}}$	$\Phi_{\text{intervals}}$

Tab. A.1. Format of the input data in the input file of the rs_diff.exe computer routine.

The second input file (if specified, i.e. when patch $\neq 0$) contains a list of $(y; z)$ point pairs, given in [mm], having the same structure as that of the input file (data_points times 2 numbers). The input files must be terminated by an extra EOL. The output file format (data_points times 4 numbers) is exactly the same as that of the input file, except for the header, which is missing here.

The algorithm implemented in the rs_diff.exe routine bases on the Rayleigh-Sommerfeld diffraction formula that can be found on page 45 of Ref. [44] as equation (3-26):

$$U(P_0) = \frac{1}{i\lambda} \iint_{\Sigma} U(P_1) \frac{e^{ikr_{01}}}{r_{01}} \cos(\mathbf{n}, \mathbf{r}_{01}) ds, \quad (\text{A.1})$$

where 'U' is the complex amplitude of the field, P_0 is a point in the output plane, P_1 is a point in the input plane, λ is the wavelength, Σ denotes the aperture over which the integration is performed, 'k' is the absolute value of the wave vector, $\mathbf{r}_{01}$ is the vector from P_0 to P_1 , $\mathbf{n}$ is the surface normal of the input plane and ds is the infinitesimal surface area in Σ . These definitions are also given in Fig. A.1.

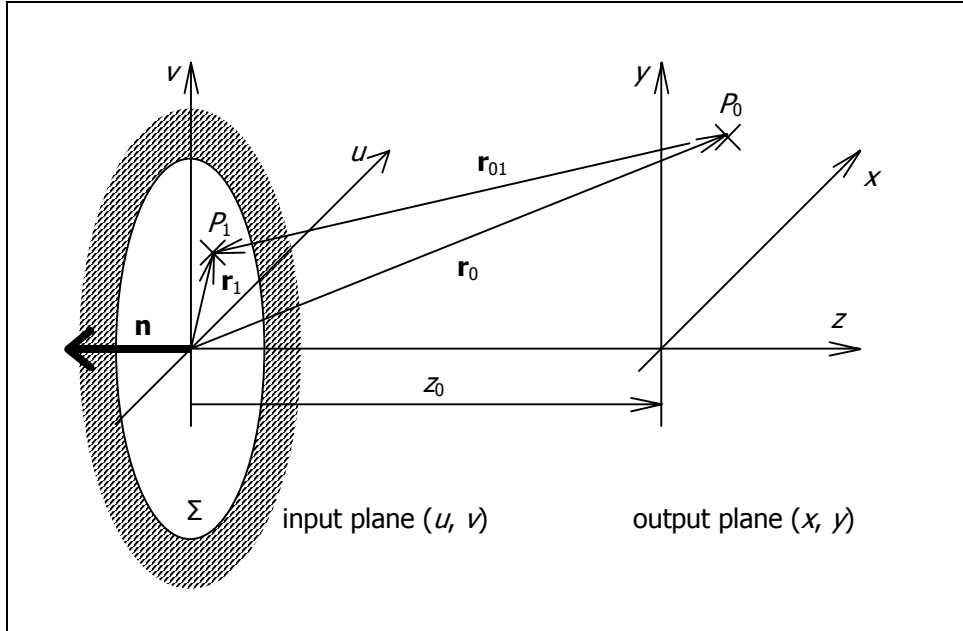


Fig. A.1. Definitions for the diffraction formula implemented in the rs_diff.exe computer routine.

Substituting the next expressions into the Rayleigh-Sommerfeld formula (A.1):

$$\begin{aligned} \mathbf{r}_{01} = \mathbf{r}_1 - \mathbf{r}_0 = (u - x, v - y, -z_0) \quad ; \quad \cos(\mathbf{n}, \mathbf{r}_{01}) = \mathbf{n} \cdot \mathbf{r}_{01} \cdot \frac{1}{r_{01}} = \frac{z_0}{r_{01}} \\ k = \frac{2\pi}{\lambda} \quad ; \quad \mathbf{n} = (0, 0, -1) \quad ; \quad ds = dudv, \end{aligned} \quad (\text{A.2})$$

we obtain the final form of the output field, which has been implemented in the rs_diff.exe routine:

$$U(P_0) = \frac{z_0}{i\lambda} \iint_{\Sigma} U(P_1) \frac{e^{i\frac{2\pi}{\lambda} \sqrt{(u-x)^2 + (v-y)^2 + z_0^2}}}{(u-x)^2 + (v-y)^2 + z_0^2} dudv. \quad (\text{A.3})$$

Fig. A.2 a) shows the sampling method I applied in the routine. The Σ aperture is divided into identical square-shaped $\Delta u \times \Delta v$ sized intervals, at the center of each one a sample $U(P_1)$ of the input field is taken. The infinitesimal “volume” contribution of this area to the integral is approximated by: $U(P_1) \cdot \Delta u \Delta v$. As was mentioned earlier, the input field is assumed to be rotationally symmetric, thus it is given only as a function of distance measured from the optical axis, see Fig. A.2 b). In two dimensions this means that the input field is known over circles of radii r_i ($i = 0 \dots$ “intervals”), concentric to the optical axis. In order to reduce computation time, the integration is performed only on one half of the Σ aperture, and the result is multiplied by two; correspondingly, the output field can be calculated only in the (y, z) plane.

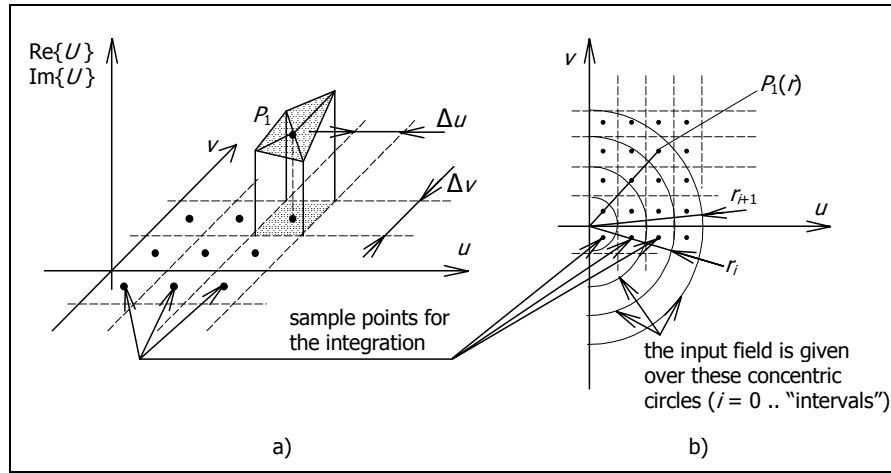


Fig. A.2. Sampling method applied in the rs_diffraction.exe computer routine.

To calculate U in each point $P_1(r)$, I apply linear interpolation on the real and imaginary values of U , using the adjacent sample $U(r_i)$ and $U(r_{i+1})$ values given by the input file, as shown in Fig. A.3.

