HOLOGRAPHY, RESEARCH AND TECHNOLOGIES

Edited by **Joseph Rosen**

INTECHWEB.ORG

Holography, Research and Technologies
Edited by Joseph Rosen

Published by InTech
Janeza Trdine 9, 51000 Rijeka, Croatia

Copyright © 2011 InTech
All chapters are Open Access articles distributed under the Creative Commons Non Commercial Share Alike Attribution 3.0 license, which permits to copy, distribute, transmit, and adapt the work in any medium, so long as the original work is properly cited. After this work has been published by InTech, authors have the right to republish it, in whole or part, in any publication of which they are the author, and to make other personal use of the work. Any republication, referencing or personal use of the work must explicitly identify the original source.

Statements and opinions expressed in the chapters are these of the individual contributors and not necessarily those of the editors or publisher. No responsibility is accepted for the accuracy of information contained in the published articles. The publisher assumes no responsibility for any damage or injury to persons or property arising out of the use of any materials, instructions, methods or ideas contained in the book.

Publishing Process Manager Katarina Lovrecic
Technical Editor Teodora Smiljanic
Cover Designer Martina Sirotic
Image Copyright Vlue, 2010. Used under license from Shutterstock.com

First published February, 2011
Printed in India

A free online edition of this book is available at www.intechopen.com
Additional hard copies can be obtained from orders@intechweb.org

Holography, Research and Technologies, Edited by Joseph Rosen
p. cm.
ISBN 978-953-307-227-2

INTECH OPEN ACCESS PUBLISHER

INTECH open

free online editions of InTech
Books and Journals can be found at
www.intechopen.com

Contents

Preface IX

Part 1 Basic Theory of Optics and Holography 1

Chapter 1 **The Fourier Transform in Optics: Analogous Experiment and Digital Calculus 3**
Petre Cătălin Logofătu, Victor Nascov and Dan Apostol

Chapter 2 **The Holographic Principle and Emergence Phenomenon 27**
Marina Shaduri

Part 2 Holographic Materials 55

Chapter 3 **Polymer Holography in Acrylamide-Based Recording Material 57**
Milan Květoň, Pavel Fiala and Antonín Havránek

Chapter 4 **Real-time, Multi-wavelength Holographic Recording in Photorefractive Volume Media: Theory and Applications 83**
Eduardo Acedo Barbosa

Chapter 5 **The Composite Structure of Hologram and Optical Waveguide 109**
Renxi Gao and Wenjun Liu

Part 3 Holographic Techniques 133

Chapter 6 **FINCH: Fresnel Incoherent Correlation Hologram 135**
Joseph Rosen, Barak Katz and Gary Brooker

Chapter 7 **Programmable Point-source Digital In-line Holography Using Digital Micro-mirror Devices 155**
Adekunle A. Adeyemi and Thomas E. Darcie

Chapter 8 **Pulsed Full-Color Digital Holography with a Raman Shifter 173**
Percival Almoro, Wilson Garcia and Caesar Saloma

Chapter 9 **Optical Holography Reconstruction of Nano-objects 191**
Cesar A. Sciammarella, Luciano Lamberti
and Federico M. Sciammarella

Part 4 Holographic Applications 217

Chapter 10 **Quantitative Analysis of Biological Cells Using Digital Holographic Microscopy 219**
Natan T. Shaked, Lisa L. Satterwhite,
Matthew T. Rinehart and Adam Wax

Chapter 11 **Digital Holography and Cell Studies 237**
Kersti Alm, Helena Cirenajwis,
Lennart Gisselsson, Anette Gjörloff Wingren, Birgit Janicke,
Anna Mölder, Stina Oredsson and Johan Persson

Chapter 12 **Fabrication of Two- and Three-Dimensional Photonic Crystals and Photonic Quasi-Crystals by Interference Technique 253**
Ngoc Diep Lai, Jian Hung Lin, Danh Bich Do,Wen Ping Liang,
Yu Di Huang, Tsao Shih Zheng, Yi Ya Huang, Chia Chen Hsu

Chapter 13 **Achieving Wide Band Gaps and a Band Edge Laser Using Face-Centered Cubic Lattice by Holography 279**
Tianrui Zhai and Dahe Liu

Chapter 14 **Accurate Axial Location for Particles in Digital In-Line Holography 293**
Zhi-Bin Li, Gang Zheng, Li-Xin Zhang, Gang Liu and Fei Xia

Chapter 15 **Hybrid Numerical-Experimental Holographic Interferometry for Investigation of Nonlinearities in MEMS Dynamics 303**
Minvydas Ragulskis, Arvydas Palevicius and Loreta Saunoriene

Chapter 16 **Vibration Measurement by Speckle Interferometry between High Spatial and High Temporal Resolution 325**
Dan Nicolae Borza

Part 5 Digital Algorithms in Holography 347

Chapter 17 **Reconstruction of Digital Hologram by use of the Wavelet Transform 349**
Jingang Zhong and Jiawen Weng

Chapter 18 **Iterative Noise Reduction in Digital Holographic Microscopy 371**
Victor Arrizón, Ulises Ruiz and Maria Luisa Cruz

Chapter 19 **Image Quality Improvement of Digital Holography by Multiple Wavelengths or Multiple Holograms 397**
Takanori Nomura

Chapter 20 **Digital Holography and Phase Retrieval 407**
Hamootal Duadi, Ofer Margalit, Vicente Mico, José A. Rodrigo, Tatiana Alieva, Javier Garcia and Zeev Zalevsky

Part 6 Non-optical Holography 421

Chapter 21 **In-line Hard X-ray Holography for Biomedical Imaging 423**
Andrzej Krol

Chapter 22 **Fundamentals and Applications of Electron Holography 441**
Akira Tonomura

Preface

Holography has recently become a field of much interest because of the many new applications implemented by various holographic techniques. This book is a collection of 22 excellent chapters written by various experts, and it covers various aspects of holography. Naturally, one book by 22 researches cannot cover all the richness of the holography world. Nevertheless, the book gives an updated picture on the hottest topics that the scientific community deals with, in the field of holography. The book contains recent outputs from researches belonging to different research groups worldwide, providing a rich diversity of approaches to the topic of holography. The aim of the book is to present a cutting edge research on holography to the reader. We are lucky that it is freely accessible on the internet and enables outstanding contributors to share their knowledge with every interested reader.

The 22 chapters of the book are organized in six sections, starting with theory, continuing with materials, techniques, applications as well as digital algorithms, and finally ending with non-optical holograms. There are two chapters in the first section of Basic Theory of Optics and Holography. The first chapter is about optical Fourier transform which is an essential tool in many holographic schemes. The second chapter discusses philosophically the role of the holographic principle in nature. The sections subsequent to the first section deal with more practical aspects of holography. The section of holographic materials contains three chapters describing holograms recorded on the following mediums:

1. Acrylamide-based recording material
2. Photorefractive media, and
3. Optical waveguides

The next section depicting holographic techniques also contains four chapters:

1. New technique of recording incoherent digital holograms,
2. Novel technique of recording in-line coherent digital holograms.
3. Original technique of recording pulsed color digital holograms.
4. The latest method of reconstructing nano-objects from optical holograms.

The subsequent section detailing holographic applications, obviously contains the largest number of chapters. The applications described in this book are only a tiny sample of the use of holography in many scientific and industrial areas. Two chapters deal with the role of holography in research of biological cells. Then the next two chapters describe the creation of holographic lattice structures for manufacturing photonic

crystals. The last three chapters in the applications section discuss the use of holography in the fields of particles tracking, MEMS, and vibration measurement, respectively. Following the extensive section of applications, a section consisting of four chapters is devoted to the growing link between holography and the world of digital computation. This link is best expressed by the digital holograms which are the type of holograms that are recorded optically and reconstructed digitally in the computer memory. Each of the four chapters in this section describes a specific digital algorithm of digital hologram reconstructions. This book rounds off with two interesting chapters on non-optical holograms: one discusses the x-ray holography for biomedical imaging, and the other introduces the topic of electron holography.

Finally, I would like to thank all of the authors for their efforts in writing these interesting chapters. Their contributions light up hidden corners in the broad topic of holography and extend the knowledge of the rapidly growing holographic community. I would also like here to cite a quote of a famous American novelist, Edith Wharton: "There are two ways of spreading light; to be the candle or the mirror that reflects it," and to say without a hesitation that this book definitely presents 22 glowing candles.

Joseph Rosen
Ben-Gurion University of the Negev,
Israel

Part 1

Basic Theory of Optics and Holography

1

The Fourier Transform in Optics: Analogous Experiment and Digital Calculus

Petre Cătălin Logofătu[1], Victor Nascov[2] and Dan Apostol[1]
[1]*National Institute for Laser, Plasma and Radiation Physics*
[2]*Universitatea Transilvania Braşov*
Romania

1. Introduction

Discrete optics and digital optics are fast becoming a classical chapter in optics and physics in general, despite their relative recent arrival on the scientific scene. In fact their spectacular blooming began precisely at the time of the computer revolution which made possible fast discrete numerical computation. Discrete mathematics in general and discrete optics in particular although predated digital optics, even by centuries, received a new impetus from the development of digital optics. Formalisms were designed to deal with the specific problems of discrete numerical calculation. Of course, these theoretical efforts were done not only for the benefit of optics but of all quantitative sciences. Diffractive optics in general and the newly formed scientific branch of digital holography, turned out to be especially suited to benefit from the development of discrete mathematics. One reason is that the optical diffraction in itself is a mathematical transform. An ordinary optical element such as the lens turned out to be a genuine natural optic computer, namely one that calculates the Fourier transform (Goodman, 1996, chapter 5).
The problem is that the discrete mathematics is not at all the same thing as continuous mathematics. For instance for the most common theoretical tool in diffractive optics, the continuous (physical) Fourier transform (CFT) we have a discrete correspondent named discrete Fourier transform (DFT). We need DFT because the Fourier transform rarely yields closed form expressions and generally can be computed only numerically, not symbolically. Of course, no matter how accurate, by its very nature DFT can be only an approximation of CFT. But there is another advantage offered by DFT which inclines the balance in favour of discrete optics. The reason is somehow accidental and requires some explanation. It is the discovery of the Fast Fourier transform, for short FFT, (Cooley & Tukey, 1965), which was followed by a true revolution in the field of discrete optics because of the reduction with orders of magnitude of the computation time, especially for large loads of input data. FFT stirred also an avalanche of fast computation algorithms based on it. The property that allowed the creation of these fast algorithms is that, as it turns out, most diffraction formulae contain at their core one or more Fourier transforms which may be rapidly calculated using the FFT. The method of discovering a new fast algorithm is oftentimes to reformulate the diffraction formulae so that to identify and isolate the Fourier transforms it contains. We contributed ourselves to the development of the field with the creation of an

improved algorithm for the fast computation of the discrete Rayleigh-Sommerfeld transform and a new concept of convolution: the scaled linearized discrete convolution (Nascov & Logofătu, 2009). The conclusion is that *we want to use DFT, even if CFT would be a viable alternative*, because of its amazing improvement of computation speed, which makes feasible diffraction calculations which otherwise would be only conjectures to speculate about.

Here is the moment to state most definitely *the generic connection between the digital holography and the DFT, more specifically the FFT*, as was outlined in the pioneer work of (Lohmann & Paris, 1967), and since then by the work of countless researchers, of which, for lack of space and because it is not our intent to write a monography about the parallel evolution of digital holography and DFT, we will mention only a few essential works that deal both subjects in connection to each other. There is, of course, a vast deal of good textbooks and tutorials dedicated to the fundamentals of the Fourier transform, continuous or discrete (Arfken & Weber, 2001; Bracewell, 1965; Brigham, 1973; Bringdahl & Wyrowsky, 1990; Collier et al., 1971; Goodman, 1996; Lee, 1978; Press et al., 2002 and Yaroslavsky & Eden, 1996), but in our opinion a severe shortcoming of the textbooks listed above is the fact that, in our opinion none offers a complete and satisfactory connection between the two formalisms, such as the expression of DFT in terms of CFT. We use DFT in place of CFT but we do not know exactly what is the connection between them! In Yaroslavsky & Eden, 1996, chapter 4 and Collier, 1971, chapter 9, a correspondence is worked out between DFT and CFT, namely that the value of the DFT is equal to the value of the CFT at the sample points in the Fourier space, but this is valid only for band-limited functions and it is not rigorous. (Strictly speaking DFT can be applied only to band-limited functions, but this is an unacceptable restriction; many of the functions of interest are not band-limited. In general we have to approximate and to compensate for the assumed approximations.) Generally those textbooks fail to link in the proper manner the fertile but inapplicable in practice in itself field of discrete optics, to continuous, physical optics, where the experiments take place and we can take advantage of the progress of the discrete optics. In our own scientific research activity in the field of digital optics we encountered the difficulty almost at every step [Apostol & al., 2007 (a); Apostol et al., 2007 (b); Logofătu et al., 2009 and Logofătu et al., 2010]. In two previous occasions (Logofătu & Apostol, 2007 and Nascov et al., 2010) we attempted to express the physical meaning of DFT, to put it in the terms of CFT using the Fourier series as an intermediary concept. Together with the present work this continued effort on our part will hopefully prove useful to all those who undertake projects in discrete optics and they are hampered by the gap between DFT and CFT, discrete mathematics, digital computers on the one hand and real physical experiments on the other hand.

With the above rich justification we did not exhaust by far the uses of DFT and the need to rigorously connect it to CFT. Apart from all virtual computation, which also requires the connection between DFT and CFT to be worked out, digital holography present special hybrid cases where a discrete and a continuous character are both assumed. For instance the recording of holograms can be made using Charged Coupled Devices (CCD) which are discrete yet they work in the real continuous physical world. The same is valid for the other end of holography, the reconstruction or the playback of the holograms. For this purpose today are used devices such as Spatial Light Modulators (SLM) of which one can also say they are hybrid in nature, digital and analogous in the same time. For such devices as CCDs and SLMs one has to switch back and forth between CFT and DFT and think sometimes in terms of one of the formalisms and some other times in terms of the other. Here we should

mention the pioneer work of Lohmann and Paris for the compensation of the "digital" effect, so to speak, in their experiments with digital holograms (Lohmann & Paris, 1967).
For physical reasons in optics, when dealing with images, a 2D coordinate system with the origin in the center of the image is used. However, FFT deals with positive coordinates only, which means they are restricted to the upper right quadrant of the 2D coordinate system. In order to work in such conditions one has to perform a coordinate conversion of the input image before the calculations, and also the final result of the calculations has to be converted in order to have the correct output image. The conversion involves permutations of quadrants and sign for the values of the field. The positive coordinates are necessary for the application of the FFT algorithm, which makes worthwhile the complication by its very fast computation time. It is possible to work only with positive coordinates because the discrete Fourier transform assumes the input and the output being infinite and periodic (Logofătu & Apostol, 2007). The disadvantage of this approach is its counterintuitive and artificial manner. The correspondence to the physical reality is not simple and obvious. In our practically-oriented paper we used a natural, physical coordinate system, (hence negative coordinates too), and we performed a coordinate conversion of the images only immediately before and after the application of the FFT. In this way, the correspondence to the physical reality is simple and obvious at all times and this gives to our approach a more intuitive character. Precisely for this reason in chapter 4 we present an alternative method for converting the physical input so that can be used by the mathematical algorithm of FFT and for converting the mathematical output of FFT into data with physical meaning, a method that do not use permutation of submatrices, which may be preferable for large matrices, a method based on the shifting property of the Fourier transform applied to DFT.
In order to keep the mathematics to a minimum the equations were written as for the 1D case whenever possible. The generalization to 2D is straightforward and the reader should keep in mind at all times the generalization to the 2D case, the real, physical case. The equations are valid for the 1D case too, of course, but the 1D case is just a theoretical, imaginary case.

2. The translation of DFT in the terms of CFT or the top-down approach

2.1 Short overview of the current situation

In our efforts to bridge the two independent formalisms of CFT and DFT first we used more of a top-down approach, working from the principles down to specific results (Logofătu & Apostol, 2007). In mathematics there are three types of Fourier transforms: (I) CFT, (II) the Fourier series and (III) DFT. Only the first type has full physical meaning, and can be accomplished for instance in optics by Fresnel diffraction using a lens or by Fraunhofer diffraction. The third type is a pure mathematical concept, although it is much more used in computation than the previous two for practical reasons. These three types of Fourier transforms are independent formalisms, they can stand alone without reference to one another and they are often treated as such, regardless of how logic and necessity generated one from another. Because only CFT has physical meaning, the other two types of transforms are mathematical constructs that have meaning only by expressing them in terms of the first. In order to be able to do this we have to present the three transforms in a unifying view. To our knowledge no mathematics or physics textbooks present such a unifying view of the three types of Fourier transforms, although all the necessary knowledge lies in pieces in the literature. An integrated unifying presentation of the three

types of Fourier transforms has then the character of a creative review, so to speak. In this paper such a unifying view is presented. The Fourier transform of type II, the Fourier series, besides its own independent worth, is shown to be an intermediary link between the Fourier transforms of types I and III, a step in the transition between them. Also some concrete cases are analyzed to illustrate how the discrete representation of the Fourier transform should be interpreted in terms of the physical Fourier transform and how one can make DFT a good approximation of CFT. In the remainder of this chapter such a unifying view is presented. The Fourier transform of type II, the Fourier series, besides its own independent worth, is shown to be an intermediary link between the Fourier transforms of types I and III, a step in the transition between them. Also some concrete cases are analyzed to illustrate how the discrete representation of the Fourier transform should be interpreted in terms of the physical Fourier transform and how one can make DFT a good approximation of CFT.

2.2 Fourier transforms

Suppose we have a function *g(t)* and we are interested in its Fourier transform function *G(f)*. Here *t* is an arbitrary variable (may be time or a spatial dimension) and *f* is the corresponding variable from the Fourier space (like time and frequency or space and spatial frequencies). We work in the 1D case for convenience but the extrapolation to 2D, (i.e. the optic case, the one we are interested), is straightforward.

The three types of Fourier transforms are defined as following: (I) continuous, i.e. the calculation of the transform *G(f)* is done for functions *g(t)* defined over the real continuum, that is the interval (–∞…+∞), and the transformation is made by integration over the same interval

$$G(f) = \mathcal{F}\{g\}(f) = \int_{-\infty}^{+\infty} g(t)\exp(-i2\pi f t)dt \tag{1}$$

$\mathcal{F}$ being the operator for Fourier transform, with *G* being also generally defined over the real continuum, (II) Fourier series, where the function to be transformed is defined over a limited range (0…*Δt*) of the continuum, and instead of a Fourier transformed function defined over (–∞…+∞) we have a series which represents the discrete coefficients of the Fourier expansion

$$G_n = \frac{1}{\Delta t}\int_{-\Delta t/2}^{\Delta t/2} g(t)\exp(-i2\pi n t/\Delta t)dt,\ n = -\infty,...,\infty \tag{2}$$

and (III) the purely discrete Fourier transform where a list of numbers is transformed into another list of numbers by a summation procedure and not by integration

$$G_q = \sum_{p=-N/2}^{N/2-1} g_p \exp(-i2\pi q p/N),\ q = -N/2,...,N/2-1 \tag{3}$$

(For clarity purposes, in this chapter we use consistently *m* and *n* to indicate the periodicity, and *p* and *q* to indicate the sampling of *g* and *G* respectively.) Although these three types of Fourier transforms can be considered independent formalisms, they are strongly connected. Indeed, the second may be considered a particularization of the first, by restricting the class

of input functions g to periodic functions only and then to take into consideration for calculation purposes only one period from the interval $(-\infty\ldots+\infty)$ over which g is defined. The third can be considered a particularization of the second, by requesting not only that the functions g to be periodic, but also discrete, to have values only at even spaced intervals δt. Therefore the third type of Fourier transform may be considered an even more drastic particularization of the most general Fourier transform I.
That is the top-down approach. It is possible another approach, a bottom-up one, in which the second type of Fourier transform is considered a generalization of the third type, or a construct made starting from the third type, and the same thing can be said about the relation between the first and the second type. In order to pass from the Fourier transform type III to type II, in the list g_m which is a discrete sampling made at equal intervals δt we make $\delta t \to 0$ and $N \to \infty$, which results in the list g_m becoming a function of continuous argument and the numbers G_n are not anymore obtained by summation as in Eq. (3) but by integration as in Eq. (2). Now we are dealing with the Fourier transform type II. Making the function g periodic, by imposing

$$g(t + m\Delta t) = g(t) \tag{4}$$

where m is an arbitrary integer, and making the period $\Delta t \to \infty$, the discrete coefficients G_n become a continuum and we are back to the Fourier transform of type I. But we will deal with this approach in more detail in chapter 3.

2.3 From continuous Fourier transform to Fourier series

Since the formalism of the Fourier transform of type I is the most general and the only one with full physical meaning, we will express the two other formalisms in its terms. As we said before, if the input function g is periodic, then the corresponding output function G degenerates into a series. Indeed, if g has the period Δt as in Eq. (4) then the corresponding G can be written as

$$\begin{aligned} G(f) &= \sum_{m=-\infty}^{+\infty} \int_{(m-1/2)\Delta t}^{(m+1/2)\Delta t} g(t)\exp(-i2\pi f t)dt = \sum_{m=-\infty}^{+\infty} \int_{-\Delta t/2}^{\Delta t/2} g(t)\exp\left[-i2\pi f(t+m\Delta t)\right]dt = \\ &= \int_{-\Delta t/2}^{\Delta t/2} g(t)\exp(-i2\pi f t)dt \sum_{m=-\infty}^{+\infty} \exp(-i2\pi m f \Delta t) \end{aligned} \tag{5}$$

Without rigorous demonstration we will state here that the infinite sum of exponentials in the uttermost right hand side of Eq. (5) is an infinite sum of delta functions called the comb function,

$$\sum_{m=-\infty}^{+\infty} \exp(-i2\pi m f \Delta t) = \frac{1}{\Delta t} \sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{\Delta t}\right) \equiv \mathrm{comb}(\Delta t f) \tag{6}$$

Indeed, one may check that the sum of exponentials is 0 everywhere except for f of the form $n/\Delta t$ when it becomes infinite. When f is of the form $n/\Delta t$ all the members of the infinite sum are equal to 1. When f differs however slightly from $n/\Delta t$, the phaser with which we can represent the exponentials in the uttermost left hand side of Eq. (6) in the complex plane runs with incremental equidistant strokes from 0 to 2 π when the index m of the sum grows

incrementally with the result that the contributions of the terms to the sum cancels each other out, although not necessarily term by term. (One should not confuse the delta function $\delta(t)$ with the sampling interval δt.) Introducing (6) into (5) one obtains

$$G(f)=\frac{1}{\Delta t}\sum_{n=-\infty}^{\infty}\delta\left(f-\frac{n}{\Delta t}\right)\int_{-\Delta t/2}^{\Delta t/2}g(t)\exp(-i2\pi n t/\Delta t)dt=\sum_{n=-\infty}^{\infty}G_n\delta\left(f-\frac{n}{\Delta t}\right) \tag{7}$$

where G_n were defined in Eq. (2). One can see from Eq. (7) that the Fourier series are a particular case of CFT, namely the Fourier transform G of a periodic function g of period Δt is a sum of delta functions of arguments shifted with $1/\Delta t$ intervals and with coefficients G_n that are the same as the coefficients defined in Eq. (2). Actually it is the coefficients from Eq. (2) that are the Fourier series, and not the function defined in Eq. (7), but the correspondence is obvious.

In optical experiments one may see a good physical approximation of the Fourier series when a double periodic mask, with perpendicular directions of periodicity, modulates a plane monochromatic light wave and the resulting optical field distribution is Fourier transformed with the help of a lens. In the back focal plane of the lens, where the Fourier spectrum is formed, we have a distribution of intensely luminous points along the directions of periodicity. The luminous points do not have, of course, a rigorous delta distribution, they are not infinitely intense and they have non-zero areas. This departure from ideal is due to the fact that neither the mask nor the plane wave are ideal. The image that is Fourier transformed is neither infinite nor rigorously periodic, since the intensity of the plane wave decreases from a maximum in the centre to zero towards the periphery. But such an experiment is a good physical illustration of the mathematics involved in Eq. (7)

The difference between CFT and the Fourier series can be seen also from the inverse perspective of the Fourier expansion of g,

$$\begin{aligned} g(t)&=\mathcal{F}^{-1}\{G\}(t)=\int_{-\infty}^{+\infty}G(f)\exp(i2\pi f t)dt= \\ &=\int_{-\infty}^{+\infty}\sum_{n=-\infty}^{\infty}G_n\delta\left(f-\frac{n}{\Delta t}\right)\exp(i2\pi f t)dt=\sum_{n=-\infty}^{\infty}G_n\exp(i2\pi n t/\Delta t) \end{aligned} \tag{8}$$

The difference is that in the case of a periodic function the Fourier expansion is a sum and not an integral anymore.

2.4 The discrete Fourier transform

For computation purposes we cannot use always a function g defined over the continuum but a sampled version instead. There are many reasons that make the sampling of g necessary. It is possible that the function g, representing a physical signal, an optical field for instance, is not known a priori and only a detected sample of it can be known. It is possible that the function g cannot be integrated because it is too complicated or it cannot be expressed in closed form functions or it causes numerical instabilities in the calculation of CFT. A very important reason might be the fact that the function g may be the result of calculations too as in the case of computer-generated Fourier holograms. In this case, g is known only as a 2D matrix of numbers. (For simplicity, however, we will continue to work

with 1D functions as long as possible.) Then we have to sample the input function, and we can do that with the help of the comb function; this type of sampling will yield the value of g at evenly spaced intervals of chosen value δt and it can be written as

$$g^{s}(t) = g(t)\delta t \sum_{p=-\infty}^{\infty} \delta(t - p\delta t) = \delta t \sum_{m=-\infty}^{\infty} g(p\delta t)\delta(t - p\delta t) = \delta t \sum_{p=-\infty}^{\infty} g_p \delta(t - p\delta t) \tag{9}$$

where g_p are the sampled values $g(p\ \delta t)$ and p is an arbitrary integer. The superscript "s" stands for "sample". One may notice that when $\delta t \to 0$ we have $g^s \to g$.

Now in order to obtain G we may apply CFT to g, but we may prefer to apply instead DFT defined in Eq. (3) for the reasons already mentioned in section 1. A sampled Fourier spectrum does not mean necessarily that the Fourier transform has to be performed discretely; we may perform it continuously and then sample the resulting continuous spectrum. But this would be a waste of effort. It is preferable to compute the Fourier spectrum discretely. But what physical correspondence has the discrete transform in reality, where the transform is continuous? To find out one needs to express DFT in terms of CFT. In other words using CFT of the sampled function we try to obtain the Fourier spectrum also under a sampled form. Here we can make use of the Fourier series, which now prove to be, as we stated before, an intermediary link between CFT and DFT.

We know from subsection 2.2 that the CFT of a periodic function is a discrete even spaced function. Since we want G to be discrete, then we have to make g periodic. The input function g is not generally periodic but in most practical cases it has values only over a finite domain of arguments. Suppose g is non-zero only for arguments in the interval $t \in (-\Delta t/2 \ldots \Delta t/2)$. We define a periodic function g^p as

$$g^{p}(t) = \sum_{m=-\infty}^{\infty} g(t + m\Delta t) \tag{10}$$

The superscript "p" means that g^p is the periodic version of g. A function that is non-zero only in the interval $(-\Delta t/2 \ldots \Delta t/2)$, does not have to be sampled to infinity, but only where has non-zero values. Sampling g over the interval $(-\Delta t/2 \ldots \Delta t/2)$ gives us the same quantity of information as sampling g^p to infinity if, for simplicity, we choose the sampling interval δt so that

$$\Delta t = N\delta t \tag{11}$$

where N is an integer. To make the input function both discrete and periodic, we have to combine the forms (9) and (10) of g together, and, taking into account (11), we obtain

$$g^{sp}(t) = \delta t \sum_{m=-\infty}^{\infty} \sum_{p=-N/2}^{N/2-1} g_p \delta(t + m\Delta t - p\delta t) \tag{12}$$

Here the superscript "sp" means that the function g^{sp} is both sampled (discrete) and periodic. We know from the general properties of CFT of its reciprocal character (Bracewell, 1965; Goodman, 1996) and that the inverse CFT is very similar to CFT itself [see Eqs. (1) and (8); DFT has a similar property]. A double CFT reproduces the input function up to an inversion of coordinates. Therefore, if the CFT of a periodic function is an even-spaced

discrete sampling, then the CFT of an even-spaced discrete sampling has to be a periodic function. Let us check this assertion by calculating the CFT of (12)

$$\begin{aligned} G^{sp}(f) &= \delta t \int_{-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \sum_{p=-N/2}^{N/2-1} g_p \delta(t + m\Delta t - p\delta t) \exp(-i2\pi f t)\, dt = \\ &= \delta t \sum_{p=-N/2}^{N/2-1} g_p \exp(-i2\pi f p \delta t) \sum_{m=-\infty}^{\infty} \exp(i2\pi f m \Delta t) \end{aligned} \tag{13}$$

We make use again of the property (6) and obtain

$$\begin{aligned} G^{sp}(f) &= \delta t \sum_{p=-N/2}^{N/2-1} g_p \exp(-i2\pi f p \delta t) \frac{1}{\Delta t} \sum_{q=-\infty}^{\infty} \delta\left(f - \frac{q}{\Delta t}\right) = \\ &= \frac{1}{N} \sum_{q=-\infty}^{\infty} \delta\left(f - \frac{q}{\Delta t}\right) \sum_{p=-N/2}^{N/2-1} g_p \exp(-i2\pi q p / N) \end{aligned} \tag{14}$$

We notice that G^{sp} is periodic with the period $\Delta f = 1/\delta t$, because adding $n/\delta t$ to the argument f this causes only a reindexation with nN of the infinite sum of delta functions that does not cause any modification to G^{sp} precisely because the sum is infinite. We also notice that the coefficients of the delta functions are, up to a multiplication constant, the DFT of Eq. (3). We may rewrite then G^{sp} as

$$G^{sp}(f) = \frac{1}{N} \sum_{n=-\infty}^{\infty} \sum_{q=-N/2}^{N/2-1} G_q \delta\left(f - \frac{q}{\Delta t} - \frac{n}{\delta t}\right) \tag{15}$$

In the expression (15) the periodic character of G^{sp}, namely of period $1/\delta t$, is more clearly visible than in Eq. (14). Also G^{sp} is sampled at intervals of $\delta f = 1/\Delta t$. Both g^{sp} and G^{sp} are discrete and periodic. This is connected with the property of the function comb that it is invariant to CFT (Bracewell, 1965; Goodman, 1996), in other words the CFT of the comb function is also the comb function. The sampling of g is made with a comb function and it was to be expected to retrieve the comb function in the expression of G^{sp}. One may say that the DFT of a sampled function g is the CFT of the comb function weighted with g^p and the result is a comb function weighted with G^p. It should be noted that G^{sp} has the same number of distinct elements as g^{sp}, N. This is to be expected since through Fourier transform no information is lost but only represented differently. We should be able to retrieve the same amount of information from G as from g, therefore the number of samples should be the same for both functions. It is also noteworthy the inverted correspondence between the sampling intervals and the periods of g^{sp} and G^{sp}. The sampling interval of G^{sp} is the inverse of the period of g^{sp}, and the period of G^{sp} is the inverse of the sampling interval of g^{sp}.

One may see now that the correspondence between CFT and DFT given in references (Collier et al., 1971; Yaroslavsky & Eden, 1996) is of a different kind than that shown above. In those references the output function G is not discrete. Only the sampled values of G correspond to the DFT of only the sampled values of g. Also our relation between CFT and DFT has the advantage of illustrating better some properties of DFT such as its cyclic or toroidal character (Collier et al., 1971).

Eqs. (12,15) represent the physical equivalent of DFT. Only for a function like g^{sp}, periodic and consisting of evenly spaced samples, can one calculate the Fourier transform as in Eq. (3), i.e. discretely, and such functions do not exist in reality, and one may wonder what is the usefulness of it all. There is usefulness inasmuch as g^{sp} relates to g and G^{sp} to G, and they are related because we built g^{sp} starting from g. But constructing g^{sp} we adapted g for the purpose of discrete computing and we departed from the original g and consequently from G. Now we have to find how close are the CFT and the physical equivalent of DFT performed on g and how we can bring them closer. But before that we think we should try the reciprocal approach, the bottom-up approach as one can call it, starting from the simplest formalism, the DFT and arriving at CFT, of course, again via the Fourier series, which seem to be the accomplished intermediary.

3. The translation of CFT in the terms of DFT or the bottom-up approach

The Fourier (or harmonic) analysis is a methodology used to represent a periodic function into a series of harmonic functions. The harmonic functions are well known elementary functions. Fourier analysis is applicable only for linear systems, where the principle of linear superposition is valid.

Let $g(t)$ be a real or complex periodic function, having the period Δt. The set of functions

$$\psi_k(t) = \exp\left(i2\pi f_k t\right),\ f_k = \frac{k}{\Delta t},\ k = -\infty,\ldots,+\infty \tag{16}$$

are the harmonic functions of g. Except for the constant function $\psi_0(t)=1$, all the other functions in the set exhibit oscillations with quantized frequencies f_k, which are integer multiples of $f_1=1/\Delta t$, called the fundamental frequency. These functions repeat periodically over the whole real axis. The real and imaginary parts of g are identical, but they are phase shifted: the real part has a phase delay of $\pi/2$ (a quarter of a period) relative to the imaginary part. This infinite set of harmonic functions is an orthonormal set over the range of $t\in[-\Delta t/2, \Delta t/2]$:

$$\left\langle \psi_m, \psi_n \right\rangle \overset{\text{def}}{=} \frac{1}{\Delta t}\cdot \int_{-\Delta t/2}^{\Delta t/2} \psi_m(t)\psi_n^*(t)dt = \frac{1}{\Delta t}\int_{-\Delta t/2}^{\Delta t/2} \exp\left[i2\pi\left(f_m - f_n\right)t\right]dt = \delta_{mn} = \begin{cases} 0, & m \neq n \\ 1, & m = n \end{cases} \tag{17}$$

The Sturm-Liouville theorem proves that a function f respecting the Dirichlet conditions can be expressed as a linear combination of the harmonic functions (Arfken & Weber, 2001, chapter 9).

$$g(t) = \sum_{k=-\infty}^{\infty} c_k \psi_k(t),\quad c_k = \left\langle g, \psi_k \right\rangle = \frac{1}{\Delta t}\int_{-\Delta t/2}^{\Delta t/2} g(t)\psi_k^*(t)dx = \frac{1}{\Delta t}\int_{-\Delta t/2}^{\Delta t/2} g(t)\exp\left(-i2\pi f_k t\right)dt \tag{18}$$

This expansion is called Fourier series and the coefficients c_k are called Fourier coefficients.

Theoretically, there is an infinite number of Fourier coefficients. However, above a certain cut-off order, their amplitudes become very small and we can neglect them. The abscissa of the spectrum is proportional to the frequency. The frequencies corresponding to the spikes are multiples of the fundamental spatial frequency $1/\Delta t$. At the same time, the multiples order is the index of the coefficient. For example, if we notice a strong spectral component at

the 10th position, we say that the 10th harmonic, of frequency $f_{10}=10\,f_1$, is one of the dominant harmonics of the spectrum. Since only a few number of spectral harmonics have significant amplitudes, we say that the given function g can be well approximated by a superposition of a few Fourier harmonics.

The bidimensional (2D) Fourier series extends the regular Fourier series to two dimensions and is used for harmonic analysis of periodic functions of two variables. If Δx and Δy are the periods of the $g(x, y)$ function along the directions defined by the x and y variables, we define two fundamental angular frequencies: $f_x=1/\Delta x$ and $f_y=1/\Delta y$. The basis of 2D Fourier series expansion is built up from 2D Fourier harmonics, which are products of two simple 1D harmonics:

$$\psi_{mn}(x,y)=\exp\left(i\,2\pi m f_x x\right)\exp\left(i\,2\pi n f_y y\right),\quad m,n=0,\pm1,\pm2,\ldots,\pm\infty \tag{19}$$

The Fourier series of the function $g(x, y)$ is double indexed, and the Fourier coefficients form a matrix.

$$\begin{aligned} g(x,y)&=\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}c_{mn}\,\psi_{mn}(x,y),\\ c_{mn}&=\left\langle g,\psi_{mn}\right\rangle=\frac{1}{\Delta x\,\Delta y}\int_{-\Delta x/2}^{\Delta x/2}dx\exp\left(-i2\pi m f_x x\right)\int_{-\Delta y/2}^{\Delta y/2}dy\,g(x,y)\exp\left(-i2\pi n f_y y\right)\end{aligned} \tag{20}$$

The continuous Fourier transform (CFT) may be understood by analyzing how the spectrum of the periodic function $g$changes as a result of enlarging its period or the gradual change of the spectrum. The larger the period Δt of g, the smaller the fundamental frequency $\delta f=2\pi/\Delta t$ is, and the quantized set of angular frequencies $f_k=k\delta f$ are bunching together. When defining the function g for an infinite period, reproducing its characteristic pattern only once, without reproducing it periodically, while outside we set it to equal zero, the function g is no more periodic, or we can say that we have extended its period to infinity, $\Delta t\to\infty$. In this limit case the spectrum is no more discrete, but it becomes continuous. Related to the continuous spectrum, we mention some facts:

a. The difference between two consecutive quantized frequencies turns infinitesimal: $f_{k+1}-f_k=1/\Delta t=\delta f\to0$, so we replaced the discrete values f_k by a continuous quantity f.
b. All the Fourier coefficient amplitudes shrank to zero. For this reason we replaced the Fourier coefficients by the quantities $\Delta t\,c_k$, which do not shrank to zero but remain finite and they became the new instruments of practical interest for describing the function g.

$$\Delta t\,c_k=\int_{-\Delta t/2}^{\Delta t/2}g(t)\psi_k^*(t)dt=\int_{-\Delta t/2}^{\Delta t/2}g(t)\exp\left(-i2\pi f_k t\right)dt,\quad \lim_{\Delta t\to\infty}\Delta t\,c_k=\int_{-\infty}^{\infty}g(t)\exp\left(-i2\pi f_k t\right)dt \tag{21}$$

c. The integer index k turns into a continuous variable when $\Delta t\to0$, hence it is more appropriate to denote the Fourier coefficients replacements $\Delta t\,c_k$ by a continuous function $G(f)$, that we call Fourier transform of the g function:

$$G(f)=\lim_{\Delta t\to\infty}\Delta t\,c_k=\int_{-\infty}^{\infty}g(t)\exp\left(-i2\pi f t\right)dt \tag{22}$$

d. The Fourier series from Eq. (18) approximates an integral, and on the limit $\Delta t \to \infty$ the series converge to that integral:

$$g(t) = \lim_{\Delta t \to \infty} \sum_{k=-\infty}^{\infty} c_k \psi_k(t) = \int_{-\infty}^{\infty} G(f) \exp(i2\pi f t) df \tag{23}$$

The rationale shown above for the transition from Fourier series to CFT is similar to the one shown elsewhere (Logofătu & Apostol, 2007). Therefore, if the function *g* is not periodic, it cannot be decomposed into a series of Fourier harmonics, but into a continuous superposition of Fourier harmonics, called Fourier integral. The Fourier integral decomposition is possible providing that the modulus of the non periodic function *g* can be integrated over the whole real axis, that is the integral $\int_{-\infty}^{\infty} |g(t)| dt$ should exist (and be finite). Very common types of functions that fulfil this condition, largely used in practical applications are the functions with finite values over a compact interval and with zero values outside that interval. We implicitly assumed that the function *g* considered above is of that type.

