Thomas Kreis

Handbook of Holographic Interferometry

Optical and Digital Methods

WILEY-VCH

WILEY-VCH Verlag GmbH & Co. KGaA

Thomas Kreis

Handbook of Holographic Interferometry

Optical and Digital Methods

Thomas Kreis

Handbook of Holographic Interferometry

Optical and Digital Methods

WILEY-VCH

WILEY-VCH Verlag GmbH & Co. KGaA

Author
Dr. Thomas Kreis
BIAS – Bremer Institut für angewandte Strahltechnik
Klagenfurter Str. 2, 28359 Bremen, Germany
e-mail: kreis@bias.de

Cover Picture
Holographic interference pattern of a deformed satellite tank. Deformation caused by variation of internal pressure. Frequency doubled Nd:YAG-laser of 532 nm wavelength used.

Library of Congress Card No.: applied for

British Library Cataloging-in-Publication Data:
A catalogue record for this book is available from the British Library

Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at <http://dnb.ddb.de>.

Printed in the Federal Republic of Germany
Printed on acid-free paper

Printing Strauss Offsetdruck GmbH, Mörlenbach
Bookbinding Litges & Dopf Buchbinderei GmbH, Heppenheim

ISBN 3-527-40546-1

Contents

Preface

The story of this book began in September 2003, when Wiley-VCH asked me to prepare a new edition of my book 'Holographic Interferometry – Principles and Methods'. Since its publication in 1996 more and more of my colleagues worldwide have used the book in their research, development, and teaching work, as was told to me and which is indicated by a continuous increase in citations. Therefore the demand for an updated new edition became apparent.

However, a new edition gives the impression that only some errors are eliminated, some presentations are streamlined and updated, and some new references are cited. On the other hand, in the last decade in holographic interferometry much substantial progress could be noticed, significant results were published, new interesting applications became possible. Especially, the field of digital holography matured from its infancy as presented in the former book to an established method, offering a lot of options not accessible before. Some of the work in this direction has been performed by my coworkers and me. The emphasis of the former book was on the computer-aided methods in holographic interferometry, so it is a logical consequence to include the methods offered by digital holography as far as they have consequences on holographic interferometry as a metrologic tool. The result is not a further edition of the old book but a totally new book, although many of the basic sections and the general organization are adopted from the former one. Therefore it was given the new title 'Handbook of Holographic Interferometry' and the subtitle 'Optical and Digital Methods'. This title on the one hand makes the big promise to deliver all information needed for solving a typical problem in holographic interferometry – e. g. a nondestructive testing task or an experimental stress analysis job – and on the other hand makes it clear-cut what is not included in this book: All the other applications of holography – e. g. in art, for displaying real or virtual scenes, for security, for general wavefront generation etc. – are not considered here. The aim of this book is to present a self-contained treatment of the underlying principles and numerical methods intended to help the physicist or engineer in planning a holographic interferometric measurement, writing the evaluation software, performing the experiments, and interpreting the results of the computer-aided evaluation. The employment of computer power in holographic interferometry should not be restricted to, for example, the determination of interference phase distributions from recorded interference patterns or the numerical generation of the interference phase distribution in digital holography; it also enables numerical feasibility studies, simulations of holographic interferograms, the optimization of a holographic setup with regard to sensitivity and accuracy, automatic control of the measurement process or further processing of the interference phase, for example by numerical strain and stress analysis, by finite element and boundary element methods or by computerized tomography. The book should provide the fundamentals for making these attempts; where solutions cannot be given – because either they do not yet exist or they would go beyond the scope of this book – it should act as an incentive for further research and development work.

The list of references by no means constitutes a complete bibliography on the subject. Basically I have included those references I have found useful in my own research over the years. Often the use of a particular paper was dictated more by chance than by systematic search. Generally I have not made any attempts to establish historical priorities. No value judgments should be implied by my including or excluding a particular work. I express my apologies to anyone who has been inadvertently slighted or whose work has been overlooked.

It is hoped that this book will help the reader to exploit the various possibilities of holographic interferometry, to make the best choice of methods, and to use successfully the tools and algorithms presented herein. The book aims to present an account of the present state of the methods of holographic interferometric metrology; it must be emphasized that there is a considerable amount of research and development in progress and the subject needs to be kept in constant review. I hope, therefore, that readers will be challenged to think, criticize, read further, and quickly go beyond the confines of this volume.

Of course, I must take the blame for any mistakes, omissions of significant facts, incomprehensible descriptions, residual errors or misprintings. Readers are cordially invited to provide me with any corrections, comments, hints, or suggestions to improve the presentation, which may be realized in the preparation of potential further editions of this book. (Dr. Thomas Kreis, Bremer Institut für Angewandte Strahltechnik – BIAS, Klagenfurter Straße 2, D 28359 Bremen, Germany. E-mail: kreis@bias.de)

Clearly this book would not have been completed without the help of many persons. In the preparation phase of the former book, Prof. Dr. W. Jüptner and Dr. W. Osten helped greatly by critical and stimulating discussions. The former and this present book draw heavily from the research done at BIAS by my past and present colleagues and collaborators, especially M. Adams, T. Baumbach, R. Biedermann, Th. Bischof, J. Geldmacher, D. Holstein, D. Kayser, V. Kebbel, E. Kolenovic, U. Mieth, U. Schnars, S. Seebacher, and M. Wrieden, which is gratefully acknowledged. S. Hotze (n. Knoll) and Chr. Kapitza produced many of the holographic interferograms used to illustrate this book, and most of the photographs have been carefully prepared by B. Schupp. Thanks are due to all colleagues, especially H.-J. Hartmann and Chr. von Kopylow, who made it possible for me to work on the book by relieving me from much of my daily routine. I sincerely appreciate the competent help of the Wiley-VCH staff, especially I want to thank Heike Höpcke, Andreas Thoss and Uwe Krieg who helped a lot in preparing this handbook. Andrew Bacon provided language polishing, making the book more readable for users. Finally thanks go to my wife Elisabeth for her understanding and acceptance when I devoted my spare time to the book.