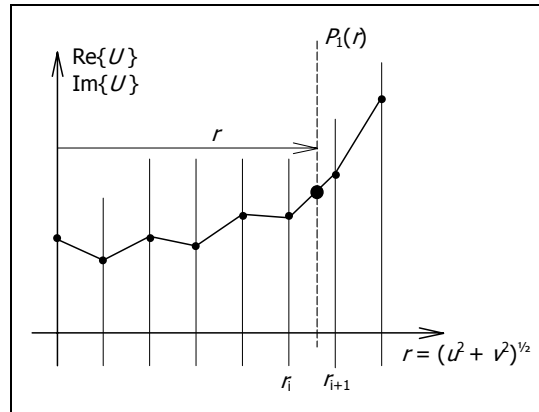


Fig. A.3. Interpolation technique applied in the rs_diffraction.exe computer routine.

The flowchart of the complete algorithm can be summarized as follows:

- 1) read the input file
- 2) **determine the representative (u, v) points of the sampling grid**
- 3) **determine the distance from the optical axis for each representative point**

- 4) **calculate the complex amplitude U for each representative point**
- 5) **execute the diffraction integration**
- 6) write the output file;

where the bold-typed items are performed for each point P_0 specified in the output plane. Programming of this routine has been carried out in C++ environment, using floating-point arithmetics. The core of the calculation, i.e. the Rayleigh-Sommerfeld integration has been optimized for minimum operation number. The result is that the complex field can be calculated in 151 seconds for one single point P_0 , in case of “intervals” = 6000, running the programme on a 366 MHz PC with one Intel Celeron processor and 64 Mbytes of RAM.

Looking at the above typical sampling value, one could ask if the sampling might be decreased so as to reduce computation time? Unfortunately, the sampling number cannot be significantly lowered (in this particular implementation of the RS formula), without resulting in large computational errors. The cause is that the RS integrand, see (A.1), rapidly oscillates as a function of distance (r) from the optical axis in the input plane. (The period of these oscillations usually shortens as P_1 goes further from the axis.) If we require ‘N’ sampling intervals inside one period of the highest spatial frequency oscillation (usually near the edge of the aperture in the input plane), the required sampling number can be estimated. For the case when $r \gg \lambda$, “intervals” can be approximated as follows:

$$\text{intervals}(r) \approx \frac{a}{\lambda \cdot \sqrt{1 + \left(\frac{z_0}{r}\right)^2}} \cdot N, \quad (\text{A.4})$$

where assigning $r := (a + \text{patch})$, one obtains a good approximation to the required value of “intervals”. For an example in the case of a focused spherical beam, such as discussed in Appendix B, we can substitute the values given in Tab. B.1, and $N := 2.52$ into (A.4). Here we get intervals = 6,000, which means that in the calculation performed in Appendix B, a possible undersampling is present, since there are only 2.52 sampling intervals (in average) inside one period of the oscillations of the integrand. Why do calculations still have very accurate results then? The answer was expressed in [22] as the phenomenon of “stationary phase”, which means that adjacent oscillations average out each others’ effect in those regions of the diffracting aperture, where the period of oscillations do not change too much. What “too much” means in reality is very hard to estimate by analytic methods. Instead, some numerical verification has to be made for each specific diffraction situation. For such a calculation an example is given in Appendix B.

APPENDIX B – Effects of sampling on the accuracy of “rs_diff.exe”

In Sections IV.6 and V.6 an implementation of the Rayleigh-Sommerfeld (RS) diffraction formula, presented in Appendix A as (A.1), was used to calculate output intensity patterns of certain optical systems along a plane surface. This implementation is actually a numerical integration realised as a computer programme written in the C++ language (see Appendix C). All numerical computations have a common weak point: sampling accuracy. This appendix was prepared to present an investigation on the effects of sampling, through giving a deeper insight into a computation that is very similar to that performed in Section V.6. In the present appendix I calculate the diffraction intensity distribution (PSF) curve of a perfect spherical wave having the following parameters (see Appendix A):

parameter	value
intervals	6,000 pcs
lambda	0.00078 mm
rfs_p = z0	21.524012 mm
rfs_a = a	6.3 mm
total_power	117.210240 AU ² ·mm ²

Tab. B.1. Input parameters to the rs_diff.exe computer routine.

The above total_power results from an on-axis input field intensity of 1 AU² (equal over a sphere centered on the focus spot and having a radius of rfs_p). The output field was calculated with different sampling for two intervals in the output plane: $6.3333 \cdot 10^{-5}$ mm between [0 mm ; 0.01995 mm], i.e. close to the axis, and $3.5 \cdot 10^{-4}$ mm for [0.0203 mm ; 0.35 mm], i.e. far from the axis. These parameters strictly correspond to those applied in Section V.6. I use this “synthetic” calculation to estimate the error that can be expected in Section V.6, since this way the pupil sampling can be adjusted freely.

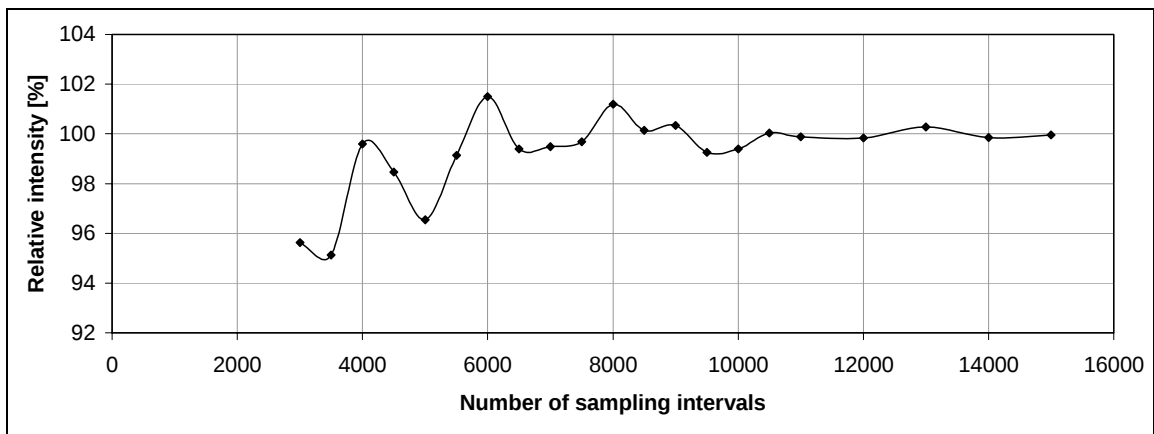


Fig. B.1. Relative error of the focal plane intensity distribution (PSF) curve at one single point (0.35 mm far from the optical axis), as a function of the number of sampling intervals in the exit pupil of the test lens (Section V.6, Fig. V.27). The reference intensity is the estimated limiting value of the depicted function as the sampling interval number tends to infinity.