Now let us consider the definition of CFT (22) and the relation used to decompose the non-periodic function *g* into the Fourier integral (23). We notice that each transform is the inverse of each other:

$$G(f) = \int_{-\infty}^{\infty} g(t) \exp(-i2\pi t) dt, \quad g(t) = \int_{-\infty}^{\infty} G(f) \exp(i2\pi f t) dt, \quad g(t) \xleftarrow{\text{Fourier}} G(f) \tag{24}$$

We say the functions *g* and *G* form a pair of Fourier transforms. The function *G* is obtained by applying the direct Fourier transform to the function *g*, while the function *g* is obtained by applying the inverse Fourier transform to the function *G*.

The bidimensional (2D) Fourier transform extends the Fourier transform to two dimensions and is used for two variables functions, which should satisfy a similar condition: the integral $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |g(x,y)| dx dy$ should exist and be finite. The Fourier integral is double:

$$\begin{aligned} G(f_x, f_y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) \exp\left[-i2\pi\left(f_x x + f_y y\right)\right] dx dy, \\ g(x,y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G(f_x, f_y) \exp\left[i2\pi\left(f_x x + f_y y\right)\right] df_x df_y \end{aligned} \tag{25}$$

While the 1D Fourier transform can be used as an illustration, or as an approximation of the 2D Fourier transform, in the special cases where the input function *g* does not depend on one or two coordinates, but three or more, although mathematically treatable, they present no interest for the physicist, because the 3D limitation of the world restricts practical interest to maximum 2D Fourier transform.

DFT has the purpose to approximate the CFT, and it is used for reasons of computation speed convenience. Although DFT is an independent formalism in itself, it was formulated so that it converges to the genuine CFT. DFT needs the function $g(t)$ as a set of a finite number *N* of samples, taken at *N* equidistant sample points, within a Δt length interval:

$$t_m = m\Delta t/N, \quad g_m = g(t_m), \quad m = -N/2, -N/2+1, \ldots, N/2-1 \tag{26}$$

In practical applications the function *g* is only given as a set of samples, and even if one knows its analytical expression, in most cases it's not possible to determine its Fourier transform by analytical calculus.

The definition of DFT can be established after a series of approximations. First, one approximates the Fourier transform by a Fourier series, which is defined as a set of coefficients associated to a set of equidistant frequencies. For this purpose we extend the domain of the sampled function to the whole real axis, making the function periodic, with the period of Δt which contains the entire initial definition domain of the function, in order to be able to expand it in Fourier series. The harmonic functions used as a decomposition basis are sampled functions too:

$$\psi_{nm} = \psi_n(t_m) = \exp\left(i2\pi f_n t_m\right) = \exp\left(i2\pi m n/N\right), \quad m, n \in Z \tag{27}$$

We modify the definition of the scalar product of these functions replacing the integral by a sum that approximates it:

$$\begin{aligned} &\frac{1}{\Delta t}\int_{-\Delta t/2}^{\Delta t/2} \psi_m(t)\psi_n^*(t)dt \approx \frac{1}{\Delta t}\sum_{k=-N/2}^{N/2-1} \psi_m(t_k)\psi_n^*(t_k)\frac{\Delta t}{N} = \frac{1}{N}\sum_{k=-N/2}^{N/2-1} \psi_{mk}\psi_{nk}^*, \\ &\left\langle \psi_m, \psi_n \right\rangle \overset{\text{def}}{=} \frac{1}{N}\sum_{k=-N/2}^{N/2-1} \psi_{mk}\psi_{nk}^* = \frac{1}{N}\sum_{k=-N/2}^{N/2-1} \exp\left(i2\pi k\frac{m-n}{N}\right) = \delta_{mn}^{(N)} = \begin{cases} 0, & m-n \neq pN, \\ 1, & m-n = pN, \end{cases} \quad p \in \mathbf{Z} \end{aligned} \tag{28}$$

There are only *N* distinct discrete harmonic functions, which are linear independent and can build up an orthonormal basis, because they repeat periodically: $\psi_{n\pm N}(t)=\psi_n(t)$. The Fourier coefficients will be calculated in the same way, approximating the integral by a sum:

$$c_n = \frac{1}{\Delta t}\int_{-\Delta t/2}^{\Delta t/2} g(t)\psi_n^*(t)dt \approx \frac{1}{\Delta t}\sum_{k=-N/2}^{N/2-1} g(t_k)\psi_n^*(t_k)\frac{\Delta t}{N} = \frac{1}{N}\sum_{k=-N/2}^{N/2-1} g_k \exp\left(-i2\pi\frac{nk}{N}\right) \tag{29}$$

There are only a limited set of *N* Fourier coefficients, because they reproduce themselves with the *N* period too, $c_{n\pm N}=c_n$. The original discrete function *g* can be expanded into a series of *N* discrete harmonic functions:

$$g_m = g(t_m) = \sum_{k=-N/2}^{N/2-1} c_k \psi_k(t_m) = \sum_{k=-N/2}^{N/2-1} c_k \psi_{km} = \sum_{k=-N/2}^{N/2-1} c_k \exp\left(i2\pi\frac{km}{N}\right) \tag{30}$$

At this point we can define the discrete Fourier transform: it is a sampled function *G* whose samples are the set of *N* Fourier coefficients approximately calculated by sums in Eq. (29): $G_n=Nc_n$, $n=-N/2, -N/2+1, \ldots, N/2-1$. The samples of *G* are obtained applying a transform to the samples of *g* and they can be inverted in order to yield back the samples of *g* from that of *G* as shown below in Eq. (31).

$$\begin{aligned} &G_n = \sum_{m=-N/2}^{N/2-1} g_m \exp\left(-i2\pi\frac{mn}{N}\right), \quad g_m = \frac{1}{N}\sum_{n=-N/2}^{N/2-1} G_n \exp\left(i2\pi\frac{mn}{N}\right), \\ &m, n = -N/2, -N/2+1, \ldots, N/2-1 \quad g \overset{\text{DFT}}{\longleftrightarrow} G \end{aligned} \tag{31}$$

The two sets of samples from g and G form a pair of discrete Fourier transforms. The transform is a linear one and can be expressed by means of a square matrix of $N\times N$ dimensions:

$$\vec{G} = \hat{W}.\vec{g}, \quad \vec{g} = \hat{W}^{-1}.\vec{G}, \quad \hat{W}_{mn} = w_N^{mn}, \quad \hat{W}_{mn}^{-1} = w_N^{-mn}/N, \; w_N = \exp(-i2\pi/N) \tag{32}$$

where for clarity we used the arrow and the triangular hat over-scripts to designate vectors and matrices respectively; also, the dot signifies dot product or matrix multiplication. To make possible the matrix multiplication we assume that the vectors are columns, matrices with N rows and 1 column, a practice we will continue throughout the subsection. Actually the convention is that in any indexed expression the first index represents the row and the second the column. The absence of the second index indicates we deal with a column or a vector. More than three indexes means we deal with a tensor and this cannot be intuitively represented easily. Of course the values G_n do not equal the corresponding samples of the continuous Fourier transform, but they approximate them. The greater the N, the better the approximation will be.

The 2D discrete Fourier transform may be obtained easily by generalizing Eqs. (30-32). Namely, 2D DFT has the form

$$G_{pq} = \sum_{m=-M/2}^{M/2-1} \sum_{n=-N/2}^{N/2-1} g_{mn} \exp\left[-i2\pi\left(\frac{pm}{M} + \frac{qn}{N}\right)\right] = \sum_{m=-M/2}^{M/2-1} \sum_{n=-N/2}^{N/2-1} g_{mn} w_M^{pm} w_N^{qn} \tag{33}$$

where $M\times N$ is the dimension of the matrix of samples g_{mn} and, consequently, the dimension of the matrix of the Fourier coefficients, or of the DFT G_{pq}, with M and N completely unrelated, and we also have the short hand notations

$$w_M = \exp(-i2\pi/M), \; w_N = \exp(-i2\pi/N) \tag{34}$$

The inverse discrete Fourier transform has, of course, the form

$$g_{mn} = \frac{1}{MN} \sum_{p=-M/2}^{M/2-1} \sum_{q=-N/2}^{N/2-1} G_{pq} \exp\left[i2\pi\left(\frac{pm}{M} + \frac{qn}{N}\right)\right] = \frac{1}{MN} \sum_{p=-M/2}^{M/2-1} \sum_{q=-N/2}^{N/2-1} g_{mn} w_M^{-pm} w_N^{-qn} \tag{35}$$

The linearity of the Fourier transform in Eq. (32) permits the matrix formulation of the direct and inverse 1D DFT. However the generalization to the 2D DFT leads us to a multidimensional matrix formulation:

$$\hat{G}_{(2)} = \hat{W}_{(4)} \cdot \hat{g}_{(2)}, \quad \hat{g}_{(2)} = \hat{W}_{(4)}^{-1} . \hat{G}_{(2)}, \quad \hat{W}_{pqmn} = w_N^{pm}/M\, w^{qn}/N, \quad \hat{W}_{mnpq}^{-1} = w_M^{-mp} w_N^{-nq} \tag{36}$$

where $\hat{g}_{(2)}$ and $\hat{G}_{(2)}$ are tensors of rank 2 (ordinary 2D matrices) and $\hat{W}_{(4)}$ and $\hat{W}_{(4)}^{-1}$ are tensors of rank 4. The direct and the inverse Fourier transforms are dot products of the tensors $\hat{W}_{(4)}$ and $\hat{W}_{(4)}^{-1}$ with the 2D matrices $\hat{g}_{(2)}$ and $\hat{G}_{(2)}$. The dot product of two tensors results in tensors with the rank equal to the sum of the tensors rank minus 2. Eqs. (33-36) are actually those with which one deals when operating 2D discrete Fourier transforms and not Eqs. (29-32). Eqs. (33-36) may seem complicated but the mastery of Eqs. (29-32) leads easily to the multidimensional forms. The term "tensor" was introduced for the sake of completeness but it does not change the simple elementary aspect of Eqs. (29-32) that are

expressed in tensor form in Eq. (36). For instance one may notice that the 2D DFT is actually two 1D DFTs applied first to the rows of the input matrix then to the resulted columns, although the order of the operations does not matter because the end result is the same.

The direct computation of all the samples G_n requires an amount of computation proportional to N^2. However, the $\hat{W}$ matrix has some special properties that enable massive reduction of the operations required to perform the matrix multiplication $\hat{W}.\vec{g}$. As far back as 1965 a method to compute the discrete Fourier transform by a very much reduced number of operations, the FFT algorithm, which allows computing the discrete Fourier transform with a very high efficiency is known. Originally designed for samples with the number of elements N being powers of 2, now FFT may be calculated for samples with any number of elements, even, what is quite astonishing, non-integer N. A fast algorithm for computing a generalized version of the Fourier transform named the scaled or fractional Fourier transform was also designed. The normalization factors from (29-32) of the direct and the inverse transforms are a matter of convention and convenience, but they must be carefully observed for accurate calculations once a convention was chosen.

Since subroutines for FFT calculations are widely available, there is no need to discuss here in detail the FFT formalism. For the interested reader we recommend (Press et al., 2002), chapter 12. We will only mention that the algorithm makes use of the symmetry properties of the matrix multiplication by the techniques called time (or space) decimation and frequency decimation, techniques that can be applied multiple times to the input in its original and the intermediary states, and with each application the computation time is almost halved. The knowledge of the FFT algorithm in detail may help the programmer also with the memory management, if that is a problem, because it shows one how to break the input data into smaller blocks, performs FFT separately for each of them and reunites them at the end.

4. Conversion of the input data for use by the FFT and conversion of the data generated by the FFT in order to have physical meaning

4.1 The transposition method

As we said, one does not need to know in detail the FFT algorithm in order to use it. The FFT subroutines can be used to a large extent as simple black boxes. There is, however, one fact about FFT that even the layman needs to know it in order to use the FFT subroutines. Namely, for mathematical convenience the DFT is not expressed in a physical manner as in Eqs. (29-32) where the current index runs not from $-N/2$ to $N/2-1$ but from 0 to $N-1$:

$$G_n = \sum_{m=0}^{N-1} g_m w^{mn}, \quad n = 0,1,\ldots,N-1 \tag{37}$$

$$g_m = \frac{1}{N}\sum_{n=0}^{N-1} G_n w^{-mn}, \quad m = 0,1,\ldots,N-1 \tag{38}$$

This shifting of the index allows the application of the decimation techniques we talked about, but also has the effect of a transposition of the wings of the input and as a consequence the, say, "mathematical" output is different than the "physical" output, the one that resembles what one obtains in a practical experiment, although the two outputs are, of

course, closely connected. The reference (Logofătu & Apostol, 2007) shows that in order for formulae (37,38) to work the wings of the input vector should be transposed before the application of the FFT procedure and then the wings of the output vector should be transposed back all in a manner consistent with the parity of the number of samples. Namely for even N the input and the output vectors are divided in equal wings. However, for odd N the right wing of the input starts with the median element, therefore is longer with one element; but in the case of the output it is the left wing which contains the median element and is longer. This transposition of the wings is the same thing as the rotation of the elements with $N/2$ when N is even, and $(N–1)/2$ when N is odd. For the case of odd N the direction of the rotation is left for the input and right for the output. For even N the direction does not matter.

In a 2D case when both M and N are even the phase change due to the "mathematical" transpositions is just an alternation of signs, in a chess board style. In other situations the phase change is more complicated. It is true that in most cases it is the amplitude spectrum that matters most, but sometimes the phase cannot be neglected and the transposition or rotation operations mentioned above have to be performed. In the 2D case the transpositions do not have to be a double series of wing transpositions for rows and columns. One can make just two diagonal transpositions of the quadrants of the input and output matrices. The division of the input and output matrices depends on the parity of M and N. For even M and N things are simple again. The matrices are divided in four equal quadrants. When one of the dimensions is odd the things get complicated, but here again we have a simple rule of thumb. If the number of rows M is odd, then the left quadrants of the input matrix have the larger number of rows (one more) while the left quadrants of the output matrix have the smaller number (one less). For odd N the lower quadrants of the input matrix have the larger number of columns (one more) while the lower quadrants of the output matrix have the smaller number (one less). And viceversa.

4.2 The sign method

In addition to the procedure with the transposition of the input before the FFT and the inverse transposition of the output after the FFT, there is another solution for reconciling the results of the mathematical calculation with the physics. It can be done by substituting the indexes of DFT with the indexes used by FFT, thus expressing DFT in the terms preferred by FFT. The expression of DFT that we use is the one from Eq. (31)

$$G_n = \sum_{m=-N/2}^{N/2-1} g_m \exp\left(-i2\pi\frac{mn}{N}\right) \tag{39}$$

We need to make a reindexation of (39) so that it looks like (37). Let the integer indexes p and q be so that

$$\begin{aligned} &p = m + N/2 \text{ or } m = p - N/2,\ p = 0,1,\ldots,N-1 \\ &q = n + N/2 \text{ or } n = q - N/2,\ q = 0,1,\ldots,N-1 \end{aligned} \tag{40}$$

which means we can introduce two similar sets of samples g' and G', having the indexes shifted by $N/2$

$$g'_p = g_{p-N/2} = g_m,\ G'_q = G_{q-N/2} = G_n \tag{41}$$

Introducing (40) and (41) in (39) we obtain

$$G_n = G'_q = \sum_{p=0}^{N-1} g'_p \exp\left[-i2\pi\frac{(p-N/2)(q-N/2)}{N}\right] = (-1)^{N/2}(-1)^q \sum_{p=0}^{N-1}\left[(-1)^p g'_p\right]\exp\left(-i2\pi\frac{pq}{N}\right) \quad (42)$$

Therefore, in order to obtain physically meaningful results all we need to do is to multiply the input data with an array consisting in alternating signs and starting with +1, perform the FFT and then multiply the result again with the same sign array and an overall $(-1)^N$ sign. Then we can identify the G'_q element of the final output with G_n, the desired element.
For odd N the sign method cannot be applied as such. Instead of signs we have exponentials with imaginary arguments. Although it has a more messy appearance the conversion method is still simple for odd N too. The new indexes are

$$\begin{aligned} &p = m + (N-1)/2 \text{ or } m = p - (N-1)/2,\ p = 0,1,\ldots,N-1 \\ &q = n + (N-1)/2 \text{ or } n = q - (N-1)/2,\ q = 0,1,\ldots,N-1 \end{aligned} \quad (43)$$

The conversion formula is

$$G_n = G'_q = \exp\left[-i\pi\frac{(N-1)^2}{2N}\right]\exp\left(i\pi q\frac{N-1}{N}\right)\sum_{p=0}^{N-1}\left[\exp\left(i\pi p\frac{N-1}{N}\right)g'_p\right]\exp\left(-i2\pi\frac{pq}{N}\right) \quad (44)$$

Instead of sign array one has to use an array with the complex unitary elements $\exp[i\pi p(N-1)/N]$ with p running from 0 to N–1. One has to multiply the input data with this array before feeding it to the FFT procedure. The outcome must be multiplied again with this array and with an overall constant factor $\exp[-i\pi(N-1)^2/(2N)]$. The generalization to 2D is straightforward in both cases. This method, although somewhat similar to the one shown in (Logofătu & Apostol, 2007) is actually better and simpler.

5. Correspondences to reality

A particular case in which we may talk in a sense of "naturally" sampled input functions is the case of binary masks (transmittance 0 or 1) that are formed out of identical squares. A 1D grating as in Fig. 1.a (mask A), or a 2D grating as in Fig. 1.b (mask B) are such examples. We chose those masks because our purpose is to compare the discrete and the continuous Fourier transforms and for those masks CFT can be calculated analytically. The masks are not sampled functions in the sense of Eq. (9), there are no delta functions in their expression. They are, however, sampled in the sense that for evenly spaced rectangular areas the transparency functions are constant, hence, the functions can be represented, as discrete matrices of samples. We chose masks with a low number of samples, only 32×32, not because DFT is difficult or time-consuming (actually, due to the existence of the FFT algorithm DFT can be done very quickly for quite a large number of samples), but to make easier the calculation of CFT, which is indeed considerably time-consuming. Also, at small number of samples the differences between the discrete and the continuous spectrum can be more easily seen. The period of the gratings, both horizontally and vertically, was chosen to be 7 squares, so that it does not divide 32; in this way the Fourier spectrum gets a little complicated and we avoid symmetry effects that may obscure the points we want to make.

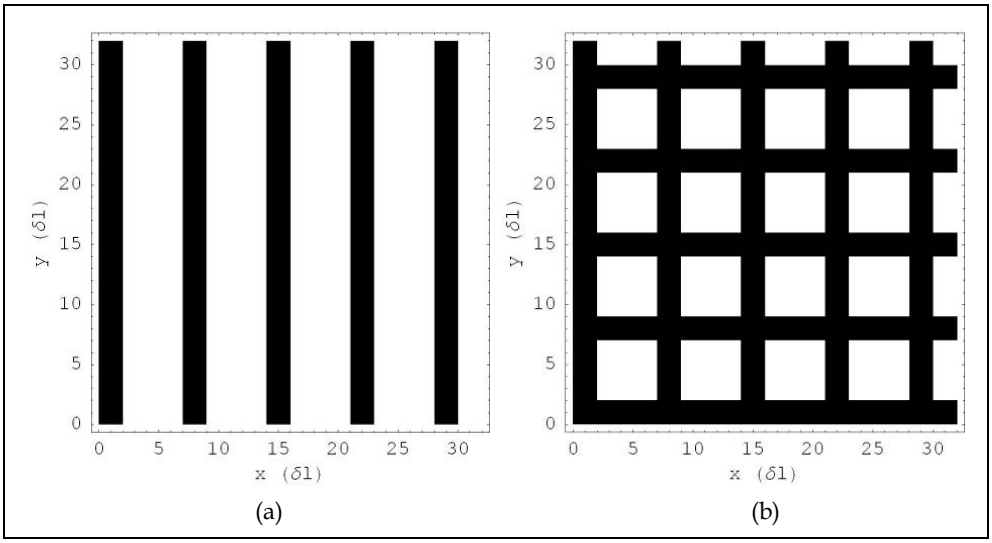


Fig. 1. Binary masks: a) 1D grating and b) 2D grating. Both gratings consist in 32×32 squares, black or white, of equal dimension δl. We labelled the two masks "A" and "B". It is assumed that outside the represented area of the masks the input light is completely blocked, i.e. the transparency function is zero.

It seems natural to represent the masks from Fig. 1, for DFT calculations, as 32×32 matrices with values 1 or 0 corresponding to the transmission coefficient of the squares of the masks. The squared absolute values of the DFT of the previously discussed type of matrix for mask A and the CFT of mask A are shown together in Fig. 2. The square absolute value is the power or the luminous intensity, the only directly measurable parameter of the light field. The CFT of mask A has the expression

$$G\left(f_x, f_y\right) = \delta l \,\mathrm{sinc}\left(f_x \delta l\right) 32\, \delta l \,\mathrm{sinc}\left(32\, f_y \delta l\right) \sum_{p=1}^{32} g_p \exp\left[-i\, 2\pi f_x \left(p-1\right)\delta l\right] \tag{45}$$

where

$$\mathrm{sinc}\left(t\right) \equiv \frac{\sin\left(\pi t\right)}{\pi t} \tag{46}$$

and obviously N is here 32. Since mask A is 1D [in a limited sense only; if it were truly 1D then the dependence on f_y of G would be of the form $\delta(f_y)$] in the calculation of $|G|^2$ in Fig. 2 and the following figures that represent $|G|^2$ for mask A, we discarded the dependence on f_y and the entire contribution of the integration over y and we used only the part corresponding to the integration over x. The DFT calculations were rescaled (multiplied with $N^{1/2}$) so that they could be compared to the CFT calculations. All the CFT spectra represented in this article are computer calculations, hence they are simulations. A correctly done experiment of Fourier optics would yield, of course, the same results.

Two types of discrepancies can be noticed in Fig. 2 between CFT and DFT. First they have different values for the same spatial frequency, sometimes there is even a considerable

difference. Second, the discrete spectrum does not offer sufficient information for the domain of spatial frequencies in between two discrete values, and the interpolation of the sampled values does not lead always to a good approximation. The structure of the continuous spectrum is much richer than that of the discrete spectrum. There are of course many reasons why the discrete Fourier spectrum $|G^{sp}|^2$ is different than the genuine spectrum $|G|^2$. One reason is the periodicity. The g input functions are not generally periodic or at least they are not infinitely periodic. The input functions illustrated in Fig. 1 are periodic in the sense they have a limited number of periods, but rigorously periodic means an infinity of periods.

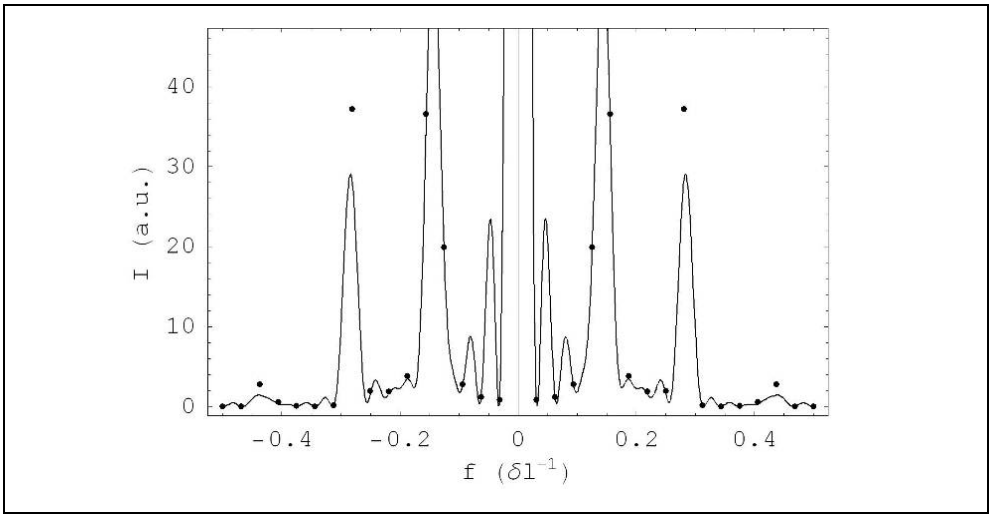


Fig. 2. The continuous (solid line) and the discrete (dots) Fourier power spectra for mask A vs. the spatial frequency shown together. For DFT calculations the "natural" sampling was used. Only the central part of the continuous spectrum for which DFT provides output values is represented. The spatial frequency is expressed in δl^{-1} units.

Corresponding to the two types of discrepancies there are also two ways of improving the discrete spectrum. One way is to increase the sampling rate. We can do that by "swelling" the sampling array, inserting more than once the value corresponding to a square. The increased sampling rate improves the agreement between DFT and CFT. We illustrated in Fig. 3 only the calculations for the same spatial frequencies as those represented in Fig. 2, not just to ease the comparison but also because the spectrum outside is very weak and, hence, negligible. One big the difference between the CFT and DFT version represented in Fig. 2 is the sinc(f_x δl) function, that is the Fourier transform of the rectangular function. Because for Fig. 3 by allotting more samples for each square we had a higher sampling rate, the rectangular shape was felt in the DFT calculations. We increased the sampling rate 10 times and, as one can see in Fig. 3, the continuous and the discrete spectrum are now closer, they are almost on top of each other. Because there are now 10 times more elements than in the "natural" sampling, and actually in this new sampling the same elements are just repeated 10 times, in order to be able to compare DFT and CFT we had to divide the DFT spectrum by 10^2.

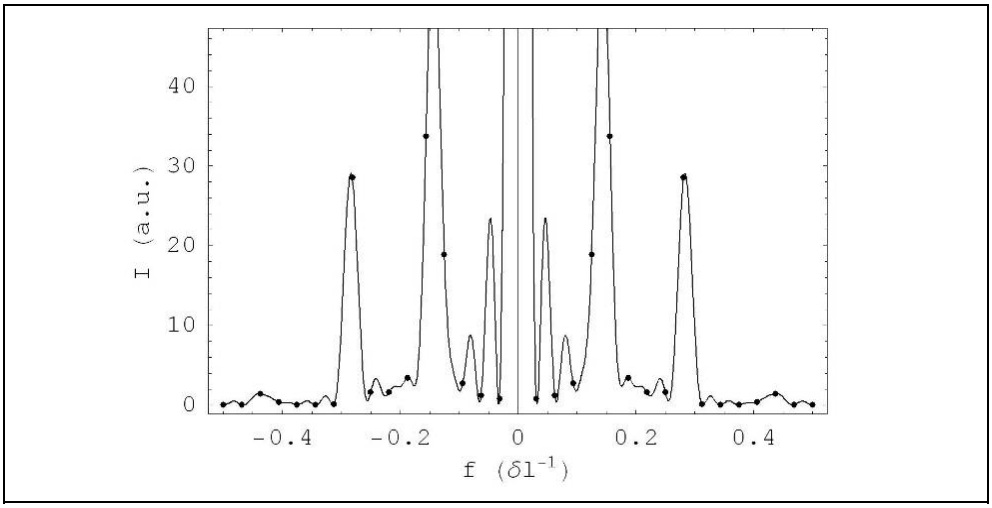


Fig. 3. The continuous (solid line) and the discrete (dots) Fourier power spectra for mask A vs. the spatial frequency shown together. For the DFT calculations was used a higher (10 times) sampling rate of mask A than that used for the DFT calculations shown in Fig. 2. Only the portion of the both spectra corresponding to the spatial frequency range of Fig. 2 is shown.

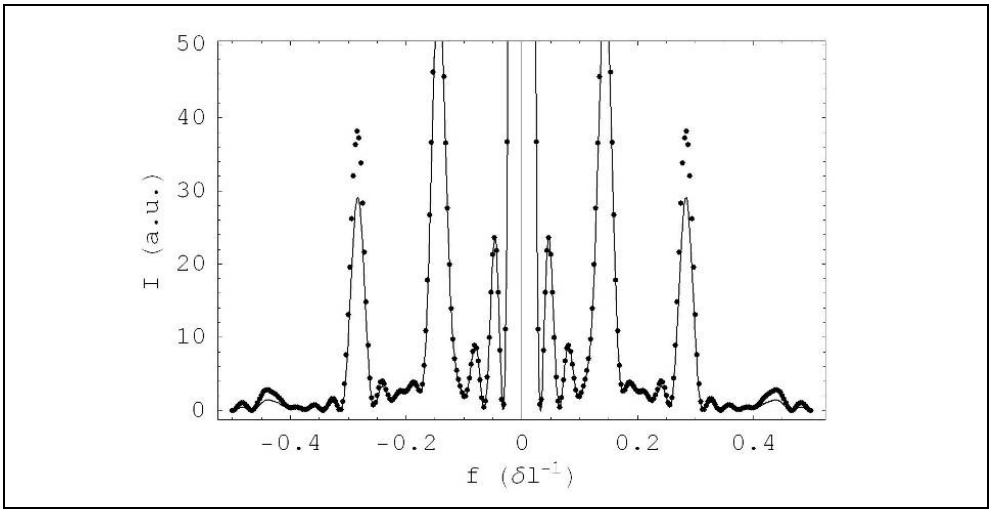


Fig. 4. The continuous (solid line) and the discrete (dots) Fourier power spectra for mask A vs. the spatial frequency shown together. The sampling of mask A was extended so that to include part of the surrounding darkness shown together, maintaining the same sampling rate as the "natural" sampling. Compared to the original sampling used for the calculations of Fig. 2, the sampling was now extended 10 times, which accounts for the higher density of dots of the DFT output, which now numbers 320 samples. Although in the figure 320 points are represented, the spatial frequency range is the same as that of Figs. 2 and 3.

Another way to improve the conformity of the DFT to CFT is to increase the sampled area. Since outside the area of mask A there is no structure, only darkness, and g is just zero, when making the sampling of g we are generally tempted to discard this surrounding darkness. But when we express DFT in terms of CFT, as we saw in section 3.2, we consider that the structure of mask A is periodically repeated back to back in what we know to be just darkness. Therefore, in order to improve the similarity of DFT to reality (which is CFT), it is a good idea to pad the original sampled function with zeros to the left and to the right to account for the surrounding darkness. We padded with zeros so that the original sampled array of values was increased 10 times. In Fig. 4 the squared absolute values of the DFT and CFT for the new sampled function are again compared, and, although their values are still different, now DFT offers more information, enough for interpolation. We did not need here another type of rescale of DFT than that done in Fig. 2 in order to have a meaningful comparison to CFT, because the new elements added to the input sampling were just zeros.
The two procedures for improving the similarity of DFT to CFT described above and illustrated in Figs. 3 and 4 may be combined and the result is shown in Fig. 5. The sampling array used for the calculation of DFT is now both "swollen" and extended, having 100 times more elements than the "natural" sampling. Now DFT is both closer to CFT and richer in information. Now DFT is both accurate and able to provide enough information for a correct interpolation.
It should be noted that in Figs. 2-5 only the discrete spectrum changes, the continuous spectrum is a constant reference.

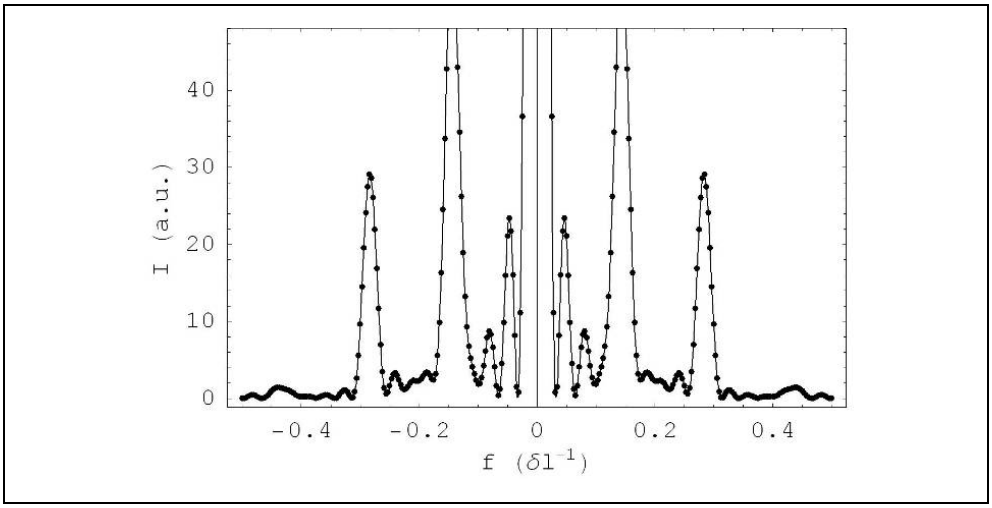


Fig. 5. The continuous (solid line) and the discrete (dots) Fourier power spectra for mask A vs. the spatial frequency shown together. For DFT calculations the sampling was both extended and its rate increased. An array of 32×10×10 points was used, but only the 32×10 points corresponding to the spatial frequency range of Figs. 2-4 were shown.

The similar procedure applied to mask A was also applied to mask B. We found appropriate to illustrate the procedure for mask B because optics is generally about images and these are 2D, not 1D, which is just a particular case, useful mostly for the easiness of the graphic representation than for practical purposes. The Fourier spectrum of a 2D mask is more

difficult to represent. We chose to represent the spectrum as levels of grey. Moreover, to simulate the vision of the eye we represented the logarithm of the luminous intensity. Another reason for using the logarithmic representation is the fact that the Fourier spectrum decreases quite sharply with the spatial frequency and only the representation of the logarithm allows the fine shades to be visible.
In Fig. 6 the represented continuous Fourier spectrum is the logarithm of the squared absolute value of the function

$$G\left(f_x, f_y\right) = \delta l^2 \operatorname{sinc}\left(f_x \delta l\right) \operatorname{sinc}\left(f_y \delta l\right) \sum_{p=1}^{32} \sum_{q=1}^{32} g_{pq} \exp\left[-i\,2\pi\left(f_x\left(p-1\right) + f_y\left(q-1\right)\right)\delta l\right] \qquad (47)$$

In Fig. 7 it is represented the discrete Fourier spectrum of the "natural" sampling of mask B, that is of a 32×32 matrix with each element having the value of the corresponding square element of mask B, 0 or 1. The comparison of Figs. 6 and 7 shows marked differences. Just as in the case of mask A we tried next to compensate for the shortcomings of the "natural" sampling by extending the sampling and increasing the sampling rate. The result is right on top of Fig. 6, so we did not consider necessary to represent it graphically.

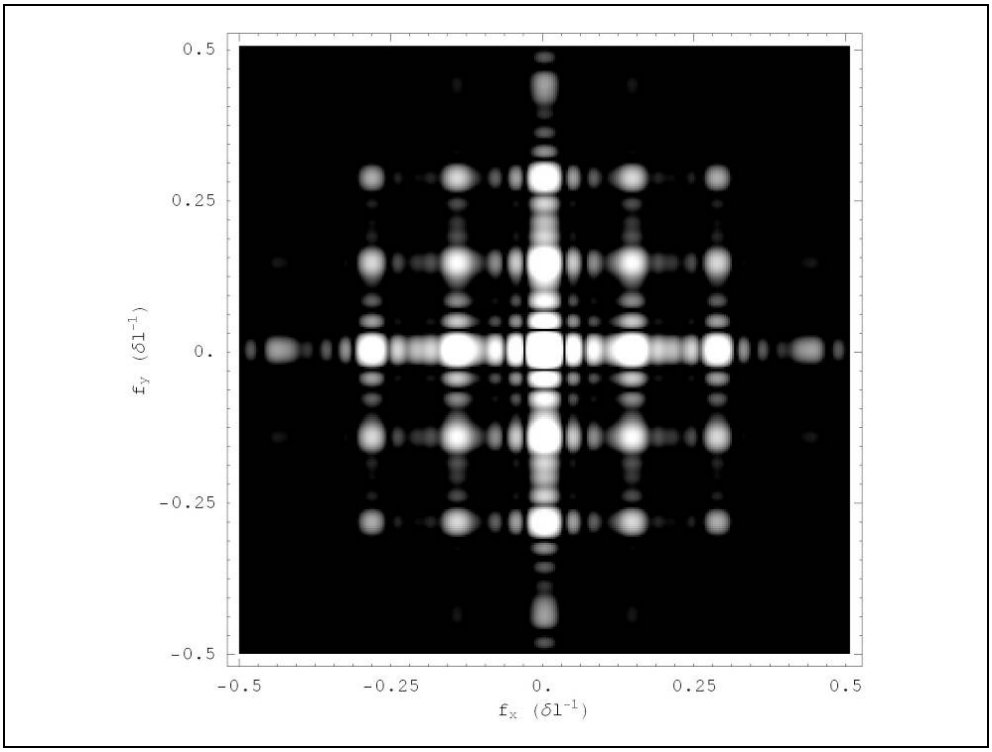


Fig. 6. The central high intensity portion of the continuous Fourier spectrum of mask B, chosen so that to match the spatial frequency ranges of the DFT of the "naturally" sampled input of mask B (see Fig. 7 below). The abscissa and the ordinate are the spatial frequencies and the light intensity of the spectrum is coded as levels of grey.

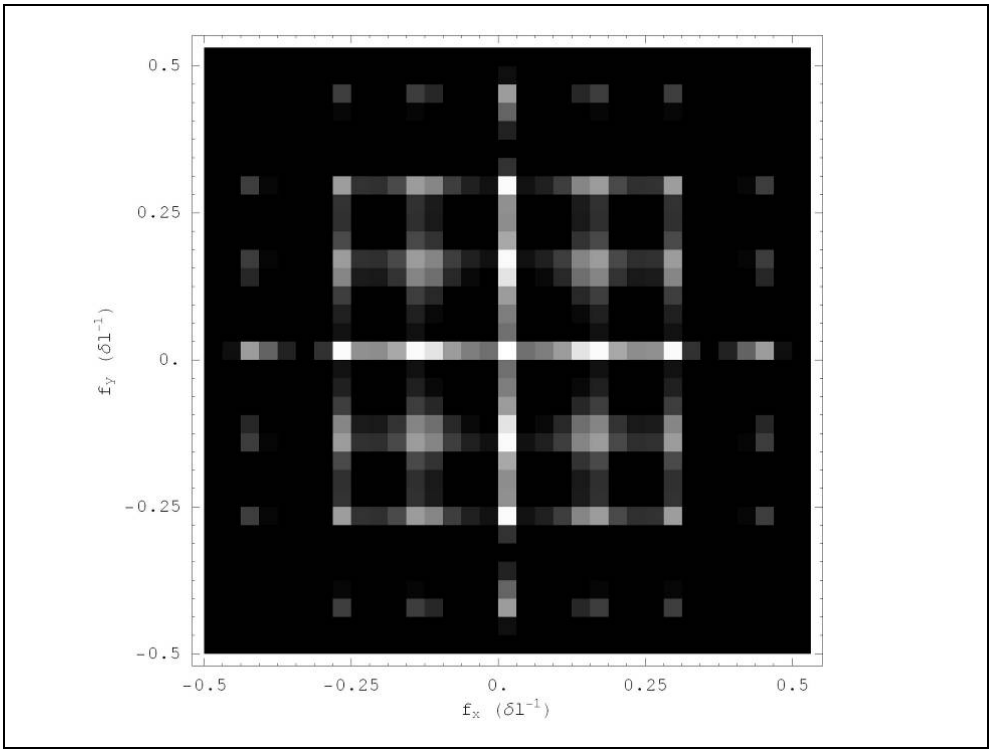


Fig. 7. The discrete Fourier spectrum of the "natural" sampling of mask B. The abscissa and the ordinate are the spatial frequencies, and the light intensity is codes as levels of grey, just as in Fig. 6.

As a side comment, one may notice that all the spectra shown in this article are symmetric. The 1D plots are symmetric with respect to the origin, and the 2D plots are symmetric with respect to the vertical axis. There is a redundancy of information. The 1D plots contain in one horizontal half all the information, and the 2D plots contain all the information in any of the 4 quadrants. This is due to the fact that we calculated the spectra of transmission masks that do not modify the phase of the input optical fields and we assumed the input wave to be a plane wave, which is actually a common situation in Fourier analysis. It is this property of the masks that causes the spectra to be symmetric. Rigorously speaking they are not symmetric, the phase differ in the two halves of the 1D plots and in the 2D plots the phase of two diagonal quadrants differs from that of the other two diagonal quadrants. But the difference is just that they have conjugate complex values. The absolute values of two conjugate complex quantities is the same, hence the 4-fold symmetry of the 2D power spectra.

