Digital Holography and Wavefront Sensing
Digital Holography and Wavefront Sensing
Principles, Techniques and Applications

Second Edition
Preface to the Second Edition

As we sat down to consider writing a new edition of *Digital Holography*, we, the original authors (U. Schnars and W. Jüptner), asked ourselves if the field had advanced sufficiently with enough new and novel developments to merit a second edition. The answer was an overwhelming YES and came from seeing the profound developments and scale of applications to which digital holography, and in the wider context 3D imaging technologies in general, are now being routinely applied. In the intervening years, the evolution of digital holography has been, both, extensive and dramatic.

Some of the areas in which we have seen considerable advances and application include computational wave field sensing and digital holographic microscopy, with a huge number of papers being published in these and related fields. To reflect these advances adequately in our book and to broaden its scope, we invited Claas Falldorf (BIAS) and John Watson (University of Aberdeen) to join us as co-authors. Claas works actively in wave field sensing using computational methods such as phase retrieval or computational shear interferometry. John is an international expert in digital holographic microscopy and, particularly, to underwater holography of aquatic organisms and particles; and also 3DTV and related fields. Both are ideal partners to support the approach and philosophy of the new edition.

Accordingly, this second edition has been significantly revised and enlarged. We have extended the chapter on Digital Holographic Microscopy to incorporate new sections on particle sizing, particle image velocimetry and underwater holography. A new chapter now deals comprehensively and extensively with computational wave field sensing. These techniques represent a fascinating alternative to standard interferometry and Digital Holography. They enable wave field sensing without the requirement of a particular reference wave, thus allowing the use of low brilliance light sources and even liquid-crystal displays (LCD) for interferometric applications. We believe that, in the coming years, computational wave field sensing will prove to be an excellent complement to Digital Holography to determine the full complex amplitude of wave fields.

All the authors wish to thank colleagues past and present (too numerous to mention) with whom they have worked over the years. As with the first edition,
several pictures and figures in this book originate from common publications with other colleagues and we thank them for permission to describe their work and to use their pictures. All of our co-workers are gratefully acknowledged.

Bremen, May 2014
Aberdeen

Ulf Schnars
Claas Falldorf
John Watson
Werner Jüptner
Preface to the First Edition

*Sag’ ich zum Augenblicke verweile doch, Du bist so schön*

J.W.v. Goethe, “Faust”

An old dream of mankind and a sign of culture is the conservation of moments by taking an image of the world around. Pictures accompany the development of mankind. However, a picture is the two-dimensional projection of the three-dimensional world. The perspective—recognized in Europe in the Middle Ages—was a first approach to overcome the difficulties of imaging close to reality. It took up to the twentieth century to develop a real three-dimensional imaging: Gabor invented holography in 1948. Yet still one thing was missing: the phase of the object wave could be reconstructed optically but not be measured directly. The last huge step to the complete access of the object wave was Digital Holography. By Digital Holography the intensity and the phase of electromagnetic wave fields can be measured, stored, transmitted, applied to simulations and manipulated in the computer: An exciting new tool for the handling of light.

We started our work in the field of Digital Holography in 1990. Our motivation mainly came from Holographic Interferometry, a method used with success for precise measurement of deformation and shape of opaque bodies or refractive index variations within transparent media. A major drawback of classical HI using photographic plates was the costly process of film development. Even thermoplastic films used as recording medium did not solve the hologram development problem successfully. On the other hand the Electronic Speckle Pattern Interferometry (ESPI) and its derivate digital shearography reached a degree mature for applications in industry. Yet, with these speckle techniques the recorded images are only correlated and not reconstructed as for HI. Characteristic features of holography like the possibility to refocus on other object planes in the reconstruction process are not possible with speckle metrology.

Our idea was to transfer all methods of classical HI using photographic plates to Digital Holography. Surprisingly, we discovered that Digital Holography offers more possibilities than classical HI: The wavefronts can be manipulated in the numerical reconstruction process, enabling operations not possible in optical holography. Especially the interference phase can be calculated directly from the holograms, without evaluation of an interference pattern.
The efficiency of Digital Holography depends strongly on the resolution of the electronic target used to record the holograms. When we made our first experiments in the 1990s of the last century, Charged Coupled Devices began to replace analogue sensors in cameras. The resolution of commercially available cameras was quite low, about some hundred pixels per line, and the output signal of cameras already equipped with CCDs was still analogue. In those days, digital sampling of camera images and running of routines for numerical hologram reconstruction was only possible on special digital image processing hardware and not, as today, on ordinary PCs. The reconstruction of a hologram digitized with $512 \times 512$ pixels took about half an hour in 1991 on a Digital Image Processing unit developed at BIAS especially for optical metrology purposes. Nevertheless we made our first experiments with these types of cameras. Today, numerical reconstruction of holograms with 1 million pixel is possible nearly in real time on state-of-the-art PCs.