First, that tendency should be determined, how the calculated intensity changes with sampling. For this purpose I computed the intensity in the output plane, 0.35 mm far from the axis with different sampling (this is the furthest position, where the error is expected to be the largest). The result can be seen in Fig. B.1. From this figure it is clear that a sampling of 12,000 is a very good approximation of the limiting value of the curve (i.e. the exact result). Since this would take an enormous amount of time to make calculations with such a high sampling, all previous RS integrations I made with intervals = 6,000. The error I made there can be well estimated by comparing the focal plane intensity distribution (PSF) and energy distribution (Encircled Energy) curves of the perfect spherical wave calculated with 6,000 and 12,000 number of interval sampling. It must be mentioned here, that in case of the 6,000 sampling, there is an approx. $360^\circ/2.5$ phase change in the input field complex amplitude between two adjacent sampling points at the edge of the exit pupil. This phase change is only approx. $360^\circ/5$ for 12,000 sampling.

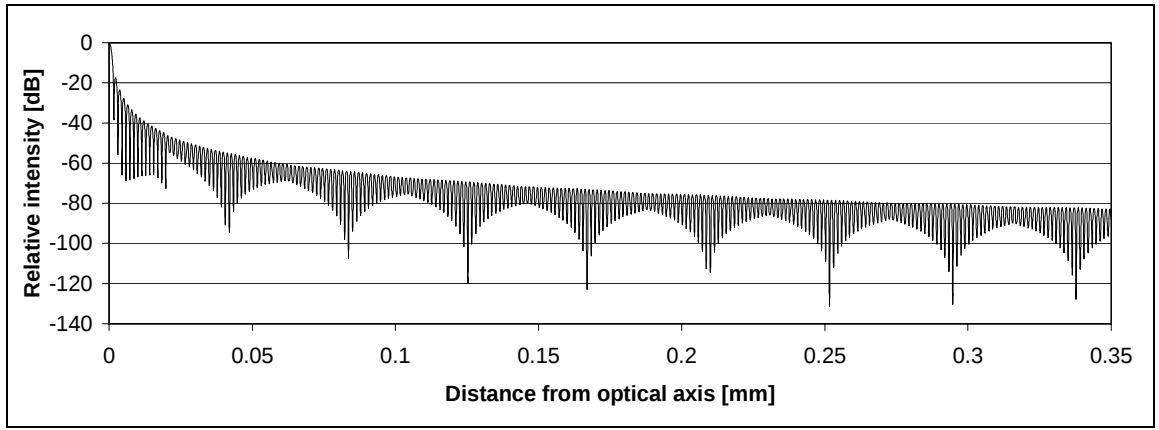


Fig. B.2. Intensity distribution (PSF) in the focal plane of the test lens, see Section V.6, Fig. V.27. Peak intensity is normalised to that of a perfect spherical wave having the same NA and total power than that applied in the calculation. The number of sampling intervals is 6,000 along the radius of the exit pupil of the test lens.

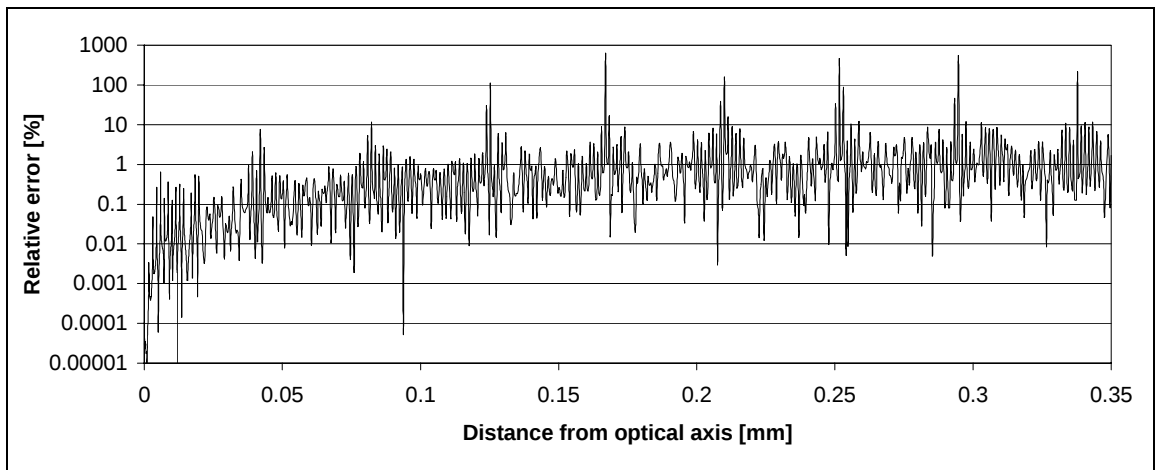


Fig. B.3. Relative error of the focal plane intensity distribution of the test lens, calculated with 6,000 sampling intervals with respect to that calculated by 12,000 sampling intervals along the radius of the exit pupil.

The results can be seen in Figs. B.2 and B.3. Fig. B.2 shows the focal plane intensity distribution in case of 6,000 sampling, while Fig. B.2 shows the relative error from the intensity curve calculated by 12,000 sampling. It is clear from this latter figure that the error can be very large, but only near those positions where the intensity is around zero, thus their contributions are negligible. The average error is however always less 1%. Finally, Fig. B.4 presents the radial energy distribution for 6,000 sampling, and Fig. B.5 shows the error relative to that calculated for 12,000 sampling. It can be seen that the highest value is one order of magnitude less than the error presented in Section V.6 , Fig. V.31 (in other words, the error curve obtained in Fig. V.31 is indeed a consequence of series truncation, introduced at the multiple diffraction order summation).

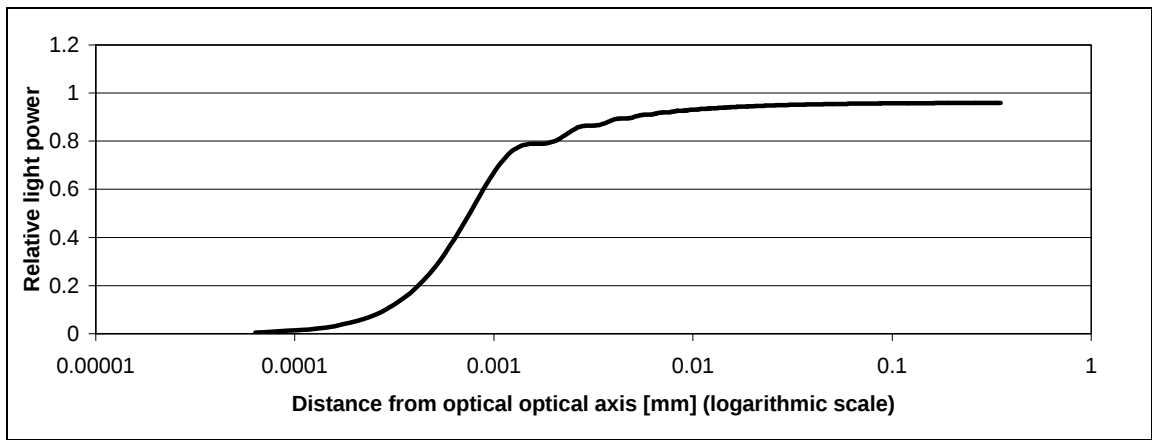


Fig. B.4. Radial energy distribution (Encircled Energy) in the focal plane of the test lens, see Section V.6, Fig. V.27. The curve has been normalised to the total power of the input beam. The number of sampling intervals is 6,000 along the radius of the exit pupil of the test lens. Horizontal axis depicts distance from the axis in the [0 mm; 0.35 mm] interval.