Thomas Kreis

Bremen, August 2004

1 Introduction

1.1 Scope of the Book

The emerging computer technology of the last decades – increasing processing speed and memory capacity, as well as CCD- and CMOS-camera targets having more and smaller pixels – makes the manifold applications of what can be called ‘computer-aided holographic interferometry’ feasible: In the planning phase of a holographic interferometric experiment the geometry of the setup can be optimized to achieve maximum sensitivity and accuracy. The load to be applied can be optimized in its type, direction, and amplitude by numerical simulation of the holographic interferograms that result from a specific load and geometry. The determination of the interference phase distribution from the recorded interference patterns by refined methods such as phase stepping or Fourier transform evaluation is only possible with powerful computers. Further processing of the interference phase distribution by solving linear equations to obtain displacement fields or by employing computer tomography to calculate refractive index fields can now be effectively carried out. Methods for numerical strain and stress analysis can be combined with computerized holographic interferometry to gain far-reaching knowledge about the behavior of the tested structure with regard to the applied load. Even structural analysis methods such as finite element methods (FEM) or boundary element methods (BEM) can be efficiently associated to holographic interferometry to assist the strain and stress calculations, to optimize the component design process, or to predict the interference patterns for a given load.

Technological progress in the computer field actually also led to an intense use of the concept of digital holography, here understood in the sense of digital recording of the holograms and the numerical reconstruction of the wave fields in a computer. This concept was principally known for a long time, but before broader application it had to wait for the advent of powerful CCD- and CMOS-arrays as well as for fast processing and storing of large data sets. Digital holography avoids many of the drawbacks of the former optical holography and holographic interferometry and furthermore offers some possibilities not given by the optical approach.

The aim of this book is to present the physical principles of holography and interferometry as far as they are needed in this context, as well as the numerical methods for reconstruction of the complex wave fields from digitally recorded holograms and for evaluation of the interference patterns, which constitute the fundamentals of computer-aided holographic interferometry. The emphasis is on quantitative measurements with a sidelook on the qualitative evaluation of holographic nondestructive testing (HNDT). The present book should provide the background needed for deriving the concepts and writing the programmes to solve problems in the above mentioned fields. To fulfill these claims but not to become too extensive and to exceed the frame set by the publishers, some topics have been intentionally omitted. The description of technical components – lasers, optics, electro-optic devices, recording media, image processing equipment and methods, computer periphery – is restricted to a very short overview; more details on these topics can be found in more specialized books. The same is true for themes such as digital image processing, or particle and flow-field measurements. These items are addressed here only briefly as far as it seems necessary for a comprehensive presentation. Other applications of holography than interferometric metrology, such as display holography, computer generated holograms, holographic optical elements, color holography, holographic data storage, etc., are excluded intentionally.

The book is organized into seven chapters and three appendices. The main body of the content is contained in Chapters 2, 3, 4, 5, and 6. Chapter 2 presents the physical prerequisites of holography, starting with the wave theory of light, describing such effects as interference, diffraction, coherence, speckle, and how these are employed in holographic recording and reconstruction of optical wave fields. The technical components employed in optical as well as in digital holography and holographic interferometry are introduced.

Chapter 3 presents the techniques of how to record holograms on CCD- and CMOS-arrays, i. e. how to solve the problem imposed by the limited resolution of these detectors. Then the reconstruction of the recorded optical wave field by the numerical Fresnel transform and by the convolution approach is considered. Not only are the reconstruction algorithms presented, but also the various options possible in digital holography, like numerical suppression of the zero diffraction order, are outlined.

In Chapter 4 the fundamentals of holographic interferometric metrology are presented. The quantitative relations for displacements or refractive index variations and the geometric and optical parameters of the holographic setup are introduced. Further discussions center on the role of the sensitivity vectors and the localization of the interference fringes.

Chapter 5 is devoted to methods for determining the interference phase distributions from optically reconstructed intensity images as well as from digitally recorded holograms. Room is given to a thorough treatment of the phase-stepping and the Fourier transform methods as well as to digital holography. Systematic and statistical errors are discussed and a number of approaches for interference phase demodulation are presented.

Chapter 6 describes the further processing of the interference phase distribution. Displacement vector fields, strain and stress distributions, vibration modes, three-dimensional object contours or refractive index fields are determined. Computerized defect detection of holographic nondestructive testing is briefly addressed.

Speckle methods for deformation measurement like ESPI or shearography are closely related to holographic interferometry. These methods first employed analogue electronics, but nowadays they are realized digitally. They differ from digital holography in recording a focused image of a speckled surface and producing correlation fringes on an intensity basis, while in digital holography the whole complex field is reconstructed from the recorded Fresnel or Fraunhofer field. The main speckle methods are discussed briefly in Chapter 7.

The appendices provide the reader with the essentials of Fourier transforms, methods for computerized tomography, and Bessel functions, as far as this seems necessary to understand and implement the related methods of the main chapters.

To make the work with this book more comfortable, the references are given in the sequence of occurrence at the end of the book. This is accompanied by an alphabetically ordered author/coauthor index. A subject index lists a number of terms; these are printed in italics in the text to make their identification easier.

1.2 Historical Developments

Holography got its name from the Greek words ‘holos’ meaning whole or entire and ‘graphein’ meaning to write. It is a means for recording and reconstructing the whole information contained in an optical wavefront, namely amplitude and phase, and not just intensity as ordinary photography does. Holography essentially is a clever combination of interference and diffraction, two phenomena based on the wave nature of light.

Diffraction was first noted by F. M. Grimaldi (1618 – 1663) as the deviation from rectilinear propagation, and the interference generated by thin films was observed and described by R. Hooke (1635 – 1703). I. Newton (1642 – 1727) discovered the composition of white light from independent colors. The mathematical basis for the wave theory describing these effects was founded by Chr. Huygens (1629 – 1695), who further discovered the polarization of light. The interference principle introduced by Th. Young (1773 – 1829) and the Huygens principle were used by A. J. Fresnel (1788 – 1827) to calculate the diffraction patterns of different objects. Since about 1850 the view of light as a transversal wave won against the corpuscular theory. The relations between light, electricity, and magnetism were recognized by M. Faraday (1791 – 1867). These phenomena were summarized by J. C. Maxwell (1831 – 1879) in his well known equations. A medium supporting the waves was postulated as the all pervading ether. The experiments of A. A. Michelson (1852 – 1931), published in 1881, and the work of A. Einstein (1879 – 1955) were convincing evidence that there is no ether.