To give some dimensional perspective to the considerations presented in this subsection, it might be instructive to give a value to δl and to specify the experimental conditions in which the Fourier transform is performed. We need very small masks in order to make the Fourier spectrum macroscopic, but also large enough so that sufficient light passes through and the Fourier spectrum is visible. 100 μm is such a value for δl. Then the masks A and B would be

squares of 3.2 mm dimension. The Fourier spectra (continuous or discrete) represented in Figs. 2-8 are segments for the 1D case and squares for the 2D case having the dimension $\Delta f = 1/\delta l = 10$ mm^{-1} in spatial frequency units. If the Fourier transform is performed by a lens of focal length $F = 1$ m and the light source is a He-Ne laser of wavelength $\lambda = 632.8$ nm, then dimensions of both spectra in the Fourier plane (back focal plane of the lens) are identical $\lambda F \Delta f = 6.328$ mm.

6. Conclusion

The problem of the relation between DFT and CFT is investigated in this article. In order to understand the physical meaning of DFT we expressed it in terms of CFT. The Fourier series was a useful tool in this endeavour, because it is an intermediary link between CFT and DFT. Namely, the two properties of both the input function and the Fourier spectrum of DFT, periodicity and discrete character, are present in the Fourier series, except that the input function is just periodic and the Fourier spectrum is just discrete. The connection between periodicity and the discrete character is stressed, namely it is shown that the periodicity of the input/output implies a discrete character of the output/input and vice-versa. For convenience the derivations were made for 1D input functions but they can be easily and straightforwardly extended to 2D input functions, if the sampling is done over two mutually perpendicular directions and the sampled area is rectangular. It is shown that DFT is the CFT of a periodic input of delta functions in which case the output is also periodic and composed of delta functions. The incongruence between DFT and CFT indicates that DFT may not be a good approximation of CFT, and some numerical examples prove it. Two masks, first 1D and the second 2D were studied with respect to the agreement of their discrete with their continuous Fourier spectra. It has been shown that if the sampling rate and the extension of the masks are properly chosen (large enough) DFT is a good approximation of CFT. No generally valid criterion for the agreement between DFT and CFT is given, only the ways of improving it are indicated and shown to be sufficient for the particular cases studied in this article.

Our previous attempts to bridge the gap between CFT and DFT, between physics and mathematics (Logofătu and Apostol, 2007; Nascov et al, 2010) were by no means a closed and shut subject but rather were intended as an opening of new avenues of research. Some details, usually left out by other authors, such as the transposition of the input data done for the application of the FFT algorithm are explained and two solutions for dealing with the problem are presented. The second solution, presented in subsection 4.2 even shows how the transposition of the input leaving the amplitude unchanged modifies the phase with a linear progressive phase function.

7. References

Apostol, D.; Sima, A.; Logofătu, P. C.; Garoi, F.; Damian, V., Nascov, V. & Iordache, I. (2007). Static Fourier transform lambdameter, *Proceedings of ROMOPTO 2006: Eighth Conference on Optics*, Vol. 6785, pp. 678521, ISSN 0277-786X, Sibiu, Romania, August 2007, SPIE, Bellingham, WA (a)

Apostol, D.; Sima, A.; Logofătu, P. C.; Garoi, F.; Nascov, V.; Damian, V. & Iordache, I. (2007). Fourier transform digital holography, *Proceedings of ROMOPTO 2006: Eighth*

Conference on Optics, Vol. 6785, pp. 678522, ISSN 0277-786X, Sibiu, Romania, August 2007, SPIE, Bellingham, WA (b)

Arfken, G. B. & Weber H. J. (2001). *Mathematical methods for physicists*, Harcourt / Academic Press, ISBN 0-12-059826-4, San Diego, chapters 14,15

Bailey, D. H. & Swarztrauber, P. N. (1991). The Fractional Fourier Transform and Applications, *SIAM Review*, Vol. 33, No. 3, pp. 389-404, ISSN 0036-1445

Bracewell, R. (1965) *The Fourier Transform and Its Applications*, McGraw-Hill, New York

Brigham, E. O. (1973) *The Fast Fourier Transform: An Introduction to Its Theory and Application*, Prentice-Hall, Englewood Cliffs NJ

Bringdahl, O. & Wyrowski, F. 1990. Digital holography - computer generated holograms, In *Progress in Optics* vol, 28, E. Wolf, (Ed.), pp. 1-86, Elsevier B.V. ISBN: 9780444884398

Collier, R. J.; Burckhardt, C. B. & Lin, L. H. (1971) *Optical holography*, Academic Press, New York

Cooley, J. W. & Tukey, J. W. (1965). An algorithm for the machine computation of the complex Fourier series, *Mathematics of Computation*, Vol. 19, No. 90, pp. 297-301, ISSN 0025-5718

Goodman, J. W. (1996). *Introduction to Fourier Optics*, McGraw Hill, ISBN 0-07-024254-2, New York

Lee, W.-H. (1978). Computer-Generated Holograms: Techniques and Applications, In *Progress in Optics* vol. 16, E. Wolf, (Ed.), pp. 119-232, Elsevier B. V., ISSN 00796638,

Lohmann, A. W. & Paris, D. P. (1967). Binary Fraunhofer holograms, generated by computer, *Applied Optics* Vol. 6, No. 10, pp. 1739-1748, ISSN 0003-6935

Logofătu, P. C. & Apostol, D. (2007). The Fourier transform in optics: from continuous to discrete or from analogous experiment to digital calculus, *Journal of Optoelectronics and Advanced Materials*, Vol. 9, No. 9, pp. 2838-2846, ISSN 1454-4164

Logofătu, P. C.; Sima, A. & Apostol, D. (2009). Diffraction experiments with the spatial light modulator: the boundary between physical and digital optics, *Proceedings of Advanced Topics in Optoelectronics, Microelectronics, and Nanotechnologies IV* Vol. 7297, pp. 729704, ISSN 0277-786X, Constanta, Romania, January 2009, SPIE, Bellingham WA

Logofătu, P. C.; Garoi, F.; Sima, A.; Ionita, B. & Apostol, D. (2010). Classical holography experiments in digital terms, *Journal of Optoelectronics and Advanced Materials*, Vol. 12 No. 1, pp. 85-93, 1454-4164

Nascov, V. & Logofătu, P. C. (2009). Fast computation algorithm for the Rayleigh-Sommerfeld diffraction formula using a type of scaled convolution, *Applied Optics* Vol. 48, No. 22, pp. 4310-4319, ISSN 0003-6935

Nascov, V.; Logofătu, P. C. & Apostol, D. (2010). The Fourier transform in optics: from continuous to discrete (II), *Journal of Optoelectronics and Advanced Materials*, Vol. 12, No. 6, pp. 1311-1321, ISSN 1454-4164

Press, W. H.; Teukolsky, S. A.; Vetterling, W. T. & Flannery, B. P. (2002). *Numerical Recipes in C++*, Cambridge University Press, ISBN 0-571-75033-4, Cambridge

Yaroslavsky, L. & Eden, M. (1996) *Fundamentals of digital optics*, Birkhäuser, ISBN 0-8176-3822-9, Boston

2

The Holographic Principle and Emergence Phenomenon

Marina Shaduri
Center of Biohолography, Ltd.
Tbilisi,
Georgia

1. Introduction

The present work was inspired by a serendipitous discovery of non-local effects in living organisms, which could not be explained by the known biological mechanisms. We have demonstrated on a large number of subjects (up to 13 000) that any small part of a human body, when exposed to pulsed electromagnetic fields, produces the interference patterns that carry diagnostically significant information; more precisely, we found that the shapes and textures of the most disorderly anatomic structures can be analyzed using minor superficial areas of the body as a source of information. This finding required a rational scientific explanation.

The studies conducted in the conditions of minimal perturbation made it possible to unveil some physical mechanisms underlying the non-local phenomena in complex systems of natural origin [Shaduri et al., 2002; 2008a]. The holographic principle offered by physicists as a solution to information-associated processes in certain (non-living) natural objects turned out to have more general scope of applicability. The real-time encoding and decoding of information have been detected in both - humans and animals [Shaduri, 2005].

Our experience makes us believe that without penetrating waves such as X-rays or ultrasound focused upon the areas of interest, it is not possible to observe internal structures of intact living body. It came as a big surprise that diverse parts of living systems may communicate not only through exchange of molecular and nervous signals, but also „wirelessly". The wireless communication had been unimaginable before Heinrich Hertz proved it experimentally in 1888. Our clinical and experimental data that suggested the existence of some previously unknown mechanisms of information transfer in biological systems were met with ferocious resistance and misunderstanding as well: the physicists, who we addressed for help, could not believe that high-resolution images of internal organs and tissues could reach the outer surface of the human body. So, a small team of biologists, medical doctors and engineers was left to investigate the phenomenon further.

We have started from the very beginning by seeking rational answers to the naïve questions about the most general principles of the genesis, organization and functioning of simple systems. Based upon the existing knowledge the answers had to be inferred to such critical questions, as: what kind of interactions might result in interconnectedness of all constituents in the space occupied by the system? What are the simplest self-organizing systems like?

Why is nature constantly in the process of creation of the new order in the universe, where as, according to the second law of thermodynamics, the complexity of isolated systems must successively decrease in time? What fundamental interactions set all the machinery of nature to the creative work? The non-local phenomenon discovered in biological systems might just be the missing piece of the puzzle.

Today we can argue that real-time holographic mechanisms are crucial for integration and self-organization of any dynamical entity defined as an emerging/developing system. A conceptually new scenario of the genesis, adaptation, integral functioning and development of natural systems is being discussed below. The presented phenomenological model is a result of 10-year-long experimental and theoretical work in the field of interdisciplinary science of *Biohолography*.

Some critics might consider our model irreverent because of the intentional simplification of certain physical interactions. However, according to the observation of Albert Einstein "A theory is the more impressive the greater is the simplicity of its premises, the more different are the kinds of things it relates and the more extended the range of its applicability."

Certain aspects of physical reality are discussed within the unifying theoretical framework. The comparative analysis and generalization of empiric data enabled us to conjoin such seemingly non-related phenomena, as inevitable aging and decentralized memory of complex systems, embryogenesis and cancer-genesis and many other manifestations of the system functioning considered so far as independent concepts.

While discussing the most important elements of our theory, we draw parallels between the manifestation of holographic principle in non-living and living systems focusing on the holographic storage of information as the factor critical for the development and evolution of any natural system. We also emphasize the universality of the emergence phenomena in observable reality; differentiate the background order of a system phase-space and the foreground events; the subject of nature-genesis is touched as well, since the peculiarities of natural systems had to be traced back to their origin in order to reveal the factors basic for the integration of separate parts into a united entity.

Our theory already helped us to implement the holography-based approach to the system-study into medical practice and also, to predict many results of our experiments with living systems. Finally, this phenomenological model, that is more evidence-based reasoning than math-based hypothesis, contradicts neither physical or life sciences, nor elementary logic.

2. Systems, information, memory

The study of complex systems is partially hampered by the lack of generally accepted definitions. Strict definitions of basic concepts are fundamental to every scientific discipline; however, the essence of many terms, such as *system*, *information* or *complexity* remains vague and ambiguous. Below some commonly used definitions and descriptions of these terms are considered.

System. Many common definitions of a *system* suggest an organized assembly of resources and processes united and regulated by interactions to accomplish a set of specific functions [Bertalanffy, 1968]. Less strict interpretation of a *physical system* usually means that certain sets of entities are understood to serve a common objective comprising a whole, in which each constituent interacts with or is related to at least one other part of the whole. Simply put, a dynamical system of natural origin encompasses numerous interdependent units/agents organized in a non-trivial way in order to compile integral whole. Certain new

qualities of systems emerge at each level of complexity. This *emergence phenomenon* is receiving renewed attention lately.

Complex dynamical systems are systems in which many factors interact creating ceaseless changes. Complex adaptive systems (CAS), such as biological ones, represent special cases of complex systems that have the capacity to learn from their experience. All such natural entities temporarily occupy a region of available space (*phase space*) where laws of nonlinear dynamics govern and new properties arise from interactions of "non-summative" agents of systems. The definitions and descriptions of systems offered in scientific literature explain neither the principles of interaction of systems' parts nor mechanisms underlying the emergence phenomena and factors that drive the entity towards a common objective.

Some characteristics and peculiarities of complex adaptive systems are widely accepted:

- Multi-level hierarchical organization and nonlinear medium; CAS comprises subsystems which, in their turn, might be regarded as individual systems;
- Self-organization - a process in which the internal organization of a system either increases its complexity (emergence phenomenon) or restores previous order after temporal perturbations (resilience); all complex adaptive systems are able to function in a non-equilibrium state maintaining the basic homeostasis;
- Adaptive behavior - the capacity of some complex systems to react to variable environmental conditions adequately.

We believe that *the concept of system memory* has to be included in the definition of adaptive systems, since two phenomena specific for CAS – permanent emergence of new order and adaptation - are unimaginable if a system is not able to keep information on its previous states either in the form of structural alterations or as some functional relationships between its modules. No process of "learning" is possible without data storage: ambient information has to be perceived, processed and stored in a retrievable form. Hence, a kind of associative memory must be an embedded feature of adaptive systems.

Information. It was shown that an amoeboid organism *Physarum polycephalum* is able to navigate mazes and solve simple puzzles [Nakagaki et al., 2000]. Besides, this slime mold consisting of protoplasmic veins and nuclei can remember former events anticipating the unfavorable conditions in the test environment. So, information about experienced states may be memorized in a decentralized/diffuse form and no specialized "storehouses" of information are necessary for adaptive behavior of primitive living systems. It seems likely, that adaptive systems function akin to communication units incorporated into a larger system of interdependencies and feedbacks; however, the term *information* (in its physical sense) is even more ambiguous than the term *system*, hence a more precise definition is needed.

Some consider the universe as being primarily comprised of information that John Wheeler summarized in the slogan "it from bit" [Wheeler, 1990]. Obviously, information cannot act by itself and it cannot fill the space-time domain. An elementary information-related act requires at least two agents: a sender and a receiver of messages. Over the past decades, it has become common to adopt a general definition of *information* in terms of "data + meaning" [Floridi, 2005]. No doubt that a signal must affect some internal mechanisms of a receiver in order to be "understood".

Shannon's model of information-exchange includes six critical elements: a source, an encoder, a message, a channel, a decoder, and a receiver [Shannon, 1948]. Any adaptive system must be able to differentiate the signals of variable environment when choosing appropriate tactics; consequently, a complex system needs some embedded mechanisms

that would encode, evaluate, decode ambient information and also keep it for future use. For adaptive systems developing in a noisy environment, any alteration of ambient noise, including its cessation, would be informative. Hence, not the signal itself but the alteration of habitual situation matters, this difference between the former and the new state being causal determinant for the behavior of any adaptive system.

3. A serendipitous discovery

There are certain issues to be clarified while studying complex systems of natural origin:

- It is recognized that not just the constituent agents of systems matter, but their relationships. Agents of CAS are mutually interdependent, but which type of physical interactions can result in such interconnectedness?
- How can systems perceive weak, e.g., non-thermal signals (aka information)?
- Any CAS has the ability to return to ordered state and also build more complex structures: where and how do the simplest systems store information on previous states?
- Multi-level systems, e.g., biological ones, are capable of division, multiplication and production of similar entities: was this ability characteristic of the simple progenitors of systems as well?

Is it possible to answer all of these questions if we assume that not only the local (e.g., molecular) events, but also the nonlinearly organized action of a relatively uniform background medium play an active role in integration and orchestrated functioning of system parts? This assumption has been made because non-local effects discovered in living systems could not be explained on the basis of local interactions and linear physical processes. We definitely needed to find the system-unifying factor(s) first.

Compromised field of research. When the air around an object is ionized by pulsed electromagnetic field, the dielectric breakdown of the gas is affected by weak fluxes and emitted charges of exposed objects so that a complex picture of brightness-variable patterns appear around the contact surfaces. The devices used for recording of resulting images are referred to as gas-discharge visualization cameras (*GDV-technology*). Application of GDV-technology (also known as *Kirlian-photography*) to the study of biological objects is considered to be non-scientific and useless - the majority of commercially available devices provide non-reproducible, non-informative and extremely variable imagery. Entire field of the bio-electro-photography was compromised, especially when certain GDV-technique users started to interpret the stimulated radiation of biological objects, e.g., the radiation of human fingertips, as a manifestation of mystic energy or "live fields" of non-physical nature.

Our own experiments with GDV-Camera produced in Russia [Korotkov, 2002] were disappointing. We found that:

- Stimulated emission of human fingertips cannot be used for diagnostic purposes if one follows the operational protocol offered by the author of the device: results of patient examination depend upon the age of subjects, duration of object exposure, the perspiration of a body and ambient humidity, as well as many other poorly controllable factors such as subject's emotional state;
- Recordings of the same finger often differ from one another considerably, even when captured within 2-3 sec interval;

- Essential components of fingertips' emission are lost: the most intense shots leave dark gaps instead of bright streamers that are either scattered upwards or shifted out of the focal plane down to the periphery of polished surface of the screen (fig. 1, A & B).

The technological process had to be changed in order to mitigate poor reproducibility of recordings; we needed to obtain more informative image of improved quality around objects of study (fig. 1, C).

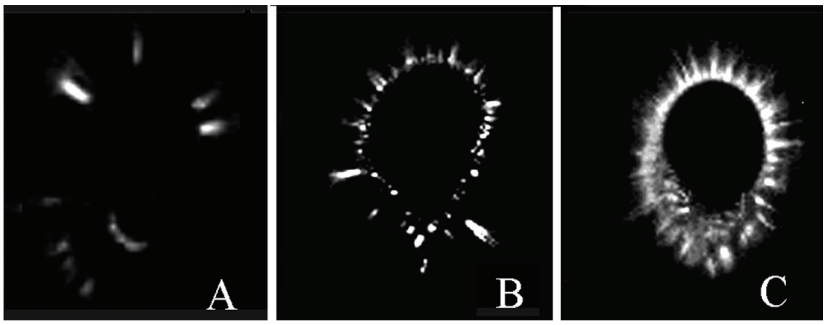
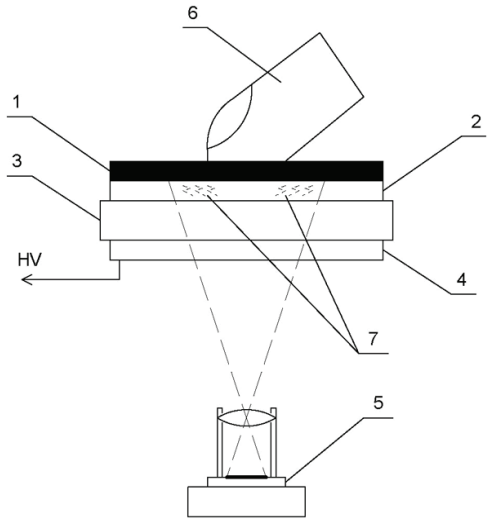


Fig. 1. Comparison of recordings obtained with Korotkov's GDV-Camera and our modality (BHT-scanner). A- a finger of a child (8y) recorded with GDV-camera; B, C – a finger of an adult (45 y) recorded using the GDV-Camera (B) and the new scanner (C).

BHT-system vs. pre-existing versions of gas-discharge devices. The improved technology and altered procedure of data acquisition enabled us to get stable, repetitive and informative imagery. New technology intended for a study of living systems' emission was named *Bioholotomography (BHT)*. The methodology is patented and used for medical diagnostics. The capturing module of the device (fig. 2) comprises: the generator of the pulsed electromagnetic field; two elastic membranes (necessary for recombination of emitted charges and the preservation of the most informative components of scattering elements in the focal plane); a transparent electrode with the rigid surface; image capture unit; fiber-optic cables to transfer the imagery to the computer, where recordings are being stored and analyzed.

Electromagnetic fields of gradually increasing frequencies (1000-5000 Hz, voltage – 17 KV) induce short-term ionization of the air around the contact area; exposed objects having oval/rounded configuration (e.g., human fingertips) produce specific patterns (coronas) that are captured and analyzed. Proprietary software controls the process of capture and enables processing of recorded data. We use various modes of recording in studies of stimulated radiation of biological objects. Continuous operational mode (30 fps) of the device is intended for experimental work that is aimed at detection of body's short-term reaction to various influences. Discrete recording of frames is used for both – experimental and clinical purposes.

The routine BHT-examination is as follows: all 10 fingers of a patient - one after another - are positioned on the insulated screen of the device and are exposed to electromagnetic fields of gradually increasing frequency; flexible membranes between a fingertip and the glass surface of the screen make it possible to retain (and recombine) scattered charges within the focal plane. Entire process of recording of all fingertips takes no longer than 3-5 minutes. Subsequent analysis of the saved data implies either "spectral" evaluation of fingertip stimulated radiation (automated) or visual interpretation of recorded imagery (performed by trained medical professionals).



1. Flexible opaque membrane;
2. Flexible transparent membrane;
3. Transparent dielectric;
4. Transparent electrode;
5. Recording camera;
6. Object (finger);
7. Gap between membranes, where the discharge of air takes place;
8. HV - high voltage supply

Fig. 2. Schema of the BHT-capturing module

As mentioned earlier, the modification of the technological process resulted in improved stability of recorded imagery thus facilitating the search of correspondence between the actual pathology and stimulated emission of certain body parts. The enhancement of data reliability yielded an unexpected result: we discovered that disordered areas of fingertip BHT-grams display some unusually shaped patterns that often repeat the characteristic contours of real anatomic structures. Moreover, we found that beside recognizable shapes of organs the fingertip *coronas* contain information on the texture and densities of the most malfunctioning organs and tissues, while properly functioning anatomic structures do not provide their signatures (fig. 3, 4).

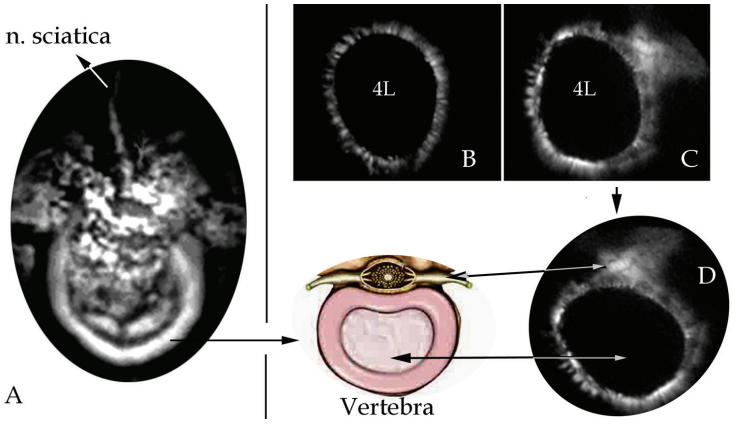


Fig. 3. Replicas of dense and soft tissues. A – a case of sciatic neuritis, a hologram; B – BHT-gram of a healthy person's finger; C – BHT-gram of the same finger on the second day of minor trauma (upper spine); D - the image is rotated to match the drawing's orientation.

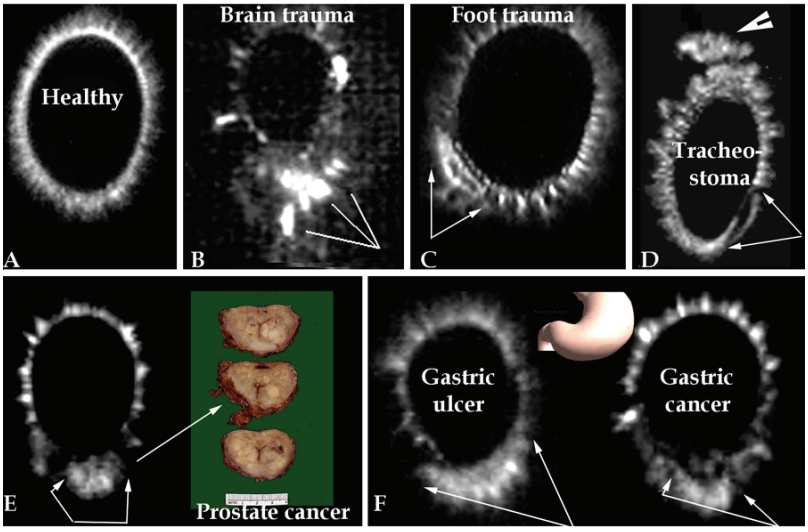


Fig. 4. Stimulated emission of fingertips: shapes and texture characteristic of the most affected organs/tissues. A – BHT-gram of a healthy person; B – recent head trauma; C – stretched ligaments of the left ankle; D – esophageal cancer, tracheostoma (arrowhead); E – prostate cancer – initial stage; F – two cases of stomach pathology: left – ulcer; right – advanced cancer.

Previously unknown peculiarity of living systems has been studied in a group of 3000 patients and also in some animals before we dared to announce our findings. Physicists gave the name to the discovered effect – *holographic diffraction*, since experimental study of this phenomenon revealed the effects that could be explained only in terms of the holographic principle.

On the holography and holographic principle. *A hologram* is encoded information about a real scene that can be decoded as a "virtual" 3-D image. Any type of waves may fix 3D information in a 2D form. Not only the amplitude, as in usual optical imaging, but also the phase of waves should be recorded in order to encode a complete spatial picture. This idea has been realized by D. Gabor, the inventor of holographic technique [Gabor, 1946] who proved that complete "structural" information about a scene/object under observation may be obtained through recording the phase-differences of interacting waves. It is known, that a periodic grating built in a nonlinear medium may generate real-time diffraction patterns [Denisyuk et al., 2000].

A rapidly developing field of the real-time holography deals with nonlinear *holographic gratings*, where the read-write processes proceed simultaneously. Holographic gratings are the diffraction gratings with a sinusoidal groove profiles. Diffracted energy and spectral performance of such holograms can be modified by changing the ratio of groove spacing and groove depth.

The holographic principle of information encoding and decoding is a natural phenomenon that should not be mixed up with the technical process of hologram-creation. The *holographic principle* states that all information lies at the boundary of a region of space [Susskind, 2008]. This principle for ordinary matter-systems and black holes has been affirmed while

studying the relationship between entropy, energy and gravitation. Some physicists argue that the holographic concept will become a part of the foundations of new physics, from which the quantum theory and relativity may be deduced as special cases [Bousso, 2002]. The processing of information takes place on various scales of physical reality. The underlying concept in holographic principle is the same as in the technical holography: in both cases a shift from one dimension to another takes place with the help of wave interactions and without significant losses of information.

All minor parts of a holographic recording contain information on the whole recorded object, although images reconstructed from the recording's different parts show entire scene from various viewing angles. We realized that interdependence of the "whole" and its parts holds in living systems as well: fingertips of a person often emit very complex patterns, which are being repeated multiple times, on various scales and from various perspectives in many areas of fingertip coronas (fig. 5a). This effect cannot be produced by molecular mechanisms. It is noteworthy, that the background noise (fig. 5b) and the fluctuations of coronas have been found to be diagnostically informative, yet for visual observation and demonstration of holographic imagery we usually increase contrast ratio of the captured stills thus reducing the noise.

Fluctuations and the *noise* in multi-scale stochastic systems are lately regarded as factors that contribute to the control, regulation and integration of intrinsic processes in dynamical systems. Moreover, the noise is recognized as one of the fundamental factors in the process of adaptation and aging phenomena, since it is closely related to stability and robustness of a system [Kirkwood, 2008]. We found that amplitudes, characteristic frequencies of the boundary emission and intensity of background noise depend upon the homeostasis of a body: in a quiet state of a healthy person and in conditions of normal exchange with environmental fields, the amplitudes of emission fluctuations are minimal (quasi-equilibrium of two nonlinear systems).

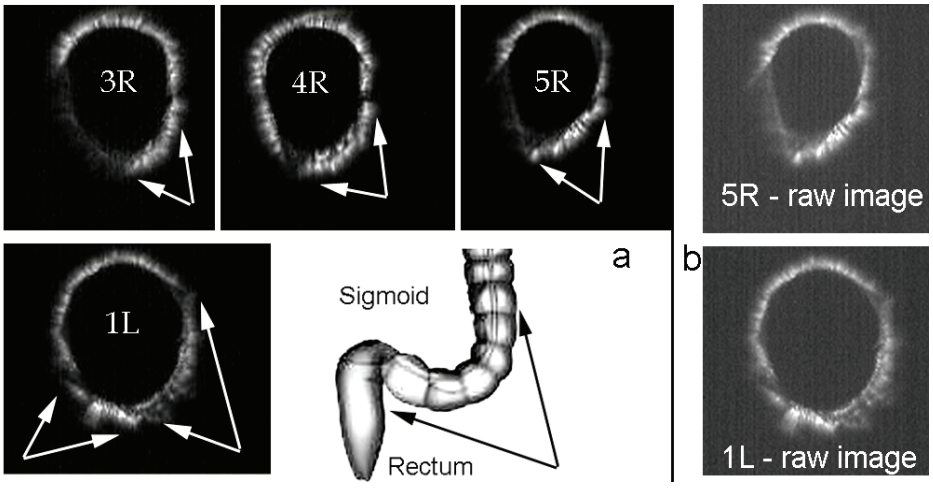


Fig. 5. Multiple replicas of large intestine: a case of colorectal cancer. a – several fingertip-coronas of the right hand (upper row) display patterns that do not differ much from one another; the corona of the left thumb shows two views of the affected area; b – the raw BHT-grams contain "noisy" components of the emission.

The real-time holography - the encoding of 3D information onto 2D surfaces (through a diffraction pattern) and its subsequent readout - seems to be the only concept explaining the appearance of signatures of malfunctioning organs/tissues on outer surfaces of a system. One can hardly imagine that in a mature organism the interference patterns can propagate reaching the surface of a body like a movie projected onto the screen. Yet, any physical phenomenon has its reasons, so we continued experiments hoping that more studies of these findings will lead us towards some rational hypotheses.

4. Holographic mechanisms in living systems - clinical study

The BHT-examination is a painless short-term procedure affecting only the periphery of a body. We believe that non-perturbing examination is extremely important in studies of system's integral functioning. It is evident, that since complex patterns of interference generated within nonlinear phase-space of the body do correspond to actual states of organs and tissues, the twists and turns of interdependent wave-forms should not be perturbed while observing the tiny patterns of interfering waves – interference patterns vanish whenever observers try to probe them. Fortunately, the optical block of the device allowed for examination of only small surfaces (e.g., human fingertips), so we avoided destructive intrusion into the main phase-space of the body affecting only its minor distal areas. In physics, the term observer effect refers to changes that the act of observation makes on the phenomenon being observed. The perturbing effects produced by observers explain the long-standing ignorance of the non-local phenomena in CAS; otherwise the physical effect of *holographic diffraction* in living systems might have been detected long ago.

As outlined above, modified gas-discharge-technology enabled us to record the patterns of malfunctioning anatomic structures by means of the harmless and easy-to-perform procedure of human fingertip examination. Since we could recognize only the signatures of misbalanced areas on fingertip BHT-grams, it was interesting to see if it is possible to perturb some parts of the body in such a way as to generate the holograms of organs and tissues deliberately. Indeed, one could try to acquire replicas of some anatomic structures either through mechanical/functional stimulation of targets or through application of perturbing signals to corresponding zones on a body surface. The most valuable for any hypothesis is its ability to predict experimental results. Our reasoning was simple: if information in living systems is really transmitted by permanently interfering and diffracting wave-fronts, then a body would emit large-scale replicas of those structures that lie in vicinity of deliberately perturbed zones. Such a selective or *Targeted BHT* is nowadays being used in the cases where certain anatomic structures have to be examined in more detail or where another problem dominates over particular areas of interest (fig. 6, 7). Hence, the Targeted BHT is based on the highly predictable reaction of the body.

Recently we developed the BHT-system which is intended for automated detection of malignant process in any part of a body. Cancer problem has been addressed because our experimental and clinical data were indicative of the dominant role of high-frequency processes in the holographic imaging. The malignant pathology is associated with the rapid proliferation of cells (processes of increased frequencies): we found that BHT-grams of the patients with malignant tumors have specific appearance. Medical doctors from Russia, Germany, France, Switzerland, New Zealand and some other countries contributed to our work sending the BHT-recordings of their patients' fingertips (as e-mail attachments) to the Center of Biohолography for expert analysis. A small-scale blinded testing in Moscow

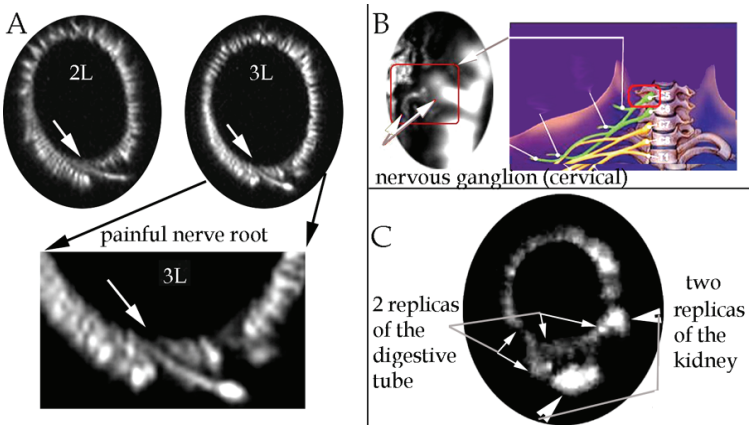


Fig. 6. Some examples of targeted BHT: A – BHT-signatures of mildly irritated nerve-roots obtained through the application of flexible magnets (10 Gauss) to lumbar spine; B – targeted BHT of cervical spine; C – simultaneous holo-imaging of two different organs in 15 min after impacts: the magnet was applied to the projection of the right kidney for 2 minutes while the person was drinking some tea (irritation of esophagus).

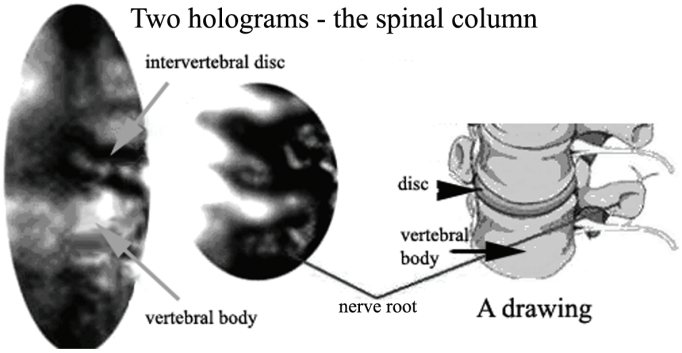


Fig. 7. Targeted imaging of vertebrae, intervertebral discs and nerve roots. BHT examination was conducted in 15 min after the 2 min-long application of EMF (3,5 mA; 0,3 Hz) to the lumbar spine of the patient with radiculopathy.

Hertzen Institute of Oncology confirmed high sensitivity of the novel *holo-imaging diagnostics* in terms of cancer detection through analysis of human fingertips' stimulated emission [Shaduri et al., 2008b]. Two cases of the early-stage cancer initially included into the control group were correctly diagnosed using BHT-technology - results of biopsy confirmed the BHT-data. Thus, it had been demonstrated that *biological holograms* are informative regarding both – the non-perturbing observation of living systems' states and the detection of malignant processes in any organ/tissue. Medical professionals have an opportunity to observe entire body and its problematic areas simultaneously and harmlessly that helps them in general diagnostics and also in evaluating the efficiency of therapy (fig.8). The arrows on fig.8 point to replicas of metastases in lymph-nodes that are prominent before chemotherapy but fairly visible after the course of therapy.

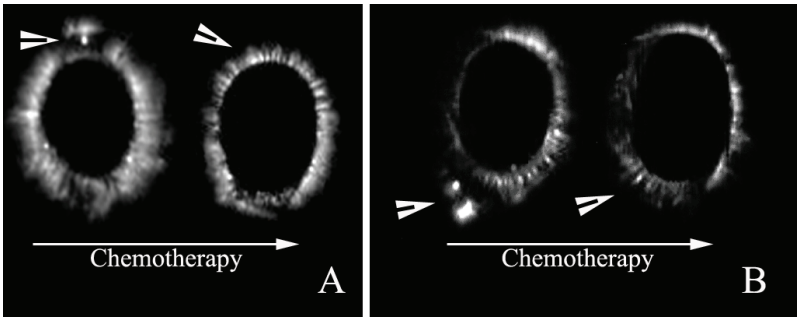


Fig. 8. The effect of chemotherapy on the fingertip BHT-grams: A – the case of advanced stage cervical cancer; B – gastric cancer.

5. Holographic Informational System (HIS) – some experimental results

It was understood that the phenomenon of non-local imaging presents a manifestation of the real-time holographic activity in CAS. We named entire set of the "wireless" mechanisms the *Holographic informational system* (*HIS*), since these integrating processes ensure the propagation of 3D imagery throughout the whole body. The role of the HIS suggests the detection and correction of any problem that affects the normal modes of vibrations within the phase-space of a consolidated system.

Certain series of experiments have been conducted in order to study the time-dependent alteration of fingertip emission in response to physiological stimulation of various functional subsystems of the body. The sensitivity of HIS has been probed while comparing the emission of non-perturbed body (first 10 seconds of the video-recording) with the changes taking place during and after various impacts (subsequent 20 seconds). Some other experiments have been aimed at relatively long-term monitoring of the HIS-activity: emission of all fingertips was examined every 10-15 minutes (for several hours) before and after consuming food or beverages, after urination, intensive breathing, etc.

Analysis of BHT-data acquired during physiological stimulation of certain organs resulted in the following conclusions:

- When a person consumes some liquids or food thus irritating (successively) the oral cavity, digestive tube and stomach, the fingertip-coronas usually display signatures of these organs in the same order. An example of the holographic imaging of digestive tube is shown on figure 9: the tube-like patterns with an opening relevant to the upper part of digestive tube are manifested on the BHT-grams within seconds after drinking some water;
- The area of radiation around contact surfaces of fingers gradually decreases in response to inflows of new substances into a system (fig. 10);
- The recognizable replicas of the stomach and duodenum usually appear on coronas in 15-20 min after the meal (fig. 11). However, this time-span depends upon the initial state of the body, as well as upon the degree of functional/structural misbalance in targeted organs;
- Replicas of certain other organs and tissues of the digestive tract (e.g., pancreas, liver and gall-bladder) can be obtained through the application of magnetic, electromagnetic or electric fields to the skin in the projections of these organs;

- Intensive breathing usually results in formation of bronchi-like patterns on coronas (fig. 12). The replicas of other parts of lungs may also be obtained: these processes of holographic imaging depend upon homeostasis in both - entire body and the respiratory system.

We found that the initial state of a system plays an important role in these experiments, since the violated homeostasis temporarily distorts and "overrides" information on less active sources of perturbations. It became clear that all participants in BHT-experiments must be in a quasi-stationary state of balance, close to a state that is usually required for measurements of the basal metabolic rate.

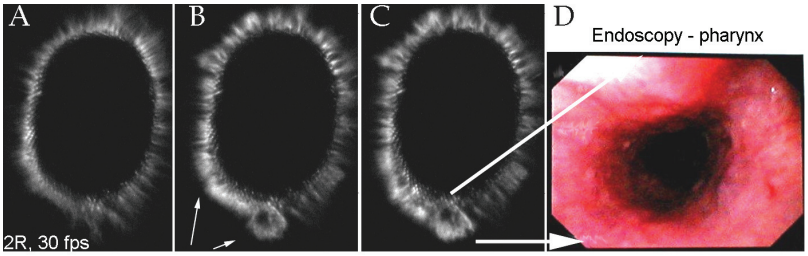


Fig. 9. Alteration of a fingertip corona in response to water consumption. Video-recording of the index finger-emission was conducted during 30 seconds (first 5 seconds - before drinking) A - background radiation. B – an immediate effect of some water intake: arrows point to the replica of esophagus (displayed upside-down); C – BHT-gram recorded just after several gulps of water; D – endoscopy of the pharynx to compare with the corresponding hologram.