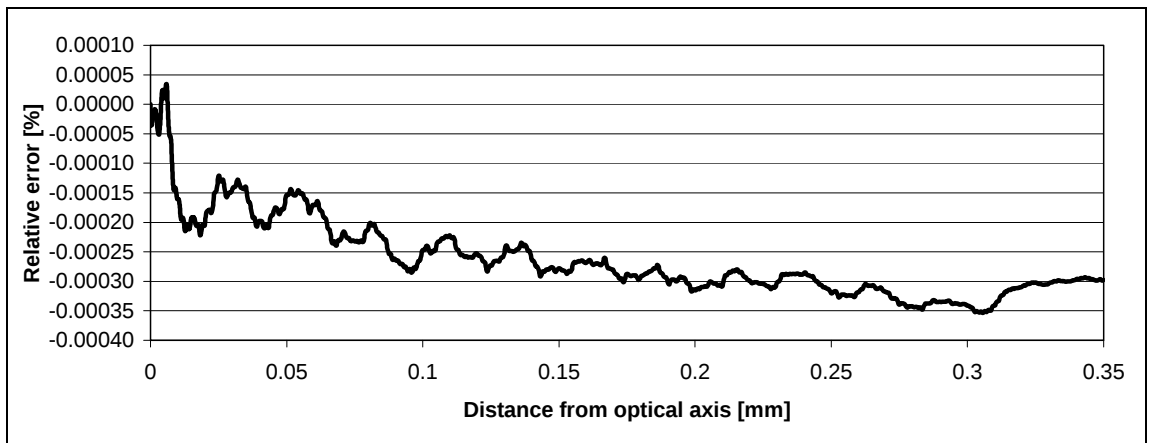


Fig. B.5. Relative error of the focal plane radial energy distribution (Encircled Energy) of the test lens, calculated with 6,000 sampling intervals with respect to that calculated by 12,000 sampling intervals along the radius of the exit pupil.

APPENDIX C – Programme listing of “rs_diff.exe”

Programme: rs diffr.exe

```

// Source code of executable 32 bit C++ program generated by ERDEI Gabor, Technical University of Budapest,
// Department of Atomic Physics, Optical Research Laboratory. All rights reserved © 19/11/1998.
// Last updated on 18/10/99 by EG.

#include <math.h> // For mathematical functions.
#include <stdio.h> // For file/CRT operations.
#include <string.h> // For string functions.
#include <malloc.h> // For MALLOC.
#include <conio.h> // For _CGETS.
#include <time.h> // For time handling functions.

#define EOL 10 // End-of-line character.
#define twopi 6.2831853071796 // Two times pi.

double radians(double x);

void main()
{
    FILE *file_id;

    int i, j, k, i0, i1, j_max,
        intervals, data_points;

    double lambda, rfs_p, a, patch,
        delta, delta2, y, z, z0, z2, r2, twopi_lambda, U_x, U_y,
        U_abs, phi, cosphi, sinphi, i_field, j2, v_y2_z2,
        U_exp, f[8], q, U_abs_ref, U_exp_ref, kk, fi_x, fi_y,
        delta_i, delta_twopi, rfs_p2, k2, total_power, t, rfs_a;

    char buffer[14], c, input_file[82] = {80}, output_file[82] = {80}, *s,
        input_file2[82] = {80};

    struct tm when;

    time_t now;

    printf("\n-----\n");
    printf("Rayleigh-Sommerfeld diffraction written by Erdei Gabor\n");
    printf("-----\n");
    printf("Technical University of Budapest\n");
    printf("Department of Atomic Physics\nOptical Research Laboratory\n");
    printf("All rights reserved, 1998.\n");
    printf("-----\n");

    printf("\nEnter input file name:\n");

    s = _cgets(input_file); // Asks for input file name.
    strcpy(input_file, s);

    if(input_file[0] == NULL) // In case the string is empty.
    {
        return;
    }

    k = 0;
    while(input_file[k] != NULL) // In strings for input of file
    { // operations '/' is used instead
        if(input_file[k] == '\\') // of '\\'. Now '\\'s are changed
        { // in output filename to '/'s.
            input_file[k] = '/';
        }

        ++k;
    }

    file_id = fopen(input_file, "rt"); // File containing input data.

    for(i = 0; i < 8; ++i) // Reading 8 input parameters
    { // from file.
        strcpy(buffer, " "); // Clearing the buffer.
        fread(&c, 1, 1, file_id); // Reading the first byte.
        k = -1; // Starting value of the pointer
        // used to address the buffer.
        while((c != 32) && (c != EOL) && (c != EOF)) // Reading a floating-pt number.
        { // 32 == ascii code of ' '.
            buffer[++k] = c;
            fread(&c, 1, 1, file_id);
        }

        f[i] = atof(buffer); // Transforming the string into
        // field.
    }

    fread(&c, 1, 1, file_id); // Reading the first byte.
    k = -1; // Starting value of the pointer
    // used to address the buffer.
    while((c != 32) && (c != EOL) && (c != EOF)) // Reading a floating-pt number.
    { // 32 == ascii code of ' '.
        input_file2[++k] = c;
        fread(&c, 1, 1, file_id);
    }

    k = 0;
    while(input_file2[k] != NULL)
    {
        if(input_file2[k] == '\\')
        {
            input_file2[k] = '/';
        }

        ++k;
    }

    fread(&c, 1, 1, file_id); // Reading the first byte.