In 1948 D. Gabor (1900 – 1979) presented holography as a lensless process for image formation by reconstructed wavefronts [1–3]. His goal was to improve electron microscopy, using this new approach to avoid the previous aberrations. However, a successful application of the technique to electron microscopy has not materialized so far because of several practical problems. The validity of Gabor’s ideas in the optical field was recognized and confirmed by, for example, G. L. Rogers [4], H. M. A. El-Sum and P. Kirkpatrick [5], and A. Lohmann [6]. But the interest in holography declined after a few years, mainly because of the poor quality of the holographic images obtained in those days. The breakthrough of holography was initiated by the development of the laser, which made available a powerful source of coherent light. This was accompanied by the solution of the twin-image problem encountered in Gabor’s in-line arrangement. E. N. Leith and Y. Upatnieks [7–9] recognized the similarity of Gabor’s holography to the synthetic aperture antenna problem of radar technology and introduced the off-axis reference beam technique. Y. N. Denisyuk combined the ideas of Gabor and Lippmann in his invention of the thick reflection hologram [10].

Now there was a working method for recording and reconstruction of complete wavefields with intensity and phase, and this also in the visible region of the spectrum. Besides the impressive display of three-dimensional scenes exhibiting effects like depth and parallax, moreover holography found numerous applications based on its unique features. Using the theory describing the formation of a hologram by interference of reference and object wave, holograms were created by calculation on a digital computer [11]. The result of this calculation was transferred to a transparency by printing or by printing on paper followed by a photographic process that might have included a reduction in scale. Now images of ideal objects not existing in reality could be generated, later on offering ways for interferometric comparison of, for example, optical components to be tested or for fabrication of diffracting elements with predescribed behavior [12,13]. The way holograms store information in a form of distributed memory has given incentive for research in holographic data storage [14]. Especially three-dimensional storage media, such as photorefractive crystals which are capable of providing Bragg selectivity became the focus of research, eventually yielding solutions to the always increasing demand for data storage capacity in the computer industry [15].

Perhaps the most important application of holography is in interferometric metrology, started by K. Stetson's discovery of holographic interferometry [16, 17]. In holographic interferometry, two or more wave fields are compared interferometrically, at least one of them must be holographically recorded and reconstructed [18]. This technique allows the measurement of changes of the phase of the wave field and thus the change of any physical quantity that affects the phase. The early applications ranged from the first measurement of vibration modes [16,17], over deformation measurement [19–22], contour measurement [23–28], to the determination of refractive index changes [29,30]. These developments were accompanied by rigorous investigations of the underlying principles, mainly performed by K. Stetson [31–35].

For certain arrangements of illumination and observation directions the resulting holographic interference fringes can be interpreted in a first approximation as contour lines of the amplitude of the change of the measured quantity. As an example, a locally higher deformation of a diffusely reflecting surface manifests in a locally higher fringe density. So such areas which give hints to possible material faults, risk of damage, or inadequate design, can easily be detected by applying a load of the same type and direction as the intended operational load, but of much less amplitude. This is the field of HNDT – holographic nondestructive testing.

Besides this qualitative evaluation of the holographic interference patterns there has been continuing work to use holographic interferometry for quantitative measurements. Beginning with manual fringe counting [36, 37], soon image processing computers were employed for quantitative evaluation, a process that consists of recording the reconstructed fringe pattern by TV camera, digitizing and quantizing it, calculating the interference phase distribution from the stored intensity values, using geometry data of the holographic arrangement to determine the distribution of the physical quantity to be measured, and the display of the results. The main one of these named tasks is the calculation of the interference phase. The first algorithms doing this resembled the former fringe counting [38]. A significant step forward in computerized fringe analysis was the introduction of the phase shifting methods of classic interferometric metrology [39,40] into holographic interferometry [41,42]. Now it was possible to measure – and not to estimate by numerical interpolation – the interference phase between the fringe intensity maxima and minima, and also the sign ambiguity was resolved. However, one had to pay for this increased accuracy by additional experimental effort. An alternative without the need for generating several phase shifted interferograms and also without requiring the introduction of a carrier [43] was presented by the author with the Fourier transform evaluation [44]. This is a flexible tool for fitting a linear combination of harmonic functions to the recorded interferogram, taking into account all intensity values even those between the fringe extrema.

While the evaluation of holographic interferograms by computer was successfully developed, there was still the clumsy work of the fabrication of the interference pattern, which was not amenable to computer. The wet chemical processing of the photographic plates, photothermoplastic film, photorefractive crystals, and other recording media showed their typical drawbacks. So the endeavor to record the primary interfering optical fields by the camera of the image processing system and to perform their superposition and thus the generation of the interferogram in the computer generated two solutions: electronic (ESPI), resp. digital (DSPI) speckle pattern interferometry and digital holography (DH) resp. digital holographic interferometry (DHI).

The imaging of diffusely scattering objects with coherent light always produces speckles, the high-contrast granular structure with which the image of the object appears to be covered. If a mutually coherent reference field is superposed to the field scattered by an object, the resulting speckle fields before and after a variation of the object can be added on an intensity basis and yield correlation fringes of the same form as in holographic interferometry [45]. It was recognized that the speckle patterns have a structure easily recordable by existing image sensors, so this metrologic method was automated by computerized recording and processing [46–48]. Due to the analog TV cameras first employed the method was called TV-holography or electronic speckle pattern interferometry (ESPI); later emphasizing the digital recording and processing the name changed to digital speckle pattern interferometry (DSPI). Its big advance was the computerized real-time fringe generation, but the method suffered from the grainy appearance on the TV screen, i. e. severe speckle noise. In the meantime a number of improvements have been achieved with the result that DSPI now is a mature technique with numerous applications in science and technology. Perhaps the most important contribution was the introduction of phase stepping to speckle interferometry by K. Creath [49] and K. Stetson and W. R. Brohinsky [50], resulting in phase stepping digital speckle pattern interferometry (PSDSPI) where optical phase distributions are calculated and compared.