Then, fully digital CCD cameras with 1 million pixels and smaller pixels than those of the previous camera generation emerged on the market. These cameras showed better performance and first applications in optical metrology became possible. Today, digital CCD cameras with 4 million pixels are standard.

The tremendous development in opto-electronics and in data processing pushed Digital Holography to new perspectives: It is applied with success in optical deformation and strain analysis, shape measurement, microscopy and for investigations of flows in liquids and gases. In this book we make the trial to describe the principles of this method and to report on the various applications. We took pains to prepare the manuscript carefully and to avoid mistakes. However, we are not perfect. Comments, suggestions for improvements or corrections are therefore welcome and will be considered in potential further editions.

Some pictures in this book originate from common publications with other co-authors. All of our co-workers, especially W. Osten, Th. Kreis, D. Holstein, S. Seebacher, H.-J. Hartmann and V. Kebbel are gratefully acknowledged.
Contents

1 Introduction ................................................................. 1

2 Fundamental Principles of Holography .............................. 5
   2.1 Light Waves ............................................................... 5
   2.2 Interference ............................................................... 8
   2.3 Coherence ................................................................. 10
      2.3.1 General ............................................................... 10
      2.3.2 Temporal Coherence .............................................. 11
      2.3.3 Spatial Coherence ............................................... 13
   2.4 Diffraction ............................................................... 15
   2.5 Speckle ................................................................. 18
   2.6 Holography .............................................................. 20
      2.6.1 Hologram Recording and Reconstruction ..................... 20
      2.6.2 The Imaging Equations ......................................... 23
   2.7 Holographic Interferometry ........................................ 25
      2.7.1 Generation of Holographic Interferograms ................. 25
      2.7.2 Displacement Measurement by HI ............................ 28
      2.7.3 Holographic Contouring ....................................... 30
      2.7.4 Refractive Index Measurement by HI ........................ 34
      2.7.5 Phase Shifting HI .............................................. 35
      2.7.6 Phase Unwrapping ............................................. 37

3 Digital Holography ......................................................... 39
   3.1 General Principles .................................................... 39
   3.2 Numerical Reconstruction ......................................... 42
      3.2.1 Reconstruction by the Fresnel Approximation ............. 42
      3.2.2 Reconstruction by the Convolution Approach ............. 49
      3.2.3 Digital Fourier Holography ................................ 52
   3.3 Shift and Suppression of DC-Term and Conjugate Image ...... 53
      3.3.1 Suppression of the DC Term .................................. 53
      3.3.2 Tilted Reference Wave ........................................ 55
      3.3.3 Phase Shifting Digital Holography ......................... 56
### 4 Digital Holographic Interferometry (DHI)

- **4.1 General Principles**
- **4.2 Deformation Measurement**
  - **4.2.1 Quantitative Displacement Measurement**
  - **4.2.2 Mechanical Materials Properties**
  - **4.2.3 Thermal Materials Properties**
  - **4.2.4 Non-destructive Testing**
- **4.3 Shape Measurement**
  - **4.3.1 Two-Illumination-Point Method**
  - **4.3.2 Two- and Multi-wavelength Method**
  - **4.3.3 Hierarchical Phase Unwrapping**
- **4.4 Measurement of Refractive Index Variations**

### 5 Digital Holographic Particle Sizing and Microscopy

- **5.1 Introduction**
- **5.2 Recording and Replay Conditions**
  - **5.2.1 In-line Recording**
  - **5.2.2 Off-axis Recording**
  - **5.2.3 Image Resolution**
  - **5.2.4 Holographic Depth-of-Field and Depth-of-Focus**
  - **5.2.5 Optical Aberrations**
- **5.3 Data Processing and Autofocusing of Holographic Images**
- **5.4 Some Applications in Imaging and Particle Sizing**
  - **5.4.1 Particle Sizing**
  - **5.4.2 Digital Holographic Microscopy (DHM)**
  - **5.4.3 Holographic Tomography**
  - **5.4.4 Phase Shifting DHM**
  - **5.4.5 Particle Image Velocimetry (PIV)**
  - **5.4.6 Underwater Digital Holography**