```



```

for(i = 1; i < data_points; ++i)
{
    outp_srf[i][0] = patch/(data_points-1)*i;
    outp_srf[i][1] = 0;
}

file_id = fopen(output_file, "w");

delta = a/intervals; delta2 = delta*delta;
twopi_lambda = twopi/lambda; delta_twopi = delta*twopi;
rfs_p2 = rfs_p*rfs_p; t = 3.3037707e-6;

U_abs_ref = 1;
U_exp_ref = radians(z0*twopi_lambda);

if(rfs_p != 0)
{
    t = 3.6824684e-6;

    U_abs_ref = twopi_lambda*rfs_p*(1-pow(pow(rfs_a/rfs_p, 2)+1, -0.5))*
    sqrt(total_power/(twopi*rfs_p2*(1-cos(atan2(rfs_a, rfs_p)))));
}

time(&now);
when = *localtime(&now);
printf("\nCalculation started on:      %s", asctime(&when));

now += (time_t)ceil(t*intervals*intervals*data_points);
when = *localtime(&now);
printf("Estimated termination on:      %s", asctime(&when));

printf("Progress report:                  0.00 percent done");

for(i = 0; i < data_points; ++i)
{
    y = outp_srf[i][0];
    z = outp_srf[i][1]+z0; z2 = z*z;
    U_x = U_y = 0;

    if(rfs_p != 0)
    {
        for(k = -intervals; k < intervals; ++k)
        {
            kk = k+0.5;
            k2 = kk*kk;
            q = kk*delta-y;
            v_y2_z2 = q*q+z2;

            j_max = (int)floor(intervals*sin(acos(kk/intervals))-0.5);
            for(j = 0; j <= j_max; ++j)
            {
                j2 = j+0.5;
                j2 *= j2;
                r2 = j2*delta2+v_y2_z2;

                i_field = sqrt(j2+k2);
                i0 = (int)floor(i_field);
                i1 = (int)ceil(i_field);
                delta_i = i_field-i0;
                fi_x = (field[i1][0]-field[i0][0])*delta_i+field[i0][0];
                fi_y = (field[i1][1]-field[i0][1])*delta_i+field[i0][1];

                q = i_field*i_field;
                phi = twopi_lambda*(sqrt(r2)-(sqrt(rfs_p2+q*delta2)-
                rfs_p));
                cosphi = cos(phi);
                sinphi = sin(phi);

                U_x += (fi_x*cosphi-fi_y*sinphi)/r2;
                U_y += (fi_x*sinphi+fi_y*cosphi)/r2;
            }
        }
    }
    else
    {
        for(k = -intervals; k < intervals; ++k)
        {
            kk = k+0.5;
            k2 = kk*kk;
            q = kk*delta-y;
            v_y2_z2 = q*q+z2;

            j_max = (int)floor(intervals*sin(acos(kk/intervals))-0.5);
            for(j = 0; j <= j_max; ++j)
            {
                j2 = j+0.5;
                j2 *= j2;
                r2 = j2*delta2+v_y2_z2;

                i_field = sqrt(j2+k2);
                i0 = (int)floor(i_field);
                i1 = (int)ceil(i_field);
                delta_i = i_field-i0;
                fi_x = (field[i1][0]-field[i0][0])*delta_i+field[i0][0];
                fi_y = (field[i1][1]-field[i0][1])*delta_i+field[i0][1];

                phi = twopi_lambda*sqrt(r2);
                cosphi = cos(phi);
                sinphi = sin(phi);

                U_x += (fi_x*cosphi-fi_y*sinphi)/r2;
                U_y += (fi_x*sinphi+fi_y*cosphi)/r2;
            }
        }
    }
}

```


APPENDIX D – Gibbs phenomenon at multiple diffraction order summation

In Section V.6, a method has been presented for the verification of the UDS surface model of surface-relief phase-modulation DOEs. The solution applied a comparison between the results of the UDS model of DOEs to those of the more traditional multiple diffraction order summation. For this purpose, the output wavefront profile and intensity distribution of some test lens (see Fig. V.27) have been calculated by both models, and a diffraction analysis were carried out to determine irradiance distributions in the focal plane of the lens (see also Appendices A, B and C). The careful reader might pose a question looking at the output wavefront and intensity curves (Figs. D.1 and D.2): are the central and subsequent oscillations a result of some real physical phenomenon or just computing artefacts? The answer is that these oscillations are a pure consequence of truncation to a finite number of diffraction orders applied at field summation. This computing artefact is the so-called Gibbs phenomenon, which presents itself as oscillations at discontinuities of functions, built up as a summation of a finite number of terms; see e.g. the Fourier series expansion of a raw-tooth function.

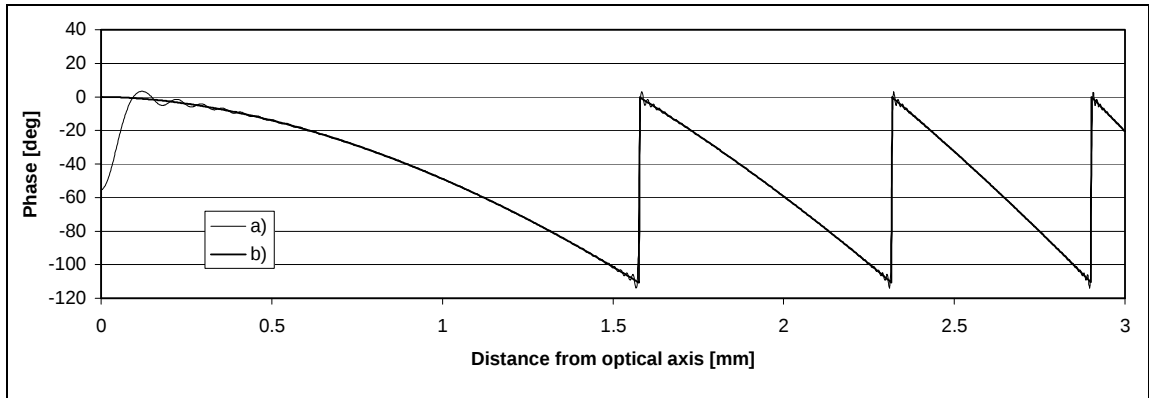


Fig. D.1. Enlarged central part of the output wavefront profile of the test lens at 780 nm (see Fig. V.28), relative to the phase of the reference wave (see definition in Appendix A). Thin line shows the phase calculated by summing up the $m = -49..51$ diffraction orders (a), thick line shows that in case of the UDS model (b). ($l = 1$, $\lambda_0 = 540$ nm, $OPL_0 = -l\lambda_0$)