While in ESPI/DSPI the object is focused onto the recording target and the fringes are correlation fringes on an intensity basis, in digital holography a Fresnel or Fraunhofer hologram is recorded. In the following I will give an admittedly “biased” outline of its development. The earliest publication on digital holography that I have found is by J. W. Goodman and R. W. Lawrence [51] and dates back to 1967, so digital holography is older than ESPI/DSPI. In this classic paper Goodman and Lawrence record a wave field using the lensless Fourier transform geometry with a vidicon whose lens assembly was removed. They write: “The output of the vidicon is sampled in a 256×256 array, and quantized to eight grey levels. To avoid aliasing errors, the object-reference-detector geometry is specifically chosen to assure that the maximum spatial frequency in the pattern of interference [the microinterference constituting the hologram, T. K.] is sampled four times per period.” The reconstruction was done on a PDP 6 computer, the squared modulus of the calculated complex distribution was displayed on a scope. The computation of the 256×256 pixel field lasted 5 minutes, “a time which compares favorably with the processing time generally required to obtain a photographic hologram in the conventional manner” as Goodman and Lawrence wrote in 1967.

A further classic paper [52] by T. S. Huang from 1971 treats both categories of digital holography: computer generated holography and computerized reconstruction from holograms. In this paper Fourier transform holograms as well as Fresnel holograms and their numerical reconstruction are discussed, also digitization and quantization effects are considered. In 1972 the work of a Soviet group around L. P. Yaroslavsky was presented [53], and in a paper published 1974 T. H. Demetrakopoulos and R. Mittra [54] consider the computer reconstruction of holograms which are recorded at acoustical or microwave frequencies. In 1980 the book of L. P. Yaroslavsky and N. S. Merzlyakov [55] was translated into English; in this book the theory of computer generation of holograms and of computer reconstruction of holograms is thoroughly treated, and the experiments performed worldwide up to this time are described, especially the work done in the Soviet Union is presented in great detail.

Then there began a long phase in which digital holography in the sense of this book was dormant. Computer generated holograms on the one hand and speckle interferometry on the other hand were fields of active research finding numerous applications. The dormancy of digital holography lasted until the beginning of the 1990s. The young scientist U. Schnars in the department led by the author in the institute (BIAS) directed by W. Jüptner was working towards his doctoral dissertation. Starting with the work of Yaroslavsky [55] and using modern CCD-cameras and computer facilities soon the first digital hologram was recorded and numerically reconstructed. All the time it was a known fact that numerically the whole complex wave field can be reconstructed from a digital hologram. But the emphasis in the first experiments [51–53] was on the intensity distribution. Now the potential lying in the numerically reconstructed phase distribution was recognized, leading in 1993 to digital holographic interferometry as a measurement tool [56]. After the first paper [56] of Schnars soon others followed [57–60] and not much later this approach to holographic metrology was taken up by other research groups. One of the first of these was the group around G. Pedrini [61–65], working during the early days of Schnars' development in a joint project with BIAS. In the context of this project [66], Schnars and I presented our first results. Other groups working in digital holography now can be found in Belgium [67–70], Brazil [71], Canada [72–79], China [80,81], Czech Republic [82], France [83–89], Hong Kong [90,91], Italy [92–95], Japan [96–110], Poland [111], Singapore [112–114], Sweden [115–119], Switzerland [120–129], Turkey [130–132], USA [81, 97, 98, 133–151], to name only some countries in alphabetical order.

Digital holography and digital holographic interferometry now are recognized metrologic methods which receive continuously increasing interest [152, 153]. This is indicated by the steadily increasing number of publications per year related to this topic or by the fact that to the knowledge of the author the "Conference on Interferometry in Speckle Light" in Lausanne/Switzerland in 2000 was the first conference with a session entitled and dedicated only to the topic "digital holography". There is reasonable hope that this technique will yield new interesting results and possibilities, maybe some which are not possible with optical reconstruction. I hope this book will contribute a little bit to the further advance of the promising techniques of digital holography and digital holographic interferometry.

1.3 Holographic Interferometry as a Measurement Tool

In holographic interferometry, two or more wave fields are compared interferometrically, at least one of them must be holographically recorded and reconstructed. The method gives rise to interference patterns whose fringes are determined by the geometry of the holographic setup via the sensitivity vectors and by the optical path length differences. Thus holographic interference patterns can be produced by keeping the optical path length difference constant and changing the sensitivity vectors, by holding the sensitivity vectors constant and varying the optical path length differences, or by altering both of them between the object states to be compared. Especially the path lengths can be modified by a number of physical parameters. The flexibility and the precision gained by comparing the optical path length changes with the wavelength of the laser light used, make holographic interferometry an ideal means for measuring a manifold of physical quantities [154–156]. The main advantages are:

- The measurements are contactless and noninvasive. In addition to an eventual loading for inducing the optical pathlength changes, the object is only impinged by light waves. The intensities of these waves are well below the level for causing any damage, even for the most delicate of biological objects.
- A reliable analysis can be performed at low loading intensities: the testing remains non-destructive.
- Not only may two states separated by a long time be compared, but furthermore the generation and evaluation of the holographic information can be separated both temporally and locally.
- Measurements can be made through transparent windows. We can therefore make measurements in pressure or vacuum chambers or protect against hostile environments. Due to the measurement of differences of the optical path lengths instead of absolute values, low quality windows do not disturb the results.
- Holographic interferometric measurements can be accomplished at moving surfaces: Short pulse illumination makes the method insensitive to a disturbing motion, vibrations can be investigated, the holographic setup can be made insensitive to specific motion components, and the rotation of spinning objects can be cancelled optically by using an image derotator.
- Deformation measurements can be performed at rough, diffusely reflecting surfaces, which occur frequently in engineering. No specular reflection of the object is required.
- The objects to be examined holographically may be of almost arbitrary shape. Using multiple illumination and observation directions or fiber optics, barely accessible areas can be studied.
- Holographic interferometry is nearly independent of the state of matter: Deformations of hard and soft materials can be measured. Refractive index variations in solids, fluids, gases and even plasmas can be determined.
- Lateral dimensions of the examined subjects may range from a few millimeters to several meters.
- The measurement range extends roughly speaking from a hundredth to several hundreds of a wavelength, for example displacements can be measured from about 0.005 μm to 500 μm.
- The achievable resolution and accuracy of a holographic interferometric displacement measurement permit subsequent numerical strain and stress calculations.

- Two-dimensional spatially continuous information is obtained: local singularities, for example local deformation extrema, cannot go undetected.
- Multiple viewing directions using a single hologram are possible, enabling the application of computerized tomography to obtain three-dimensional fields.