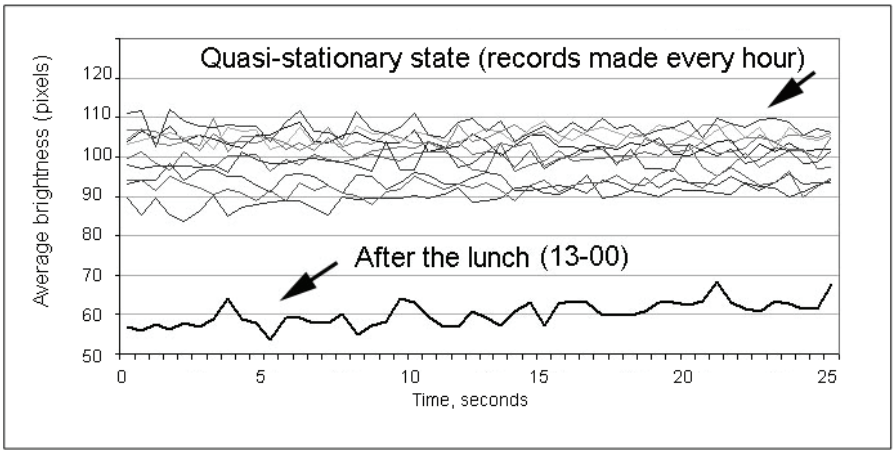


Fig. 10. Variation of a person's emission during the day. Results of a finger video-BHT conducted each hour during a day (from 10 a.m. till 21 p.m.). In 30 min after the lunch the intensity of superficial emission decreases significantly, while as all other curves occupy a relatively narrow range of average brightness.

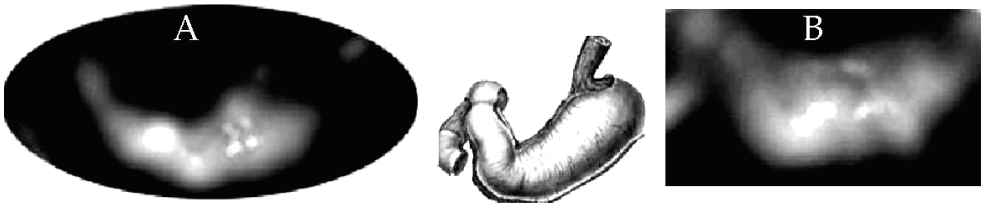


Fig. 11. Replicas of the stomach obtained using the Targeted BHT: A – a hologram of the healthy stomach is displayed on the 15-20-th min after the intake of some cereal (100 g); B – a case of gastritis: large-scale replicas of the stomach appeared earlier than in healthy subjects (on the 10-th min after the meal).

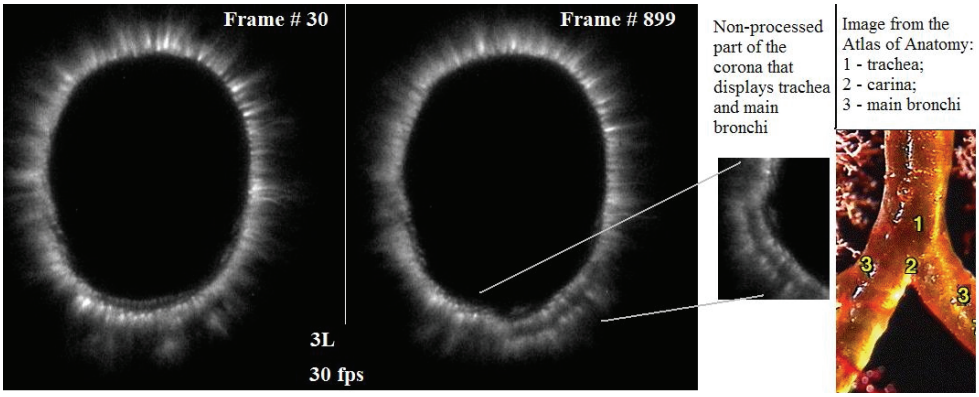


Fig. 12. Alteration of a finger corona in response to intensive breathing (started on the 11-th second of video recording).

While conducting BHT-examination of patients, we realized that internal bleeding or discharge from the body greatly alters emission of fingertips. It was interesting to check whether the emission would react to a small lesion of the skin. The figure 13 presents results of such an experiment: scarification of previously massaged ring finger of the left hand was conducted on the 11-th second of BHT video-recording. One can see bright shots (indicated by arrows) that appear almost instantly on the middle fingertip-corona of the right hand in response to the skin puncture.

Thus, we studied some non-local effects of substantial impacts upon a system, though an important question remained unanswered: can the HIS mechanisms differentiate signals of various types and strength? It is known that some biological effects are associated with ambient fields so weak that their intensities are below the "thermal threshold" [Comare, 2006]. Stating that there are no known biological mechanisms that might perceive such weak signals, mainstream science does not offer any explanation of the sensitivity of organisms towards the fields of non-thermal intensity. Is it possible to study the perception of feeble ambient fields by means of BHT-technology?

Figure 14 shows how a weak magnetic field and the inputs of chemical substances (water) into a body alter the total area of fingertips' emission. One can see that the areas of fingertip-coronas decrease slightly when video-BHT is being conducted without any impact upon a

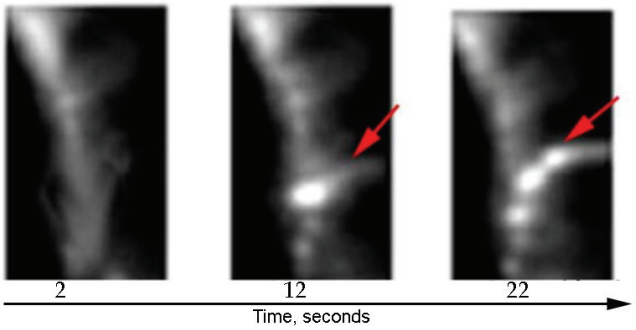


Fig. 13. HIS reaction to the skin damage. The puncture of the left hand finger alters the emission of the right-hand middle finger. See the text for more details.

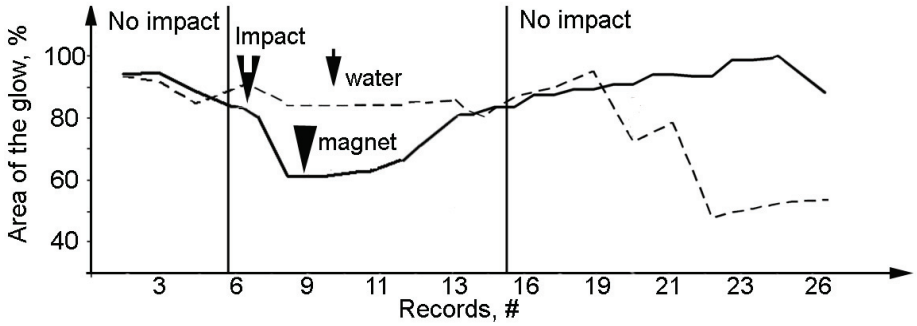


Fig. 14. Alteration of BHT-grams during the intake of some water (100 ml) and application of magnetic field (10 Gauss) to the left arm. Emission of the right-hand index fingertip is recorded with 3 sec-long intervals. Area of the emission is calculated as the % of the first 5 stills (no iinfluence).

body; however, fluctuations are suppressed as soon as a person starts to consume some water (dotted line in the middle part of the chart). In several seconds after the first gulps of water the areas of coronas begin to shrink in a step-like manner while disturbances (information) propagate towards the lower levels of the system-hierarchy. Hence, not only the mean intensity of coronas decreases when new molecular substances enter a system, as shown on figure 10, but also their areas become smaller.

Application of a weak magnet (*Bioflex*, 10 Gauss, diameter 20 cm) to the arm of the opposite hand affects the emission in a distinct manner: the area of fingertip emission decreases much sooner than in the cases of substantial inputs, this effect being compensated in several seconds. The area of emission returns to initial values even earlier than the magnet is taken off (and earlier than the effect of water consumption becomes prominent). It is evident that there is a major difference between reactions of the HIS to magnetic fields and to substantial inputs. Dynamics of both processes is time-dependent and is indicative of a state which is far from equilibrium.

Several series of experiments have been conducted using coherent and non-coherent light, electric, electromagnetic and acoustic signals directed to various areas of the body surface.

Below we present results of the experiment, where beams of red laser (50 Hz) have been flashed to the palm of the right hand from the distance 0, 5 m (fig. 15). Immediate decrease of the left hand finger emission and subsequent modulation of fluctuations are indicative of the strong non-local reaction of HIS mechanisms towards the short-term superficial impacts of light. These effects might be much more spectacular, though our observations refer to time averages in non-ideal experimental conditions: for experimental work we use the certified device developed for clinical purposes.

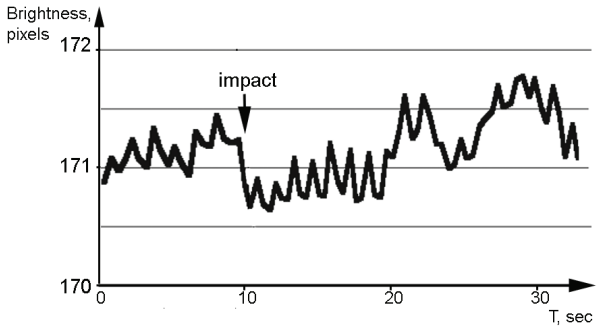


Fig. 15. The effect of laser beams on a finger emission (see the text).

While conducting the routine clinical work we realized that problems associated with significant misbalance in the region of high frequency fluctuations (as e.g., in the cases of viral infections or toxic states) are manifested as major gradients of emission intensity and the fragmentation of corona patterns. One can see (fig.16) that in the cases of cellular misbalance certain areas of BHT-grams present the chains of bright "grains"/speckles.

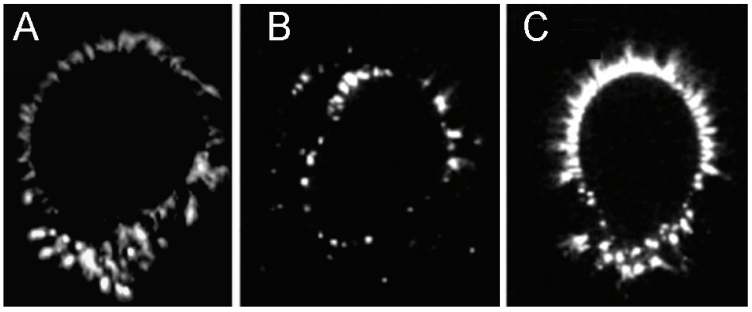


Fig. 16. Fractal appearance of coronas: 3 cases of cellular/metabolic disorder. A – malignant transformation of blood cells – the recording made several months prior to clinical manifestation of acute leucosis; B – a case of AIDS; C – reactivation of *Herpes Zoster*.

The fragmentation of fingertip coronas is nowadays used as an indication of increased activity on the cellular/intra-cellular levels of system-hierarchy. We found that fractal areas and bright speckles on BHT-grams appear in response to various weak impacts, e.g., alteration of ambient fields. The idea that geomagnetic and geo-electric fields are fundamental parts of information flow in the biosphere is not new [Cole & Graf, 1974]. Indeed, alteration of the earth fields results in a wide range of patterning defects and

abnormal reactions in plants and animals [Conley, 1970; Braun, 1973]. Our own experiments have demonstrated that even partial isolation of a body from natural geomagnetic fields causes the distortion and gradual fragmentation of BHT-grams. These results have been interpreted as the transmission of ambient information down to the cellular level of the system-hierarchy. BHT-examination has been conducted in the normal ambience and every 10-15 min after entering the room where magnetic fields comprised only 30 - 45% of natural values. Major gradients of brightness and abnormal shaping of BHT-grams, as well as their increasing fragmentation during first 30 - 45 min of experiments have been determined in all replicate studies (25 participants). We found that the coronas always restore normal appearance after the first stage of chaos-augmentation, albeit the second stage of the relaxation requires longer time (80 – 110 min) than the first one.

It has been interesting to check the reaction of HIS to the changes of boundary temperature as well. Figure 17 presents results of the experiment, where the emission of the left hand fingertip-corona has been recorded while another hand was affected by heated water. This experiment yielded the counter-intuitive results, since water of comfortable temperature caused the large-scale periodic fluctuations of emission much earlier than the cold or very hot water. It is likely that afferent nervous impulses, which represent linearly propagating signals, temporarily suppress the activity of nonlinear mechanisms of the HIS disrupting the phase-conjugated vibrations.

One can infer from the presented data that the wireless mechanisms of signal transfer react more efficiently and rapidly to weak signals, than to more powerful ones: the HIS uses only nonlinear mechanisms of week interactions, while the nervous system operates through the fixed and linear guides of impulses, so intensive firing of linear signals seems to arrest the fluctuations for a while.

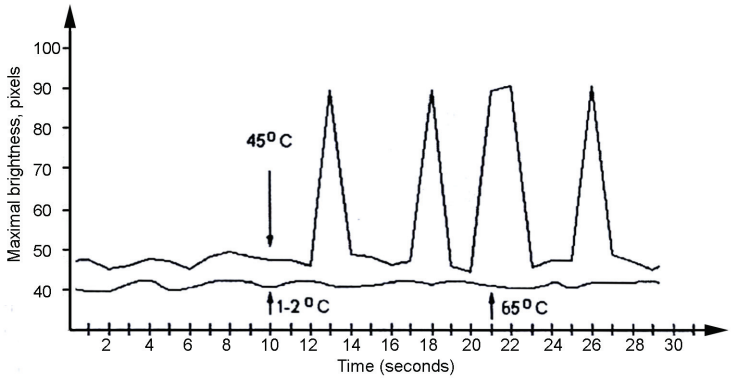


Fig. 17. The alteration of a body emission in response to distant thermal signals. See the text for explanations.

6. Dynamical systems in nature: our model and its discussion

The need to unveil the physical basis of our accidental finding pushed us towards various trends of modern sciences. Numerous inconsistencies, ambiguities and over-complicated explanations slowed down our research. We have been faced with the fact that open questions dominate over comprehensible answers when it comes to the most fundamental

physical issues. As an example of poor reliability of some widely accepted physical constants and ideas, one might name the recent data of Randolf Pohl of the Max Planck Institute in Germany, who reported that a primary building block of the visible universe, the proton, is smaller than previously thought [Pohl et al., 2010]. The value of gravitational force is also being challenged by two different methods as announced in the journal Nature [2010], etc. Hence, we had to seek out the "pearls" hidden in the oceans of excessive information in order to find, understand and reevaluate (with the new discovery in mind) all the facts and hypotheses that could integrate various notions and observations into a united, viable and verifiable system of knowledge. Drawing on our experimental data and on the knowledge accumulated in various trends of science we came to a conclusion that all dynamical systems of natural origin pass through similar stages while emerging and developing. Generic scenario of system's "life-cycling" is discussed below. The offered model is based on the assumption that the space occupied by an integral entity (more exactly, the nonlinear phase-space of all dynamic systems, the universe being one of them) plays an active role in all emergence phenomena.

The structure and functional role of the space is being actively debated lately [Verlinde, 2010; Makela, 2010]. The hypothesis on background independence where geometry and gravity are regarded as emergent concepts is being studied in quantum theories of gravity [Marcopoulou, 2007]. It is thought that the nonlinearity of space may explain many enigmas in modern physics. According to our model, the background phase-space is an active and self-organizing domain of physical reality – it acts as a "self-healing network" that controls and regulates all local perturbations by imposing physical forces and constraints on passive matter (on particles and their composites). The same phase-space acts also as a 3D medium for real-time holographic processes within the entire system.

The principle of holography with the new discovery in mind. The holographic principle, which is rooted in the "holomovement" hypothesis of David Bohm [1983], has been reevaluated and further developed by Gerard 't Hooft [1993] and Leonard Susskind [2008]; physicists ascertain that total information about the (D+1)-dimensional space may be found on its boundary/surface (the D-dimensional space). Jacob Bekenstein was the first to recognize that a universal relation exists between geometry and information [Beckenstein, 1972]. Later Bousso [2002] offered the general concept of covariant entropy bound showing its close relation to the holographic principle.

Essential arguments contributing to the support of the holographic hypothesis are:

- The theorem of Bell [1964], who showed that entanglement in a quantum system (EPR phenomenon) may persist over long distances;
- Data of Aspect [1982], Gisin [1999] and some other physicists who demonstrated the existence of quantum non-locality experimentally; many scientists plan to use entangled states for quantum cryptography [Ekert, 1991], entanglement-assisted communication [Bennet et al., 1997], fast quantum computations [Shor, 1997], etc.;
- The entropy bound: an upper limit on the entropy or information that can be contained within a given finite region of space-time [Beckenstein, 1981; Busso, 2002];
- The "Maldacena duality": a conformal field theory defined on the boundary of a maximally symmetric manifold with constant negative curvature [Maldacena, 1998];
- Fractal distribution of galaxies: Mureika (2007) did show that such a distribution is a signature of holography (*fractal holography*);
- Wide-angle power correlation suppression in the cosmic microwave radiation that has been found to be compatible with the idea of holography (Huang et al. 2006);

- The holonomic model of Karl Pribram [Pribram, 1991]. The author argues that memories may be enfolded within every region of the brain rather than being localized;
- The *holo-diffraction* phenomenon: the first experimental evidence of the non-local effects manifested in living systems [Shaduri, 2005].

We have demonstrated that information about any significant disorder in a body may be detected through the analysis of minor „terminals" of a system such as human fingertips. The *non-local imaging* of internal anatomic structures may be explained only if information is transmitted throughout entire body using the mechanisms of wave interactions: all agents and functional units of complex systems have to be permanently covered by a high-bandwidth holographic network.

In modern communication systems the most promising developments are envisioned in the area of the self-healing, self-configuring and scalable wireless networks that are capable to control and repair many deviations from normal modes of the network functioning. Highly resilient systems that use interconnected meshes of self-healing rings are being studied worldwide. Hence, humans re-invented a technology that actually presents the oldest and the most general mechanism of permanent communication between all constituents of natural systems.

We believe that the holographic grating of a system preserves the memories on the former and present states during whole lifecycle of an individual system. Thanks to the holographic principle of system functioning it is possible to readout the life story of a CAS, since its experience is being remembered in the form of updatable interference patterns. For instance, the BHT-grams of humans sometimes display holograms of former structural lesions such as bones fractured in early childhood. We suggest that the orchestrated phase-space acts as a master reference for local disturbances opposing all new formations and restoring the harmonic order whenever possible. Such a permanent feedback between the *reference* domain and various *object-associated events* had to appear in early universe.

Some arguments in favor of our model.

Theoretical physics and cosmology. NASA's Wilkinson Microwave Anisotropy Probe (WMAP) has found the sound waves in the early universe, where there is a primary "note" and a series of harmonics, or overtones [NASA, 2008]. So, some resonances of compression waves in the fabric of the space-time were already present in the first trillionth of a second after the "birth" of our universe. The most important (for our model) is the discovery of primordial fluctuations as it supports the hypothesis on the organizing role of system phase-space vibrations. Recently Verlinde proposed a theory which is actively discussed in scientific literature: the author derived Newton's Second Law, Newtonian gravity and general relativity, all by considering various holographic "screens" in space and suggesting the gravity to be an emergent property of the deeper underlying structure of the universe [Verlinde, 2010].

Waves as particles. A nested hierarchy of vibrating "lattices" of the non-linear medium may be viewed as a system of standing waves. It is acknowledged that *standing waves* arise when counter-propagating flows interact within a bounded area. Each and every possible standing wave has its specific energy. The localized standing waves that retain their forms over a long period of time are named *solitary waves, or solitons.* Solitary waves are robust and exhibit particle-like behavior; they can pass through each other or bounce back from each other emerging asymptotically unaltered from the collision [R.Rajaraman, 1982]. Hence, solitons behave as *wavicles* (both - waves and particles). Cohen et al. described the so called *holographic solitons* that consist of two mutually coherent field components. According to the

authors, the two components of the holographic solitons interfere, induce a periodic change in the refractive index, and simultaneously Bragg-diffract from the grating [Cohen et al., 2002]. The wavicles (solitons) and composite particles are extremely important, since a holographic grating may store information and continue to build new orders with their help.

Why did not everything just stay in a wave-form within the early universe? Could the universe develop without particle-like structures? The answer is definitely negative: the "solid agents" are crucial for the evolution and growth/inflation of any natural system – they help to preserve structural information on former states of a macro-system. We suggest that the smallest solitons act as the building blocks for more complex composites and also as the scattering elements distributed across the space. This assumption helps to approach many theoretical problems from a new viewing angle: the emergence of mass (constellations of fundamental particles), the appearance of the diffraction grating in the universal medium, the arrow of time, the generic phenomenon of continuous evolution and some other "enigmas" can be reconsidered in terms of real-time holographic processes.

We also assumed that the lowermost boundary of the universe might be formed through the generation of higher-order harmonics - the only option to reduce free energy and create an order out of the ultimate chaos in a bounded region of space (see below – "Genesis of complex systems"). Just imagine that we took out all the mass possessing matter from the universal system. The vibrating particle-free medium left after this "thought-experiment" would never evolve/change: no displaced mass - no fringes within the phase-space grating; no fringes/memories - no new order and no emergence phenomena; no changes of a system-state - no time arrow, etc.

Emergence of the holographic grating. Our scenario suggests that the lowest threshold of the observable reality is distributed within the space as a kind of „pinch harmonics". This idea is not original: Misner, Thorne and Wheeler [1973] speculated that gravitational collapse at the Planck scale of distance is taking place everywhere and all the time in the geometry of the space-time; the space-time being made of Planck-size black holes has been assumed by some other physicists as well [Crane, 1993].

Indeed, the fragmentation/quantization of the dynamic background of our universe might result in emergence of the ground-state order where the Planck-scale solitons stabilize and solidify the *lowermost boundary of entire space*. Such a uniform medium with finite-value frequencies and permanent interaction between the incident and back-reflected wave-fronts can exhibit the properties of a holographic grating. Any local instability or displacement of solid inclusions from their quasi-stationary positions would perturb the resonances of the "humming" medium: the effects produced by perturbed lattices of the phase-space would then be manifested as physical forces imposed on passive particles. It should be emphasized that the displacement of particles towards the lower-energy states (relative immobilization) is similar to recording of information in the form of new structural combination and new interference patterns in the background ranked by energy states; this process is exactly the same as used in data holographic storage.

The self-healing network of a system-medium can deal with chaotic processes using the mechanisms similar to those known in non-linear optics (e.g., the generation of complex conjugate waves propagating backwards through a distorting medium). We have postulated that the universal medium is organized, fine-tuned and ranked by energy states, so any new boundary and any linear displacement in such a medium would be opposed by the background activity. Just imagine a small partition of such a nonlinear phase-space enclosed

within a spherical volume. The isolated phase-space would be forced to reorganize its medium according to new conditions: excessive energy "trapped" within it would generate new sub-harmonics (sub-lattices) and, correspondingly, new band-gaps, so that the spectrum of a bounded area would be squeezed becoming narrower and also more coherent compared to the host system's spectrum. Obviously, the boundary vibrations of a distinct phase-space have to match a normal mode of the host-system – otherwise it would be impossible to preserve the structural stability.

Physical forces and phase-space. Any excitement in the range of high frequencies greatly affects all vibrations of a living system (fig. 13; fig. 16): major gradients of the emission brightness and density are characteristic of cellular/intracellular disorders. We have also demonstrated that superficial signals propagate towards the lower levels of the hierarchy, where high frequency vibrations are able to displace some small molecules and atoms from their meta-stable positions. Besides, we found that a system relaxes only after the high frequency agents become involved in *signal-processing*. One can infer from these data that high frequency vibrations play crucial role in the self-organization of the "whole". Both - upstream and downstream propagation of perturbations through all the lattices (energy states) of medium-grating are necessary for normal functioning of multi-level systems; however, micro-scale events are associated with strong coupling of high-frequency processes (minimal degrees of freedom), whereas events on upper levels of system-hierarchy are driven by medium lattices of lower frequencies (weak forces, higher degrees of freedom).

Weak interactions are considered a separate type of the four fundamental interactions. Yet our scenario that is consonant with some of the new theories suggests that all types of physical interactions present the manifestation of the *geometry-dependent* interplay between the active phase-space and the complementary but passively driven agents of a system. By turning the logic around and postulating the existence of local equilibrium conditions, Jacobson [1995] came to a conclusion that the space-time might be comprised of waves propagating as adiabatic compression waves (analogous to sound in a liquid). As mentioned earlier, the Newton's universal law of gravitation arises naturally and unavoidably in a theory in which space is emergent through a holographic scenario [Verlinde, 2010; Makela, 2010]. These and similar theories suggest that gravitation is emergent property of more general description of observable reality.

A subject that to our knowledge was never discussed in scientific literature concerns the concept of nature. The universe and nature are not synonyms: we believe that nonlinear phase-space with its self-healing grating must be distinguished from the "foreground" processes (local/linear events); nature is an emergent realm encompassing all the diverse products yielded by permanent interaction of the reference and "object" domains. Hence, according to our model, nature presents a by-product of perturbations within the ordered medium; its masterpieces would never emerge without the feedback between the holographic grating of the self-organizing phase-space and particle-associated events.

7. Emergence and evolution of complex systems

Genesis of a new system through the "lasing"of an old system's part. An isolated partition of a relatively complex system behaves in the same way as bounded partitions of other phase-spaces – it utilizes excessive kinetic energy by creating new boundaries and resonances through successive fractalization (an example is the binary cleavage of a

fertilized ovum in mammals). The lesser the diameter of a sphere, the greater surface area it has relative to its volume, so the generation of smaller compartments enlarges the total surface area reducing the entropy of the whole (important in terms of thermodynamics and information theory).

The fragmentation of an isolated phase-space of a larger system and the narrowing of spectra may result in appearance of a new entity with its individual holographic grating: high frequency vibrations within a bounded fraction of medium can destroy all the complex substances of parental system driving the area towards the ground-state order – normal modes of the medium vibrations; as soon as all the constituents within the isolated area reach a state of ultimate excitement, the "random lasing" can take place [Polson & Vardeny, 2004]. The term *random lasing* defines amplification of vibration intensity in a randomly scattering medium that eventually leads to a coherent or quasi-coherent state. A team of researchers [Song; Xiao et al., 2010] have shown that dense tissues of bones present an ideal biological material for random lasing. A perfect order and coherent spectra may be obtained out of ultimate disorder if energy is trapped within (or pumped into) a minute area of the biological tissue. Hence, in order to become an individual system, a minuscule partition of the nonlinear phase-space of a host-system has to be isolated and transformed into an ordered medium through the effect of lasing. This process is opposite to emergence and development. The "antidevelopment" through lasing effects is necessary - any new system should empty its holographic "diary" before it starts to record personal history.

We argue that the lasing is crucial for emergence of any new system within the phase-space of a complex host-system. When the low-frequency vibrations of the parental phase-space are cut out by a boundary (closed for particles), the generation of new lattices-harmonics and corresponding band-gaps squeezes the spectrum of bounded area creating conditions favorable for emergence of a new holographic grating. The lasing of medium takes place during the transformation of normal biological tissues into a *neoplastic system* (*malignantly growing system*) as well.

Once created, an individual system may divide, shrink or degrade but it would never conquer additional space without regular exchange of particles with the host-system. The growth takes place in those multi-level systems that are unable to match ambient vibrations without enlargement of their surface area: they continue to grow up to the moment when near-zero effective force is achieved on the boundary between the outer and internal phase-spaces. As the number of oscillating particles increases within a new entity, its free energy decreases gradually being spent on formation of new resonances and new molecular structures. The time a maturing system takes to transfer energy/excitation from the upper level lattices of low frequencies to the bottom lattices of high-frequency grows in parallel with the number of resonances and mass. We already mentioned that minimum 30 min is required for weak ambient signals to reach the cellular level of the system-hierarchy in adults. We also found that BHT-grams of children and pregnant women, as well as fingertip coronas of the patients with cancer, are more disordered and variable than BHT-grams of other groups of examinees.

Evolution of systems. The evolution of natural systems reached a point of bifurcation at some moment: a multi-level hierarchy could be built only through the exchange of particles with the environment. The separate "units" of simple structure such as elementary particles could interact with one another but they were unable to generate new resonances within their phase-spaces. Certain interactions between elementary particles and their composites resulted in the appearance of complex flexible molecules that matched some profiles of

medium lattices. A tuned entity of flexible molecules and medium grating became able to withstand intrusion of small portions of chemical substances into the bounded area – new particles started to play the stabilizing role and also contributed to the holographic storage of data.

Some flexible chains of molecules and atoms that matched the sinusoidal profiles and grooves of medium (vibrating in synchrony with corresponding lattices) survived the degradation of the "whole": they started to act independently of their host-systems either through transition into other systems or by building their own phase-spaces in appropriate medium. In this way certain macromolecular structures (e.g., chains of nucleotides) made it possible to exchange and/or spread information on fundamental modes of original phase-spaces. It is known that the main blueprint of genetic instructions – DNA – is structurally flexible polymer: it exists in many possible conformations; the carbonyl centers of its esters may give rise to 120° angles and DNA also contains grooves/spaces between the twin helical strands of complementary base pairs. Besides, DNA may carry out low-frequency collective motion [Urabe et al., 1983]. Hence, these molecules are able to provide information on the profiles of those vibrating lattices of medium that a new system can use while organizing its individual phase-space.

The encoding of system-memory about individual phase-space-modes through its transformation into the genetic (solid) form has started a new era in the history of CAS. The further cooperation of distributed mechanisms of HIS with some linear conductors of signal transmission (nerves) greatly improved the control and regulation of internal processes, and also facilitated exchange of information with system ambience. The unity of linear and nonlinear information-managing mechanisms created a new platform for accelerated development.

Associative memory of complex systems. We believe that genetic memory is used as a code for new phase-space grating; however, it cannot contribute to the adaptive behavior of systems: any CAS should create its own archive/database in order to behave adaptively. The "fringes" of memory (immobilized particles) in the holographic grating of a system medium can appear in response to any significant perturbation of a system state. The medium resonances and scattering/diffracting particles comprise a kind of system-archive where information about personal experience of a system is retained in a readily updatable form. An update of saved information might be compared to search-engines that browse within the memory of a computer using keywords as guides. 3D-holograms are sensitive to selected wavelengths/frequencies, so perceived packets of signals may find targets due to this sensitivity. Interaction of detuned or linear wave-fronts with the holographic grating and simultaneous diffraction makes it possible to read out the phase-conjugated waves.

The higher the coincidence between previously experienced and new states of the body and the higher the dominant frequency of new signals, the better and faster is the process of "data-recollection" (aka the reconstruction of saved holograms). With this conjecture in mind one can explain certain poorly understood peculiarities of adaptive systems. For instance, it is occasionally claimed that a complex adaptive system has individual strategy based on particular goals and predictions of the future. A goal to which all agents of CAS thrive is nothing but the desired equilibrium (albeit, animals and humans may act voluntarily). Some scholars refer to goal-directed units as basic agents of CAS. This opinion contradicts a widely accepted point of view that behavior of a system is not predicted by the behavior of its parts. Obviously, neither a system nor its parts would have any particular "goals" while the archive of personal memory is empty; however, any adaptive system has a

kind of a *free will* in the sense that even the most primitive organisms are able to choose particular scenario(s) drawing on their former experience: since a sequence of numerous successive holographic recordings can be updated almost instantly, an adaptive system may have some time to change its behavior faster than undesirable events cross the brink of unavoidability. Such a capability of CAS to foresee the final results of particular states (*premonition*) has nothing to do with paranormal phenomena.

According to the described model, any adaptive system conducts the real-time holography permanently, during the whole life-cycle. CAS may differentiate signals operating across multiple wave-bands; however, not all the signals and states of a system are being "remembered": the *self-healing network* may restore all resonances after some non-essential or already experienced (weak) impacts; so only relatively powerful, significant and previously non-experienced states are being recorded and archived. The described type of the non-local memory cannot be fully erased until the energy is sufficient to avoid irreversible collapse of near-boundary lattices; each new input into the holographic database of a system brings closer the moment, when the HIS becomes less sensitive to new inputs due to the growing number of immobilized particles. Hence, data storage yields not only beneficial (e.g., experience) but also unwanted effects such as aging and inevitable collapse/death. The life-cycle of a system ends naturally when the outer lattice of the HIS ceases its tuned vibrations. This "gravitational" collapse takes place if free energy falls below the threshold due to critical number of fringes (manifested also as wrinkles on a body surface) and the deterioration of exchange via the shrunken boundary. Thus, the life-cycling and the decentralized memory of complex systems are mutually interdependent concepts.

Cancer: undesired system-genesis. Oncogenesis suggests the process of malignant transformation when a subset of the cells, called cancer stem cells, rapidly replicates themselves [Cho & Clarke, 2008]. We argue that cancer is a new system that emerges within a host-body as a distinct phase-space, so there should be a kind of similarity between the first stages of oncogenesis, embryogenesis and nature-genesis. Consider a region in a body that contains or accumulates excessive energy but has a restricted access to ambient substances (poorly permeable boundary). Energy may be increased due to several factors: viral activity; the presence of certain molecules that either resonate with very high frequencies of ambient fields or affect polypeptides/nucleic acids in such a way that ground-state vibrations become altered; repeated mechanical irritation may also create conditions beneficial for cancer. Definitely, the boundary of an emerging system must be closed for substances at the first stage of oncogenesis. Such a situation might take place within the tissues of increased density or low excitability (e.g., fibroids, scars, etc.). In the case the energy is high enough to cause the random lasing of entire area, a new individual system – a neoplasm - might emerge. The smaller the area occupied by new phase-space and the narrower the range of its characteristic frequencies more chances appear for a neoplasm to start aggressive growth.

Not every area that accumulates energy would produce malignant tumors in biological systems. Some possible scenarios of the host phase-space-partitions are given below:

1. The exchange of substances becomes possible earlier than the cellular constellation reaches coherent state through the lasing: the cells grow non-aggressively yielding a benign tumor;
2. The gain medium accomplishes the lasing but boundaries remain closed for particles: entire area „melts down" yielding a cyst(s);

3. Exchange of substances with the surroundings begins just after the coherent state is achieved through the lasing: the neoplasm builds its individual holographic grating and starts malignant growth.

BHT-data substantiate our model of cancer-emergence. The lasing of medium and quasi-coherent radiation may be suspected when BHT-grams contain extremely bright lines/inclusions within dark areas. The narrow focused „beams" and bright "bulbs" that often produce diffuse or branching *illumination* are characteristic for aggressively developing neoplasm in early stage cancer (fig. 18, A, B). The metastatic transformation of cells also provides bright inclusions within dark and poorly structured replicas of degrading tissues (fig. 18, C).

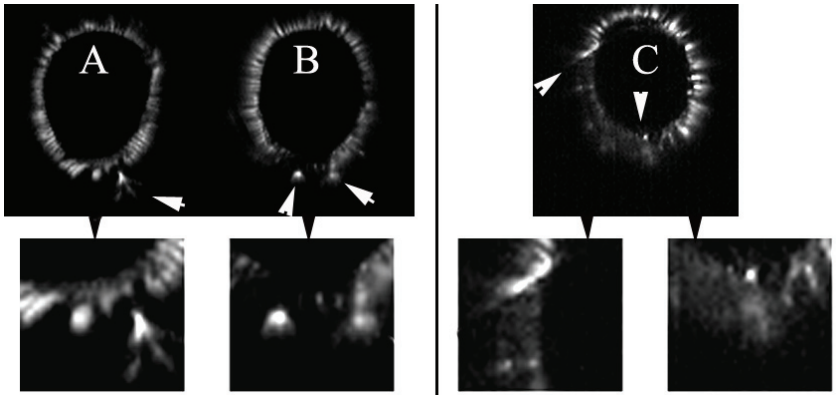


Fig. 18. BHT-signatures of lasing in biological tissues (white arrows). A, B – early stage of breast cancer (two recordings made within 3 months); C – metastases in gastric cancer.

The proposed model of the cancer-genesis is indirectly substantiated by a team of researchers, who have discovered that the malignant human tissue can be distinguished from nonmalignant samples through the assessment of their emission spectra [Polson & Vardeny, 2004]. A more detailed version of our theory of cancer-genesis will be published elsewhere. We hope that the generic model of cancer would help medical professionals in their struggle against this undesired manifestation of the universal phenomenon of system-emergence.

8. Concluding remarks

We have outlined our phenomenological model of system-emergence and development. The discovery of non-local effects in living systems has triggered the search for physical mechanisms that might make it possible to integrate the separate constituents of a system and their relationships into an orchestrated entity. We found that real-time holographic processes within nonlinear phase-spaces of natural systems could explain our experimental evidence. The principle of holography was already known to be manifested on other scales of observable reality that facilitated our theoretical research. Obviously, the holographic replicas of anatomic structures could not be formed in a mature organism: we recognized that emergence of the holography-based mechanisms is closely related to the very first

stages of natural system-formation. Hence, it became necessary to study the generic process of system-genesis, more precisely, the origin of natural systems in our universe.

To expand the understanding of the most general mechanisms of system-genesis and explain the phenomenon of emergence we addressed certain problems in theoretical physics. While tracing back the origin of self-organizing systems in our universe only the most reliable hypotheses, observations and experimental data were considered. Our quest encountered the problems of nature-genesis and wave-interactions at some point, and, eventually, we came to the conclusion that the only way to overcome these stumbling-blocks is to turn the common logic around postulating that the universal space presents the nonlinear reference-medium for local (particle-associated) disturbances and linear processes. The hidden activity of the phase-space - the rear side of a "coin" - turned out to be much more important than visible and measurable manifestations of the physical reality. Indeed, "the obscure connection is more powerful than the apparent one" as has been put by Heraclitus. Hence, the active role of periodically organized nonlinear space shared by all constituents of a system became a cornerstone in our model - it facilitated further search for reasonable explanations of system-peculiarities. Recently we found that some physicists regard the microstructure of the universal phase-space as an answer to the deepest questions in science [Smolin, 2010], namely why space, time and gravity exist at all. So our reasoning is not just a *vox clamantis in deserto*.

To summarize, we argue that the most fundamental aspect of physical reality is the interaction between the nonlinear phase-space of the universe and the local "foreground" events that allows the permanent creation of new order and emergence of complex structures through the holographic scenario. We believe that this interaction drives the evolution of the universal macro-system producing the "daughter-units" (new systems) of various types, aromas and flavors during its finite lifecycle.

9. Acknowledgements

We would like to express our sincere gratefulness to Professor Jacob Beckenstein from the Hebrew University of Jerusalem for inspiring comments on our discovery; to our mentor and the author of the term "holo-diffraction", late Professor Vladimer Chavchanidze – the former head of the first Institute of Cybernetics (Georgia); to the staff of the Center of Biohолography (Tbilisi, Georgia) and to the Luxemburg-based company "Advanced Bioresearch & Technology" for efficient help in our R&D; also, many thanks to all the medical professionals and scientists from various countries who contributed to our work.