### 6 Special Techniques

- **6.1 Applications Using Short Coherence Length Light**
  - **6.1.1 Light-in-Flight Measurements**
  - **6.1.2 Short-Coherence Tomography**
- **6.2 Endoscopic Digital Holography**
- **6.3 Optical Reconstruction of Digital Holograms**
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.4</td>
<td>Comparative Digital Holography</td>
<td>131</td>
</tr>
<tr>
<td>6.4.1</td>
<td>Fundamentals of Comparative Holography</td>
<td>131</td>
</tr>
<tr>
<td>6.4.2</td>
<td>Comparative Digital Holography</td>
<td>132</td>
</tr>
<tr>
<td>6.5</td>
<td>Encrypting of Information with Digital Holography</td>
<td>135</td>
</tr>
<tr>
<td>6.6</td>
<td>Synthetic Aperture Holography</td>
<td>137</td>
</tr>
<tr>
<td>6.7</td>
<td>Holographic Pinhole Camera</td>
<td>138</td>
</tr>
<tr>
<td>7</td>
<td>Computational Wavefield Sensing</td>
<td>141</td>
</tr>
<tr>
<td>7.1</td>
<td>Overview</td>
<td>141</td>
</tr>
<tr>
<td>7.2</td>
<td>Phase Retrieval</td>
<td>142</td>
</tr>
<tr>
<td>7.2.1</td>
<td>Projection Based Methods</td>
<td>144</td>
</tr>
<tr>
<td>7.2.2</td>
<td>Gradient Search Methods</td>
<td>154</td>
</tr>
<tr>
<td>7.2.3</td>
<td>Deterministic Methods</td>
<td>156</td>
</tr>
<tr>
<td>7.3</td>
<td>Shear Interferometry for Wavefield Sensing</td>
<td>162</td>
</tr>
<tr>
<td>7.3.1</td>
<td>Wavefront Reconstruction</td>
<td>164</td>
</tr>
<tr>
<td>7.3.2</td>
<td>Computational Shear Interferometry</td>
<td>177</td>
</tr>
<tr>
<td>7.4</td>
<td>Shack-Hartmann Wavefront Sensing</td>
<td>183</td>
</tr>
<tr>
<td>8</td>
<td>Speckle Metrology</td>
<td>185</td>
</tr>
<tr>
<td>8.1</td>
<td>Electronic Speckle Pattern Interferometry (ESPI)</td>
<td>185</td>
</tr>
<tr>
<td>8.2</td>
<td>Digital Shearography</td>
<td>189</td>
</tr>
<tr>
<td>8.3</td>
<td>Digital Speckle Photography</td>
<td>193</td>
</tr>
<tr>
<td>8.4</td>
<td>Comparison of Conventional HI, ESPI and Digital HI</td>
<td>194</td>
</tr>
</tbody>
</table>

Appendix A: The Fourier Transform | 199  |

Appendix B: Phase Transformation of a Spherical Lens | 203  |

Appendix C: Simple Reconstruction Routines | 207  |

References | 211  |

Index | 223  |
Chapter 1
Introduction

The recording and storage of full-parallax 3D images was and is a recurring goal of science and engineering since the first photographs were made. To accomplish this, the whole (“holos” in Greek) optical information emanating from a source needs to be written (“graphein” in Greek), recorded or captured on a sensing device for later recreation or reconstruction of the original object. This is the technique we now know as holography.

The history of holography started in principle when Lord Rayleigh experimentally created a Fresnel lens [268] and showed the generation of an interference pattern by the superposition of a spherical wave with a planar wave. In holography the planar wave is regarded as the reference wave and the spherical wave represents the object. The Fresnel lens in this sense can be regarded as the hologram of a point source. However, it was Denis Gabor who recognized that the same procedure carried out over a number of point’s leads to the ability to optically reconstruct their position in space. Consequently, he coined the name “holography” since he was able to reconstruct the amplitude and phase of a wave [68–70].

A holographically stored image or hologram in the classical sense is a photographically, or otherwise, recorded interference pattern between a wave field scattered from an object and a coherent background denoted as the reference wave. A hologram is usually recorded on a flat two-dimensional surface, but contains the entire information about the three-dimensional wave field. This information is encoded in the form of interference fringes, usually not visible to the human eye due to their high spatial frequencies. The object wave can be recovered by illuminating the hologram with the original reference wave. This reconstructed wave is optically indistinguishable from the original object wave by passive means. An observer sees a three-dimensional image with all effects of perspective, parallax and depth-of-focus.

In his original set-up, Gabor illuminated the hologram with a parallel beam of light incident on a predominantly transparent object. The axes of both the object wave and the reference wave were parallel. The reconstruction of this hologram results in a real image superimposed on the undiffracted part of the reconstruction.
wave and a so called ‘twin image’ (or virtual image) lying on the same optical axis, i.e. an in-line hologram. Significant improvements of this in-line geometry were proposed by Leith and Upatnieks [141, 142], who introduced an off-axis reference wave at an oblique angle; this wave does not pass through the object. This approach spatially separates the two images and the reconstruction wave and allows the capture of opaque objects.