2 Optical Foundations of Holography

This chapter discusses the physical basis of holography and holographic interferometry. The primary phenomena constituting holography are interference and diffraction, which take place because of the wave nature of light. So this chapter begins with a description of the wave theory of light as far as it is required to understand the recording and reconstruction of holograms and the effect of holographic interferometry. In holographic interferometry the variation of a physical parameter is measured by its influence on the phase of an optical wave field. Therefore the dependence of the phase upon the geometry of the optical setup and the different parameters to be measured is outlined.

2.1 Light Waves

2.1.1 Solutions of the Wave Equation

Light is a transverse, electromagnetic wave characterized by time-varying electric and magnetic fields. Since electromagnetic waves obey the Maxwell equations, the propagation of light is described by the wave equation which follows from the Maxwell equations. The *wave equation* for propagation of light in vacuum is

$$\nabla^2 \boldsymbol{E} - \frac{1}{c^2}\frac{\partial^2 \boldsymbol{E}}{\partial t^2} = 0 \tag{2.1}$$

where $\boldsymbol{E}$ is the *electric field strength*, ∇^2 is the *Laplace operator*

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{2.2}$$

(x, y, z) are the Cartesian spatial coordinates, t denotes the temporal coordinate, the time, and c is the propagation speed of the wave. The *speed of light* in vacuum c_0 is a constant of nature

$$c_0 = 299\,792\,458\ \mathrm{m\,s^{-1}} \qquad \text{or almost exactly} \quad c_0 = 3 \times 10^8\ \mathrm{m\,s^{-1}}. \tag{2.3}$$

Transverse waves vibrate at right angles to the direction of propagation and so they must be described in vector notation. The wave may vibrate horizontally, vertically, or in any direction combined of these. Such effects are called *polarization* effects. Fortunately for most applications it is not necessary to use the full vectorial description of the fields, so we can assume a wave vibrating in a single plane. Such a wave is called *plane polarized*. For a plane polarized wave field propagating in the z-direction the *scalar wave equation* is sufficient

$$\frac{\partial^2 E}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 E}{\partial t^2} = 0. \tag{2.4}$$

It is easily verified that

$$E(z,t) = f(z-ct) \qquad \text{or} \qquad E(z,t) = g(z+ct) \tag{2.5}$$

are also solutions of this equation, which means that the wave field retains its form during propagation. Due to the linearity of (2.4)

$$E(z,t) = a\,f(z-ct) + b\,g(z+ct) \tag{2.6}$$

is likewise a solution to the wave equation. This *superposition principle* is valid for linear differential equations in general and thus for (2.1) also.

The most important solution of (2.4) is the *harmonic wave*, which in real notation is

$$E(z,t) = E_0\cos(kz-\omega t). \tag{2.7}$$

E_0 is the *real amplitude* of the wave, the term $(kz-\omega t)$ gives the *phase* of the wave. The *wave number* k is associated to the *wavelength* λ by

$$k = \frac{2\pi}{\lambda}. \tag{2.8}$$

Typical figures of λ for visible light are 514.5 nm (green line of argon-ion laser) or 632.8 nm (red light of helium-neon laser). The *angular frequency* ω is related to the *frequency* ν of the wave by

$$\omega = 2\pi\nu \tag{2.9}$$

where ν is the number of periods per second, that means

$$\nu = \frac{c}{\lambda} \qquad \text{or} \qquad \nu\lambda = c. \tag{2.10}$$

If we have not the maximum amplitude at $x=0$ and $t=0$, we have to introduce the *relative phase* ϕ

$$E(z,t) = E_0\cos(kz-\omega t+\phi). \tag{2.11}$$

With the *period* T, the time for a full 2π-cycle, we can write

$$E(z,t) = E_0\cos\left(\frac{2\pi}{\lambda}z - \frac{2\pi}{T}t + \phi\right). \tag{2.12}$$

Figure 2.1 displays two aspects of this wave. Figure 2.1a shows the temporal distribution of the field at two points $z=0$ and $z=z_1>0$, and Fig. 2.1b gives the spatial distribution of

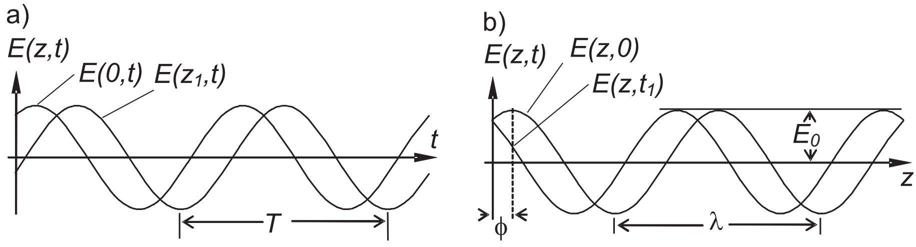

Figure 2.1: Spatial and temporal distribution of a scalar harmonic wave.

two periods for time instants $t = 0$ and $t = t_1 > 0$. We see that a point of constant phase moves with the so called *phase velocity*, the speed c.

The use of trigonometric functions leads to cumbersome calculations, which can be circumvented by using the complex exponential which is related to the trigonometric functions by *Euler's formula*

$$e^{i\alpha} = \cos\ \alpha + i\ \sin\ \alpha \tag{2.13}$$

where $i = \sqrt{-1}$ is the imaginary unit. Since the cosine now is

$$\cos\ \alpha = \frac{1}{2}(e^{i\alpha} + e^{-i\alpha}) \tag{2.14}$$

the harmonic wave (2.11) is

$$E(z,t) = \frac{1}{2}E_0\ e^{i(kz - \omega t + \phi)} + \frac{1}{2}E_0\ e^{-i(kz - \omega t + \phi)}. \tag{2.15}$$