10. References

Aspect, A. Grangier, Ph. & Roger, G. (1982). Experimental Realization of Einstein-Podolsky-Rosen-Bohm Gedankenexperiment: A New Violation of Bell's Inequalities, *Phys. Rev. Lett*, Vol. 49, pp. 91-94

Bekenstein, J. (1972). Black holes and the second law, *Lettere al Nuovo Cimento*, Vol. 4, pp. 737-740

Bekenstein, J. (1981). Universal upper bound on the entropy-to-energy ratio for bounded systems, *Physical Review D*, Vol. 23, No. 2, (January 15, 1981), pp. 287-298

Bell, J. (1964). On the Einstein-Podolsky-Rosen Paradox, *Physics*, Vol. 1, pp. 195-200

Bennett, C.; Fuchs, C. & Smolin, J. (1997). Entanglement-Enhanced Classical Communication on a Noisy Quantum Channel, In: *Quantum Communication, Computing, and Measurement*, Hirota, O.; Holevo A. & Caves C. (Ed.), pp. 79-88, Plenum, ISBN 0-306-45685-0, New York

Bertalanffy, L. (1968). *General System theory: Foundations, Development, Applications*, George Braziller(Ed.), revised edition 1976, ISBN 0-8076-0453-4, New York

Bohm, D. (1983). *Wholeness and the Implicate Order*, Routledge, ISBN 0-7100-0971-2, London

Bousso, R. (2002). The holographic principle, *Reviews of Modern Physics*, Vol. 74, pp. 825-870

Brown, F. & Chow, C. (1973). Interorganismic and environmental influences through extremely weak electromagnetic fields, *Biological Bulletin*, Vol. 144(3), pp. 437–461

Cho, R. & Clarke, M. (2008). Recent advances in cancer stem cells. *Curr. Opin. Genet. Dev*, Vol. 18, No 1, pp. 48–53

Cohen, O.; Uzdin, R.; Carmon, T. & Odulov, S. (2002). Holographic solitons. *Optics Letters,* Vol. 27, No 22, pp. 2031-2033

Cole, F. & Graf, E. (1974). Precambrian ELF and abiogenesis. In: *ELF and VLF electromagnetic field Effects*, Persinger, M. (Ed.), pp 275-310, Plenum Press, New York;

COMARE (Committee on Medical Aspects of Radiation in the Environment), (2006). Eleventh report, *Health Protection Agency for the Committee on Medical Aspects of Radiation in the Environment*, ISBN 0-85951-578-8

Conley, C. (1970). *A review of the biological effects of very low magnetic fields*, NASA Technical Note, TN D-5902

Crane, L. Categorical Physics. arXiv: hep-th/9301061v1

Denisyuk, A.; Andreoni, M.; Bondani, M. & Potenza, M. (2000). Real-time holograms generated by second-harmonic cross correlation of object and reference optical wave fields, *Optic letters*, Vol. 25, No 12, pp. 890-892

Ekert, A. (1991). Quantum cryptography based on Bell's theorem, *Phys. Rev. Lett*. Vol. 67, pp. 661-663

Floridi L. (2005). Is Information Meaningful Data?. *Philosophy and Phenomenological Research*, Vol. 70, No 2, pp. 351 – 370

Gabor, D. (1946). Theory of Communication, *J. IEE*, Vol. 93, Part III, No. 26, pp. 429-457

Gisin, N. & Gisin, B. (1999). Bell inequality for arbitrary many settings of the analyzers. *Phys. Lett. A,* Vol. 260, pp. 323-333

Huang, Z.; Wang, B.; Abdalla, E. & Su, R. (2006). Holographic explanation of wide-angle power correlation suppression in the cosmic microwave background radiation, *Journal of Cosmology and Astroparticle Physics*, Vol. 05, pp. 13-23

Jacobson, T. (1995). Thermodynamics of Spacetime: The Einstein Equation of Space. *Phys. Rev. Lett.,* Vol. 75 No 7, pp. 1260-1263

Kirkwood, T. (2008). A systematic look at an old problem. *Nature*, 451, pp. 644-647

Korotkov K. (2002). *Human Energy Field: study with GDV Bioelectrography*. Backbone Publishing Co., NewYork, USA

Makela, J (2010). Notes Concerning "On the Origin of Gravity and the Laws of Newton" by E. Verlinde, arXiv:1001.3808v1 [gr-qc]

Maldacena, J. (1998). The large N limit of superconformal field theories. *Advances in Theoretical and Mathematical Physics*, Vol. 2, pp. 231–252

Marcopoulou, F. (2007). New directions in Background Independent Quantum Gravity, arxiv.org/pdf/gr-qc/0703097v1

Misner, C.; Thorne K. & Wheeler, J. (1973). *Gravitation*, ISBN 0-7167-0344-0, W. H. Freeman and Company, New York, USA

Mureika, J. (2007). Fractal holography: A geometric re-interpretation of cosmological large scale structure, *Journal of Cosmology and Astroparticle Physics*, Vol. 05, 021;

NASA (2008). *WMAP Reveals Neutrinos, End of Dark Ages, First Second of Universe*, at: http://www.nasa.gov/topics/universe/features/wmap_five.html

Nature - online journal (2010), 23 August. doi:10.1038/4661030a,

Pohl, R.; Antognini, A.; Nez, F. et al. (2010). The size of the proton. *Nature*, 466, pp. 213-216

Polson, R. & Vardeny Z. (2004). Random lasing in human tissues, *Appl. Phys. Lett.*, Vol. 85, pp. 1289-1292

Pribram, K. (1991). *Brain and perception: holonomy and structure in figural processing*, Hillsdale, N. (Ed.), ISBN 0898599954, Lawrence Erlbaum Associates, Mahwah, New Jersey, USA

Rajaraman, R. (1982). *Solitons and Instantons*, North-Holland Publishing Company, ISBN-10 /ASIN: 0444870474, Amsterdam, Holland

Shaduri, M.; Tshitshinadze, G. & Davitashvili, T. (2002). Investigation of biological systems' holographic properties. *Bulletin of the Georgian Academy of Sciences*, No 2, pp. 264-267;

Shaduri, M. (2005). Secondary holodiffractional radiation of biological systems. *Kybernetes – The International Journal of Systems & Cybernetics,* Vol. 34, № 5, pp. 666-680

Shaduri, M. (2008a). Principle of holography in complex adaptive systems. *Kybernetes – The International Journal of Systems & Cybernetics*, Vol. 37, № 6, pp 732-738

Shaduri, M.; Benford, M.; Bouchoucha, M.; Sukhin, D. & Lebedev, V. (2008b). Holo-imaging - the principle of holography and its practical application. *Proceedings of the Int. conf. Actual problems of modern physics*, pp. 94-97, Krasnodar, Russia

Shannon, C. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, Vol. 27, pp. 379–423

Shor, P. (1997). Polynomial time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Comput*. № 26, pp. 1484-1509

Smolin, L. (2010). Newtonian gravity in loop quantum gravity, arXiv:1001.3668v2 [gr-qc]

Song, Q.; Xiao, S.; Xu, Z.; Liu, J.; Sun, X.; Drachev, V.; Shalaev, V.; Akkus, O. & Kim, Y. (2010). Random lasing in bone tissue. *Optics Letters*, Vol. 35, Issue 9, pp. 1425-1427

Susskind, L. (2008). *The Black Hole War - My Battle with Stephen Hawking to Make the World Safe for Quantum Mechanics*, ISBN 978-0-316-01640-7, Little, Brown and Company, New York, USA

't Hooft, G. (1993), *Dimensional reduction in quantum gravity*, In: Aly A, Ellis J & Randjbar--Daemi S. (Ed.), Salam--festschrifft World Scientific, Singapore

Urabe, H.; Tominaga, Y. & Kubota, K. (1983). Experimental evidence of collective vibrations in DNA double helix (Raman spectroscopy). *Journal of Chemical Physics*, № 78, pp. 5937-5939

Verlinde, E. (2010). On the Origin of Gravity and the Laws of Newton, arXiv:1001.0785v1

Wheeler, J. (1990). Information, physics, quantum: The search for links, In: *Complexity, Entropy, and the Physics of Information*, Zurek, W. (Ed.), ed., pp. 3-28, Redwood City, CA: Addison-Wesley, USA

Part 2

Holographic Materials

3

Polymer Holography in Acrylamide-Based Recording Material

Milan Květoň[1], Pavel Fiala[2] and Antonín Havránek[3]
[1,2]*Czech Technical University in Prague, Faculty of Nuclear Sciences and Physical Engineering, Department of Physical Electronics*
[3]*Charles University, Faculty of Mathematics and Physics, Department of Macromolecular Physics*
Czech Republic

1. Introduction

Photopolymers are attractive recording materials for different holographic applications such as image holography, holographic memories, diffractive optical elements, or holographic interferometry (Hariharan (1996)). They differ from traditionally used silver-halide emulsions or dichromated gelatin (Bjelkhagen (1996)) in their self-developing characteristics. A hologram in a photopolymer is formed already during exposure of the interference field and no additional wet developing process is needed. The first photopolymer recording media were prepared by Close et al. (1969). Since then, numerous systems have been examined but only a small number become suitable for holographic applications. Companies such as DuPont, InPhase, or Polygrama produce commercial photopolymer systems. However, the exact chemical composition is not known and is proprietary. This fact complicates both experimental and theoretical studies of recording processes which is necessary for understanding of behavior of the recording material. Next to the commercial systems, photopolymers can be also prepared in a laboratory. Its chemical composition is known and can be modified, if we need to study the response on material components. This makes the laboratory prepared photopolymers interesting for the research even if their sensitivity, resolution, refractive index change, or stability are often lower in comparison with commercial systems.

In our holographic laboratory, we prepare, test, and apply different types of recording media. If we use traditional recording materials, we often get on their limits. For example, a thick layer of silver-halide emulsion is difficult to fabricate and process. But only gratings formed in thick layers have high angular selectivity and diffraction efficiency which is required for high capacity holographic recording media. Photopolymers can be prepared in thicknesses ranging from 10-1000 μm. They are also much more convenient for holographic interferometry due to their self-developing ability. However, recording mechanisms running during holographic exposure are still not fully understood and so they can not be effectively used. For our experimental study on photopolymers, the acrylamide-based material was chosen for the following reasons. It is composed of available chemicals, it can be prepared in our laboratory and its composition can be easily modified to produce samples for different tests. The used photopolymer system with acrylamide monomer, triethanol amine initiator,

and polyvinyl alcohol binder was described in the late 1980s by Calixto (1987). Martin et al. (1994) optimized the composition for recording in the green region of the visible spectrum by the addition of a xanthene dye. Acrylic acid as the second co-monomer was introduced into the recording system by Xia et al. (2002). The characteristics of the acrylamide-based photopolymer have been already studied by several authors: Babeva et al. (2008); Martin et al. (1997); Naydenova et al. (2004); Neipp et al. (2003), etc. Within our research on the acrylamide-based recording material, we have modified its chemical composition to obtain more sensitive recording material with higher final value of the refractive index modulation in comparison with other acrylamide-based photopolymers. We have also extended the standard real-time measurement method. Next to the real-time measurement of a growth of the first harmonic grating, we have simultaneously measured a growth of the second harmonic grating that gives a better idea about the non-linear recording process. Some of the results have been already published (Květoň et al. (2007; 2008)).

Next to the experimental works, theoretical models of polymer recording have been developed. Zhao & Mouroulis (1994) proposed one dimensional diffusion model. The model assumes that a growing polymer chain is located at its place of origin. The diffusion equation, which describes polymerization and diffusion processes of monomers, was solved and the model predicts failure of the reciprocity law for intensity and time and the low spatial frequency cut-off. Colvin et al. (1997); Kwon et al. (1999) revised the model and specified more exactly process of the linear radical chain polymerization. With the basic diffusion model, the high spatial frequency cut-off is not possible to explain. Therefore, a non-local response function was introduced to the diffusion equation by Sheridan & Lawrence (2000). The extended model assumes that a polymer chain grows away from its place of initiation and so its size limits the resolution of the recording medium. The non-local polymerization driven diffusion model has been continuously improved, additional processes running during the holographic exposure have been included. A detail review of the non-local modelling was elaborated by Gleeson & Sheridan (2009). Our contribution to the topic of polymer recording is the immobilization-diffusion theory (ID-theory). The theory is based on the fact that also a growing polymer can move and its mobility successively decreases with its polymerization degree (Havránek & Květoň (2008); Havránek et al. (2007)).

Having in mind the goal of the paper, i.e. to explain the process of polymer recording and present our experimental findings obtained with acrylamide-based photopolymer, the paper is organized as follows. The theory of polymer recording is introduced in section 2. The recording process is explained with the ID-theory and the basic simulations are presented. In section 3, the chemical composition of the acrylamide-based photopolymer and its preparation procedure is described. The principles of the characterization method, our setup, and processing of measured data are given in section 4. In the following section 5, our main experimental findings are presented. The material response depends on both exposure conditions and chemical compositions. Hence, an influence of different parameters is shown and discussed. Also the measurements of non-linearities, which form during the recording process, are studied. Finally, the paper is concluded.

2. Theory of polymer recording

The process of a diffraction grating record is complex and involves several mechanisms such as absorption of light, polymerization, and diffusion. Monomer and initiator are the most important components of the photopolymer recording material. The material is sensitive to a particular wavelength of light. If the material is exposed with the interference field,

absorption of light occurs and the polymerization reaction of monomer molecules is initiated. The polymerization rate alternates according to the intensity of illumination and so does the concentration of monomers. In intensity maxima, the amount of polymers rapidly increases and additional monomers flow from surroundings to homogenize the monomer concentration gradient. The mobility of a long polymer chain is very low in comparison with a monomer. Therefore polymers remain at their place of origin where their concentration increases and hence does the refractive index. The recording process is irreversible and a permanent refractive index grating is formed. In Fig. 1, the recording process is illustrated.

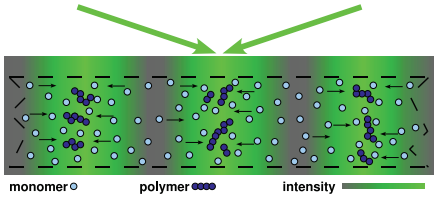


Fig. 1. Record of the harmonic interference field in a photopolymer material.

The briefly described recording process will be further formulated with our ID-theory and some basic simulations of the theoretical model will be presented.

2.1 Immobilization-diffusion theory

The main idea of the new theory, which we denoted immobilization-diffusion theory, is the fact that the mobility of polymer chains decreases as they grow. The mobility of dimers, trimers and other short oligomers is comparable with that of monomers whereas the mobility of polymer chains with higher polymerization degree (number of monomer units of which the polymer is composed) is negligible. The standard diffusion theories neglect the mobility of short oligomers and suppose that the position of a polymer is fixed from the very beginning of its growth. The typical S-shape form of the measured refractive index modulation grow-curves is in contradiction with this simplified assumption. The ID-theory has been created to explain the S-shape of the measured grow-curves. The ideas of the ID-theory will be described at first for one isolated polymer chain and then it will be extended to the real situation when polymers start to grow with some space and time distribution.

2.1.1 Single polymer case

Schematic view of the recording medium with the coordinate system used in further calculations is given in Fig. 2. The number of monomers in a unit volume of the medium is denoted as the monomer concentration $C_m = C_m(x, y, z)$. Only the x-direction dependence $C_m = C_m(x)$ is supposed in the forthcoming calculations. Neglecting the finite dimensions of the exposed film in z-direction leading to ignoring the z-dependence of C_m is not a crude approximation. But the dependence of C_m on the y-coordinate may be important as the thickness of the film ($>10\,\mu$m) is substantially larger than monomer dimensions and the illumination profile may be changed substantially along the y-coordinate. In more refine calculations, the dependence of C_m on y probably cannot be neglected.

Now, we shall study how moves the polymer that starts to grow at the plane $x = 0$. The probabilities to move in $+x$ and $-x$ directions are the same and so, as the mobility of a polymer decreases with increasing polymerization degree k, the polymer moves closer and closer to the plain $x = 0$ at which it starts to grow. The start of the polymerization process

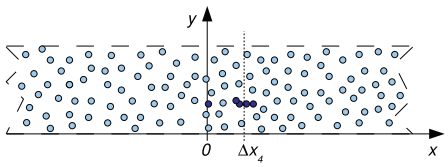


Fig. 2. A schematic view of the recording medium with one quadrimer which actual distance from the plane $x = 0$, where the polymer started to grow, is Δx_4.

means that two monomers join and then other monomers join the developing polymer chain. Polymer chain with low number of monomers is called oligomeric chain or oligomer. Mobility of the oligomer decreases with increasing number of joined monomers and it concentrates near the plain $x = 0$. So the homogeneity of the monomer concentration C_m in the film is disrupted and the modulation of it occurs.
The polymer grows as the monomer units join it. In the intervals Δt_k, monomers are added to the growing polymer chain, k is the polymerization degree. The polymer loses its mobility with increasing k; its probable deviation (actual value of this deviation for quadrimer $k = 4$ is denoted Δx_4 in Fig. 2) from the initial position ($x = 0$) decreases. We assume that

$$\Delta x_k = \frac{\Delta x_1}{k^m}, \; \mathrm{k} = 1, 2, 3 \cdots \tag{1}$$

Exponent m describes the rate of decrease and so the polymer immobilization. The exponent m has values between 0.5 and 1 if viscous random motion of a linear polymer is supposed. Higher values of m are obtained if some branching or cross-linking occurs. The probable deviation $\Delta x_1 = \Delta l_0/3$ where Δl_0 is the mean distance which a monomer molecule must undergo to reach another molecule at the beginning of the polymerization process. The factor $1/3$ is due to the fact that deviations only in the x-direction are concerned. During the polymerization process, the mean distance

$$\Delta l_k = \frac{1}{\sqrt{2} S C_{m,k}} \tag{2}$$

increases as the monomer concentration after k steps $C_{m,k}$ decreases as monomers are consumed by the polymer. The equation (2) is set in analogy to the derivation of the mean free path of molecules in the kinetic theory of gasses (see e.g. Reif (1965)). S is the effective cross-section of the radical through which the polymerization proceeds. The mean distance Δl increases as monomers are consumed by the polymer.
The next important point of the theory is that the deviation Δx_k is identified with the probable deviation δ_k of the Gaussian distribution function (e.g. Reif (1965));

$$|\Delta x_k| = \delta_k \doteq \frac{2}{3}\sigma. \tag{3}$$

Thus the probability $p_k(x)$ to meet the polymer of polymerization degree k at the distance $|x|$ from the plane $x = 0$ can be expressed as

$$p_k(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right) = \frac{2k^m}{\sqrt{2\pi}\Delta l_0} \exp\left(-\frac{2k^{2m}x^2}{\Delta l_0^2}\right), \tag{4}$$

where the mean free path of a monomer molecule Δl_0 was introduced in the second equation. The density $C_{p,k}$ of monomers incorporated in the growing polymer is obtained if the probability p_k is multiplied with the polymerization degree k;

$$C_{p,k}(x) = kp_k(x) = \frac{\sqrt{2}k^{m+1}}{\sqrt{\pi}\Delta l_0} \exp\left(-\frac{2k^{2m}x^2}{\Delta l_0^2}\right). \tag{5}$$

The drift of the polymer density towards the plane $x = 0$ with an increasing polymerization degree k is demonstrated in Fig. 3 for three values of k and three values of the exponent m. The stabilization of polymers in the vicinity of $x = 0$, where they have started to grow, is evident. The rate of stabilization increases with increasing m.

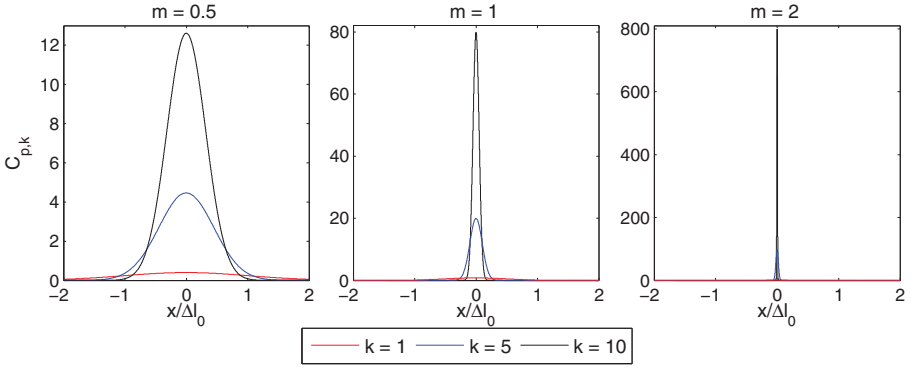


Fig. 3. Stabilization of the polymer near its initial position for three values of immobilization exponent m.

Now we shall study the time dependence of the process. As the free path Δl_k increases with increasing step k (see Eq. (2)) also the time interval Δt_k between two subsequent addings of monomers increases. If the time interval at initial concentration of free monomers C_{m0} is Δt_0, the interval after k steps is

$$\Delta t_k = \Delta t_0 \frac{P_0}{P_0 - k} \tag{6}$$

where P_0 is the total number of monomers in the area from which monomers arrive to our selected single polymer. The total time from the start of polymerization to the k step of it is the sum of the intervals given by Eq. (6);

$$t_k = \Delta t_0 \sum_{j=1}^{k} \frac{P_0}{P_0 - j}. \tag{7}$$

If we plot the density of polymers for $x = 0$, i.e. $C_{p,k}(0)$, against t_k we obtain the typical S-shaped grow-curve. Such a curve is given in Fig. 4 for $P_0 = 100$ and $m = 1$. The dots in the graph represent the individual k steps. The t_k axis is given in units Δt_0. The intervals between dots are the Δt_k values and they increase with increasing k in accordance with Eq. (6).
In this simplified model, the S-shape of the grow-curve may be well explained. At the beginning of the polymerization process (small k values), the immobilization of oligomers prevails and the slope of the grow-curve rises. At the end of the process, nearly all

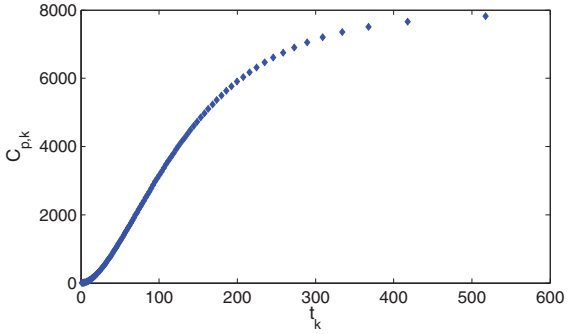


Fig. 4. The dependence of the density $C_{p,k}$ on time t_k for $m = 1$. Time and density are given in arbitrary units.

the monomers are consumed, the Δt_k intervals dramatically grow and the slope of the curve approaches zero value. Somewhere between, the two processes compensate and the grow-curve has its inflexion point.

2.1.2 Continuous spatial and time distribution of the growing polymers

The number of polymers, which start growing in the place of coordinate $x = u$ at time $t = T$, is introduced by the initiation rate $\mathcal{R}_i$. It is assumed that $\mathcal{R}_i$ can be expressed as

$$\mathcal{R}_i(u, T) = AI(u)C_m(T), \tag{8}$$

where A is a constant characterizing the efficiency of initiation, $I(u)$ defines the intensity of illumination and $C_m(T)$ is the concentration of monomer units at time T. We use this relation to include the effect that monomers are exhausted at later times of the exposure. The concentration of free monomers decreases according to the exponential law which can be expressed as

$$C_m(T) = C_{m0} \exp(-T/\tau), \tag{9}$$

where τ is the relaxation time, $C_m(T)$ is the actual and C_{m0} the initial density of free monomers. It is assumed that the concentration of monomers decreases homogeneously (concentration gradient does not arise) due to a high mobility of monomers and short diffusion distances in the recording medium. Next, we introduce the actual degree of conversion

$$d = \frac{C_{m0} - C_m(T)}{C_{m0}}. \tag{10}$$

The degree of conversion d is the ratio of monomers remaining free at time T to the initial number of monomers. The degree of conversion d corresponds in the continuous case to the polymerization step k of the single polymer case. Only the reduction to the number P_0 from Eq. (6) in the single polymer case must be done. Then we obtain for the time dependences of d and reduced k the expressions

$$k' = \frac{k}{P_0} = d(T) = 1 - \exp(-T/\tau). \tag{11}$$

The spatial and time density of monomers incorporated in growing polymers $C_p(x,t)$ may now be expressed as

$$C_p(x,t) = \int_0^t \int_{-\infty}^{\infty} k'(T)\mathcal{R}_i(T,u)p\left(k'(T), x-u\right)\mathrm{d}u\mathrm{d}T. \tag{12}$$

The relaxation time τ and distribution of the harmonic interference field $I(u)$ are given with the expressions

$$\tau = \frac{1}{m\rho\sqrt{I_0}} \quad \text{and} \quad I(u) = I_0\left(1 + V\cos\left(\frac{2\pi}{\Lambda}u\right)\right). \tag{13}$$

m is the immobilization coefficient, ρ is a coefficient that associates initiation, polymerization, and termination rate constants, I_0 is the overall intensity, V is the visibility or contrast of the inference fringes and Λ is the period of the interference field. If we introduce the relaxation time τ and the intensity distribution $I(u)$ in Eq. (12) and assume that the polymer density $C_p(x,t)$ is proportional to the distribution of the refractive index, we obtain after evaluation of the first harmonic amplitude the refractive index modulation

$$n_1(t) \propto \exp\left(-2m\sqrt{I_0}t\right)\left(1 - \exp\left(m\sqrt{I_0}t\right)\right)^2. \tag{14}$$

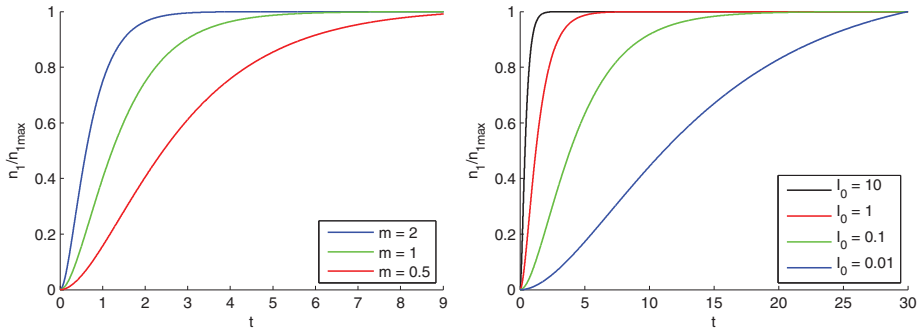
The proportionality constant depends especially on the initial concentration of monomers, the visibility and the intensity of light. For more details see (Havránek et al. (2007; 2008)).

2.1.3 Results of model calculations

The time dependence of the refractive index modulation $n_1(t)$, which describes the grating formation process, is named the grow-curve. The teoretical forms of the grow curves predicted by Eq. (14) are given in Fig. 5. The refractive index modulation at first rises with an increasing grow rate, goes through an inflexion point and with a decreasing grow rate, it achieves its maximum value at which all the monomers are consumed by polymers. The rate of this process depends on the immobilization parameter m and also on the intensity of illumination I_0 and visibility V.

The dependence of grow-curves on exponent m is illustrated in Fig. 5 (a) where the calculated reduced curves for three different values of the immobilization parameter m are given. It can be seen, that with increasing m, the curves become more abrupt and their points of inflexion occur at lower reduced time values. The grow-curves corresponding to higher values of m indicate that some faster mechanism of polymer growth, such as branching or network formation, probably occurs. The ID-theory prediction of the intensity dependence of grow-curves is given in Fig. 5 (b). Four curves, which mutual relative intensity differs ten times, are illustrated. The rate of growth increases with an increasing intensity of illumination and an earlier saturation occurs. Presented curves are normalized for a better comparison of the form of grow-curves with different exposure intensities. These curves correspond also with the results which were obtained with simulations of the diffusion model at different recording intensities.

The ID-theory coming out from very natural assumptions is able to explain the experimentally observed S-type form of the grow-curves. The predicted intensity dependence of theoretically obtained curves fits well with main features of experimentally observed curves. Branching or cross-linking during the polymerization process is introduced through the immobilization exponent m. The presented form of the theory is an elementary version with many approximations and simplifications. A dependence of the refractive index modulation on the



(a) different exponent m (b) different intensity I_0

Fig. 5. The normalized grow-curves for different values of the immobilization exponent m and recording intensity I_0. Time and intensity are given in relative units.

spatial period Λ and a formation of higher harmonics are not considered. A refinement of the theory will be possible after thorough analyzes of experimental results. Preliminary results will be shown and discussed further.

3. Recording material

For the measurements, the laboratory prepared acrylamide based photopolymer recording material is used. It is composed of acrylic monomers, polyvinyl alcohol binder, initiator and sensitizing dye. The prepared dry layers are typically 80 μm thick and are deposited on glass plates. The absorption maximum of the material is at a 537 nm wavelength (green light), its optimal exposure dose ranges from 20 to 40 mJ/cm^2, and the diffraction efficiency of a formed grating (volume phase transmission grating) reaches almost 100 %.

The chemical composition is based on the formula introduced by Martin et al. (1994). Within our experimental studies, concentrations of components and thickness of the material have been varied to obtain the optimal response for our experimental applications. All used chemicals are commercially available and they have been used without previous purification. The material is composed of the binder – polyvinyl alcohol (PVA), solid monomer – acrylamide (AA), liquid monomer – acrylic acid (AAC), crosslinker – N,N'-methylene-bis-acrylamide (BAA), initiator – triethanolamine (TEA), and sensitizing dye – erythrosine B (EB).

The chemical composition of the recording material, which is typically used for the measurements, is described in Tab. 1. The Solution I is prepared at first. PVA is dissolved in demineralized water and then 17 % water solution of PVA is produced. A normal white light room illumination can be used as PVA is not sensitive to a visible light. Then the Solution II is prepared. Components from the list (except Solution I) are put together and mixed well at first. It takes some time to dissolve the crosslinker. To this mixture, the appropriate amount of Solution I is added and mixed again. The viscous solution contains many air bubbles which might degrade the optical quality of the layer. Hence the solution is evacuated to release all unwanted bubbles. Red safe light must be used during the preparation procedure of the Solution II to avoid initiation of the polymerization reaction and thus a damage of the material.

The prepared solution is poured over a levelled glass plate, spread with a stick or other coating tool and gravity settled. The amount of the solution is a function of the desired thickness of the sensitive layer. For example, if we want to obtain a 80 μm dry layer, we are to take 0.04 cm^3 of the solution on each 1 cm^2 of a glass plate. After the deposition, the layer is dried in dark under normal laboratory conditions (temperature 22 °C and 40 % relative humidity) for about 24 hours. The dry layer of the recording material is not in the solid state but it can be characterized as a tough gel permeable for small molecules such as monomers. As the monomer concentration is relatively high, precipitation of it on the surface of the layer may occur during storage and so the material becomes unusable for the measurements. The precipitation can be avoided, if a thin cover glass is attached on the layer with a cyanoacrylate glue just after drying. Another way is covering the surface with a clear lacquer.

4. Measuring method and experimental setup

The material is tested with the continuous measurement of the diffraction efficiency of a forming grating during holographic exposure. Martin et al. (1997); Moreau et al. (2002); Rhee et al. (1995), and others have been already used methods with the similar principle. Usually, only the time measurements of the diffraction efficiency were obtained that is not sufficient for a material characterization. The diffraction efficiency is a parameter describing the quality of a diffraction grating, but the recording material is better characterized with the refractive index modulation. Refractive index is more closely connected with the chemical composition of a recording material and its changes can be used for describing the processes running within the recording material.

A quantity of the refractive index modulation is determined from the measurement of the diffraction efficiency with Eq. (18) derived in the coupled wave theory (Kogelnik (1969)). The efficiency is measured with a wavelength which is different from the recording one. It is chosen from the non-absorbing band of the recording material and so the measurement does not influence the recording process. As the recording process can be non-linear and higher harmonics can arise, the detection method has been recently extended to the measurement of the diffraction efficiency of the second harmonic grating. Both efficiencies are measured at once and the amplitudes of refractive index gratings are determined.

In Fig. 6, a schema of the setup, which enables recording of the diffraction grating and measurement of its diffraction efficiency, is given. The beam from the recording laser is

Solution I	
Polyvinyl alcohol ($M_w = 89000 - 98000$)	5 g
Water	30 cm^3
Solution II	
Triethanolamine	4.2 cm^3
Water	2.8 cm^3
Erythrosine B ($2{,}5\times10^{-3}$ M water sol.)	2.2 cm^3
Acrylamide	1.9 g
N,N′-methylene-bis-acrylamide	0.22 g
Acrylic acid	1.84 cm^3
Solution I	20 cm^3

Table 1. Typical chemical composition of the acrylamide-based recording material

expanded, filtered, and collimated through the spatial filter and collimating lens. The formed plane wave is divided into two parts. One half of it comes through the material directly and the second half is reflected by the mirror. Both then overlap and form the harmonic interference field which is finally recorded. The recording material is placed symmetrically into the interference field which is perpendicular to the surface of the recording material where a volume phase transmission grating is formed.

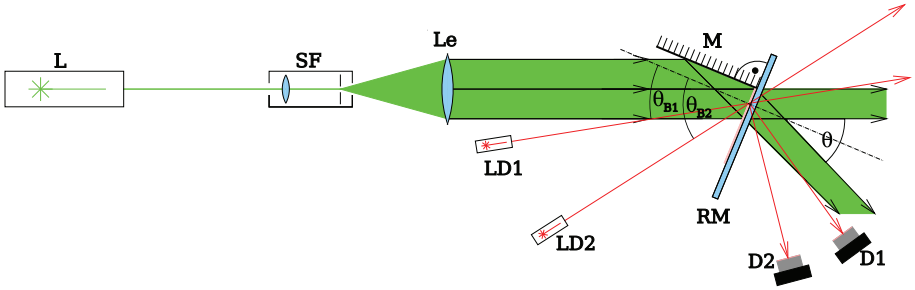


Fig. 6. A schema of the experimental setup. Legend: L - recording laser (532 nm), SF - spatial filter, Le - collimating lens, M - plane mirror, RM - recording material, LD1, LD2 - laser diodes (656 nm), D1, D2, D3 - photo-detectors.

A narrow beam of the detection laser diode (LD1), which wavelength is out of the absorption band, is incident at the Bragg angle θ_{B1} on the recording material. When the beam is passing through the grating, a diffracted beam is generated. The power of the beam is continuously measured with the detector (D1). The second harmonic grating is detected with the laser diode (LD2) which is adjusted at the Bragg angle θ_{B2} of the second harmonic grating. The power of the diffracted beam is detected with the detector (D2). The exposure and the measurements proceed simultaneously.

The interference field is unslanted in the recording medium. The period of the interference field Λ is given by the expression

$$\Lambda = \frac{\lambda}{2\sin\frac{\theta}{2}}, \tag{15}$$

where λ is the wavelength of recording light and θ is the angle between incident waves (in the air). The first harmonic grating has the same spatial period as the interference field and the period of the second harmonic grating is $\Lambda/2$.

The volume phase transmission grating is very selective with respect to the reconstruction angle. The maximum diffraction efficiency is obtained at the Bragg angle. Bragg angles can be evaluated from Eq. (15) where the reconstruction wavelength λ_r, period of the first harmonic grating Λ and period of the second harmonic grating $\Lambda/2$ are substituted;

$$\theta_{B1} = \operatorname{asin}\frac{\lambda_r}{2\Lambda} \qquad \text{and} \qquad \theta_{B2} = \operatorname{asin}\frac{\lambda_r}{\Lambda}. \tag{16}$$

The diffraction efficiency is a ratio of the diffracted power to the incident power. It can be evaluated from the real-time measurement of power $P_{d1}(t)$, power $P_{d2}(t)$, of the diffracted beam, respectively and power P_{i1}, P_{i2} of the incident beam, respectively.

$$\eta_1(t) = \frac{P_{d1}(t)}{P_{i1}} \qquad \text{and} \qquad \eta_2(t) = \frac{P_{d2}(t)}{P_{i2}}. \tag{17}$$

First harmonic grating

The efficiency η_1 and the refractive index modulation n_1 relate through the equation derived in the coupled wave theory.

$$\eta_1 = \sin^2 \frac{\pi n_1 d}{\lambda_r \cos\theta_{B1}}, \tag{18}$$

where d is the thickness of a recording layer, λ_r is the reconstruction wavelength and θ_{B1} is the Bragg angle. The refractive index modulation is evaluated as an inverse function to (18) and hence it is given with the expression

$$n_1 = \frac{\lambda_r \cos\theta_{B1}}{\pi d} \operatorname{asin}\sqrt{\eta_1}, \tag{19}$$

for $n_1 \in \langle 0, \lambda_r/(2d)\cos\theta_{B1})$. The function (18) is not a uniquely invertible function and so when the overmodulation occurs, $n_1 \in \langle \lambda_r/(2d)\cos\theta_{B1}, \lambda_r/d\cos\theta_{B1})$, the refractive index modulation equals

$$n_1 = \frac{\lambda_r \cos\theta_{B1}}{\pi d} \left(1 - \operatorname{asin}\sqrt{\eta_1}\right). \tag{20}$$

Second harmonic grating

The relation between the efficiency η_2 and the second harmonic amplitude n_2 was derived in the extended coupled wave theory (Benlarbi & Solymar (1980a;b)). The analytical solution of the coupled wave equation was revised and used for calculations of the second harmonic grating strength by Zhao & Mouroulis (1995). The equation is given by

$$\eta_2 = \sin^2 \left[\frac{\pi d}{\lambda_r \cos\theta_{B2}} \left(n_2 - \frac{n_1^2}{n_0 \sin^2\theta_{B2}} \right) \right], \tag{21}$$

where θ_{B2} is the Bragg angle of the second harmonic grating and n_0 is the mean value of the refractive index. The second harmonic amplitude of the refractive index modulation is again evaluated as an inverse function to Eq. (21) and hence it is given with the expression

$$n_2 = \frac{\lambda_r \cos\theta_{B2}}{\pi d} \operatorname{asin}\sqrt{\eta_2} + \frac{n_1^2}{n_0 \sin^2\theta_{B2}} \tag{22}$$

for $n_2 \in \langle 0, \lambda_r/(2d)\cos\theta_{B2} + n_1^2/(n_0\sin^2\theta_{B2})\rangle$. It can be seen, that the value n_1 must be known for determination of the second harmonic amplitude n_2. Usually, the second harmonic grating is not overmodulated, because the value n_2 is relatively small and hence only the unique inversion is assumed.

In the Kogelnik's coupled wave theory, higher diffraction orders are neglected due to violation of the Bragg condition. In reality, very weak higher diffraction orders may arise even in a thick recording medium. We analyzed possible higher diffraction orders of the first harmonic grating and compared them with diffraction orders of the second harmonic grating. It was found that the diffraction angles, which correspond to higher orders of the first harmonic grating, are completely different from the diffraction angles of the second harmonic grating. Therefore the measurement of the diffraction efficiency of the second harmonic grating is not influenced with the measurement of the first harmonic grating. Both measurements are independent and thus they can be carried out simultaneously.

4.1 Experimental setup

The exposure and detection setup is placed on the optical table which is isolated and damped from surrounding vibrations (see Fig. 7). For the exposure, the diode-pumped, frequency-doubled Nd:YAG laser (L), that emits at a 532 nm wavelength, is used (Coherent model 532-400). The beam from the laser passes through the electronically controlled shutter (S). The attenuator (A), which reduces power of the beam, is the next optical component. With the attenuator, the required optical intensity incident on the recording material is adjusted. The beam then passes through the spatial filter (SF) where it is focused and spatially filtered. The divergent wave from the filter is collimated with the lens (Le) to form a quasi plane wave. One half of it is incident directly on the recording material placed in the mount (H) and the other half is reflected with the mirror which is firmly fasten to the mount. Both waves overlap and create a harmonic interference field in the place where the sample is located.