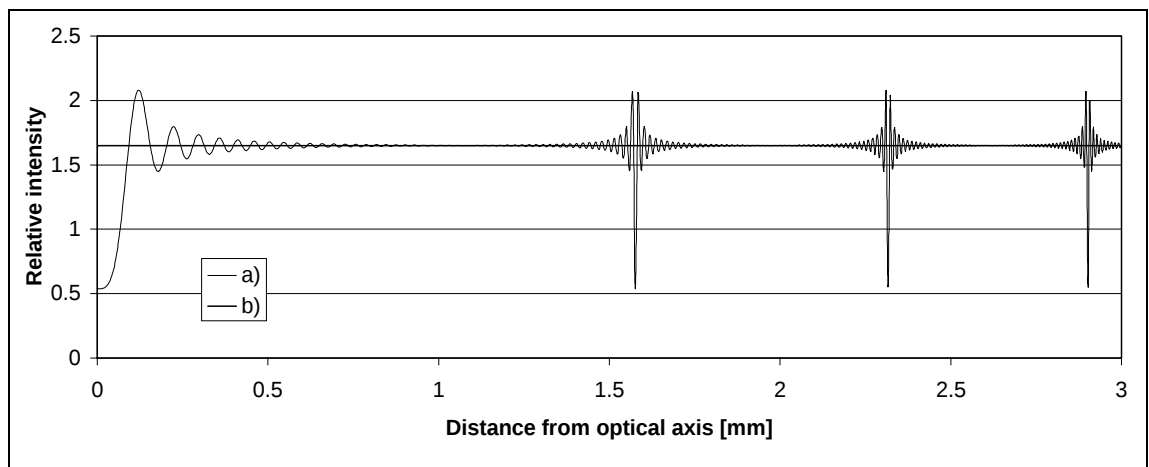


Fig. D.2. Enlarged central part of the intensity distribution over the exit pupil of the test lens (see Fig. V.29). Thin line shows intensity calculated by summing up the $m = -49..51$ diffraction orders (a), thick line shows that in case of the UDS model (b). ($l = 1$, $\lambda_0 = 540$ nm, $OPL_0 = -l\lambda_0$)

The discontinuity itself at the center of curve b) in Fig. D.1 cannot be seen in the calculated wavefront profile, since it is rotationally symmetric to the axis. However, if one takes a look at Fig. V.3, case $OPL_0 = -l\lambda_0$, in Subsection V.2.2, this will be clear that theoretically, the wavefront profile has a discontinuity as the distance, measured from the optical axis, tends to zero.

The first reason for the large extent of the central oscillation is that its period has a monotonic connection to the slope of the original curve, see b) in Fig. D.1. As the slope increases, the period gets smaller; this can well be followed in curve a) Fig. D.1, where the slope of curve b) is zero at the axis.

The second reason is the slow convergence of multiple diffraction order summation. In Figs. D.3 and D.4 the output wavefront and intensity profiles are depicted with 601 summed diffraction orders. Though the number of terms included has been multiplied by six, the distance of the first zero-place of the wavefront curve reduced only to its 35%.

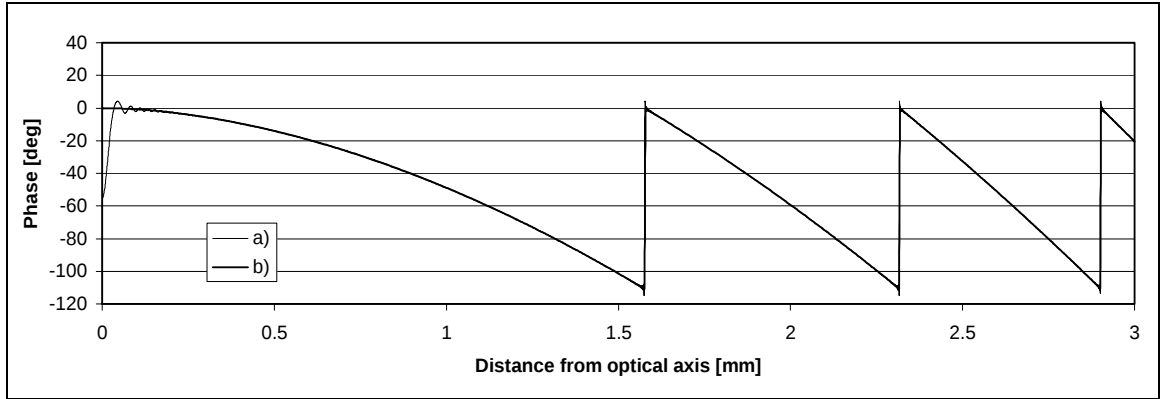


Fig. D.3. Enlarged central part of the output wavefront profile of the test lens at 780 nm (see Fig. V.28), relative to the phase of the reference wave (see definition in Appendix A). Thin line shows the phase calculated by summing up the $m = -299..301$ diffraction orders (a), thick line shows that in case of the UDS model (b). ($l = 1$, $\lambda_0 = 540$ nm, $OPL_0 = -l\lambda_0$)

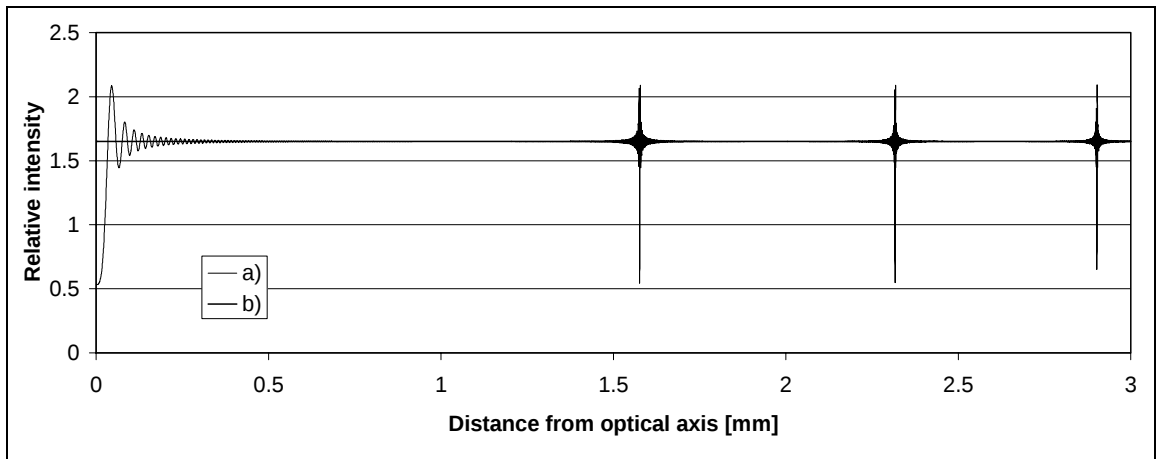


Fig. D.4. Enlarged central part of the intensity distribution over the exit pupil of the test lens (see Fig. V.29). Thin line shows intensity calculated by summing up the $m = -299..301$ diffraction orders (a), thick line shows that in case of the UDS model (b). ($l = 1$, $\lambda_0 = 540$ nm, $OPL_0 = -l\lambda_0$)

The third reason is that in the above figures the $H'(H)$ function of the test lens has been calculated by using the $OPL_0 = -l\lambda_0$ assumption, which reflects most fabrication techniques (see Subsection V.2.2). In case $OPL_0 = -0.5 \cdot l\lambda_0$ is used instead of the above formula, the $H'(H)$ function would be continuous at the center, thus no oscillation would occur at multiple diffraction order summation. This can be seen in Figs. D.5 and D.6, where the $H'(H)$ function of the test lens has been re-calculated with $OPL_0 = -0.5 \cdot l\lambda_0$.