The second term on the right-hand side is the complex conjugate of the first term and can be omitted as long as it is understood that only the real part of $E(z,t)$ represents the physical wave. Thus the harmonic wave in complex notation is

$$E(z,t) = \frac{1}{2}E_0\ e^{i(kz - \omega t + \phi)}. \tag{2.16}$$

A *wavefront* refers to the spatial distribution of the maxima of the wave, or other surfaces of constant phase, as these surfaces propagate. The wavefronts are normal to the direction of propagation. A *plane wave* is a wave which has constant phase in all planes orthogonal to the propagation direction for a given time t. For describing the spatial distribution of the wave, we can assume $t = 0$ in an arbitrary time scale. Since

$$\boldsymbol{k} \cdot \boldsymbol{r} = \text{const} \tag{2.17}$$

is the equation for a plane in three-dimensional space, with the *wave vector* $\boldsymbol{k} = (k_x, k_y, k_z)$ and the spatial vector $\boldsymbol{r} = (x, y, z)$, a plane harmonic wave at time $t = 0$ is

$$E(\boldsymbol{r}) = E_0\ e^{i(\boldsymbol{k} \cdot \boldsymbol{r} + \phi)}. \tag{2.18}$$

This wave repeats after the wavelength λ in direction $\boldsymbol{k}$, which can easily be proved using $|\boldsymbol{k}| = k = 2\pi/\lambda$ by

$$E\left(\boldsymbol{r} + \lambda\frac{\boldsymbol{k}}{k}\right) = E(\boldsymbol{r}). \tag{2.19}$$

The expression

$$E(\boldsymbol{r},t) = E_0\, \mathrm{e}^{\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{r} - \omega t + \phi)} \tag{2.20}$$

describes the temporal dependence of a plane harmonic wave propagating in the direction of the wavevector or

$$E(\boldsymbol{r},t) = E_0\, \mathrm{e}^{\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{r} + \omega t + \phi)} \tag{2.21}$$

if the wave propagates contrary to the direction of $\boldsymbol{k}$.

Another waveform often used is the *spherical wave* where the phase is constant on each spherical surface. The importance of spherical waves comes from the *Huygens principle* which states that each point on a propagating wavefront can be considered as radiating itself a spherical wavelet.

For a mathematical treatment of spherical waves the wave equation has to be described in polar coordinates (r, θ, ψ), transformed by $x = r\sin\theta\cos\psi$, $y = r\sin\theta\sin\psi$, $z = r\cos\theta$. Due to the spherical symmetry, a spherical wave is not dependent on θ and ψ. Then the scalar wave equation is

$$\frac{1}{r}\frac{\partial^2}{\partial r^2}(rE) - \frac{1}{c^2}\frac{\partial^2 E}{\partial t^2} = 0. \tag{2.22}$$

The solutions of main interest are the harmonic spherical waves

$$E(r,t) = \frac{E_0}{r}\, \mathrm{e}^{\mathrm{i}(kr - \omega t + \phi)}. \tag{2.23}$$

One observes that the amplitude E_0/r decreases proportionally to $1/r$. Furthermore at a long distance from the origin the spherical wave locally approximates a plane wave.

The complex amplitudes of wavefronts scattered by a surface are generally very complicated, but due to the superposition principle (2.6) they can be treated as the sum of plane waves or spherical waves. There are still other solutions to the wave equation. An example are the *Bessel waves* of the class of *nondiffracting beams* [157]. But up to now they have not found applications in holographic interferometry, so here we restrict ourselves on the plane and on the spherical waves.

2.1.2 Intensity

The only parameter of light which is directly amenable to sensors – eye, photodiode, CCD-target, etc. – is the *intensity* (and in a rough scale the frequency as color). Intensity is defined by the energy flux through an area per time. From the Maxwell equations we get

$$I = \varepsilon_0 c E^2 \tag{2.24}$$

where we only use the proportionality of the intensity I to E^2

$$I \sim E^2. \tag{2.25}$$

It has to be recognized that the intensity has a nonlinear dependence on the electric field strength. Since there is no sensor which can follow the frequency of light, we have to integrate over a *measuring time* T_m, the momentary intensity is not measurable. So if $T_m \gg T = 2\pi/\omega$, omitting proportionality constants we define

$$I = E_0\, E_0^* = |E_0|^2 \tag{2.26}$$

where * denotes the complex conjugate. The intensity of a general stationary wave field is

$$I(\boldsymbol{r}) = \langle E\, E^* \rangle = \lim_{T_m \to \infty} \frac{1}{T_m} \int_{-T_m/2}^{T_m/2} E(\boldsymbol{r}, t') E^*(\boldsymbol{r}, t')\, dt'. \tag{2.27}$$

This intensity is the limit of the *short time intensity*

$$I(\boldsymbol{r}, t, T_m) = \frac{1}{T_m} \int_{t-T_m/2}^{t+T_m/2} E(\boldsymbol{r}, t') E^*(\boldsymbol{r}, t')\, dt' \tag{2.28}$$

which is a sliding average of a temporal window centered around t with width T_m. The measuring time T_m always is large compared with the period of the light wave but has to be short in the time scale of the investigated process.

2.2 Interference of Light

2.2.1 Interference of Two Waves with Equal Frequency

The *interference* effect which occurs if two or more coherent light waves are superposed, is the basis of holography and holographic interferometry. So in this *coherent superposition* we consider two waves, emitted by the same source, which differ in the directions $\boldsymbol{k_1}$ and $\boldsymbol{k_2}$, and the phases ϕ_1 and ϕ_2, but for convenience have the same amplitude E_0 and frequency ω and are linearly polarized in the same direction. Then in scalar notation

$$\begin{aligned} E_1(\boldsymbol{r}, t) &= E_0\, \mathrm{e}^{\mathrm{i}(\boldsymbol{k_1} \cdot \boldsymbol{r} - \omega t + \phi_1)} \\ E_2(\boldsymbol{r}, t) &= E_0\, \mathrm{e}^{\mathrm{i}(\boldsymbol{k_2} \cdot \boldsymbol{r} - \omega t + \phi_2)}. \end{aligned} \tag{2.29}$$

For determination of the superposition of these waves we decompose the vectors $\boldsymbol{k_1}$ and $\boldsymbol{k_2}$ into components of equal and opposite directions, Fig. 2.2:, $\boldsymbol{k}' = (\boldsymbol{k_1} + \boldsymbol{k_2})/2$ and $\boldsymbol{k}'' = (\boldsymbol{k_1} - \boldsymbol{k_2})/2$. If θ is the angle between $\boldsymbol{k_1}$ and $\boldsymbol{k_2}$ then

$$|\boldsymbol{k}''| = \frac{2\pi}{\lambda} \sin\frac{\theta}{2}. \tag{2.30}$$

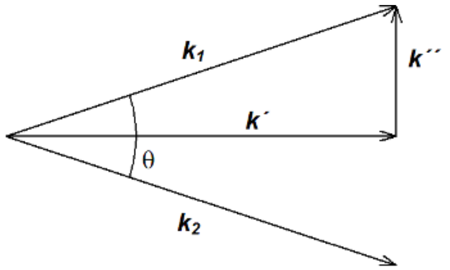

Figure 2.2: Decomposition of wave vectors.