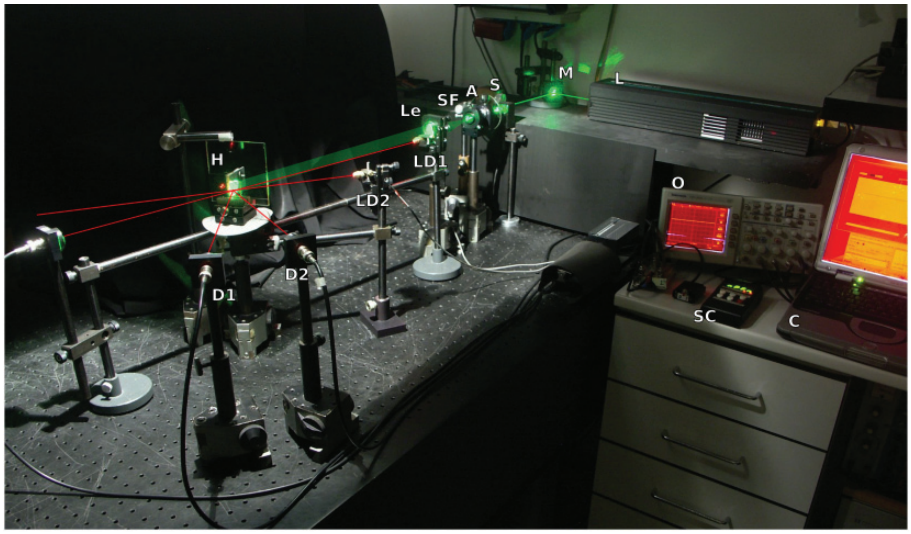


Fig. 7. Experimental setup used for the real-time measurements.

The detection part of the setup is composed of laser diodes (LD1, LD2), optical detectors (D1, D2) and oscilloscope (O). The standard collimated laser diodes at a 656 nm wavelength and 5 mW output power are used. The power of diffracted and passing beams is detected with PIN photodiodes. The measured optical power ranges from 0 to 5 mW. The photodiodes are wired in a reverse bias voltage mode to provide a high linearity. When illuminated, it generates a current passing through a load resistance on which the voltage is measured with the oscilloscope. The digital storage oscilloscope is used for the measurement of signals from photodiodes.

The exposure and measurement is operated with the shutter-control (SC). It operates the shutter which opens and closes the recording beam and also triggers the oscilloscope measurement. The acquired data are stored in the memory of the scope and can be transferred through the serial line to a computer for further processing.

4.2 Measurement and data processing

Before the measurement, exposure parameters such as overall recording intensity, period of the interference field, and exposure time are adjusted. The intensity is accurately set with the attenuator (A).The spatial period of the interference field is adjusted by rotating with mount (H) around its vertical axis. The value of the period is determined from the angle between recording waves with Eq. (15). Finally, the exposure time is set on the shutter-control which operates the exposure and measurement.

Laser diodes must be adjusted at Bragg angles. Firstly, laser diode beams are roughly directed at calculated Bragg angles (Eq. (16)). Secondly, a test sample of the grating is recorded and angles of detection beams are precisely adjusted at maximum diffraction efficiency which is measured with the detector. The measured voltage is linearly proportional to the optical power on the detector; the constant of proportionality has been determined experimentally.

Before the exposure, the sample of the recording material is placed to the mount and fixed with a small clamp against shifting. The size of each sample is typically $2 \times 2\,\mathrm{cm}^2$ and xylen is used as an immersion liquid to prevent reflections on boundaries between the sample and the mount. Other losses caused by Fresnel reflections on boundaries are eliminated during evaluation of the diffraction efficiency. Noise of the recording material and possible higher diffraction orders are negligible.

The shutter-control starts the exposure and measurement. The grating continuously grows up which is measured with detectors in the real-time. The diffraction efficiency is calculated with Eq. (17) from the acquired data. For evaluation of the refractive index modulation (Eq. (19)), the thickness of the sample layer must be known. It is measured with a mechanical profilometer after each exposure. In Fig. 8, an example of measured diffraction efficiency $\eta_1(t)$ and evaluated refractive index modulation $n_1(t)$ of the forming grating are presented. As was already mentioned, the plot of the refractive index modulation on time is the grow-curve.

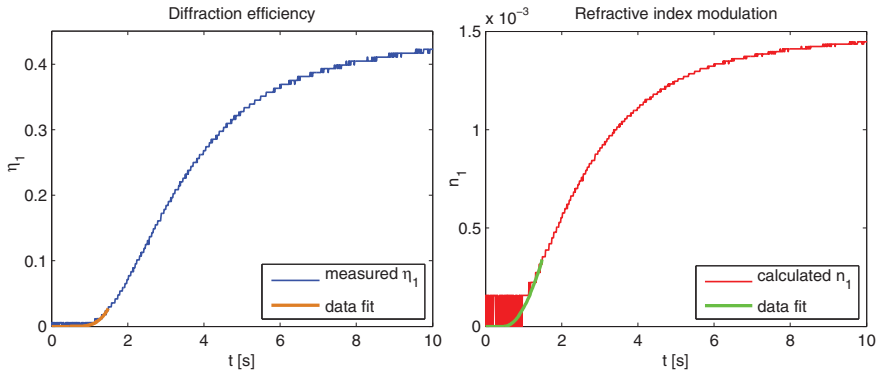


Fig. 8. Time variations of the diffraction efficiency and refractive index modulation. Exposure parameters: $\Lambda = 700\,\mathrm{nm}$ and $I_1 = I_2 = 1.3\,\mathrm{mW/cm}^2$.

The time resolution of the data measured with the oscilloscope is excellent. The typical curve is composed of 2500 points (see Fig. 8). The noise at low values of the diffraction efficiency is due to the reverse bias mode of the photodiode and also due to the low voltage resolution of the digital oscilloscope. This relatively low noise is even more amplified when the refractive index modulation is evaluated. The main cause of this problem is Eq. (19). If we analyze the

differential of Δn_1, it can be seen that it depends non-linearly on the differential of $\Delta\eta_1$:

$$\Delta n_1 = \frac{\mathrm{d}n_1}{\mathrm{d}\eta_1}\Delta\eta_1 = \frac{\lambda\cos\theta_{B1}}{2\pi d}\,\frac{1}{\sqrt{\eta_1 - \eta_1^2}}\Delta\eta_1. \tag{23}$$

A small deviation $\Delta\eta_1$ at low values of efficiencies ($\eta_1 \to 0$) is very much amplified and it results in a large deviation of Δn_1. It complicates the interpretation of beginning parts of the grow-curve; mainly the initiation time of the growth is difficult to find.
The problem with a noise for low signals can be reduced when the measurement is performed twice with the same exposure conditions. At first, the whole measurement is done with the reverse biased photodiode. At second, the photodiode is wired in the unbiased mode and the measurement is performed with a new sample and the same exposure conditions. In this mode, the photodiode has a high sensitivity and a low noise. However, the detection range is limited only to low values of the optical power as the saturation of the photodiode occurs. The initial part of the curve is now measured very precisely and hence the initiation time can be found. Finally, we replace the measured noise data from the reverse biased mode with the data obtained in the unbiased mode before the saturation of the detector and the whole curve of the refractive index modulation is evaluated without the noise. In Fig. 8, the preciously measured data (data fit) are also presented.

5. Experimental results and findings

Conventional recording materials such as silver-halide emulsions or dichromated gelatin are characterized with measurements of the diffraction efficiency which are obtained after processing. In the optimization procedure, several test gratings must be recorded, processed, and measured to find optimal values of exposure parameters. In photopolymers, the real-time detection method enables to obtain more characteristics at once (the grow-curve). However, the behavior of a photopolymer material is more complicated. The reciprocity law (a relation connecting the optimal value of the exposure energy with exposure time and recording intensity) is not valid and also the performance of the material strongly depends on other parameters such as the spatial period of the interference field or chemical composition of the recording system.
Within our work on this topic, a large number of measurements have been carried out. The detection method and experimental setup have been continuously improved and also the accuracy of the measurements. The selected results, which are here presented, describe our main findings. Some more experimental results have been already published in our recent papers (Květoň et al. (2007; 2008)).

5.1 The grow-curve

The grow-curve describes the process of the diffraction grating formation through the parameter n_1. The typical grow-curve is illustrated in Fig. 9 where the formation process is divided into four recording phases:

I – induction period. At the beginning of exposure, a short induction period of the length t_i is observed; no refractive index modulation is detected. In spite of no detectable change of the refractive index, some initiation processes may be running in the recording material.

II – exposure growth. At time t_i, the growth of the grating begins. This phase is interrupted at exposure time t_e when the recording laser is switched off. The rate of a grating formation

depends on exposure parameters and a chemical composition of the material. Typically, the formation rate increases in the beginning of this phase. The grow-curve goes through its inflexion point and the formation rate decreases at the end of the exposure phase.

III – post-exposure growth. An additional increase of the refractive index modulation occurs. The growth is not interrupted immediately at t_e even if the exposure is switched off. It runs out successively and saturates as the formation rate decreases to zero. This phase also strongly depends on exposure parameters and a material composition.

IV – final period. In the final period, the refractive index modulation remains stable or it may slowly decrease in time. The decrease is caused by degradation processes and its rate depends on the value of n_1 and chemical composition of the material. In fact, the degradation process may already be running from the beginning of the grating growth, but it is observed only at later exposure times when the grow rate decreases.

The phases of the grow-curve have been analyzed further in a more detail. They depend on both the exposure conditions and chemical composition.

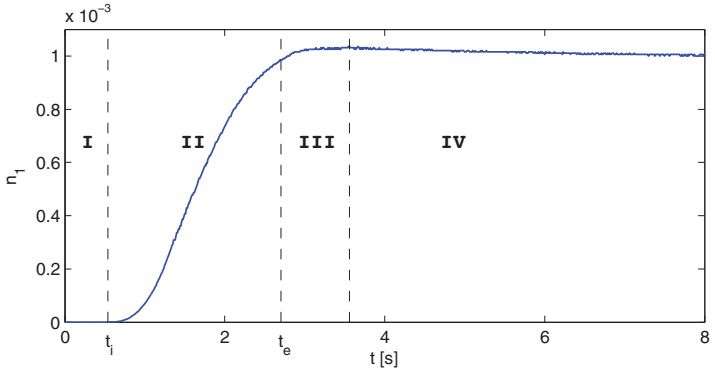


Fig. 9. Phases of the grow-curve. I – induction period, II – exposure growth, III – post-exposure growth, IV – final period.

Volume phase transmission gratings with the diffraction efficiency close to 1 are produced in the recording material given in Tab. 1. This efficiency corresponds to the value of n_1 of the order 10^{-3} ($n_{1max} = 4 \times 10^{-3}$). However, the maximum value is strongly dependent on exposure conditions and chemical composition, already mentioned above. Also the laboratory conditions, which occur during the preparation procedure and measurements (e.g. humidity, temperature), influence the performance of the material (grow rate, n_{1max}). So if we compare measurements at different exposure conditions, they have been performed on samples of the recording material prepared at the same laboratory conditions to avoid misinterpretations.
The optimal exposure energy is a parameter of a special interest in recording materials. It is defined as the amount of radiation received by a surface per unit area and it is calculated as a product of the incident intensity and time. Waves can be incident at angles different from the normal to the surface and so cosine factors must be considered. The interference field is formed by two waves of intensities I_1 and I_2 which are incident at angles θ_1 and θ_2 and so the exposure energy is

$$E = (I_1 \cos\theta_1 + I_2 \cos\theta_2)\, t. \tag{24}$$

In conventional recording materials, the reciprocity law is often assumed

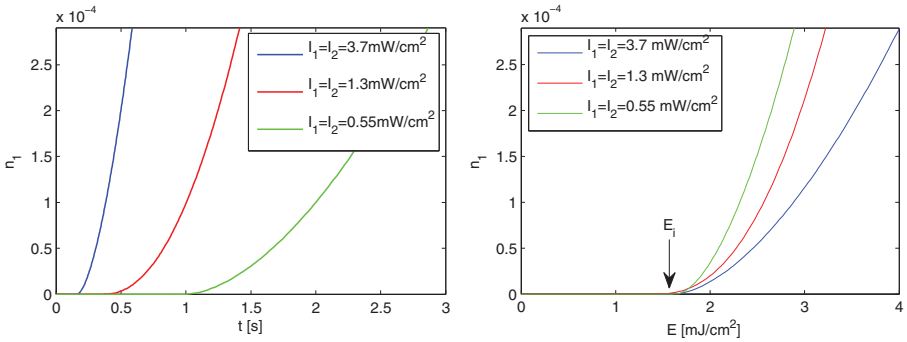
$$E_{opt} = It_e, \tag{25}$$

where intensities and cosine factors are included in the overall recording intensity termed I. The reciprocity law says that the product of the overall recording intensity I and the exposure time t_e has the same efficiency independently on the ratio of I to t_e. There are some small deviations from the reciprocity law for very low and high intensities even in conventional materials. The deviations from the reciprocity are more prominent in the case of photopolymers. At lower recording intensities, E_{opt} is usually lower.

5.2 Influence of exposure parameters

5.2.1 Induction period

The length of the induction period t_i depends mainly on the recording intensity. In the case of a higher recording intensity, the induction period is shorter than that in the case of a lower intensity, as can be seen in Fig. 10 (a).



(a) grow-curves – induction periods (b) E-n_1 curves – initiation energies

Fig. 10. Measurements of induction periods and the calculated E-n_1 curves for different recording intensities (Λ=0.7 μm).

Measurements for different values of the intensity and the spatial period of the recording interference field were done and the results of induction times t_i are given in Tab. 2. For the same exposure intensities, the value of t_i is very similar for all periods of the interference field. Relatively small differences are due to the short spatial period cut-off. In the case of $\Lambda_1 = 0.45\,\mu$m, a grating with a very low value of the refractive index modulation is formed, in comparison with $\Lambda_3 = 2\,\mu$m. Therefore, it is assumed that the first detectable changes are slightly shifted in time.

In Fig. 10 (b), exposure energy variations of n_1 are given for the same measurements. A linear relation between the exposure energy and time is given by Eq. (24) where the constant of proportionality is the overall intensity. It can be seen that the dose of the induction exposure energy E_i is very similar for all recording intensities. The values of E_i were evaluated also for other spatial periods and recording intensities and the results are summarized in Tab. 2 (the values in brackets).

At the beginning of polymerization, the effect of inhibition with polymerization suppressors is described (Odian (2004)). We assume that the induction period is caused by oxygen as

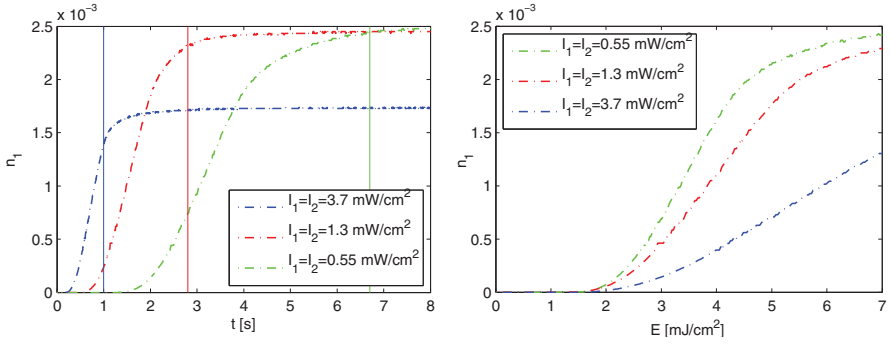
Intensity $I_1 = I_2$	t_i [s] (E_i [mJ/cm^2])					
[mW/cm^2]	$\Lambda_1 = 0.45\,\mu$m		$\Lambda_2 = 0.7\,\mu$m		$\Lambda_3 = 2\,\mu$m	
0.55	1.70	(1.51)	1.70	(1.73)	1.40	(1.53)
1.3	0.70	(1.47)	0.63	(1.51)	0.55	(1.42)
3.7	0.30	(1.79)	0.25	(1.71)	0.22	(1.60)

Table 2. Measured induction times t_i and calculated initiation energies E_i (given in brackets) for different exposure intensities and spatial periods of the recording interference field.

it is present in the recording layer due to the preparation procedure. It reacts preferably with primary radicals and thus inhibits the standard polymerization reaction of monomers. The dose of E_i is the energy which must be delivered to the photopolymer system to consume all molecules of oxygen. Each sample of the recording material contains a similar amount of oxygen and therefore the same dose of E_i applies. After this induction period, the polymerization of monomers normally proceeds and the change of the refractive index is detected. The theoretical models do not include the effect of inhibition and so the grow-curves start to rise from the very beginning.

5.2.2 Exposure growth

After the induction period, the grating starts to grow. At the beginning of this phase, the rate of growth rises up. When the grow-curve reaches its inflexion point, the rate of growth starts to decrease and a saturation of the grow-curve occurs (even if the exposure continues). In the case of a higher recording intensity, a higher recording rate is detected than that in the case of a lower recording intensity (see Fig. 11 (a)). The effect of intensity can by theoretically explained by the both diffusion models. The polymerization rate depends on the recording intensity. If a higher recording intensity is applied, the polymerization proceeds at the higher polymerization rate and so the grating growth is faster than that in the case of a lower recording intensity.



(a) grow-curves (b) E-n_1 curves

Fig. 11. Comparison of measured grow-curves and calculated E-n_1 curves for different recording intensities (Λ=2 μm).

In Fig. 11 (a), the exposures are interrupted at 1, 2.8, and 6.7 s (the times of interruption t_e are marked with appropriate vertical lines). At these times, the exposure energies are similar in all three case, but the values of n_1 differ. To show the reciprocity law Eq. (25) failure, we have

plotted the measured refractive index modulation n_1 versus exposure energy E. It follows, that a lower recording intensity requires less energy to achieve the same value of n_1. So the exposure process is more effective for low recording intensities. It seems better to use longer exposure times at lower intensities to obtain a higher index modulation and hence a more efficient grating. However, the exposure time is a crucial parameter. A fine interference field is not stable enough for long lasting exposures and also some other degradation effects, which will be discussed further, may negatively influence the formation process.
The grating growth depends strongly on the spatial period of the interference field. The grow-curves of different spatial periods and the same recording intensities are compared in Fig. 12. In the case of a long period ($\Lambda = 2\,\mu$m), the grating grows up rapidly and a high value of the refractive index n_1 is reached. On the other hand, a short spatial period ($\Lambda = 0.45\,\mu$m) is difficult to record. It is caused by a limited resolution of the acrylamide based material that ranges from 0.5 to 5 μm. For spatial periods longer than 2 μm, a relief grating is also formed, higher diffraction orders are produced and the detection method becomes inconvenient.

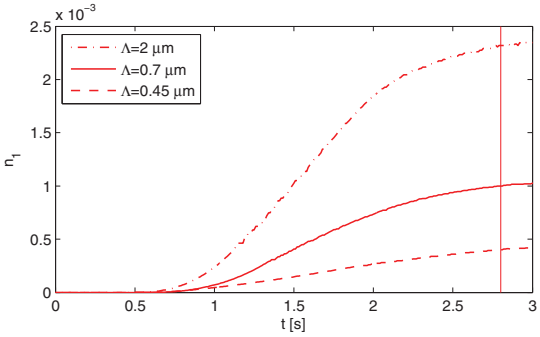


Fig. 12. The grow-curves for different spatial periods of the recording interference field ($I_1 = I_2 = 1.3\,\mathrm{mW/cm^2}$, $t_e = 2.8\,$s).

5.2.3 Post-exposure growth

The grating growth continues even if the exposure is interrupted. The post-exposure growth is a part of the grow-curve between the point of interruption t_e and the maximum of the grow-curve. In Fig. 13, the exposures of similar samples have been interrupted at different exposure times (2, 4, and 7 s). The highest value of the post-exposure increase ($\Delta n_1 = 1.4 \times 10^{-4}$) was obtained when the exposure was interrupted at t_e=2 s (red curve).
The magnitude of the post-exposure growth depends on the grow rate of the grating at the interruption time. In the case of a higher grow rate (usually at early stages), the post-exposure growth is larger than that in the case of lower grow rates. To verify this observation, the post-exposure growth was measured for different spatial periods and intensities of the interference field (see Fig. 14). At first, the samples were exposed with the same exposure intensities ($I_1 = I_2 = 3.7\,\mathrm{mW/cm^2}$, $t_e = 1\,$s) but three different spatial periods of the interference field were used. In Fig. 14 (a), it can be seen that the additional increase decreases with shortening of the spatial period. At second, the post-exposure increase for different recording intensities was measured (see Fig. 14 (b)). The spatial period was set to 2 μm and the dose of exposure energy was similar in all three tested samples. The largest additional

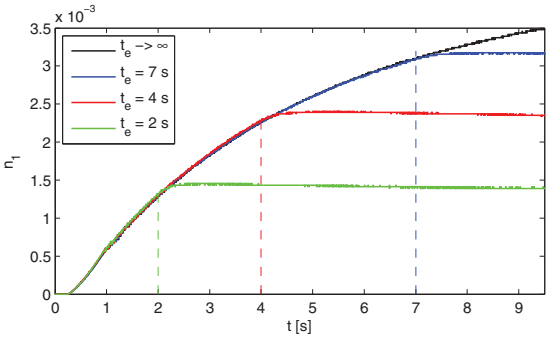
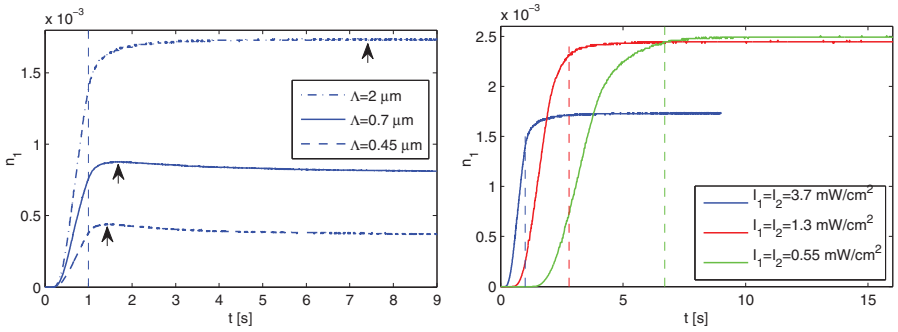


Fig. 13. Post-exposure growth – the exposure was interrupted at $t_e = 2,\ 4,$ and $7\,s$ ($\Lambda = 0.7\,\mu$m and $I_1 = I_2 = 3.7\,\mathrm{mW/cm^2}$).



(a) (b)

Fig. 14. Post-exposure growth – influence of spatial periods (a) and recording intensities (b).

increase is observed for the highest recording intensity. A similar behavior was found also for other tested spatial periods.

The effect of the post-exposure growth is usually explained with the diffusion model. At the interruption time t_e, a gradient of monomer concentration is formed in the recording material. After the exposure, radicals recombine very rapidly and so the post-exposure growth is ascribed only to the diffusion of monomers from dark to bright places. This diffusion process continues until the gradient vanishes.

From our complementary measurements (electron paramagnetic resonance) follows that the life-time of radicals is relatively long. Hence the polymerization process continues even after the exposure until the radicals recombine. It seems more probable that the post-exposure growth is caused by remaining radicals. Our conclusion is also supported with the recent measurements of the diffusion process in acrylamide photopolymers performed by Babeva et al. (2008). It has been found that the diffusion coefficient is much higher than that assumed in previous studies. The diffusion of monomers is very fast and so the magnitude of the concentration gradient is irrelevant. These results are in agreement with the ID-theory where no concentration gradient is supposed.

5.2.4 Final period

The refractive index modulation remains stable or slowly decreases after the exposure. Sometimes, a decrease of n_1 is observed already during long lasting exposures. A decrease of the refractive index modulation is assigned to the degradation process. The degradation process starts already at the beginning of the formation process but the formation process is much more dominant and therefore, a decrease of n_1 is not observed. The degradation process shows out at later times when the value of n_1 is high and the formation rate starts to decrease. In contradiction to the exposure and post-exposure growth, the degradation does not depend much on the recording intensity. A comparison of the degradation processes of gratings recorded with different intensities is given in Fig. 15.

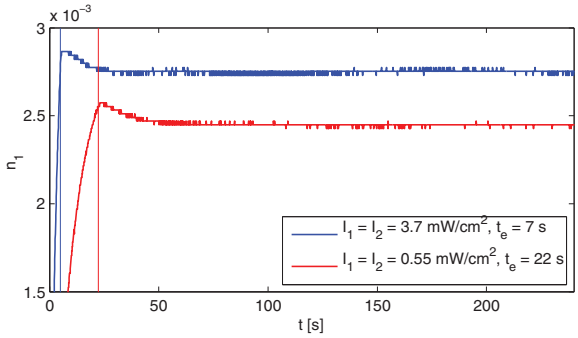


Fig. 15. Degradation process of the diffraction grating ($\Lambda = 0.7\,\mu$m).

The degradation rate depends on the spatial period of a grating formed, as may be also seen in Fig. 14 (a). The relative decrease of n_1 is 17% for $\Lambda = 0.45\mu$m, 7% for $\Lambda = 0.7\mu$m, and 1% for $\Lambda = 2\mu$m, 9 s after the interruption of the exposure. The gratings with larger spatial periods are more stable and also their values of the refractive index modulation are higher.

The degradation process is not included in the diffusion model describing formation of the grating in photopolymers; it is assumed that a grating formed is stable due to immobility of polymers. This is not true because the opposite diffusion of polymers occurs due to their concentration gradient. The mobility of polymers is much lower than that of monomers but it should not be automatically neglected. This is taken into account in the ID-theory. In the next section, it will be shown that the polymer mobility can be decreased by addition of a crosslinker and so the stability of the grating is increased.

In the case of DuPont or Polygrama photopolymers, additional processes are involved to fix or even improve characteristics of diffraction gratings formed. The grating is stabilized by the post-exposure with a homogeneous illumination. In the case of acrylamide-based photopolymers, the post-exposure illumination speeds up the degradation effect. Experiments of post-exposure illumination were carried out. The value of the refractive index modulation dropped to its half value after the exposure but the final grating was stable and no additional decrease of n_1 was observed. This effect is not very clear and more experiments have to be done to clarify it.

5.3 Influence of the chemical composition

Monomers and crosslinker are fundamental for the process of the grating formation. Acrylamide (AA), acrylic acid (AAC), and methylene-bis-acrylamide (BAA) influence the

recording process in different ways. Each molecule differs in polymerization rate, size, structure, or diffusion ability. A high value of the refractive index change is desired and so a high concentration of monomers in the recording layer is necessary. Unfortunately, the crosslinker BAA is very badly soluble in water and hence layers with a high concentration can not be prepared. A very high concentration of AA produces opaque layers due to its precipitation. Finally, gratings formed in layers composed of only AAC are not stable enough and degrade rapidly in time.
Samples of the recording material with a different composition of monomers were prepared and tested to show the influence of different monomers on the recording process. In each sample of the recording layer, the concentration of C=C double bonds of all monomers was similar. Molecules of AA and AAC include only one double bond, BAA has two double bonds which both may radically polymerize. The concentrations of remaining components were exactly the same in each sample (see optimized material composition in Tab. 1). Contents of monomers in the recording layer are summarized in Tab. 3 and the results of the measurements are given in Fig. 16.

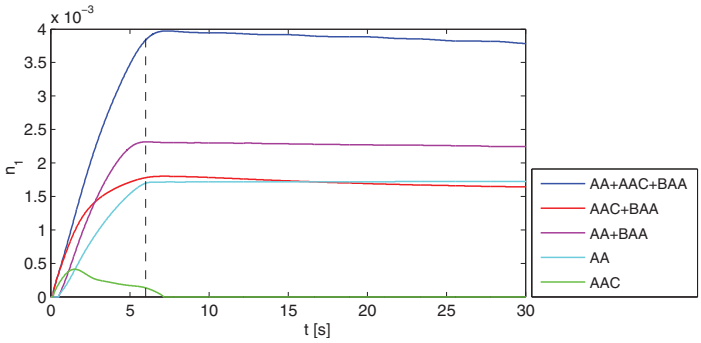


Fig. 16. Grow-curves of materials with different contents of monomers ($I_1 = I_2 = 3.7\,\mathrm{mW/cm^2}$, $\Lambda = 0.7\,\mu\mathrm{m}$, and $t_e = 6\,\mathrm{s}$).

If only AA (cyan line) is present in the layer, the grating forms slowly and a low value of the refractive index modulation n_1 is reached. The formed grating is stable and even a slow increase of n_1 is observed in the phase IV. Addition of crosslinker (AA+BAA – magenta line) increases the rate of grating growth. A higher value of n_1 is obtained but the refractive index modulation slightly decreases after the exposure proceeds.
If only AAC (green line) is added, the grating starts to form early but the rate of grating formation is low. The maximum value of n_1 is low and a decrease of n_1 can be observed already during the exposure. When the exposure is interrupted, the modulation decreases very quickly to zero and the grating is completely erased. If a small amount of crosslinker is added (AAC+BAA – red line), the rate of the grating formation increases. The length of the induction period is shorter than that in the case of the material with AA+BAA (magenta line). The rate of formation changes from high on the beginning to low at the end of the exposure process. After exposure, a small decrease of n_1 is measured.
The grating, which is formed in the recording material with both monomers and crosslinker (AA+AAC+BAA - blue line), starts to grow very early. Its grow rate is high and large value of n_1 is reached at time t_e. The post-exposure growth can be observed but also the degradation process. The results of all materials are summarized in Tab. 3.

Monomer	Ratio AA:AAC:BAA	t_i [s]	$n_1(6\,\mathrm{s}) \times 10^{-3}$	$n_1(30\,\mathrm{s})/n_1(6\,\mathrm{s})$
AA	100:0:0	0.27	1.72	1.02
AA+BAA	95:0:5	0.25	2.33	0.97
AAC	0:100:0	0.23	0.13	0
AAC+BAA	0:95:5	0.2	1.82	0.92
AA+AAC+BAA	47.5:47.5:5	0.17	3.85	0.99

Table 3. Characteristics of recording materials with different compositions of monomers.

It was demonstrated that addition of the crosslinker increased the grow rate. This effect has been predicted with the ID-theory. The crosslinker concentration is characterized with the immobilization exponent m; the exponent increases with an increasing amount of the crosslinker. Branching and networking of polymers occur even when a small amount of the crosslinker is added. Stabilization of the branched polymer is faster and its opposite diffusion is much more difficult.

5.4 Detection of the second harmonic grating

The recording process is non-linear and so higher harmonics are formed even if the recording material is exposed with the harmonic interference field. Next to the first harmonic grating with the same elementary period as the interference field, a second harmonic grating is formed with the half period. The detection method of the second harmonic grating has been introduced in section 4. The first and second harmonic amplitude of the refractive index are plotted together in one figure for a comparison (see Fig. 17). Typically, the first harmonic grating starts to grow first. The second harmonic appears later when the first reaches a specific value. Both amplitudes then continue in growth. The maximum of the first amplitude is reached at a different time (usually much earlier) than that of the second harmonic grating. When the exposure goes on, the first harmonic amplitude n_1 may start to decrease due to the degradation process but the growth of the second harmonic continues.

5.4.1 Influence of the spatial period

First and second harmonic amplitudes were measured for different spatial periods of the recording interference field (see Fig. 17). In the case of a short period ($\Lambda = 0.7\,\mu$m), n_1 is lower than that for longer periods and values of n_2 are almost undetectable. Such a grating has nearly ideal harmonic profile. In the case of longer spatial periods ($\Lambda = 2\,\mu$m), higher values of n_1 are obtained but also a relatively high n_2 is formed. The recording process is more non-linear.

A lower value of n_2 relates to lower resolution of the material at short spatial periods. The period of the second harmonic frequency is two times shorter than that of the first harmonic. So in the case of spatial period $0.7\,\mu$m, the second harmonic period is only $0.35\,\mu$m which is out of the resolution of the medium and hence the second harmonic grating is not formed.

5.4.2 Influence of the recording intensity

A significant non-harmonic profile is produced when the recording interference field with a large spatial period is applied. The period of the interference field was set to $2\,\mu$m and measurements with different exposure intensities were carried out. The results are presented in Fig. 18. Dependences of n_1 and n_2 on the exposure energy E are given for a better comparison of all displayed curves.

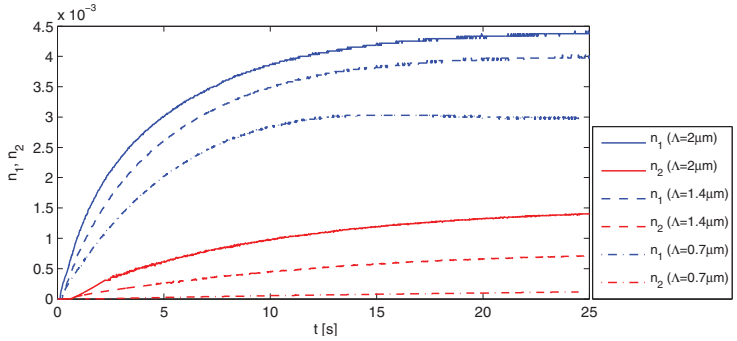


Fig. 17. Grow-curves of n_1 and n_2 obtained with the exposure of different spatial periods ($I_1 = I_2 = 3.7\,\mathrm{mW/cm^2}$).

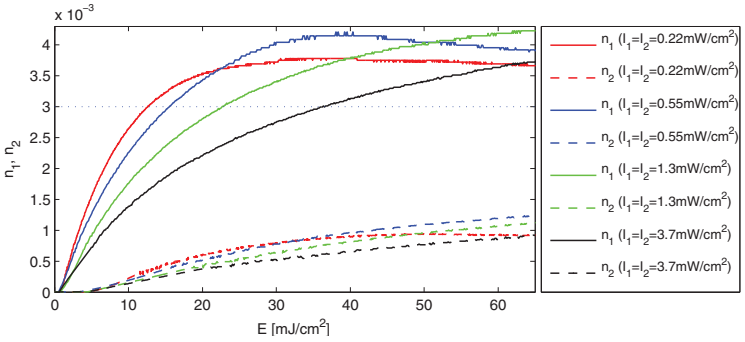


Fig. 18. Dependences of the first and second harmonic amplitudes on the exposure energy for different intensities of the recording interference field (spatial period was 2 μm). The reference value of n_1 is marked with the dotted line.

The saturation values of n_1 and n_2, respectively are very similar for all recording intensities. Usually, the exposure is interrupted when n_1 reaches a sufficiently high value. So we chose a number $n_1 = 3 \times 10^{-3}$ as the reference value for the examination of non-linearities at different intensities. The quantum of the exposure energy, when the grow-curve reaches the value 3×10^{-3}, is labelled $E_{n_1=3\times10^{-3}}$. The energy varies for different intensities. At this quantum, n_2 is evaluated and is given for a comparison in Tab. 4. In the case of lower recording intensities, n_1 reaches the number 3×10^{-3} at lower exposure energies and a lower amplitude of the second harmonic grating is formed.

It was also found that the post-exposure growth of the second harmonic amplitude is low. The degradation process occurs similarly as in the case of the first harmonic amplitude, but the value of n_2 is lower than that of n_1 and hence a thorough analysis of the post-exposure effects of n_2 is less significant. In general, the non-linear recording process causes undesirable effects. The formation of the second harmonic grating reduces the concentration of monomers available for the growth of the first harmonic grating. Also higher diffraction orders may arise due to non-linearities (thin gratings). The growth of n_2 proceeds at later exposure times and so it can be partially eliminated if the exposure is interrupted earlier.

$I_1 = I_2$ [mW/cm^2]	$E_{n_1=3\times10^{-3}}$ [mJ/cm^2]	$n_2(E_{n_1=3\times10^{-3}})$ $\times10^{-4}$	$n_2 : n_1$
0.22	12.7	3.2	0.11
0.55	15.6	3.8	0.12
1.3	22.7	4.9	0.16
3.7	36.4	6.1	0.2
9.2	77.6	7.1	0.24
36	807	9.9	0.33

Table 4. Comparison of non-linearities for different exposure intensities. Second harmonic amplitudes were evaluated at exposure energies when $n_1 = 3 \times 10^{-3}$.

6. Conclusion

The paper was devoted to the study of acrylamide-based recording material for optical holography. During the holographic exposure, the refractive index of the material changes and a volume phase grating is formed without an additional developing process. The formation process of a grating was described theoretically with the ID-theory. It was shown that the grow-curve obtained from both modelling and measurements could be used for characterization of the self-developing photopolymer recording materials. We measured samples of our laboratory prepared photopolymer with the real-time detection method at different exposure conditions and chemical compositions. Obtained grow-curves were compared and confronted also with the theoretical results. The measurements were performed on the acrylamide-based photopolymer, but the results are valid for other types of photopolymer recording materials. In our laboratory, the photopolymers produced by Polygrama were also tested with our experimental setup and the results were published by Lédl & Květoň (2008). Recently, we have used our results and experience for modification of the acrylamide-based material and prepared samples of the optimized photopolymer which has been successfully used as a recording medium for holographic interferometry and holographic memories (Květoň et al. (2010)).

7. References

Babeva, T., Naydenova, I., Martin, S. & Toal, V. (2008). Method for characterization of diffusion properties of photopolymerisable systems, *Optics Express* 16: 8487–8497.

Benlarbi, B. & Solymar, L. (1980a). An analysis of volume phase gratings at higher Bragg angle incidence, *Int. J. Electronics* 48: 361–368.

Benlarbi, B. & Solymar, L. (1980b). Higher-order modes in non-sinusoidal voluma phase holograms, *Int. J. Electronics* 48: 351–359.

Bjelkhagen, H. I. (ed.) (1996). *Selected Papers on Holographic Recording Materials*, Vol. MS 130, first edn, SPIE, Bellingham USA.

Calixto, S. (1987). Dry polymer for holographic recording, *Applied Optics* 26: 3904–3910.

Close, D. H., Jacobson, A. D., Margerum, J. D., Brault, R. G. & McClung, F. J. (1969). Hologram recording on photopolymer materials, *Applied Physics Letters* 14: 159–160.

Colvin, V. L., Larson, R. G., Harris, A. L. & Schilling, M. L. (1997). Quantitative model of volume hologram formation in photopolymers, *J. Applied Physics* 81: 5913–5923.

Gleeson, M. R. & Sheridan, J. T. (2009). A review of the modelling of free-radical photopolymerization in the formation of holographic gratings, *J. Optics A* 11: 24008–24019.

Hariharan, P. (1996). *Optical Holography: Principle, techniques, and applications*, second edn, Cambridge University Press, Cambridge UK.

Havránek, A. & Květoň, M. (2008). Polymer holography - new theory of image growth, *Macromolecular Symposia* 268: 43–47.

Havránek, A., Květoň, M. & Havránková, J. (2007). Polymer holography II - the theory of hologram growth: Polymer growth detected by holographic method, *Polymer Bulletin* 58: 261–269.

Kogelnik, H. (1969). Coupled wave theory for thick hologram gratings, *The Bell System Tech. J.* 48: 2909–2947.

Květoň, M., Havránek, A., Fiala, P. & Richter, I. (2007). Polymer holography I – method and experiment: Polymerization and networking as a method of permanent holographic record formation, *Polymer Bulletin* 58: 253–259.

Květoň, M., Havránek, A., Fiala, P. & Richter, I. (2008). The experimental study of acrylamide based photopolymer recording materials for applications in optical holography, *Macromolecular Symposia* 268: 115–119.

Květoň, M., Lédl, V., Havránek, A. & Fiala, P. (2010). Photopolymer for optical holography and holographic interferometry, *Macromolecular Symposia* 295: 107–113.

Kwon, J. H., Hwang, H. C. & Woo, K. C. (1999). Analysis of temporal behavior of beams diffracted by volume gratings formed in photopolymers, *J. Optical Society of America B* 16: 1651–1657.

Lédl, V. & Květoň, M. (2008). Characterization and application of new photopolymer recording media for usage in holography, *in* P. Tománek, D. Senderáková & M. Hrabovský (eds), *Photonics, Devices, and Systems IV*, number 71381E in *Proc. SPIE Vol.7138*, SPIE, Prague CZ.

Martin, S., Feely, C. A. & Toal, V. (1997). Holographic recording characteristics of an acrylamide-based photopolymer, *Applied Optics* 36: 5757–5768.

Martin, S., Leclere, P., Renotte, Y., Toal, V. & Lion, Y. (1994). Characterization of an acrylamide-based dry photopolymer holographic recording material, *Optical Engineering* 33: 3942–3946.

Moreau, V., Renotte, Y. & Lion, Y. (2002). Characterization of DuPont photopolymer: determination of kinetic parameters in a diffusion model, *Applied Optics* 41: 3427–3435.