One early application of classical holography is Holographic Interferometry (HI), developed in the late 1960s by Stetson, Powell [190, 222] and others. HI made it possible to map the displacements of rough surfaces with an accuracy of a fraction of a micrometer. It also enabled interferometric comparisons of stored wave fronts existing at different times.

The development of computer technology allowed transferring either the recording process or the reconstruction process into the computer. The first approach led to Computer Generated Holography (CGH), which artificially generates holograms by numerical methods followed by their optical reconstruction. This technique is not considered here and the interested reader is referred to the literature; see e.g. Lee [140], Bryngdahl and Wyrowski [20] or Schreier [204].

Numerical hologram reconstruction was initiated by Goodman and Lawrence [74] and Yaroslavskii et al. [132]. They sampled optically enlarged parts of in-line and Fourier holograms recorded on a photographic plate. These digitized conventional holograms were reconstructed numerically. Onural and Scott [146, 167, 168] improved the reconstruction algorithm and applied this method to particle measurement. Haddad et al. described a holographic microscope based on numerical reconstruction of Fourier holograms [78].

A big step forward in the 1990s was the development of direct recording of Fresnel holograms with Charged Coupled Devices (CCD’s) by Schnars and Jüptner [197, 198]. This method enabled full digital recording and processing of holograms, without any photographic recording as intermediate step. The name which has been originally proposed for this technique was ‘direct holography’ [197], emphasizing the direct way from optical recording to numerical processing. Later on the term Digital Holography has been accepted in the optical metrology community for this method. Although this name is sometimes also used for Computer Generated Holography, the term Digital Holography is used within the scope of this book as a designation for digital recording and numerical reconstruction of holograms.

The dramatic developments in optics, electronics and computing widened the possibilities to capture, by computer means, phase information as well as amplitude. Computational wave front sensing [57] liberates the measurement procedures from a number of restrictions concerning coherence of the light or environmental requirements. It was shown that in some cases the light generated by the display of a smartphone has sufficient coherence to enable the recording of holograms [56]. Even the twin-image problem of in-line holography can be solved when phase shifted holograms are recorded according to Yamaguchi [255].

Schnars and Jüptner applied DH to interferometry and demonstrated that digital hologram reconstruction offers much more possibilities than conventional (optical) processing: The phase of the stored light waves can be calculated directly from
digital holograms, without the need for generating phase shifted interferograms [195, 196], see example in Fig. 1.1. Other methods of optical metrology, such as shearography or speckle photography, can be derived numerically from digital holograms [199]. Using mathematical reconstruction, the choice of interferometric technique (hologram interferometry, shearography or other) can be left until after hologram recording.

The use of electronic devices such as CCDs to record interferograms was already established in Electronic Speckle Pattern Interferometry (ESPI, also named TV-holography). Proposed independently by Butters and Leendertz [23], Macovski et al. [150] and Schwomma [205], two speckle interferograms are recorded in different states of the object under investigation. The speckle patterns are subtracted electronically. The resulting fringe pattern has some similarities to that of conventional or digital HI. Digital Holographic Interferometry (DHI) and ESPI are competing methods: image subtraction in ESPI is easier than the numerical reconstruction of DHI, but the information content of digital holograms is higher. ESPI and other methods of speckle metrology are also discussed in this book in order to compare them with Digital Holographic Interferometry.

The main disadvantage of ESPI is the loss of phase information of the original wave in the correlation process [46, 147, 148]. The interference phase has to be recovered with phase shifting methods [35, 223, 224]. However, all the information can be reconstructed by evaluating phase shifted shearograms without the ESPI approach [57] leading to a wave sensing method with low demands on the coherence and the environmental requirements.

Since its inception Digital Holography has been extended, improved and applied to several measurement tasks. Some of these advances include:
• improvements of the experimental techniques and of the reconstruction algorithm [37, 39, 40, 75, 113, 122, 123, 126, 136, 182, 184, 203],
• applications in deformation analysis and shape measurement [34, 119, 171, 181, 200, 206, 246],
• the development of phase shifting digital holography [47, 103, 137, 255–258, 264],
• the development of Digital Holographic Microscopy [38, 43, 48, 94, 114, 183, 235–237, 253]
• applications in particle tracking and sizing and underwater holography [5, 6, 85, 127, 176, 214, 227],
• measurement of refractive index distributions within transparent media due to temperature or concentration variations [105, 106, 177, 252],
• applications in encrypting of information [95, 135, 231, 232],
• the development of digital light-in-flight holography and other short-coherence-length applications [25, 100, 162–164, 179, 189],
• the development of methods to reconstruct the three-dimensional object structure from digital holograms [63, 64, 96, 154, 233, 263]
• the development of comparative Digital Holography [173, 174]