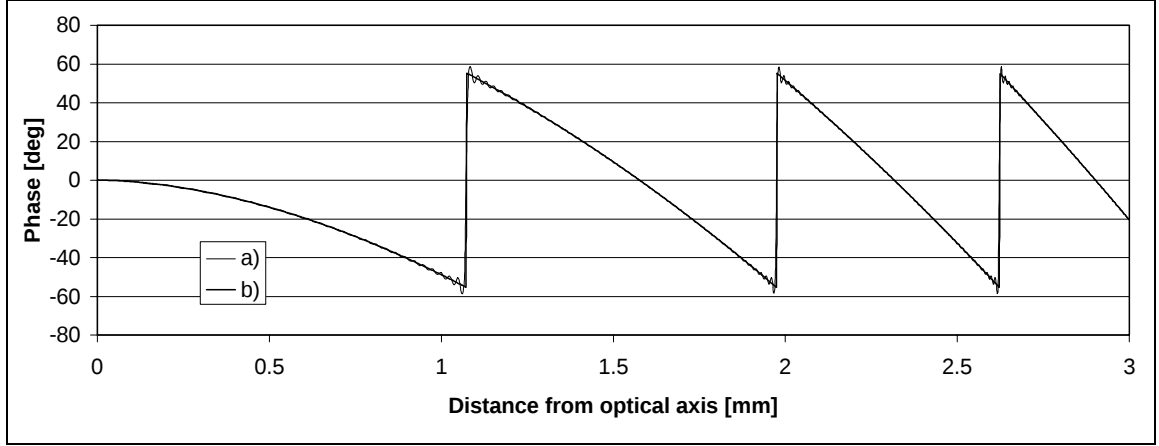


Fig. D.5. Enlarged central part of the output wavefront profile of the test lens at 780 nm (see Fig. V.28), relative to the phase of the reference wave (see definition in Appendix A). Thin line shows the phase calculated by summing up the $m = -49..51$ diffraction orders (a), thick line shows that in case of the UDS model (b). ($l = 1$, $\lambda_0 = 540$ nm, $OPL_0 = -0.5 \cdot l\lambda_0$)

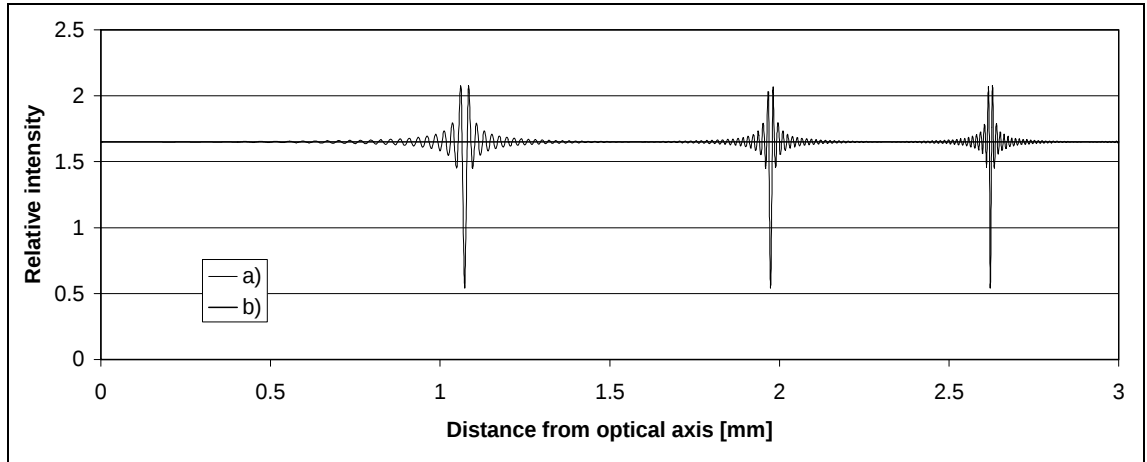


Fig. D.6. Enlarged central part of the intensity distribution over the exit pupil of the test lens (see Fig. V.29). Thin line shows intensity calculated by summing up the $m = -49..51$ diffraction orders (a), thick line shows that in case of the UDS model (b). ($l = 1$, $\lambda_0 = 540$ nm, $OPL_0 = -0.5 \cdot l\lambda_0$)

Concerning the above demonstrative calculations, one remark has to be made. In order that I can freely change the number of summed up orders and the parameter OPL_0 , I used a synthesized wavefront and a mathematical programme to perform multi-order summation, and not OSLO's ray-tracing routines. The main reason behind this was that OSLO's routines may result in imprecise results for extremely large diffraction orders (such as e.g. the 300th).

APPENDIX E – Glossary on definitions and abbreviations

Technical acronyms	
2D	two dimensional
3D	three dimensional
AU	arbitrary unit
CAD	computer assisted design
CCD	charged coupled device
CCL	compiled command language
CGH	computer generated hologram
DOE	diffractive optical element
EOL	end-of-line character
FFT	fast Fourier transform
FWHM	full width at half maximum
MTF	modulation transfer function
NA	numerical aperture
OPD	optical path difference
OPL	optical path length
PC	personal computer
PMMA	poly-methyl methacrilate
PSF	point spread function
RAM	random access memory
RMS	root-mean-square
TE	transverse electric (mode)
TEM	transverse electro-magnetic (mode)
UDS	used defined surface

General abbreviations	
e.g.	for example
i.e.	that is
viz.	namely
cf.	compare with

All physical units are given in the SI (Système Internationale) metrological system throughout the dissertation.

Physical terminology	
β	propagation constant
$f_{\#}$	f-number
I	intensity
$\mathbf{k}$	wave-vector
milli (m)	10^{-3}
mikro (μ)	10^{-6}
nano (n)	10^{-9}
mega (M)	10^6
n	refractive index
n_e	extraordinary refractive index
n_o	ordinary refractive index
P	power
RS	Rayleigh-Sommerfeld (formula)
U	field complex amplitude
T	power transmission
λ	wavelength

Mathematical terminology	
bold character	vector-type quantity
italic character	variable-type quantity
normal character	constant-type quantity
$:=$	assign a value
$\equiv$	defininition
$\cdot$	scalar product
‰	one piece per one thousand
d	infinitesimally small interval
Δ	finite but very small interval
$\lim$	limes (limiting value of a function)
Re	real part (of a complex number)
Im	imaginary part (of a complex number)
$a < b$	'a' less than 'b'
$a > b$	'a' more than 'b'
$\approx$	approximately equals
$x \text{ [dB]}$	$10 \cdot \log_{10}(x)$ for power-type quantities
$[a ; b]$	closed interval for x , $a \leq x \leq b$
i	complex unit
$\text{floor}(x)$	closest 'i' integer to x , $i \leq x$
$\text{sinc}(x)$	$\sin(\pi \cdot x) / (\pi \cdot x)$
e	2.71828182845904523536029...
π	3.14159265358979323846264...