Naydenova, I., Jallapuram, R., Howard, R., Martin, S. & Toal, V. (2004). Investigation of the diffusion processes in a self-processing acrylamide-based photopolymer system, *Applied Optics* 43: 2900–2905.

Neipp, C., Gallego, S., Ortuno, M., Marquez, A., Belendez, A. & Pascual, I. (2003). Characterization of a pva/acrylamide photopolymer. Influence of a cross-linking monomer in the final characteristics of the hologram, *Optics Communications* 224: 27–34.

Odian, G. (2004). *Principles of Polymerization*, fourth edn, Wiley, New Jersey USA.

Reif, F. (1965). *Statistical Physics in Berkeley Physics Course – volume 5*, first edn, McGraw Hill, New York USA.

Rhee, U., Caulfield, H. J., Vikram, C. S. & Shamir, J. (1995). Dynamics of hologram recording in DuPont photopolymer, *Applied Optics* 34: 846–853.

Sheridan, J. T. & Lawrence, J. R. (2000). Nonlocal-response diffusion model of holographic recording in photopolymer, *J. Optical Society of America A* 17: 1108–1114.

Xia, C., Zhu, J., Wang, K., Wang, B. & Liu, L. (2002). Optimization of high-quality photopolymer for holographic recording, *in* D. Hsu, J. Chen & Y. Sheng (eds), *Holography, Diffractive Optics, and Applications*, SPIE, Shanghai China, pp. 239–242.

Zhao, G. & Mouroulis, P. (1994). Diffusion model of hologram formation in dry photopolymer materials, *J. Modern Optics* 41: 1919–1930.

Zhao, G. & Mouroulis, P. (1995). Second order grating formation in dry holographic photopolymers, *Optics Communications* 115: 528–532.

4

Real-time, Multi-wavelength Holographic Recording in Photorefractive Volume Media: Theory and Applications

Eduardo Acedo Barbosa
Laboratório de Óptica Aplicada, Faculdade de Tecnologia de São Paulo, CEETEPS – UNESP, Pça Cel Fernando Prestes, 30, 01124-060, São Paulo – SP Brazil

1. Introduction

Due to very interesting properties like reversible recording with unlimited number of recording-erasure cycles, high image resolution and high storage capability, thick photorefractive media have been applied in many areas like information storage (Chuang & Psaltis, 1997), pattern recognition (Alam & Khoury, 2000), phase conjugation (Chang et al, 2001), optical image processing (Poon & Banerjee, 2001) and optical metrology (Georges et al, 2001; Frejlich & Garcia, 2000).
Photorefractive crystals of the sillenite family ($Bi_{12}SiO_{20}$, $Bi_{12}TiO_{20}$ and $Bi_{12}GeO_{20}$) have characteristics which make them very useful for image processing and for optical metrology, particularly for holographic interferometry (Barbosa & Muramatsu, 1997; Kukhtarev et al, 1993). Those crystals present a relatively short response time, which is very desirable for the field of non-destructive testing, and the typically low diffraction efficiency of these materials can be overcome by taking advantage from their anisotropic diffraction properties (Kamshilin & Petrov, 1985). The $Bi_{12}TiO_{20}$ (BTO) crystals have the additional advantage of a relatively low optical activity for the red light, which allows the highest possible diffraction efficiency for this wavelength range (Frejlich, 2006).
Holographic contouring is generally based on two-exposure recording, i.e. in each exposure the holographic image is recorded with one setup configuration. During the holographic readout the reconstructed waves referent to each object state interfere and thus generate a contour fringe pattern. In two-wavelength techniques (Kuchel & Tiziani, 1981; Millerd & Brock, 1997) the laser is tuned between the exposures, while in other methods the beam illuminating the object is slightly tilted between the first and the second recording (Pedrini et al, 1999; Yamaguchi et al, 2001).
The study and application of holographic processes in sillenite crystals by using multi-wavelength, large free spectral range diode lasers provided the generation of real-time contour interferograms with only one exposure, thus enabling much faster and easier testing. It was shown that the diffracted light intensity strongly depends on the phase difference between the reference and the object beams and on the laser FSR. The holographic image thus appears covered of interference fringes corresponding to the contour lines of the object surface (Barbosa, 2005). By phase-shifting the reference beam, the fringes run along

the surface, thus allowing for quantitative three-dimensional scanning of the object by phase shifting methods. This chapter shows the basic theory and the consequent applications of multi-wavelength holography in thick photorefractive materials. The text is organized as follows: in section 2 the fundamentals of the holographic recording in photorefractive crystals are presented; section 3 shows the basic principles which originate the longitudinal lasers modes and the contour fringe formation when holographic interferometers are illuminated by multimode lasers; section 4 treats on the theory and experiments about two-laser holography, and section 5 describes two applications of multi-wavelength and two-laser holography.

2. Holographic recording in thick photorefractive materials

When two coherent, monochromatic waves of wavelength λ interfere at a photorefractive material, the resulting light interference pattern spatially rearranges the charge carriers (electrons or holes) inside the medium by diffusion or drift. The charge carriers are optically excited from filled donors and are recombined to empty traps thus migrating from illuminated to dark regions of the material, resulting in a space-dependent charge distribution. According to Poisson's equation this charge grating generates a spatially periodic electric field E_{sc}, and through the electrooptic effect (Pockels effect) this space charge field produces a refractive index grating of amplitude Δn. The induced birefringence grating in turn depends not only on the photorefractive medium constants, like e.g. its refractive index and its electro-optic response, but also on experimental parameters, like the angle between the interfering beams and the orientation of the holographic vector with respect to the cristalographic axes, in the case of photorefractive crystals. For instance, in a very common configuration of the $LiNbO_3$ crystal one gets $\Delta n = n_0 r_{33} E_{sc}/2$, while for the sillenite crystals ($Bi_{12}SiO_{20}$, $Bi_{12}TiO_{20}$ and $Bi_{12}GeO_{20}$) cut in the electrooptic transverse configuration one gets $\Delta n = n_0^3 r_{41} E_{sc}/2$, where r_{33} and r_{41} are electrooptic coefficients, and n_0 is the bulk refractive index (Frejlich, 2006).

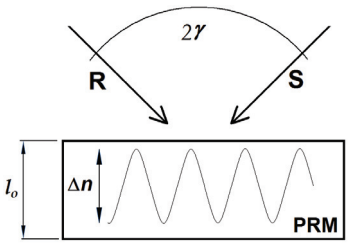


Fig. 1. Incidence of the reference and the object beams onto the photorefractive medium

Due to its simplicity and versatility, the most used optical scheme in holographic interferometry is the dual-beam interferometer which allows simultaneous holographic recording and readout. In this configuration the readout occurs by self diffraction, with the reference wave acting also as the reconstructing wave. Let 2γ be the angle between the object and the reference beams interfering at the photorefractive medium (PRM) of thickness l_0, as shown in Figure 2. According to Kogelnik's coupled-wave theory (Kogelnik, 1969), the diffraction efficiency η of a phase transmission grating is given by

$$\eta = \frac{\sin^2\left(\sigma^2 + \xi^2\right)^{1/2}}{1 + \left(\xi/\sigma\right)^2}, \tag{1}$$

where $\sigma \equiv \pi d\Delta n/(\lambda \cos\alpha)$. The *Bragg factor* ξ is a measure of how out of Bragg regime the readout process occurs, such that $\xi = 0$ for perfect Bragg conditions. This parameter will be studied in more detail in the next sections.

In this chapter the holographic process will be analyzed regarding the parameters and the properties of the sillenite crystals. Those materials present typically low values of refractive index modulation Δn (in the order of 10^{-5}), hence $\sin\sigma \approx \sigma$. In the most experiments of homodyne holographic interferometry the perfect Bragg condition is automatically fullfiled, so that equation (1) becomes simply $\eta \cong \sigma$. This latter parameter in turn was calculated considering the cut of the sillenite crystals in the [110] electrooptic transverse configuration to be (Mallick & Rouède, 1987)

$$\eta_0 = \left(\frac{n_0^3 r_{41} E_{sc}}{2\lambda \cos\gamma} \frac{\sin \rho l_0}{\rho l_0}\right)^2 \tag{2}$$

where ρ is the crystal rotatory power. In this case, the anisotropic diffraction properties of the sillenite crystals and their optical activity can be explored in order to make the transmitted and the diffracted waves orthogonally polarized at the crystal output, provided the input wave polarization is selected to be parallel to the [001] axis half-way through the crystal (Kamshilin & Petrov, 1985). Thus, by blocking the transmitted object wave by a polarizer the holographically reconstructed object image only can be visualized.

In multi-wavelength holography the Bragg conditions are often not obeyed; in some experiments of wavefront reconstruction by two-laser holography the parameter ξ can be significantly larger than σ, so that from equation (1) the diffraction efficiency assumes the form

$$\eta = \eta_0 \frac{\sin^2 \xi}{\xi^2} \tag{3}$$

Equation (3) shows that the diffraction efficiency η equals the perfect Bragg value η_0 for $\xi = 0$ and becomes smaller than unit in off-Bragg regimes, falling down to 0 for $\xi = \pi$ rad.

3. Holographic imaging and wavefront reconstruction with a single multimode laser

3.1 Longitudinal modes of large bandwidth, large free spectral range lasers

The resonator of the diode lasers can be regarded basically as a Fabry-Perot étalon or interferometer with the gain medium inside. The electromagnetic field originated from the gain medium oscillates between the two plane and parallel mirrors of the étalon, being partially reflected, partially transmitted every time it bounces one of the mirrors. The spectrum of a radiation field transmitted through an étalon filled with a passive (i.e. non-amplifying) medium indicates the wavelengths which may oscillate inside the laser resonator. From the basic Fabry-Perot theory the relation between those wavelengths and the resonator L is given by (Hecht, 1998)

$$L = q\frac{\lambda}{2}, \tag{4}$$

where q = 1,2,3… Hence, when such an étalon is the cavity of a inhomogeneously broadened laser (Koechner, 1998), the waves obeying equation (4) comprise the *longitudinal modes* of this laser. The free spectral range (FSR) $\Delta\nu$ is the spectral interval between two adjacent modes and can be written with the help of equation (4) and the relation $c = \nu\lambda$ as

$$\Delta\nu = \nu_{q+1} - \nu_q = \frac{c}{2L}, \tag{5}$$

where c is the light velocity inside the medium. The corresponding wavelength difference is then given by

$$\Delta\lambda = \frac{\lambda^2}{c}\Delta\nu = \frac{\lambda^2}{2L} \tag{6}$$

3.2 Holographic recording and readout with a multimode laser; contour fringe formation

Let us now consider both the reference and the object waves originated from a multimode laser with N longitudinal modes. At the holographic medium input, the reference (R_N) and the object (S_N) beams can be expressed as

$$\begin{aligned} R_N &= R_0 \sum_{q=-(N-1)/2}^{q=(N-1)/2} A_q e^{i[(k+q\Delta k)\Gamma_R + \varphi_q]} \\ S_N &= S_0 \sum_{q=-(N-1)/2}^{q=(N-1)/2} A_q e^{i[(k+q\Delta k)\Gamma_S + \varphi_q]} \end{aligned} \tag{7}$$

where $k \equiv 2\pi/\lambda$, $\Delta k = 2\pi\Delta\lambda/\lambda^2$ is the wavenumber gap between two adjacent modes, A_q is a real coefficient associated to the amplitude of each longitudinal mode and ϕ_q is the phase of the q-th mode at the laser output. Γ_S and Γ_R are the optical paths of the object and the reference beams at the crystal, respectively.

In the absence of an externally applied electric field on the photorefractive crystal the charge transport occurs only by diffusion. The space charge electric field E_{sc} is then given by

$$E_{sc} \cong imE_D \equiv iE_D \frac{2R_N^{*}S_N}{I_0} \tag{8}$$

where m is the modulation of the interference pattern, $I_0 = |R_N|^2 + |S_N|^2$ and E_D is the electric field generated by diffusion. The symbol * denotes complex conjugation.

Each mode (i.e., each wavelength) contributes to the holographic recording with its own hologram and therefore the resulting grating is a superposition of all holograms. Since different modes are not mutually coherent, the phase difference $\phi_q - \phi_p$ $(q \neq p)$ has a random behaviour in time. Thus, the interference of different modes does not contribute to the holographic recording, being only responsible for a background light which lowers the grating amplitude Δn and consequently decreases the overall hologram diffraction

efficiency. Hence, from equations (7) and (8) the diffraction efficiency can be expressed through the following matrix product:

$$\eta = \eta_0 \left| ie^{ik(\Gamma_S - \Gamma_R)} \sum_{p=-(N-1)/2}^{p=(N-1)/2} \sum_{q=-(N-1)/2}^{q=(N-1)/2} \delta_{p,q} A_p A_q e^{i\Delta k(p\Gamma_S - q\Gamma_R)} \right|^2 \Rightarrow$$

$$\eta = \eta_0 \left| \sum_{n=-(N-1)/2}^{n=(N-1)/2} A_n^2 \, e^{in\Delta k(\Gamma_S - \Gamma_R)} \right|^2 \tag{9}$$

where $\eta_0 \equiv \left(\frac{n_0^3 r_{41} E_D}{2\lambda \cos\gamma} \frac{R_0 S_0}{I_0} \frac{\sin \rho l_0}{\rho l_0} \right)^2$ is the unmodulated diffraction efficiency.

In self-diffraction, each mode is diffracted by all the holograms. Thus, for N longitudinal laser modes, there are N recorded gratings and N^2 diffracted waves. The diffracted waves originated from the same mode add coherently and are responsible for the fringe formation while the diffracted beams originated from different modes add incoherently (addition of wave intensities). This can be mathematically expressed by writing the amplitude of the diffracted wave as a function of the square root of the diffraction efficiency. Let us consider the example in which the multiple holograms result from the interference of waves originated from a three-mode diode laser. When the volume hologram is illuminated by the reference wave $R = R_0 A_0 e^{ik\Gamma_R}$ (which correspond to the 0-th mode of the laser) the diffracted wave amplitude can then be written as

$$\begin{aligned} E_D(q=0) &= \sqrt{\eta} R_0 e^{ik\Gamma_R} \\ &= \sqrt{\eta_0} \left(\chi_{-1,0} A_{-1} e^{-i\Delta k(\Gamma_S - \Gamma_R)} + A_0 + \chi_{1,0} A_1 e^{i\Delta k(\Gamma_S - \Gamma_R)} \right) R_0 e^{ik\Gamma_R} \end{aligned} \tag{10a}$$

Correspondingly, the diffracted waves due to the incidence of the modes -1 and 1 can be written as

$$E_D(q=-1) = \sqrt{\eta_0} \left(A_{-1} e^{-i\Delta k(\Gamma_S - \Gamma_R)} + \chi_{0,-1} A_0 + \chi_{1,-1} A_1 e^{i\Delta k(\Gamma_S - \Gamma_R)} \right) R_0 e^{i(k-\Delta k)\Gamma_R} \tag{10b}$$

$$E_D(q=1) = \sqrt{\eta_0} \left(\chi_{-1,1} A_{-1} e^{-i\Delta k(\Gamma_S - \Gamma_R)} + \chi_{0,-1} A_0 + A_1 e^{i\Delta k(\Gamma_S - \Gamma_R)} \right) R_0 e^{i(k+\Delta k)\Gamma_R} \tag{10c}$$

The terms $\chi_{p,q} \equiv \sin \xi_{p,q} / \xi_{p,q}$ take into account the deviations from the perfect Bragg regime. The three diffracted waves from equation (10a) add coherently, and the same occurs with the ones of equations (10b) and (10c). On the other hand, the waves resulting from equation (10a), equation (10b) and equation (10c) add incoherently.

The intensity of the waves diffracted by the volume grating can be obtained in a much more convenient form by making some simplifying assumptions without loss of generality. The parameters of the multimode diode lasers typically lead to values of $\chi_{p,q}$ very close to 1. Moreover, let us consider the particular case for which A_q=*1* in order to simplify the following analysis of surface profilometry. Hence, when the reference beam of intensity I_R is diffracted by the hologram, the holographic reconstruction of the object beam of intensity I_S is obtained from equations (8), (9) and (10) as

$$I_D = \eta\, I_R = \eta_0 \left\{ \frac{\sin[N\Delta k(\Gamma_S - \Gamma_R)/2]}{\sin[\Delta k(\Gamma_S - \Gamma_R)/2]} \right\}^2 I_R \tag{11}$$

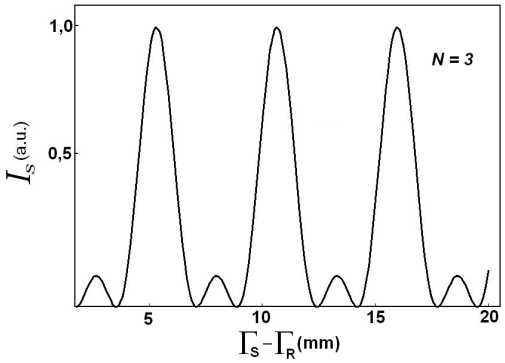


Fig. 2. Intensity of the diffracted wave for N = 3 longitudinal modes

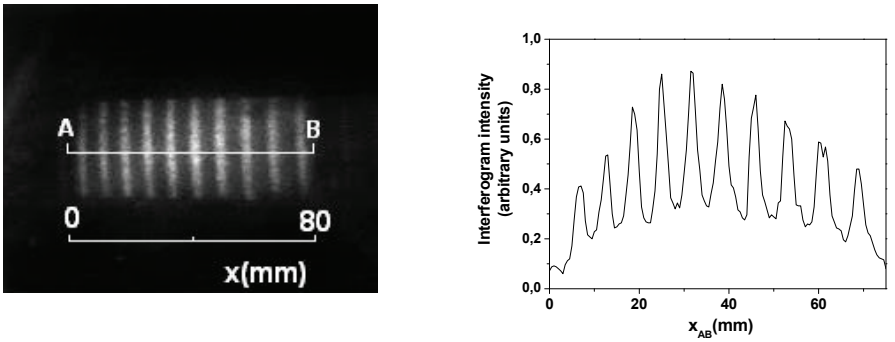
Figure 2 shows the behaviour of I_S from equation (11) as a function of $\Gamma_S - \Gamma_R$, for $\Delta k = 1.39$ rad/mm and N = 3. The intensity maxima correspond to the bright fringes on the reconstructed object image. Since the optical path Γ_S depends on the object surface, and considering that the reference beam is a plane wave, one comes to the conclusion that the contour interference fringes are originated from the intersection of the surface with parallel planes of constant elevation. The distance between those planes is closely related to the resulting synthetic wavelength, as will be seen with more detail in the next sections. The analysis of such interference pattern allows the profilometric measurement of three-dimensional surfaces. From equation (11) one determines the difference on the optical paths between any pair of adjacent fringes as a function of the wavelength difference $\Delta\lambda$:

$$\Gamma_{S,O} - \Gamma_{S,P} = \frac{2\pi}{\Delta k} = \frac{\lambda^2}{\Delta\lambda} \equiv \lambda_S \tag{12}$$

The term λ_S is called as the *synthetic wavelength*, widely used in heterodyne interferometry. Notice that the synthetic wavelength is directly related to the laser FSR for longitudinal modes given by equation (6), such that $\lambda_S/2 = L$. Hence, when long resonator (e.g. L > 20 cm) multimode lasers are employed the contour fringes cannot be visualised for surface depths smaller of tens of centimeters, since the free-spectral range of such lasers is much smaller than that of short-length ($L \cong 2$ mm, typically) diode lasers.

From equation (11) one sees clearly that the bright contour fringes resulting from multi-wavelength holography do not necessarily present the same form of the usual double-exposure, cos²-fringe patterns. In fact, the width $\delta\Gamma_S$ of the bright fringe, defined as twice the Γ_S difference from the fringe peak to the first zero, can be easily obtained with the help of equation (15) as $\delta\Gamma_S = 2/N\Delta k$. This expression evidences that the higher is the number N of longitudinal modes, the narrower is the bright fringe. Figure 3a shows the holographic image of a flat metallic bar, 30º-tilted with respect to the front face of the crystal with the expected vertical and parallel contour fringes, while figure 3b shows the interferogram intensity profile along the line A-B from figure 3a. The narrow fringes in the intensity profile

from figure 3b in comparison with the usual cos²-fringes suggests the oscillation of 4 laser modes.



(a) (b)

Fig. 3. (a) – Multi-wavelength contour interferogram generated on a flat plate; (b) – intensity profile of the fringe pattern.

The group of phase-shifting techniques for quantitative fringe evaluation is based on the intensity change of the interferogram by phase shifting one of the interfering beams in the holographic setup. In the current case, this is carried out in a four-frame process, through which the reference beam is sequentially 0-, $\pi/2$-, π- and $3\pi/2$-phase shifted. Thus, from equation (14) the intensity at a point (x,y) can be conveniently written as

$$I_{D,l}(x,y) = \eta_0 \left\{ \frac{\sin[N(\Delta k \Gamma_S(x,y) + l\pi/2)/2]}{\sin[(\Delta k \Gamma_S(x,y)/2 + l\pi/2)/2]} \right\}^2 I_R \,, \tag{13}$$

where l = 0,1,2 and 3. Considering N = 2 longitudinal modes of the laser, the relative phase $\phi_S(x,y) \equiv \Delta k \Gamma_S(x,y)/2$ of the surface can be obtained as a function of the interferogram intensities I_{D0}, I_{D1}, I_{D2}, and I_{D3} according to (Barbosa et al, 2005)

$$\phi_{4-step}(N=2) = \arctan\left(\frac{I_{D3} - I_{D1}}{I_{D0} - I_{D2}} \right) \tag{14}$$

From equation (14) one gets a phase map in a grey level intensity, with the range $-\pi < \phi < \pi$ corresponding to 256 grey levels, from black ($\phi = -\pi$) to white ($\phi = \pi$).
The arctan function has a period of 2π rad and therefore it is not possible do directly determine the whole phase which exceed this range of $-\pi$ to π rad. For this reason the phase obtained from the equations above is said to be "wrapped". In order to unwrap the phase map – or make its deconvolution -obtained through equation (14) several methods were applied in the literature, like the cellular-automata (Ghiglia et al) and the branch-cut methods (Gutmann & Weber, 1999).

3.3.1 Example of shape measurements

Profilometry by four stepping method – figure 4 shows a typical dual-beam optical setup employed for holographic profilometry experiments. The reference and the object beams

interfere at the BTO crystal after passing through the polarizer P1. The object is imaged onto the BTO crystal by lens L1, while the reconstructed object wave is collected by lens L2 to build the holographic object image at the CCD camera. The crystal is cut in the [110] transverse electro-optic configuration in order to take advantage from the anisotropic diffraction properties of the sillenite crystals as described in section 2. The transmitted object beam is thus blocked by analizer P2 so that only the holographic image is displayed in a computer monitor for further processing. The 90º-prism PR is attached to a micrometer and translated in order to introduce the phase shift in the reference beam for the four-stepping procedure.

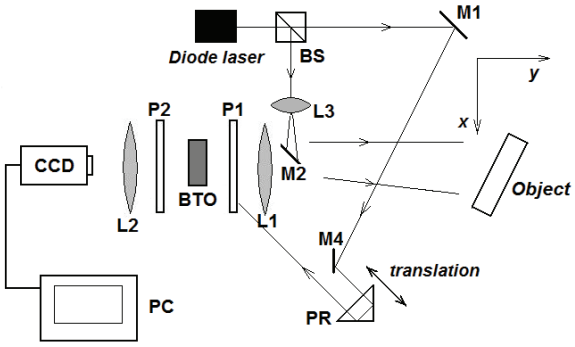


Fig. 4. Holographic setup. L1 to L3, lenses; M1 to M4, mirrors; BS, beam splitter; P1 and P2, polarizers; BTO, holographic storage medium; PR, 90º-prism; CCD, camera; PC, computer

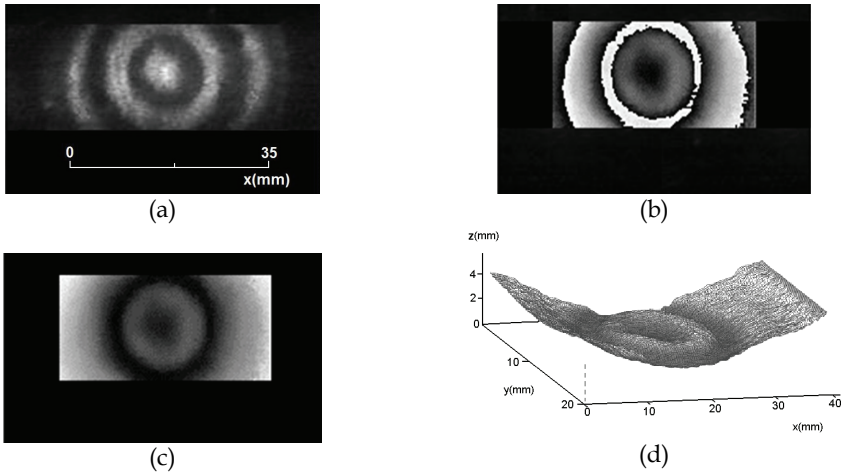


Fig. 5. Profilometric analysis of a still loudspeaker. (a) - Interferogram; (b) - phase map; (c) - unwrapped phase pattern and (d) - three-dimensional plot (Barbosa et al, 2005).

Since both the beam which illuminates the object surface and the beam scattered by it propagate in nearly opposite directions, the planes of constant elevation are nearly perpendicular to both beams and the contour interval (i.e., the distance between two

adjacent planes of constant elevation) is thus $\Delta z \cong \lambda_S/2$. Figure 5a shows the holographic image of a partially illuminated still loudspeaker for N = 3. The spatial frequency of the contour fringes is due to the laser free spectral range of $\Delta\nu = 53\,\mathrm{GHz}$, corresponding to $\Delta z = 2.66\ mm$, for a central emission wavelength of 670 nm. In the four-stepping process the translation stage was displaced by $\Delta z/4 = 0.66$ mm , resulting the phase map shown in figure 5b. The unwrapped phase pattern obtained through the cellular-automata method is shown in figure 5c, while the three-dimensional reconstruction of the plate is depicted in figure 5d, with an enlarged scale in the z-direction.

3.4 Methods for enhancing the measurement sensitivity

It has been mentioned in the previous sections that the contour interval Δz equals the laser resonator length L when the illuminating and the scattered beams propagate in opposite directions. If this property on the one hand make the diode lasers quite suitable in the profilometry of objects with several millimetres to few centimetres of size, on the other hand it may be a limitation remarkably in the measurement of nearly flat, low-derivative surfaces, which require smaller Δz values in order to minimize the measurement noise. The next sections present the most employed methods to overcome this limitation and to obtain interferogram spatial frequencies which are higher than the ones usually obtained through a single multimode laser. Those procedures aim more accurate visual inspections and the decrease of the measurement noise.

3.4.1 Recording with two phase-shifted reference beams

Holographic processes using lasers emitting more than two longitudinal modes results in bright interference fringes which are narrower than the $\cos^2$-fringes obtained in two-wavelength techniques. One can take advantage from this property in order to obtain holographic images by using optical setups with two properly shifted reference beams. In multi-wavelength holographic processes generated by more than two wavelengths (N>2 in equation 11), the resulting bright fringes are sufficiently narrow to allow the superposition of two interferograms and generate a third one. Let us consider the incidence of two reference beams 1 and 2 with intensities I_{R1} and I_{R2}, such that I_{R1}=I_{R2}=I_R/2. The resulting holographic image of the object in this case is the incoherent sum of two interferograms, each one generated by the interference of the object wave with one of the reference waves. The intensity of the interferograms superposition can be written as (Barbosa et al, 2005)

$$I_D^{'} = 4\eta_0\{4\cos^2[\Delta k_{eff}(\Gamma_S - \Gamma_{R1})/2] + 1\}I_R\,, \tag{15}$$

where $\Delta k_{eff} \equiv 2\pi/\lambda_{eff}$ can be regarded as an effective wavelength interval and $\lambda_{Seff} = \lambda^2/(2\Delta\lambda) = \lambda_S/2$ is the effective synthetic wavelength. The fact that $\lambda_{Seff} = \lambda_S/2$ implies that the interferogram generated by the superposition of both fringe patterns has twice their spatial frequency. Figure 6a shows the intensity distribution of the interferogram generated by a single reference beam for N = 3 laser modes as a function of the object coordinate Γ_S (taking Γ_R arbitrarily as zero) according to equation (11), while figure 16b shows the fringe pattern generated by two reference beams according to equation (15) with optical paths adjusted so that $\Gamma_{R2} - \Gamma_{R1} = q\lambda_{Seff}$, (q = 1,2, 3...). Both figures have the same Γ_S-scale, showing that the second interferogram has twice the spatial frequency than that of the

first one. Consequently, the profilometry sensitivity is twice enhanced when the optical paths of the two reference beams are properly adjusted. Figure 6c shows the holographic image the loudspeaker studied in section 3.3.1 with two reference beams. The higher spatial frequency of the resulting interferograms allows for a more detailed surface visualisation, leading to a more accurate qualitative analysis.

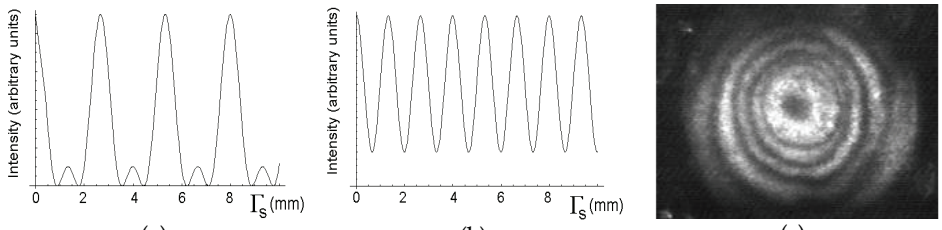


Fig. 6. - Interferogram intensity distribution as a function of Γ_S for holographic recording with (a) - one reference beam and (b) - two reference beams; (c) – Holographic image of the loudspeaker generated by a two-reference beam setup (Barbosa et al, 2005)

3.4.2 Laser mode selection by extra-cavity Fabry-Perot etalons

This section shows how the interferogram spatial frequency can be increased by placing an extracavity Fabry-Perot étalon (FBE) at the laser output. If the FBE is placed at the laser output and conveniently aligned a small fraction of the output beam is reflected back into the laser diode. Only the modes which are re-injected will be amplified while the oscillation of the others will vanish. This increases the effective free spectral range of the emission, thus raising the spatial frequency of the contour interferogram. After positioning the FBE at the laser output the synthetic wavelength Λ_S upon mode selection is obtained as (Barbosa et al, 2007):

$$\Lambda_S = \frac{1}{m}\,\frac{\lambda^2}{\Delta\lambda} = \frac{\lambda_S}{m}\,, \tag{16}$$

Equation (16) shows that, by conveniently adjusting the FBE spacing, the synthetic wavelength Λ_S achieved by mode selection can be progressively scaled to smaller values. This scaling is limited by the number of longitudinal modes of the laser emission. If $m = 1$, all modes are reflected back into the laser resonator, so there is no mode selection and thus no changes in the interferogram are observed. If $m > 1$ only some modes are reinforced, and the resulting free spectral range of the laser illuminating the holographic setup increases, thus increasing the interferogram spatial frequency and enhancing whether the measurement sensitivity or the profilometry resolution.

Figure 7a shows the holographic image of a flat tilted plate with a multimode diode laser without FBE. The holographic medium is a BTO crystal cut in the electrooptic transverse configuration. In this case, $\Lambda_S = \lambda_S = 5.30$ mm and $\Delta z = 2.65$ mm. Figure 7b compares the z-coordinate of a cross section of the plate reconstructed from the unwrapped with the expected profile of the plate. Figure 7c shows the contour fringe pattern of the same plate, at the same position, with the FBE adjusted for $m = 3$, leading to a synthetic wavelength

$\Lambda_S = 1.76$ mm and a contour interval $\Delta z = 0.88$ mm, while figure 7d shows the resulting intensity profile of the same cross section of figure 7b.
By comparing figures 7b and 7d one comes to the conclusion that smaller contour intervals provide less noisy measurements as expected. The object profile of figure 7d presents smaller rms deviations from the expected plate shape if compared to the curve of figure obtained without mode selection. For both cases, the measurement fluctuations are ~1/20th of a fringe.

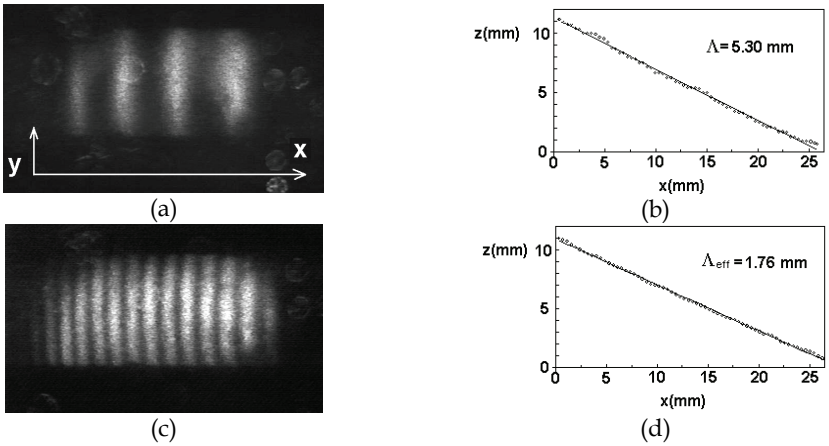


Fig. 7. Profilometry of a tilted plate. (a) – Interferogram for $\lambda_S = 5.30$ mm and $\Delta z = 2.65$ mm; (b) – corresponding profile; (c) - Interferogram for $\lambda_S = 1.76$ mm and $\Delta z = 0.88$ mm; (d) – profile of the same section (Barbosa et al, 2007).

4. Two-laser holography in volume media

In this section the theory and the experiments concerning the employ of two tunable lasers in holographic wavefront reconstruction are exposed and discussed. In order to obtain smaller values of the contour interval Δz and improve the measurement sensitivity, the central wavelengths of both lasers are tuned allowing the simultaneous illumination of the object with two slightly different wavelengths thus providing a very wide range of Δz values. From the typical values of contour intervals in the order of few millimetres obtained through single-multimode lasers, experiments using two-laser schemes can decrease these values down to 40 μm. Since the fringe evaluation techniques exposed in section 3.3.1 achieve measurement precisions up to 1/100 of a fringe in low optical noise experiments, the two-laser holographic processes represent a great sensitivity and precision improvement in wavefront reconstruction through heterodyne holography.
The two-laser optical setup using also a BTO crystal as storage medium shown in figure 8 is essentially the same of that shown in figure 4, except by the arrangement at the beam splitter BS which couples the beams coming from lasers 1 and 2. The arm of the interferometer containing the transmitted reference beam is sent to a spectrum analyzer with a 1200-lines/mm diffraction grating for wavelength monitoring. Mirror M4 can be translated and tilted in order to spatially couple the beams originated from both lasers.

Consider that the diode lasers emit at the wavelengths λ_1 and λ_2 such that $|\lambda_1 - \lambda_2| << \lambda_1$. The following analysis takes into account the influence of the lasers detune and the misalignment of the beams on the generation of interference contour fringes on the object image. If $\Gamma_{S1,2}$ and $\Gamma_{R1,2}$ are respectively the optical paths of the object and the reference beams with wavelength $\lambda_{1,2}$ from the coupling beam splitter to the crystal, the waves impinging the BTO crystal can then be written as (Barbosa & Carvalho, 2007)

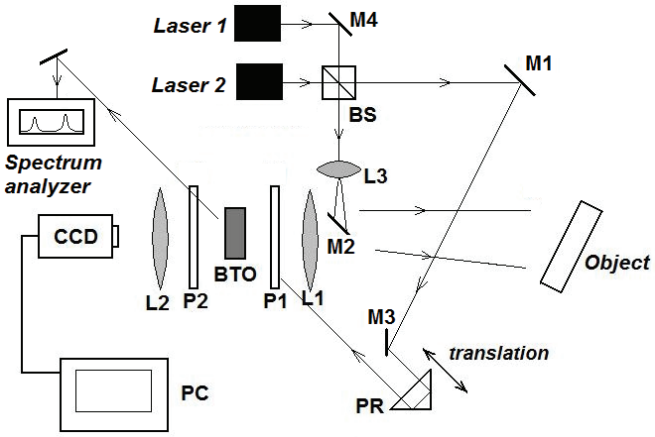


Fig. 8. Two-laser holographic interferometer

$$R = R_0\left(e^{i(k_1\Gamma_{R1}+\phi_1)} + e^{i(k_2\Gamma_{R2}+\phi_2)}\right)$$
$$S = S_0\left(e^{i(k_1\Gamma_{S1}+\phi_1)} + e^{i(k_2\Gamma_{S2}+\phi_2)}\right) \quad (17)$$

where $k_{1,2} = 2\pi/\lambda_{1,2}$. The distinction between $\Gamma_{S(R)1}$ and $\Gamma_{S(R)2}$ takes into account an eventual (and sometimes quite useful) beam misalignment at the beam splitter output. Due to the interaction of waves with two different wavelengths, there are two recorded holograms in the crystal. According to equations 10a-c, the diffraction of reference wave 1 by both holograms provides the holographic reconstruction of the object wavefront given by

$$E_{D1} = \sqrt{\eta_0} R_0\left[e^{ik_1\Gamma_{S1}} + \chi_{1,2}\ e^{i(k_1\Gamma_{R1}+k_2\Gamma_{S2}-k_2\Gamma_{R2})}\right] \quad (18)$$

The holographic reconstruction of the object occurs correspondingly for the diffraction of reference wave 2, resulting in the wave amplitude E_{D2}:

$$E_{D2} = \sqrt{\eta_0} R_0\left[e^{ik_2\Gamma_{S2}} + \chi_{1,2}\ e^{i(k_2\Gamma_{R2}+k_1\Gamma_{S1}-k_1\Gamma_{R1})}\right] \quad (19)$$

where $\chi_{1,2} = \sin\xi_{1,2}/\xi_{1,2}$, $\xi_{1,2}$ being given by (Kogelnik, 1969)

$$\xi_{1,2} = 2\pi l_0 \sin\gamma\left[\frac{\Delta\alpha}{\sqrt{\lambda_1\lambda_2}} + n_0 \tan\gamma\frac{\Delta\lambda}{\lambda_1\lambda_2}\right] \quad (20)$$

The angle $\Delta\alpha$ accounts for the misalignment between the beams at the BS output and $\Delta\lambda \equiv \lambda_1 - \lambda_2$. In single multimode lasers, $\Delta\lambda$ is the wavelength gap between two adjacent modes. $\Delta\alpha$ is small enough to allow the superposition of the waves as they propagate from the BS to the holographic medium. In equations (18) and (19) the phases ϕ_1 and ϕ_2 at the lasers output were set arbitrarily to zero.
Both waves E_{D1} and E_{D2} are incoherent and combine at the crystal output as a sum of their intensities, thus providing the intensity of the holographic object image:

$$\begin{aligned} I_D \propto |E_{D1}|^2 + |E_{D2}|^2 &\propto R_0^2\eta_0\left[1+\chi^2+2|\chi|\cos\left(k_1\Gamma_{S1}-k_1\Gamma_{R1}-k_2\Gamma_{S2}+k_2\Gamma_{R2}\right)\right] \\ &= R_0^2\eta_0\left(1+\chi^2\right)\left[1+V\cos\left(k_1\Gamma_{S1}-k_1\Gamma_{R1}-k_2\Gamma_{S2}+k_2\Gamma_{R2}\right)\right] \end{aligned} \tag{21}$$

where $V \equiv 2|\chi|/\left(\chi^2+1\right)$ is the interferogram visibility. Notice that since the output power is the same for both lasers, V refers only to the Bragg regime degree. By analyzing the phase in the sinusoidal term of equation (21), one retrieves information about the object surface, the position of the constant elevation planes and their relative distances as well as the interferogram localization.