A number of alternative concepts for wavefield sensing are based on computational methods. Here, in contrast to Digital Holography, determination of the complex amplitude of a wave field is treated as an inverse problem. The recorded intensities are interpreted as an effect which has been caused by an unknown wavefield that has undergone various manipulations. Examples include intensities corresponding to different propagation states or superposition of a wavefield with a shifted (or propagated) copy of itself. The great benefit is that no particular reference wave is required to measure the complex amplitude. In many situations, this makes computational methods not only more robust and flexible than Digital Holography but also enables application to wave fields with low spatial and/or temporal coherence. However, solving the inverse problem requires application of sophisticated numerical methods. In most cases there is no way to directly track back to the complex amplitude from the recorded intensities alone. It is also not possible to record on film material in order to optically reconstruct the investigated wave field. The evaluation procedure can therefore be regarded as an integral part of the measurement process.

As an introduction to the field, we will review the three methods of phase retrieval [71, 193, 260], shear interferometry [55–57] and Shack-Hartmann wavefield sensing in Chap. 7, which have been constantly developed since the early 1970s.
Chapter 2
Fundamental Principles of Holography

2.1 Light Waves

The behaviour of light can be modelled either as a propagating electromagnetic (e-m) wave or as a stream of massless particles known as photons. Although the models are seemingly contradictory both are necessary to fully describe the full gamut of light phenomena. Whichever model is most appropriate depends on the phenomenon to be described or the experiment under investigation. For example, interaction of light with the atomic structure of matter is best described by the photon model: the theory of photon behaviour and its interactions is known as quantum optics. The phenomenon of refraction, diffraction and interference, however, are best described in terms of the wave model i.e. classical electromagnetism.

Interference and diffraction form the basis of holography. An e-m wave is described in terms of the propagation through space of mutually perpendicular electric and magnetic fields. These fields oscillate in a plane that is perpendicular to the direction of travel i.e. they are described as transverse waves, as depicted in Fig. 2.1. Light waves can be described either by the electrical or by the magnetic field, but in optics convention is to describe the e-m wave in terms of the electric vector.

Light propagation is described by the wave equation, which follows from Maxwell’s equations. The wave equation in a vacuum is

\[ \nabla^2 \vec{E} - \frac{1}{c^2} \frac{\partial^2 \vec{E}}{\partial t^2} = 0 \]  

(2.1)

Here \( \vec{E} \) is the electric field and \( \nabla^2 \) is the *Laplace operator* defined as

\[ \nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \]  

(2.2)
and $c$ is the speed of light in vacuum:

$$c = 2.9979 \times 10^8 \text{ m/s}$$  \hspace{0.5cm} (2.3)

The electrical field $\vec{E}$ is a vector quantity and can vibrate in any direction perpendicular to the direction of propagation. However, in many applications the wave vibrates only in a single plane. Such light is called linear polarized light. In this case it is sufficient to consider the scalar wave equation

$$\nabla^2 E - \frac{1}{c^2} \frac{\partial^2 E}{\partial t^2} = 0$$  \hspace{0.5cm} (2.4)

It can be easily verified that a linearly polarized, harmonic plane wave with amplitude

$$E(x, y, z, t) = a \cos\left(\omega t - \vec{k} \vec{r} - \phi_0\right)$$  \hspace{0.5cm} (2.5)

is a solution of the above wave equation.

$E(x, y, z, t)$ is the modulus of the electrical field vector at the point with spatial vector $\vec{r} = (x, y, z)$ at the time $t$. The quantity $a$ is the amplitude of the wave. The wave vector $\vec{k}$ describes the propagation direction of the wave:

$$\vec{k} = k\vec{n}$$  \hspace{0.5cm} (2.6)

$\vec{n}$ is a unit vector in the propagation direction. Points of equal phase are located on parallel planes that are perpendicular to the propagation direction. The modulus of $\vec{k}$ is the wave number and is described by

$$|\vec{k}| \equiv k = \frac{2\pi}{\lambda}$$  \hspace{0.5cm} (2.7)

The angular frequency $\omega$ corresponds to the frequency $f$ of the light wave by

$$\omega = 2\pi f$$  \hspace{0.5cm} (2.8)

Frequency $f$ and wavelength $\lambda$ are related through the speed of light $c$.
The spatially varying term

\[ u = \frac{C_0}{k f} \]  

is the phase, with phase constant \( \phi_0 \). It has to be pointed out that this definition is not standardized. Some authors designate the entire argument of the cosine function, \( \omega t - k f - \phi_0 \), as phase. The definition Eq. (2.10) is favourable to describe the holographic process and therefore used in this book.