In the same way we define the mean phase $\phi = (\phi_1 + \phi_2)/2$ and the half phase difference $\Delta\phi = (\phi_1 - \phi_2)/2$. Now the superposition gives the field

$$\begin{aligned}(E_1 + E_2)(\boldsymbol{r}, t) &= E_0\,\mathrm{e}^{\mathrm{i}(\boldsymbol{k_1}\cdot\boldsymbol{r} - \omega t + \phi_1)} + E_0\,\mathrm{e}^{\mathrm{i}(\boldsymbol{k_2}\cdot\boldsymbol{r} - \omega t + \phi_2)} \\ &= E_0\big\{\mathrm{e}^{\mathrm{i}(\boldsymbol{k}'\cdot\boldsymbol{r} + \boldsymbol{k}''\cdot\boldsymbol{r} - \omega t + \phi + \Delta\phi)} \\ &\quad + \mathrm{e}^{\mathrm{i}(\boldsymbol{k}'\cdot\boldsymbol{r} - \boldsymbol{k}''\cdot\boldsymbol{r} - \omega t + \phi - \Delta\phi)}\big\} \\ &= E_0\,\mathrm{e}^{\mathrm{i}(\boldsymbol{k}'\cdot\boldsymbol{r} - \omega t + \phi)}\big\{\mathrm{e}^{\mathrm{i}(\boldsymbol{k}''\cdot\boldsymbol{r} + \Delta\phi)} + \mathrm{e}^{\mathrm{i}(-\boldsymbol{k}''\cdot\boldsymbol{r} - \Delta\phi)}\big\} \\ &= 2\,E_0\,\mathrm{e}^{\mathrm{i}(\boldsymbol{k}'\cdot\boldsymbol{r} - \omega t + \phi)}\cos(\boldsymbol{k}''\cdot\boldsymbol{r} + \Delta\phi).\end{aligned} \tag{2.31}$$

In this field the exponential term is a temporally varying phase but the cosine term is independent of time. Thus we get the temporally constant intensity

$$\begin{aligned} I(\boldsymbol{r}) &= (E_1 + E_2)(E_1 + E_2)^* \\ &= 4\,E_0^2\,\cos^2(\boldsymbol{k}''\cdot\boldsymbol{r} + \Delta\phi). \end{aligned} \tag{2.32}$$

This means the intensity is minimal where $\cos^2(\boldsymbol{k}''\cdot\boldsymbol{r} + \Delta\phi) = 0$. These are the loci where

$$\boldsymbol{k}''\cdot\boldsymbol{r} + \Delta\phi = (2n+1)\frac{\pi}{2} \qquad n \in \mathbb{Z}. \tag{2.33}$$

Here the wavefronts are said to be *anti-phase*, we speak of destructive interference. The intensity is maximal where

$$\boldsymbol{k}''\cdot\boldsymbol{r} + \Delta\phi = n\,\pi \qquad n \in \mathbb{Z}. \tag{2.34}$$

Here the wavefronts are *in-phase*, we have constructive interference.

The resulting time independent pattern is called an *interference pattern*, the fringes are called *interference fringes*. For plane waves they are oriented parallel to $\boldsymbol{k}'$ and have a distance of $\pi/|\boldsymbol{k}''|$ in the direction $\boldsymbol{k}''$. This is shown in moiré analogy in Fig. 2.3.

2.2.2 Interference of Two Waves with Different Frequencies

In the following we investigate the interference of two waves where not only the propagation directions and the phases but additionally the frequencies $\nu_i = \omega_i/(2\pi)$ are different.

$$\begin{aligned} E_1(\boldsymbol{r}, t) &= E_0\,\mathrm{e}^{\mathrm{i}(\boldsymbol{k_1}\cdot\boldsymbol{r} - 2\pi\nu_1 t + \phi_1)} \\ E_2(\boldsymbol{r}, t) &= E_0\,\mathrm{e}^{\mathrm{i}(\boldsymbol{k_2}\cdot\boldsymbol{r} - 2\pi\nu_2 t + \phi_2)}. \end{aligned} \tag{2.35}$$

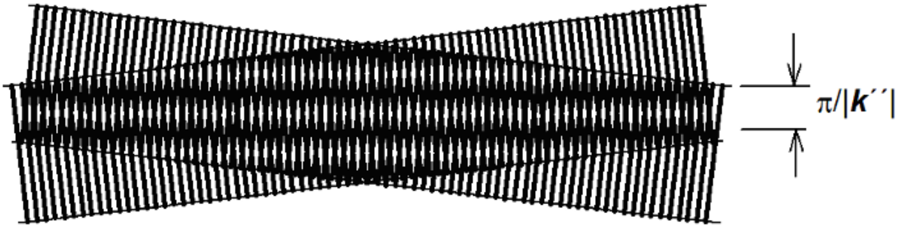

Figure 2.3: Interference fringes constant in time.

Besides the definitions of $\boldsymbol{k}'$, $\boldsymbol{k}''$, ϕ and $\Delta\phi$ now let $\nu = (\nu_1 + \nu_2)/2$ and $\Delta\nu = (\nu_1 - \nu_2)/2$. Then we have