APPENDED JOURNAL PAPERS

G. Erdei, G. Szarvas and M. Barabás, “Extensions to a lens design program for the ray-optical design of waveguide lenses and prism couplers”, <i>Jour. of Mod. Opt.</i> , Vol. 44, No. 2, pp. 415-430, 1997.....	112
G. Erdei, G. Szarvas, E. Lőrincz and P. Richter, “Optimisation method for the design of beam shaping systems”, to be published in <i>Optical Engineering</i> , February 2002.*	128
G. Erdei, G. Szarvas, E. Lőrincz and P. Richter, “Cascading low-quality beam-shapers to improve overall performance”, to be published in <i>Optical Engineering</i> , February 2002.*	137
H. Sauer, P. Chavel and G. Erdei, “Diffractive optical elements in hybrid lenses: modeling and design by zone decomposition”, <i>Applied Optics</i> , Vol. 38, No. 31, pp. 6482-6486, 1999.	145

* These articles are *uncorrected* non-editable proofs received from the Journal, thus may contain spelling and other errors.

ABSTRACT

Currently, an intense research takes place in the field of optical design, both on industrial as well as academic platforms. With my Ph.D. dissertation I wish to contribute to this activity, through the improvement of commercial optical design programmes. My research work encompasses the following branches of modern optical design: integrated optics, beam shaping and hybrid diffractive/conventional optics. For the modelling of the above devices I developed new techniques to make their design process (analysis and synthesis) more powerful (i.e. accurate, fast, convenient, providing additional degrees of freedom, allowing easy integration into complex systems) in comparison with former methods. Making a model in my case partly implies the creation of a new physical approach to describe some optical phenomenon, as well as the *efficient* implementation (usually on the basis of exact ray-tracing) of one that already existed. This way I could enrich the selection of “engineering tools” (i.e. models), by which modern optical devices can be more easily incorporated in the process of optical design. My results can be summarised as follows. (1) For the analysis and optimisation of integrated optical systems and beam-shapers I developed efficient methods, by which their design process can be performed relatively easily. (2) I made theoretical considerations for the structural modification of these devices to achieve better performance with respect to former solutions. (3) I designed such devices for improved characteristics. (4) I developed a method that alleviates the diffraction analysis of hybrid diffractive/conventional optical systems. In addition, I performed laboratory experiments and computer simulations to check the accuracy of my results.

Keywords

optical design, ray-tracing, diffraction, integrated optics, optical anisotropy, prism coupler, beam shaping, diffractive optics

Új modellezési eljárások kifejlesztése és alkalmazása a modern optikai tervezésben

Rövid összefoglaló

Napjainkban az optikai tervezés mind ipari, mind akadémiai körökben intenzív kutatás tárgyát képezi. Ph.D. disszertációmmal ehhez a tevékenységhez kívánok hozzájárulni, a kereskedelemben kapható optikai tervező programok továbbfejlesztése révén. Kutatásaimat a modern optikai tervezés következő részterületein folytattam: integrált optika, nyalábformálás, hibrid diffraktív/hagyományos optikák. A fent felsorolt eszközök modellezésére új eljárásokat fejlesztettem ki, hogy ezáltal a tervezési folyamatot a korábbi módszerekhez képest hatékonyabbá – azaz pontosabbá, gyorsabbá, kényelmesebbé, több szabadsági fokot nyújtóvá, bonyolult rendszerekbe könnyebb integrálhatóságot eredményezővé – tegyem. A modellalkotás esetében részint új, valamilyen optikai jelenséget leíró fizikai megközelítés kidolgozását, részint pedig egy már létező elvnek megfelelően működő *hatékony* tervezési módszer (rendszerint a valós sugárátvezetésen alapuló) megvalósítását jelenti. Ily módon a „mérnöki eszközök” (modellek) tárházát igyekeztem gazdagítani, hogy a különleges optikai elemeket könnyebb legyen az optikai tervezés folyamatába beilleszteni. Eredményeim a következő pontokban összegezhetők. (1) Integrált optikai, valamint nyalábformáló rendszerek analízisére és optimalizációjára hatékony módszereket fejlesztettem ki, amelyek használatával a tervezési folyamat leegyszerűsíthető. (2) Elvi megfontolásokat tettem ezen eszközök felépítésének módosítására, hogy alkalmazásukkal a korábbi megoldásoknál jobb optika paraméterek legyenek elérhetőek. (3) Javított optikai jellemzőkkel rendelkező ilyen eszközöket terveztem. (4) Kifejlesztettem egy módszert, amely megkönnyíti hibrid diffraktív/hagyományos optikai rendszerek diffrakció alapú kiértékelését. Továbbá, eredményeim pontosságának ellenőrzésére laboratóriumi méréseket és számítógépes szimulációkat végeztem.

Kulcsszavak

optikai tervezés, sugárkövetés, diffrakció, integrált optika, optikai anizotrópia, prizmás becsatoló, nyalábformálás, diffraktív optika

Développements et applications de nouvelles techniques de modélisation pour la conception optique contemporaine

Résumé

Actuellement, un intense travail de recherche a lieu dans le domaine de la conception optique, tant au niveau industriel qu'académique. Cette thèse de doctorat s'inscrit dans ce contexte général en abordant l'extension et l'utilisation de programmes commerciaux de conception optique dans de nouvelles branches de l'ingénierie optique, incluant l'optique intégrée, la mise en forme de faisceau et les systèmes optiques hybrides diffractifs/conventionnels. Afin de modéliser ces nouveaux systèmes, j'ai développé de nouvelles techniques permettant de rendre leur analyse et leur conception plus 'puissantes' que par les méthodes précédemment en usage, par la précision, la rapidité de calcul, la convivialité ou les degrés de liberté à la disposition du concepteur. Ces travaux ont parfois nécessité la mise en place de nouvelles descriptions des phénomènes physiques mis en jeu, mais ils ont surtout été axés sur la recherche d'une mise en œuvre *efficace* (usuellement basée sur le tracé exact de rayons) d'approches déjà connues. J'ai ainsi enrichi la panoplie des *outils pour l'ingénieur*, permettant une incorporation plus simple de composants optiques modernes dans la phase de conceptions des optiques. Mes résultats peuvent être résumés en quatre points: (1) J'ai développé des méthodes efficaces pour l'analyse et l'optimisation de systèmes optiques intégrés avec lesquelles leur conception peut être réalisée avec plus de facilité. (2) J'ai réalisé des investigations théoriques de modifications structurelles de tels systèmes afin d'obtenir de meilleures performances qu'auparavant. (3) J'ai conçu de tels dispositifs à caractéristiques améliorées. (4) J'ai développé une méthode qui allège l'analyse des systèmes optiques hybrides diffractifs/conventionnels. De surcroît, j'ai réalisé des expériences et des simulations pour tester la justesse de mes résultats.

Mots-clefs

conception optique, tracé de rayons, diffraction, optique intégrée, biréfringence, coupleur à prisme, mise en forme de faisceau, optiques diffractives