The vacuum wavelengths of visible light are in the range of 400 nm (violet) to 780 nm (deep red). The corresponding frequency range is \( 7.5 \times 10^{14} \) Hz to \( 3.8 \times 10^{14} \) Hz. Light sensors such as the human eye, photodiodes, photographic film or CCD’s are not able to detect such high frequencies due to technical and physical reasons. The only directly measurable quantity is the intensity. It is proportional to the time average of the square of the electrical field:

\[
I = \varepsilon_0 c \frac{\langle E^2 \rangle_t}{\langle E^2 \rangle_t} = \varepsilon_0 c \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} E^2 dt \quad (2.11)
\]

\( \langle E^2 \rangle_t \) denotes the time average over many light periods. The constant factor \( \varepsilon_0 c \) results if the intensity is formally derived from the Maxwell equations. The constant \( \varepsilon_0 \) is the vacuum permittivity. Note: we are using the term intensity here. In photometry and radiometry intensity has a different meaning (radiant power per solid angle, unit W sr\(^{-1}\)).

For a plane wave Eq. (2.5) has to be inserted:

\[
I = \varepsilon_0 c a^2 \langle \cos^2 \left( \omega t - k f - \phi_0 \right) \rangle_t = \frac{1}{2} \varepsilon_0 c a^2 \quad (2.12)
\]

According to Eq. (2.12) the intensity is proportional to the square of the amplitude.

The expression (2.5) can be written in complex form as

\[
E(x, y, z, t) = a \text{Re} \left\{ \exp \left( i \left( \omega t - k f - \phi_0 \right) \right) \right\} \quad (2.13)
\]

where ‘Re’ denotes the real part of the complex function. For computations the real part ‘Re’ can be omitted (in accordance with the superposition principle). However, only the real part represents the physical wave:

\[
E(x, y, z, t) = a \exp \left( i \left( \omega t - k f - \phi_0 \right) \right) \quad (2.14)
\]
One advantage of the complex representation is that the spatial and temporal parts factorize and Eq. (2.14) can be written as:

\[ E(x,y,z,t) = a \exp(i\varphi) \exp(i\omega t) \]  

In many calculations of optics only the spatial distribution of the wave is of interest. In this case only the spatial part of the electrical field, its complex amplitude, need be considered:

\[ A(x,y,z) = a \exp(i\varphi) \]  

Equations (2.15) and (2.16) are not just valid for plane waves, but apply in general to three-dimensional waves whose amplitude, \( a \), and phase, \( \varphi \), are functions of \( x, y \) and \( z \).

In complex notation the intensity is now simply calculated by taking the square of the modulus of the complex amplitude

\[ I = \frac{1}{2} \varepsilon_0 c |A|^2 = \frac{1}{2} \varepsilon_0 c A^* A = \frac{1}{2} \varepsilon_0 c a^2 \]  

where \( * \) denotes complex conjugation. In many practical calculations where the absolute value of \( I \) is not of interest the factor \( \frac{1}{2} \varepsilon_0 c \) can be neglected, and the intensity simplifies to \( I = |A|^2 \).

### 2.2 Interference

The superposition of two or more waves in space is named interference. If each single wave described by \( \vec{E}_i(\vec{r}, t) \) is a solution of the wave equation, the superposition

\[ \vec{E}(\vec{r}, t) = \sum_i \vec{E}_i(\vec{r}, t) \quad i = 1, 2, \ldots \]  

is also a solution. This is because the wave equation is a linear differential equation.

In the following, interference of two monochromatic waves with equal frequencies and wavelengths is considered. The waves shall have the same polarization directions, i.e. scalar formalism can be used. The complex amplitudes of the respective waves are represented by;

\[ A_1(x,y,z) = a_1 \exp(i\varphi_1) \]  
\[ A_2(x,y,z) = a_2 \exp(i\varphi_2) \]

The resulting complex amplitude is then calculated by the sum of the individual amplitudes:
According to Eq. (2.17) the intensity can be written as

\[ I = |A_1 + A_2|^2 = (A_1 + A_2)(A_1 + A_2)^* \]

\[ = a_1^2 + a_2^2 + 2a_1a_2 \cos(\phi_1 - \phi_2) \]

\[ = I_1 + I_2 + 2\sqrt{I_1I_2} \cos \Delta \varphi \quad (2.22) \]

where \( I_1, I_2 \) are the individual intensities and the phase difference between the two waves is

\[ \Delta \varphi = \phi_1 - \phi_2 \quad (2.23) \]