$$\begin{aligned}(E_1 + E_2)(\boldsymbol{r}, t) &= E_0 \{ \mathrm{e}^{\mathrm{i}(\boldsymbol{k}' \cdot \boldsymbol{r} + \boldsymbol{k}'' \cdot \boldsymbol{r} - 2\pi\nu t - 2\pi\Delta\nu t + \phi + \Delta\phi)} \\ &\quad + \mathrm{e}^{\mathrm{i}(\boldsymbol{k}' \cdot \boldsymbol{r} - \boldsymbol{k}'' \cdot \boldsymbol{r} - 2\pi\nu t + 2\pi\Delta\nu t + \phi - \Delta\phi)} \} \\ &= E_0 \, \mathrm{e}^{\mathrm{i}(\boldsymbol{k}' \cdot \boldsymbol{r} - 2\pi\nu t + \phi)} \{ \mathrm{e}^{\mathrm{i}(\boldsymbol{k}'' \cdot \boldsymbol{r} - 2\pi\Delta\nu t + \Delta\phi)} \\ &\quad + \mathrm{e}^{\mathrm{i}(-\boldsymbol{k}'' \cdot \boldsymbol{r} + 2\pi\Delta\nu t - \Delta\phi)} \} \\ &= 2\, E_0 \, \mathrm{e}^{\mathrm{i}(\boldsymbol{k}' \cdot \boldsymbol{r} - 2\pi\nu t + \phi)} \cos(\boldsymbol{k}'' \cdot \boldsymbol{r} - 2\pi\Delta\nu t + \Delta\phi)\end{aligned} \tag{2.36}$$

and the intensity is

$$\begin{aligned} I(\boldsymbol{r}, t) &= 4\, E_0^2 \, \cos^2(\boldsymbol{k}'' \cdot \boldsymbol{r} + \Delta\phi - 2\pi\Delta\nu t) \\ &= 2\, E_0^2 \big[1 + \cos(2\boldsymbol{k}'' \cdot \boldsymbol{r} + 2\Delta\phi - 4\pi\Delta\nu t) \big] . \end{aligned} \tag{2.37}$$

If the frequency difference is small enough, $\nu_1 \approx \nu_2$, a detector can register an intensity at $\boldsymbol{r}$ oscillating with the *beat frequency* $2\Delta\nu = \nu_1 - \nu_2$. The phase of this modulation is the phase difference $2\Delta\phi = \phi_1 - \phi_2$ of the superposed waves. Contrary to the frequencies of the optical waves the beat frequency can be measured electronically and further evaluated as long as it remains in the kHz or MHz range. The measurement of the beat frequency $\Delta\nu$ enables one to calculate the motion of a reflector via the *Doppler shift* or to determine the phase difference $\Delta\phi$ between different points of an object where the intensity oscillates with the same constant beat frequency.

2.2.3 Interference of Two Waves with Different Amplitudes

If we have plane linearly polarized waves of the same frequency, but different direction and phase and moreover different amplitudes

$$\begin{aligned} E_1(\boldsymbol{r}, t) &= E_{01} \, \mathrm{e}^{\mathrm{i}(\boldsymbol{k_1} \cdot \boldsymbol{r} - \omega t + \phi_1)} \\ E_2(\boldsymbol{r}, t) &= E_{02} \, \mathrm{e}^{\mathrm{i}(\boldsymbol{k_2} \cdot \boldsymbol{r} - \omega t + \phi_2)} \end{aligned} \tag{2.38}$$

we get the intensity

$$\begin{aligned} I(\boldsymbol{r},t) &= \left(E_{01}\,\mathrm{e}^{\mathrm{i}(\boldsymbol{k_1}\cdot\boldsymbol{r}-\omega t+\phi_1)} + E_{02}\,\mathrm{e}^{\mathrm{i}(\boldsymbol{k_2}\cdot\boldsymbol{r}-\omega t+\phi_2)}\right) \\ &\quad \times \left(E_{01}\,\mathrm{e}^{-\mathrm{i}(\boldsymbol{k_1}\cdot\boldsymbol{r}-\omega t+\phi_1)} + E_{02}\,\mathrm{e}^{-\mathrm{i}(\boldsymbol{k_2}\cdot\boldsymbol{r}-\omega t+\phi_2)}\right) \\ &= E_{01}^2 + E_{02}^2 + E_{01}E_{02}\left\{\mathrm{e}^{\mathrm{i}(\boldsymbol{k_1}\cdot\boldsymbol{r}-\boldsymbol{k_2}\cdot\boldsymbol{r}+\phi_1-\phi_2)}\right. \\ &\quad \left. + \,\mathrm{e}^{\mathrm{i}(\boldsymbol{k_2}\cdot\boldsymbol{r}-\boldsymbol{k_1}\cdot\boldsymbol{r}+\phi_2-\phi_1)}\right\} \\ &= E_{01}^2 + E_{02}^2 + 2E_{01}E_{02}\cos(2\boldsymbol{k}''\cdot\boldsymbol{r}+2\Delta\phi). \end{aligned} \tag{2.39}$$

This result can be written as

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos(2\boldsymbol{k}''\cdot\boldsymbol{r}+2\Delta\phi) \tag{2.40}$$

or using the identity $\cos\alpha = 2\cos^2(\alpha/2) - 1$ for comparison with (2.32) as

$$I = E_{01}^2 + E_{02}^2 + 4E_{01}E_{02}\cos^2(\boldsymbol{k}''\cdot\boldsymbol{r}+\Delta\phi) - 2E_{01}E_{02}. \tag{2.41}$$

The special case $E_{01} = E_{02} = E_0$ gives (2.32).

In general the result of superposing two waves consists of one part that is the addition of the intensities and another part, the interference term, (2.40). Up to now we only have investigated *parallelly polarized waves*. The other extreme are *orthogonally polarized waves*. These waves do not interfere, their superposition only consists of the addition of the intensities

$$I = I_1 + I_2. \tag{2.42}$$

For other angles between the polarization directions the field vector has to be decomposed into components of parallel and orthogonal polarizations, the result contains interference parts as well as an addition of intensities.

Reasons for the additive intensity term not only may be mutually oblique polarization directions or different intensities, but also an insufficient coherence of the interfering waves. Because in the superposition of incoherent light we always observe a pure addition of the intensities but no interference, the additive term often is called the *incoherent part*, or we speak of *incoherent superposition*.

The *visibility* or *contrast* of the interference pattern is defined by

$$V = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}. \tag{2.43}$$

If two parallel polarized waves of the same intensity interfere, we have the maximal contrast of $V = 1$; we have minimal contrast $V = 0$ for incoherent superposition. For example, if the ratio of the intensities of interfering waves is 5:1, the contrast is 0.745.

2.3 Coherence

With sunlight or lamplight we rarely observe interference. Only light of sufficient coherence will exhibit this effect. Roughly speaking coherence means the ability of light waves to interfere. Precisely, coherence describes the correlation between individual light waves. The two aspects of the general spatio-temporal coherence are the temporal and the spatial coherence.