The resulting intensity is the sum of the individual intensities plus the interference term \( 2\sqrt{I_1I_2} \cos \Delta \varphi \), which depends on the phase difference between the waves. The intensity reaches its maximum when the phase difference between consecutive points is a multiple of \( 2\pi \)

\[ \Delta \varphi = 2n\pi \quad \text{for } n = 0, 1, 2, \ldots \quad (2.24) \]

This is known as constructive interference. The intensity reaches its minimum when

\[ \Delta \varphi = (2n + 1)\pi \quad \text{for } n = 0, 1, 2, \ldots \quad (2.25) \]

And this is known as destructive interference. The integer \( n \) is the interference order. An interference pattern therefore consists of a series of dark and light lines, “fringes”, across the field-of-view as a result of this constructive and destructive interference. Scalar theory can also be applied to waves with different polarization directions, if the components of the electric field vector are considered.

The superposition of two plane waves which intersect at an angle \( \theta \) with respect to each other results in an interference pattern with equidistant spacing, as seen in Fig. 2.2. The fringe spacing \( d \) is the distance from one interference maximum to the next and can be calculated from geometrical considerations. Figure 2.2 shows that

\[ \sin \theta_1 = \frac{\Delta l_1}{d}; \quad \sin \theta_2 = \frac{\Delta l_2}{d} \quad (2.26) \]

The quantities \( \theta_1 \) and \( \theta_2 \) are the angles between the propagation directions of the wavefronts and the vertical direction of the screen. The length \( \Delta l_2 \) is the path difference between wavefront W2 and wavefront W1 at the position of the interference maximum P1 (W2 has to travel a longer path to P1 than W1). At the neighboring maximum P2 the conditions are exchanged: now W1 has to travel a longer path; the path difference of W2 with respect to W1 is \( -\Delta l_1 \). The variation
between the path differences at neighboring maxima is therefore $\Delta l_1 + \Delta l_2$. This difference is equal to one wavelength. Thus the interference condition is:

$$\Delta l_1 + \Delta l_2 = \lambda$$  \hspace{1cm} (2.27)

Combining Eq. (2.26) with Eq. (2.27) gives the fringe spacing as:

$$d = \frac{\lambda}{\sin \theta_1 + \sin \theta_2} = \frac{\lambda}{2 \sin \frac{\theta_1 + \theta_2}{2} \cos \frac{\theta_1 - \theta_2}{2}}$$  \hspace{1cm} (2.28)

The approximation $\cos(\theta_1 - \theta_2)/2 \approx 1$ and $\theta = \theta_1 + \theta_2$ can be applied to give

$$d = \frac{\lambda}{2 \sin \frac{\theta}{2}}$$  \hspace{1cm} (2.29)

Instead of the fringe spacing $d$, the fringe pattern can also be described in terms of the spatial frequency $f$, which is just the reciprocal of $d$, i.e.

$$f = d^{-1} = \frac{2}{\lambda} \sin \frac{\theta}{2}$$  \hspace{1cm} (2.30)

### 2.3 Coherence

#### 2.3.1 General

Generally the resulting intensity of two different sources, e.g. two electric light bulbs directed on a screen, is additive. Instead of dark and bright fringes as expected by Eq. (2.22) only a uniform brightness according to the sum of the individual intensities is visible.
2.3 Coherence

In order to observe interference fringes, the phases of the individual waves have to be correlated. The ability of light to form interference patterns is called coherence and is investigated in this chapter. The two aspects of coherence are temporal and spatial coherence. Temporal coherence depends on the correlation of a wave with itself at different instants in time [121], whereas spatial coherence is based on the mutual correlation of different parts of the same wavefield in space.

2.3.2 Temporal Coherence

The phenomenon of interference between two coherent beams of light can be described in terms of a two beam interferometer such as the Michelson-interferometer, as shown in Fig. 2.3. Light emitted by the source S is split into two waves of reduced amplitude by the beam splitter BS. These waves travel to the mirrors M1 and M2 respectively, and are reflected back into their incident directions. After passing the beam splitter again they are superimposed at a screen. Usually the superimposed waves are not exactly parallel, but are incident at a small angle. As a result a two-dimensional interference pattern becomes visible.

The optical path length from BS to M1 and back to BS is $s_1$, and the optical path length from BS to M2 and back to BS is $s_2$. Experiments show that interference can only occur if the optical path difference $s_1 - s_2$ does not exceed a certain length $L$. If the optical path difference exceeds this limit, the interference fringes vanish and just a uniform brightness is visible on the screen. The qualitative explanation for this phenomenon is that interference fringes can only develop if the superimposed waves have a well defined (constant) phase relationship between them. The phase difference between waves emitted by different sources varies randomly and thus the waves do not interfere. The atoms within the light source emit wave trains with a finite length $L$. If the optical path difference exceeds $L$, the recombined waves do not overlap after passing the different ways and interference is not observed.

![Fig. 2.3 Michelson’s interferometer](image-url)
The critical path length difference or, equivalently, the length of a wave train is the coherence length \( L \) of the wave. The corresponding time over which the wave train is emitted is its coherence time,

\[
\tau = \frac{L}{c}
\]  

(2.31)

According to the laws of Fourier analysis a wave train with finite length \( L \) corresponds to light with finite spectral width \( \Delta f \), where

\[
L = \frac{c}{\Delta f}
\]  

(2.32)

The coherence length is therefore a measure for the spectral linewidth of the source at a specific frequency, \( f \). Light with a long coherence length accordingly has a correspondingly small linewidth and is therefore highly monochromatic.

Typical coherence lengths of light radiated from thermal sources, e.g. conventional electric light bulbs, are in the range of some micrometers. That means, interference can only be observed if the arms of the interferometer have nearly equal path lengths. On the other hand lasers have coherence lengths from a few millimetres (e.g. a multi-mode diode laser) to several 100 m (e.g. a stabilized single mode Nd:YAG-laser) up to several hundred kilometres for specially stabilized gas lasers used for research purposes.

The fringe visibility

\[
V = \frac{I_{\text{max}} - I_{\text{min}}}{I_{\text{max}} + I_{\text{min}}}
\]

(2.33)

is a measure of the contrast of a particular interference pattern, where \( I_{\text{max}} \) and \( I_{\text{min}} \) are two neighbouring intensity maxima and minima. They are calculated by inserting \( \Delta \varphi = 0 \) and \( \Delta \varphi = \pi \) respectively into Eq. (2.22). In the ideal case of infinite coherence length the visibility is given by,

\[
V = \frac{2\sqrt{I_1 I_2}}{I_1 + I_2}
\]  

(2.34)

To consider the effect of finite coherence length the complex self-coherence function \( \Gamma(\tau) \) is introduced:

\[
\Gamma(\tau) = \langle E(t + \tau)E^*(t) \rangle
\]

\[
= \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} E(t + \tau)E^*(t)dt
\]  

(2.35)
$E(t)$ is the electrical field (to be precise: the complex analytical signal) of one interfering wave while $E(t + \tau)$ is the electrical field of the other wave. The latter is delayed in time by $\tau$. Equation (2.35) represents the autocorrelation of the corresponding electric field amplitudes. The quantity

$$\gamma(\tau) = \frac{\Gamma(\tau)}{\Gamma(0)}$$

(2.36)

is the normalized self-coherence function; the absolute value of $\gamma$ defines the degree of coherence.

With finite coherence length the interference equation (2.22) has to be replaced by

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2} |\gamma| \cos \Delta \varphi$$

(2.37)

The maximum and minimum intensity are now calculated by

$$I_{\text{max}} = I_1 + I_2 + 2\sqrt{I_1 I_2} |\gamma|$$
$$I_{\text{min}} = I_1 + I_2 - 2\sqrt{I_1 I_2} |\gamma|$$

(2.38)

Inserting these quantities into Eq. (2.33) yields

$$V = \frac{2\sqrt{I_1 I_2}}{I_1 + I_2} |\gamma|$$

(2.39)

For two partial waves with the same intensity, $I_1 = I_2$ Eq. (2.39) becomes

$$V = |\gamma|$$

(2.40)

$|\gamma|$ is equal to the visibility and is therefore a measure of the ability of the two wave fields to interfere. When $|\gamma| = 1$ we have ideally monochromatic light or, likewise, light with infinite coherence length; when $|\gamma| = 0$ for the light is completely incoherent. Partially coherent light therefore lies in the range $0 < |\gamma| < 1$.

### 2.3.3 Spatial Coherence

Spatial coherence describes the mutual correlation of spatially separated parts of the same wavefield. This property can be measured, for example, a Young’s interferometer, Fig. 2.4. Here, an extended light source emits light from a large number of elementary point sources. An aperture with two transparent holes is mounted between the light source and the screen. The aim of the experiment is to determine the mutual correlation (degree of coherence) of the light incident on the aperture at the spatially separated positions given by the holes. If the light at these