4.1 Interferogram positioning and distance between planes

Consider the incidence of the collimated illuminating beam onto the analyzed surface impinging points A and B as shown in figure 9. The angle between the incident beam and the normal direction with respect to the photorefractive sample front face (plane xz) is $\alpha+\Delta\alpha$ for beam 1 and α for beam 2. The phase difference φ between two points A and B on the object surface is obtained with the help of equation (21) :

$$\phi = k_1\Gamma_{S1B} - k_2\Gamma_{S2B} - k_1\Gamma_{S1A} + k_2\Gamma_{S2A}, \tag{22}$$

where $\Gamma_{S1(2)A}$ and $\Gamma_{S1(2)B}$ are the optical paths of the object beam at wavelength $\lambda_{1(2)}$ through A and B, respectively. In the equation above, the influence of the beams misalignment in the reference-beam arm on the fringe generation can be considered negligible. According to figure 9, the phases of equation (22) can be explicitly written as a function of the unit propagation vectors of the illuminating and the scattered beams $\hat{k}_1$, $\hat{k}_2$ and $\hat{k}$ as $\Gamma_{1SA} = \hat{k}_1.\vec{r}_{1A} + \hat{k}.\vec{r}_A$, $\Gamma_{2SA} = \hat{k}_2.\vec{r}_{2A} + \hat{k}.\vec{r}_A$, $\Gamma_{1SB} = \hat{k}_1.\vec{r}_{1B} + \hat{k}.\vec{r}_B$ and $\Gamma_{2SB} = \hat{k}_2.\vec{r}_{2B} + \hat{k}.\vec{r}_B$.

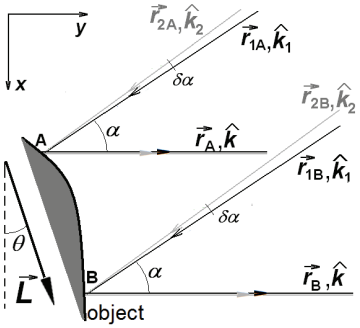


Fig. 9. Incidence of the waves onto the surface.

Vectors $\vec{r}_{1A(B)}$, $\vec{r}_{2A(B)}$ represent the displacement of the waves with wavelengths λ_1 and λ_2, respectively, through point *A* (B), while $\vec{r}_A$ and $\vec{r}_B$ are the displacement vectors from points *A* and *B* to the BTO crystal front face. Vector $\vec{L}$ is the position vector of point *B* with respect to *A*. By examining figure 9 one obtains the relations $\hat{k}_1.(\vec{r}_{1A}-\vec{r}_{1B})=\hat{k}_1.\vec{L}$, $\hat{k}_2.(\vec{r}_{2A}-\vec{r}_{2B})=\hat{k}_2.\vec{L}$ and $\hat{k}.(\vec{r}_B-\vec{r}_A)=\hat{k}.\vec{L}$. From equation (22) and the relations above φ is written as (Barbosa, 2010)

$$\varphi = 2\pi\left(\frac{\hat{k}_1}{\lambda_1}-\frac{\hat{k}_2}{\lambda_2}-\frac{\hat{k}}{\lambda_S}\right).\vec{L} \tag{23}$$

where $\lambda_S \equiv \lambda_1\lambda_2/(\lambda_2-\lambda_1)$ is the synthetic wavelength for the two-laser holographic process. Depending on the lasers tuning, the synthetic wavelength can be whether positive or negative. According also to figure 9, the vectors in equation (23) can be written as a function of α and θ as $\hat{k}_1=(\sin\alpha,-\cos\alpha)$, $\hat{k}_2=(\sin(\alpha+\delta\alpha),-\cos(\alpha+\delta\alpha))$, $\hat{k}=(0,1)$ and $\vec{L}=L(\sin\theta,\cos\theta)$.

The phase obtained in equation (23) describes the generation of contour fringes due to two-laser holography. With the help of this equation it is also possible to determine important parameters of the surface measurement, like the direction of the constant elevation planes, the distance between two adjacent planes and the position of the contour interferogram with respect to the object surface.

If points A and B lay on the same plane of constant elevation (i.e., on the same bright/dark fringe), thus $\varphi=0$ and from equation (23) one obtains the direction of the elevation planes given by angle θ:

$$tg\theta = \frac{\sin\alpha-\Im\cos\alpha}{2\cos^2\left(\frac{\alpha}{2}\right)+\Im\sin\alpha}, \tag{24}$$

where $\Im \equiv \lambda_S\,\Delta\alpha/\lambda_2$. Equation (24) shows that the direction of the planes of constant elevation with respect to the *x*-axis depend not only on the illumination angle, but also on the lasers detune and the misalignment $\Delta\alpha$. Notice that, when $\Delta\alpha=0$, thus $\theta=\alpha/2$, i.e. the constant elevation planes are perpendicular to the bisector formed by the illuminating beam and the beam scattered by the object, a well-known result in conventional two-colour holography and used in the previous sections. As $\Delta\alpha$ increases, however, the planes rotate and may significantly change the contour interference pattern, remarkably for larger synthetic wavelengths.

It should be also noticed that the wavelengths detune and the misalignment $\Delta\alpha$ cannot be set at arbitrary values. As exposed in section 2, the limitations imposed by the Bragg regime must be necessarily considered and thus there must be a trade-off between all the parameters in order to obtain acceptable interferogram visibilities. Concerning the Bragg regime in the recording and readout, the terms $\Delta\alpha$ and $\lambda_2-\lambda_1$ are related to each other according to equation (20). The interferogram maximal visibility $V_{\max}=1$ obtained in perfect Bragg processes is achieved for $\xi=0$ and $\chi=1$. This tolerance for slight off-Bragg regimes may result in higher spatial frequency interferograms and make the conditions for beam alignment less severe. The term $\Im$ appears recurrently throughout this analysis and will deserve special attention, since it is closely related to the Bragg condition.

With the help of the first-order expansion $\chi = \sin\xi/\xi \cong 1 - \xi^2/6$, $\Im$ can be written as

$$\Im \cong \frac{\sqrt{6(1-\chi)}}{2\pi\ l_0 \sin\gamma}\lambda_S - \frac{\tan\gamma}{n_0} \tag{25}$$

where $(\lambda_1/\lambda_2)^{1/2}$ was considered to be ≈ 1. The term $\Im$ determines how the lasers can be tuned and how the beams can be aligned in order to combine desired values of contour intervals without significantly decreasing the interferogram visibility.

The distance between two consecutive planes of constant elevation can be obtained with the help of equation (23). Let us call $\vec{R}$ as the *sensitivity vector,* representing the distance between two consecutive planes. Since the phase difference between them is $\varphi = 2\pi$ rad, one obtains the relation

$$\left(\frac{\hat{k}_1}{\lambda_1} - \frac{\hat{k}_2}{\lambda_2} - \frac{\hat{k}}{\lambda_S}\right).\vec{R} = 1 \tag{26}$$

Since $\vec{R} \perp \vec{L}$, from equation (23) the modulus of $\vec{R}$ is obtained as

$$R = \left| \frac{\lambda_S}{\Im \sin(\theta - \alpha) - 2\cos\left(\frac{\alpha}{2}\right)\cos\left(\theta - \frac{\alpha}{2}\right)} \right| \tag{27}$$

In order to illustrate the influence of the lasers wavelengths on the position of the constant elevation planes, let us analyze the experimental situation in which the wavelengths λ_1 and λ_2 of lasers 1 and 2 behave in such a way that $\lambda_1 > \lambda_2$ is their whole tunability ranges. The aim of this analysis is to study the alignment requirements through which the constant elevation planes are parallel to the *x*-axis in figure 9, what is equivalent to say that the planes are parallel to the front face of the holographic medium. This configuration is particularly interesting in the study of surface geometries which present symmetry with respect to the *y*-axis, e.g. spherical or cylindrical surfaces, since the qualitative analysis may become much more intuitive.

Since $\lambda_1 > \lambda_2$, one gets $\lambda_S < 0$ for any tuning configuration. Hence, from the definition of $\Im$ one concludes that equation (25) is only fulfilled if $\Delta\alpha \geq 0$, and consequently $\Im \leq 0$. This imposes limitations to the setup geometry.

The requirement that the planes of constant elevation must be parallel to the *x*-axis and perpendicular to the *y*-axis is achieved for $\theta = 0$, or $\vec{L}$ perpendicular to the *yz*-plane. This is equivalent to say that vector $\Delta\hat{k} = \hat{k}_1/\lambda_1 - \hat{k}_2/\lambda_2$ must be parallel to the *y*-axis. From equation (24) this requirement is fulfilled when

$$\Im = \tan\alpha \tag{28}$$

In the illumination scheme of figure 10a the desired plane direction cannot be achieved because in this case $\alpha > 0$, which is in disagreement with equation (28). Notice that vector $\Delta\hat{k}\ (\perp \vec{L})$ cannot be made perpendicular to the *y*-axis, unless $\lambda_2 > \lambda_1$, which is impossible to achieve in those experimental conditions. Figure 10b shows the contour interferogram on a holographic image of a spherical surface generated through this illumination scheme with a positive illumination angle and a negative synthetic wavelength. The excentrical,

elliptical-like contour fringes clearly evidence that the interferogram was a result of the intersection of the surface with elevation planes tilted with respect to x-axis ($\theta \neq 0$). This interferogram assimetry cannot be removed upon laser tuning or further beam alignment unless the fringe visibility strongly decreases. In these figures, λ_S = -480 μm.
On the other hand, in the arrangement of figure 10c the illumination angle was inverted with respect to the y-axis, while the relative angular positions of vectors $\hat{k}_1$ and $\hat{k}_2$ remained unchanged. In this second case it becomes clear that for certain values of the wavelengths the vector $\Delta\hat{k}$ can be made parallel to the y-direction, as evidenced by the circular and concentric contour fringes of the same spherical surface shown in figure 10d.

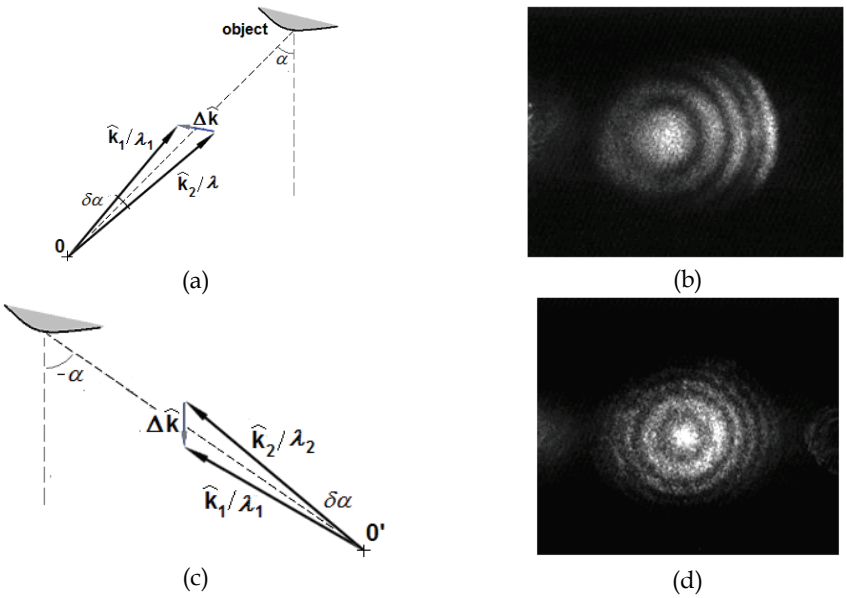


Fig. 10. (a) - Illumination scheme for $\alpha > 0$ showing the vector $\Delta\vec{k}$; (b) - contour interferogram of the surface obtained with positive α; (c) - illumination scheme for $\alpha < 0$; (d) – resulting contour interferogram of the surface with circular concentric fringes (Barbosa, 2010).

4.2 Interferogram localization

In this section, it is assumed that the interferogram is not formed at the object surface, but somewhere else near this region. Consider a point Q belonging to the region where the fringe pattern is formed; the total optical wave at Q is the contribution from the waves scattered by the points at the vicinity of a point P which in turn lays on the surface. The region surrounding P is the base of a cone with the apex at Q and whose area is directly proportional to the optical aperture of the imaging lens. The condition for fringe formation at point Q requires that the fringe visibility is a maximum, which implies that the phase variation at this point is a minimum. The interferogram intensity at a point P on the object can be attributed to its phase with respect to a point O on a reference constant elevation plane. Since points O and P define a vector $\vec{H}$ parallel to the sensitivity vector, it is convenient to choose a new coordinate system with axes x' and y' perpendicular and

parallel to vector $\vec{H}$, respectively, as shown in figure 11. Hence, the contour fringe at Q with phase δ is formed if $\partial\delta/\partial x' = 0$ (Vest, 1979).
Considering points P, O and Q in the new coordinate system as having coordinates (x', H), $(x', 0)$ and (X', Y'), respectively, the phase at Q is determined analogously as in the previous section according to the scheme of figure 11:

$$\delta \cong 2\pi\left(\frac{\hat{k}_2'}{\lambda_2} - \frac{\hat{k}_1'}{\lambda_1} - \frac{\hat{k}_3'}{\lambda_S}\right).\vec{H} \tag{29}$$

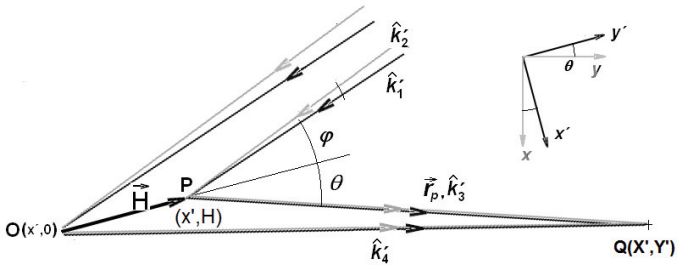


Fig. 11. Incident collimated beams and beams scattered by the surface at point P.

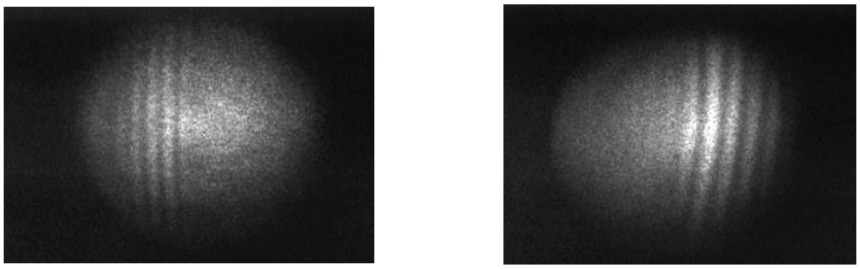
where $\hat{k}_1'$ and $\hat{k}_2'$ are the propagation vectors of the illuminating beams. Equation (29) was obtained considering a small incidence angle α of the illuminating beam onto the object surface (as in the experimental setup) and the fact that $\hat{k}_3' - \hat{k}_4'$ and $\vec{r}_p$ are nearly orthogonal. $\hat{k}_3'$ and $\hat{k}_4'$ are the propagation vectors of the scattered beams (see figure 11), and $\vec{r}_p$ the position vector of P relatively to Q. According to figure 11 the position vector $\vec{H}$ and the normalized propagation vectors are written as $\vec{H} = (0, H)$, $\hat{k}_1' = (\sin\alpha', -\cos\alpha')$, $\hat{k}_2' = \left(\sin(\alpha' + \Delta\alpha), -\cos(\alpha' + \Delta\alpha)\right)$ and $\hat{k}_3' = (X' - x', Y' - H)\kappa^{-1}$, where $\kappa \equiv \sqrt{(X' - x')^2 + (Y' - H)^2}$, $\varphi = \alpha - \theta$ and $H = H(x')$.
After some algebraic manipulation the position Y' where the fringe pattern is formed can be obtained from equation (29) as

$$Y' = \left\{1 + \frac{\frac{\sin 2\theta}{2}\left(\sin\theta\frac{\partial H}{\partial x'} - \cos\theta\right)}{\left[\Im\sin(\alpha - \theta) + 2\cos\left(\frac{\alpha}{2}\right)\cos\left(\theta - \frac{\alpha}{2}\right)\right]\frac{\partial H}{\partial x'}}\right\} H(x') \tag{30}$$

Equation (30) provides the relative position of the interferogram with respect to the object surface. When the expression in the brackets equals 1, then $Y' = H(x')$, and the fringe pattern occurs right on the object surface. As the expression in the brackets assumes values smaller than 1, the fringe pattern is formed behind the object; otherwise, the interferogram is formed between the object surface and the imaging lens. Equation (30) shows also that the slope $\partial H/\partial x'$ of the studied surface and the illumination angle α determine whether the fringes are formed behind or in front of the surface.

Since $\left|\Im\sin(\alpha-\theta)\right| < 2\cos(\alpha/2)\cos(\theta-\alpha/2)$ in typical experimental conditions, one notes also from equation (30) that the fringe pattern is formed behind the surface ($Y' < H(x')$) for positive surface slopes ($\partial H/\partial x' > 0$). Correspondingly, if $\partial H/\partial x' < 0$, one gets $Y' > H(x')$, and the interferogram is formed between the surface and the imaging lens.

The holographic images of the "V"-shaped surface constituted of two concurrent planes shown in figure 12 confirm those predictions, for $\lambda_S = -420$ μm, $\Delta\alpha = 2\times10^{-4}$ rad and $\alpha = 5\times10^{-2}$ rad. The two halves of the surface present slopes of different signals. In figure 12a, while the interferogram on the left side with negative slope presents a good visibility, the fringe pattern on the right side with positive slope appears completely blurred. On the other hand, in figure 12b, when the interferogram of the right half of the surface is focused, the contour fringes of the left half cannot be visualized.



(a) (b)

Fig. 12. Two-wavelength holographic imaging of the V-shaped surface by focusing the fringe pattern of (a) – the left half ($x' > 0$ and $\partial H/\partial x' < 0$) of the surface, and (b) - the right half of the surface ($x' < 0$ and $\partial H/\partial x' > 0$) (Barbosa, 2010)

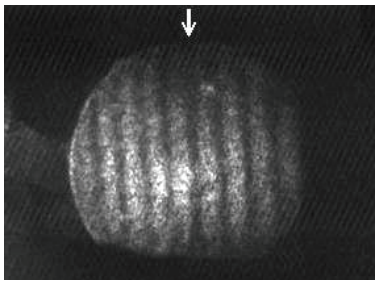


Fig. 13. Two-wavelength contour interferogram of the whole V-shaped surface for $\theta = 0$. The arrow shows the limit between the negative and the positive slope of the surface (Barbosa, 2010).

The limitation of focusing the whole contour interferogram of a surface with slopes of different signals can be overcome by adjusting the setup in order to set the direction of the constant elevation planes as $\theta = 0$. According to equations (30), in this case one gets $Y' = H$, and the fringe pattern is formed right on the surface, regardless its geometry. This configuration does enable and easier and more intuitive visual inspection of the surface shape. The contour interferogram of the same "V"-shaped object of Figure 12 was studied under the same value of $\Im$; the illumination angle α in turn was set at negative values in

order to obtain null values of θ, according to equations (24) and (28). Figure 13 shows the resulting contour fringe pattern with the arrow indicating the edge position which separates the negative slope from the positive slope of the surface. As expected, the contour fringes pattern can be clearly visualized along both sides of the surface, regardless the aperture of the optical system.

5. Further applications

5.1 Holographic refractometry

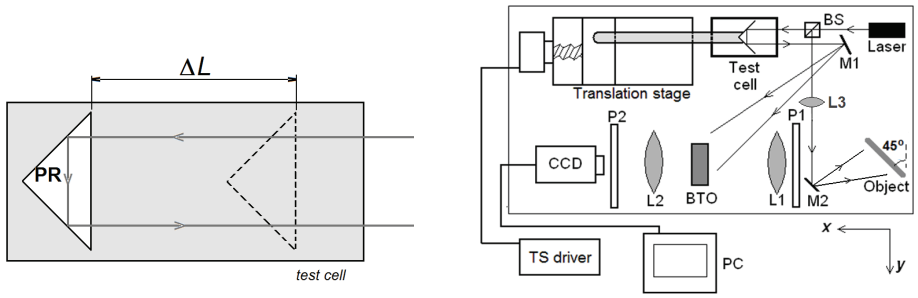
Many refractometry methods and devices have been developed for material characterization. Refractometry is useful for identifying unknown liquid compounds, to evaluate their purity, or to determine the solute concentration in solutions by comparison with standard curves. Most of the techniques make use of purely geometrical optics phenomena like ray deviation or total internal reflection (McAlister et al, 1956). Interferometric (Santiago Neto et al, 2004), holographic (Gustafsson & Sebesta, 2004) and other methods combining ray optics with diffractive components (de Angelis et al, 1996) have been also presented. Multi-wavelength holography with photorefractive materials can be successfully employed for liquid and amorphous material characterization through refractometry. The principle of holographic refractometry is based on the measurement of fringe displacement generated by the variation on of the optical path inside the test cell containing the sample.

The optical setup for liquid characterization is shown in figure 14. In this setup a simple flat plate tilted with respect the holographic medium is the object, as accomplished in the experiments described in section 3. However, in this holographic refractometer the reference beam passes through a glass cell filled with the liquid with a moveable reflector inside. By translating the reflector, the optical path inside the liquid is changed, thus changing the optical path of the reference beam. According to equation (11), this shifts the parallel and straight contour fringes formed on the holographic image of the tilted plate. This can be described in more detail with the help of figure 14a, which shows the reference beam propagation through the test cell. Inside the glass cell and immersed in the liquid, the prism 90°-prism PR is used as a retro-reflector. If PR is translated by ΔL, the optical path change into the reference beam is $2n\Delta L$. According to equation (11), this leads to the following expression for the refractive index:

$$n = \frac{\lambda_S p}{2\Delta L}, \tag{31}$$

where p is the number of running fringes across the object image. This experimental configuration features some important characteristics. The reduced rectangular cross section of the test cell requires relatively small amounts of liquid even for somewhat large optical paths of the reference beam; the small diameter of the collimated reference beam propagating through the liquid allows a high light intensity at the holographic medium in spite of moderate scattering and absorption losses in the liquid. This makes the requirements for flatness and thickness uniformity of the cell walls much less severe. Moreover, the scalableness of the measurement precision is other interesting property of this refractometer: several reflector configurations can be employed in order to enlarge the effective optical path of the reference beam through the liquid without significantly enlarging the liquid quantity.

The optical scheme designed for a compact and portable holographic refractometer prototype using a BTO crystal as the storage medium and a single multimode diode laser is depicted in figure 14b (Barbosa et al, 2010). The prism PR is supported by a mount from the top of the test cell, in a way that there is no mechanical contact between the PR mount and the cell walls. A 5-μm resolution step motor and a translation stage move the reflector mount along the *x*-axis in a closed loop. The synthetic wavelength in this case was measured to be λ_S = 6.6280 ± 0.0005 mm and the test cell of this prototype permitted a maximum PR translation of ~70 mm. Hence, when the test cell is empty and the PR suffers a complete translation, one observes the displacement of 21 contour fringes through a reference point on the object image. If the test cell is filled e.g. with pure water (n_{water}=1.3317) then one counts 28 running fringes for nearly the same prism displacement, and so forth.



(a) (b)

Fig. 14. (a) - 90°-prism retro reflector PR in test cell; (b) - Optical setup for the holographic refractometer : BTO, $Bi_{12}TiO_{20}$ crystal; BS, beam splitter; M1 and M2, mirrors; P1 and P2, polarizers; L1 to L3, lenses; translation stage TS driver , CCD camera and PC computer for image acquisition (Barbosa et al, 2010).

The setup of figure 14b allows the use of a second laser for two-laser holographic recording measurements. The consequent generation of significantly smaller synthetic wavelengths in this case enables the use of much smaller liquid quantities without decreasing the measurement precision, which is a very desirable condition for the analysis of turbid liquids. As an example, the refractive index measurement of a sample of a highly scattering artificial melon juice is briefly described: taking into account scattering and absorption, the total loss on the reference beam due to the juice was 7.5 %, corresponding to a loss coefficient for red light of ≈ 0.13 cm^{-1}, according to Beer´s law. By properly tuning the lasers to obtain a wavelength difference of $\lambda_1 - \lambda_2$ = 0.56 nm a synthetic wavelength of $\lambda_S = 788.81 \pm 0.3$ μm was generated. This λ_S-value was small enough to allow a fringe displacement of 18 fringes upon a translation of only (if compared to the 70-mm translation in the single multimode laser case) 5.33 ± 0.01 mm of the prism reflector. According to equation (31), this successful measurement resulted in a refractive index of 1.331 ± 0.002 .

5.2 Lens characterization

Several methods were proposed and developed for lens characterization, whether by purely geometrical optics or by interferometry. Moiré interferometry (Keren et al, 1988), diffraction gratings (de Angelis et al, 1997), lithographically recorded Fresnel Zone holograms (DeBoo

& Sasian, 2003) and digital holography (Anand & Chhaniwal, 2006), among other methods, were successfully applied for lens characterization. The common feature of those works is the fact that they are restricted to focal length measurement only. In order to provide a more complete study of the lens, this section is devoted to describe how the wavefront reconstruction methods exposed in section 3 and the refractometry techniques analyzed in the previous section can be combined in order to integrally characterize several types of spherical and toric lenses. For this reason, this technique of lens characterization is able to measure all their refractive and geometric parameters through two-laser holography. This is accomplished by reconstructing three wavefronts: the surfaces of both sides of the lens – which provides their radii of curvatures - and the shape of the wave that propagates through the lens, impinges a opaque and diffusely scattering plane surface and travels back again through the lens. This latter measured wavefront is used to determine the lens refractive index by adjusting the experimental data to the corresponding wavefront calculated through exact ray tracing. This calculation is performed as a function of the radii of curvature, the refractive index and the thickness of the lens.

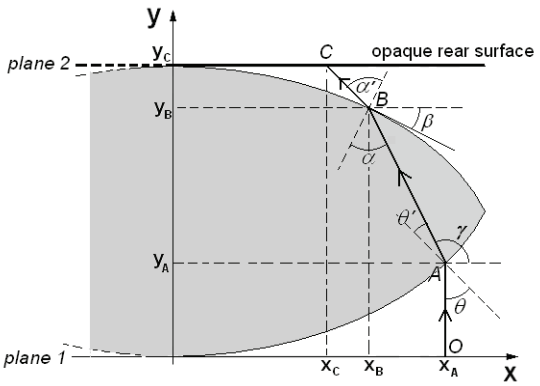


Fig. 15. Beam propagation through the spherical test lens

Figure 15 shows the scheme of the test lens with refractive index n, thickness t and radii of curvature R_F and R_R and the optical path of the incident beam. The beam travels from plane 1 to plane 2 and back to plane 1 through the spherical lens. The functions which describe the y-coordinate of the front and the rear surfaces in the xy-plane are given by $y_F(x) = R_F - \sqrt{R_F^2 - x^2}$ and $y_R(x) = t - R_R + \sqrt{R_R^2 - x^2}$, respectively. In the experiments the opaque plane coinciding with plane 2 and perpendicular to the optical axis y was placed right behind the rear face of the test lens. The entrance pupil of the imaging lens is much smaller than the distance between the imaging lens and the test lens, thus, the total optical path Γ of the wave corresponds to twice the path Γ_{1-2} introduced when the wave propagates from plane 1 to plane 2 only.

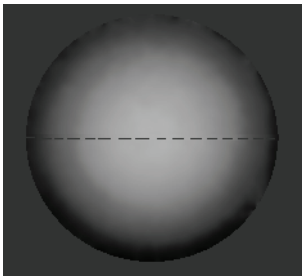
Since the input is a plane wave, the incident beam is parallel to the y-direction and impinges the front surface of the lens at point A of coordinates (x_A, y_A). The total optical path of the light wave is then written as a function of the segments OA, AB and BC according to

$$\Gamma_{1\to 2} = (OA + nAB + BC) \tag{32}$$

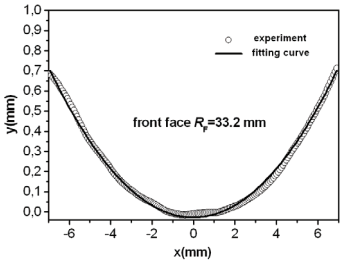
where $OA = y_A = R_F - \left(R_F^2 - x_A^2\right)^{1/2}$. The calculation of AB and BC is lengthy but conceptually simple: by taking the derivative dy_F/dx for $x = x_A$ the tangential and the normal curves to the front face are determined at point (x_A, y_A). This provides angle θ and through Snell's law angle θ' (see figure 15) can be determined. The curve describing the beam propagating inside the lens can be written as a function of the angle $\gamma = \theta - \theta' + \pi/2$ shown in figure 15 according to $y_h(x) = x\tan\gamma - x_A\tan\gamma + y_A$. The coordinates (x_B, y_B) of point B can be obtained from the relation $y_h(x) = y_R(x)$. Hence, AB is determined as $AB = \left[\left(x_B - x_A\right)^2 + \left(y_B - y_A\right)^2\right]^{1/2}$. By proceeding analogously at the lens rear surface one obtains straightforwardly the coordinates (x_C, y_C) of point C, providing the length BC, given by $BC = \left[\left(x_C - x_B\right)^2 + \left(y_C - y_B\right)^2\right]^{1/2}$. The total optical path of the beam in the xy-plane is thus given by

$$\Gamma_{1\to 2} = y_A + n\left[\left(x_B - x_A\right)^2 + \left(y_B - y_A\right)^2\right]^{1/2} + \left[\left(x_C - x_B\right)^2 + \left(y_C - y_B\right)^2\right]^{1/2} \tag{33}$$

By comparing the experimentally reconstructed wavefront profiles with the optical path calculated in equation (33), the refractive index of the lens can be determined. The two-laser experimental apparatus is essentially the same as that of figure 8. The following figures refer as to the analysis of a 5.9-mm thick spherical lens made of BK7 glass with a diameter of 17 mm. In this technique the thickness is the only parameter obtained by conventional, i.e., non-optical means. In order to determine the radii of curvatures, each surface of this lens is covered by opaque films. The reconstructed wavefront after applying phase unwrapping procedures obtained with the contour interval Δz = 0.32 mm is shown in figure 16a, while the averaged y-coordinate (circles) of the lens surface taken from figure 16a along radial directions and the fitting curve $y(x) = R - (R^2 - x^2)^{1/2}$ (solid black curve) are shown in figure 16b. The radius of curvature of each surface is determined by fitting the experimental data with $y(x)$, providing the radius of curvatures R_F = 33.2 mm. By performing analogously with the rear face, one obtains R_R = 32.9 mm.



(a)



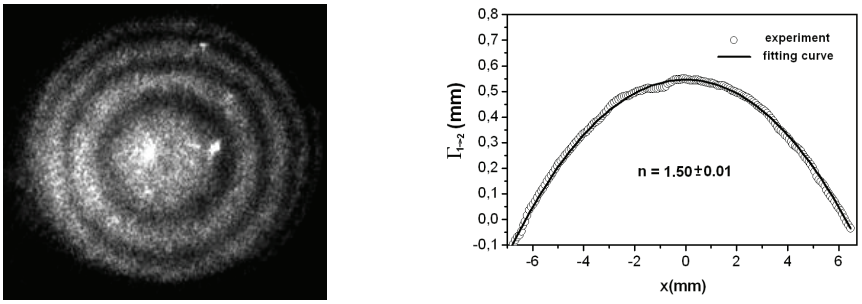
(b)

Fig. 16. Reconstruction of the front face of the lens; (a) – unwrapped phase; (b) - y-coordinate (circles) averaged along radial directions, and fitting curve (black curve) providing RF =33.2 mm (Barbosa et al, 2009)

In order measure the refractive index, the films are removed and the opaque surface is positioned behind the lens, so that the illuminating beam can propagate through the lens and hit the plate. The resulting contour fringe pattern of one of the four frames for Δz = 0.24 mm is shown in figure 17a, while figure 17b shows the averaged y-coordinate of the reconstructed wavefront. The thick black line refers to the measured coordinates while the thin grey one is the fitting of the experimental data with the Γ-values given by equation (35). The best fitting in this case is obtained for $n = 1.50 \pm 0.01$. The measurement error defined as $\delta(\%) = 100|n - n_E|/n_E$ was estimated to be ~ 0.7 %, considering the expected value for red light as n_E =1.514.
Through the well-know formula

$$\frac{1}{f} = (n-1)\left[\frac{1}{R_F} - \frac{1}{R_R} + \frac{t}{R_F R_R}\frac{(n-1)}{n}\right] \tag{34}$$

the focal length f can be determined. By using the experimental results R_F = 33.2 mm, R_R = 32.9 mm, t = 5.95 mm and n =1.50, the experimental focal length is determined to be 32.1 mm. The nominal focal length of the lens provided by the manufacturer is 32.4 mm, resulting in an experimental error of ~ 0.9 %.



(a) (b)

Fig. 17. (a) - Contour fringe pattern for Δz = 0.24 mm when the illuminating beam traveled through the lens; (b) - averaged y-coordinate of the reconstructed wavefront providing $n = 1.50 \pm 0.01$ (Barbosa et al, 2009).

6. Conclusion

Multi-wavelength holography has shown to be a powerful tool for optical metrology and wavefront reconstruction. Good visibility interferograms can be obtained in real-time processes, thus allowing faster and easier measurements with simple and compact optical setups. The profilometry techniques shown in this chapter presented the advantage of being nondestructive, thus allowing for the analysis of delicate surfaces, organic tissues or smalls objects, whose analysis by conventional, contacting means would not be viable. The whole-field character of holographic imaging provides also the additional advantage of the analysis of the whole surface, instead of lengthy point-by-point analyses.
Applications in refractometry and lensometry have shown the potentialities of multi-wavelength holography in these areas. The lens characterization method allows the

measurement of spherical, toric and other aspherical components with good precision and accuracy, while the holographic refractometry method allows the precise refractive index measurement of liquids with a measurable range which by far exceeds that of most of the commercial refractometres.

This work studied also the holographic recording and readout concerning the volume character of the photorefractive materials. Through this study, illumination procedures were successfully demonstrated in order to achieve good visibility interferograms. The anisotropic diffraction characteristics of the sillenite photorefractive materials, their high optical quality and high resolution capability are fundamental for the visualization of high-quality images.

7. References

Alam, M. S., Khoury, J. (2000). Photorefractive material based pattern recognition using fringe-adjusted correlation, Proceedings of the SPIE Conference on Photorefractive Fiber and Crystal Devices: Materials, Optical Properties, and Applications, pp. 14-17, ISBN 9780819437556, San Diego, California, 2000, SPIE, San Francisco

Anand, A. , Chhaniwal, V. (2007). Measurement of parameters of simple lenses using digital holographic interferometry and a synthetic reference wave, *Appl. Optics* 46, 2007, 2022-2026, ISSN 1559-128X

Barbosa, E.A., Muramatsu, M. (1997). Mapping of vibration amplitudes by time average holography in $Bi_{12}SiO_{20}$ crystals, *Optics and Laser Technology* 29, June of 1997, 359-364, ISSN 0030-3992

Barbosa, E.A. (2005). Holographic Imaging with multimode, large free spectral range lasers in photorefractive sillenite crystals, *Applied Physics B: Lasers and Optics* 80, 2005, 345-350, ISSN 0946-2171

Barbosa E. A., Filho, A. A. V., Gesualdi, M. R. R., Curcio, B. G., Muramatsu, M., Soga, D. (2005). Single-exposure, photorefractive holographic surface contouring with multi-wavelength diode lasers, *Journal of the Optical Society of America A* 22, 2005, 2872-2879, ISSN 1084-7529.

Barbosa, E. A., Lima, E. A., Gesualdi, M. R. R., M. Muramatsu (2007). Enhanced multi-wavelength holographic profilometry by laser mode selection, *Optical Engineering* 46, 2007, 075601-075607, ISSN 0091-3286

Barbosa, E. A., Carvalho, J. F. (2007). Surface analysis by two-diode laser photorefractive holography, *Applied Physics B: Lasers and Optics* 87 ,2007, 417-423, ISSN 0946-2171

Barbosa, E. A, de Sousa, C. B. F., Maffei, W. M. (2009). Measurement of low-derivative surface lenses by two-laser holography with $Bi_{12}TiO_{20}$ crystals, Applied Optics 48, 27, 2009, 5114-5120, ISSN 1559-128X

Barbosa, E.A. (2010). Positioning and localization of two-wavelength interferograms for wavefront reconstruction with volume holographic media, *Optics Express* 18, 9, April 2010, 8743-8758, ISSN 1094-4087

Barbosa, E.A., Silva, D. M., Preto, A. O., Verzini, R. (2010). Design, construction and performance of a real-time holographic refractometry setup for liquid analysis, accepted for publication in *Review of Scientific Instruments*, 2010.

Chang, C., Chen T.C., Hon-Fai Yau, Ye P.X., Self-pumped and mutually pumped phase comjugation using pentagon-shaped $BaTiO_3$ crystal, *Optical Materials* 18 (2001) 143-145, ISSN 0925-3467

Chuang E., Psaltis D. (1997). Storage of 1000 holograms with use of a dual-wavelength method, Applied Optics 36 (1997) 8445-8454, ISSN 1559-128X

De Angelis, M., de Nicola, S., Ferraro, P., Finizio, A., Pierattini, G. (1996), A reflective grating interferometer for measuring refractive index of liquids , *Optics Communications* 5, 1996, 761-765 , ISSN 0030-4018

De Angelis, S. De Nicola, P.Ferraro, A. Finizio, G. Pierattini (1997). A new approach to high accuracy measurement of the focal lengths of lenses using digital Fourier transform, *Optics Communications*, 136, 1997, 370-374, ISSN 0030-4018

DeBoo, B., Sasian J. (2003). Precise Focal-Length Measurement Technique with a Reflective Fresnel-Zone Hologram, *Appl. Optics* 42 (2003) 3903-3909, ISSN 1559-128X

Frejlich, J., Garcia, P. M. (2000). Advances in real-time holographic interferometry for the measurement of vibrations and deformations, *Optics and Lasers in Engineering* 32, December 2000, 515-527, ISSN 0143-8166

Frejlich, J. (2006). Photorefractive Materials: Fundamental Concepts, Holographic Recording and Materials Characterization, Wiley, ISBN 9780471748663, New York

Georges , M.P. , Scauflaire, V.S., Lemaire, P.C. (2001). Compact and portable holographic camera using photorefractive crystals: Application in various metrological problems, *Appl. Phys. B: Lasers and Optics* 72 (2001) 761-765, ISSN 0946-2171

Ghiglia, D. C., Mastin, G.A.,. Romero, L. A. (1987). Cellular-automata method for phase unwrapping, *Journal of the Optical Society of America* A, 1987, 210-219, ISSN 1084-7529.

Gustafsson, M., Sebesta, M. (2004). Refractometry of Microscopic Objects with Digital Holography, *Applied Optics* 43, 25, 2004, 4796 – 4801, ISSN 1559-128X

Gutmann, B., Weber, H. (1999). Phase Unwrapping with the Branch-Cut Method: Clustering of Discontinuity Sources and Reverse Simulated Annealing, *Applied Optics* 38, 26, 1999, 5577-5593, ISSN 1559-128X

Hecht, E. (1987). *Optics*, 2nd Ed. Addison-Wesley, ISBN 020111609X, NewYork

Kamshilin, A. A., Petrov, M. P. (1985). Continuous reconstruction of holographic interferograms through anisotropic diffraction in photorefractive crystals, *Optics Communications* 53, October 1984, 23-26, ISSN 0030-4018

Keren, E., Kreske K. M. , Kafri, O. (1988). Universal Method for determining the focal length of optical systems by moiré deflectometry, *Appl. Optics* 27 ,1988, 1383-1385, ISSN 1559-128X

Koechner, W. (1998), *Solid State Lasers Engineering*, Springer Verlag, ISBN 0-387-53756-2, Berlin

Kogelnik, H. (1969). Coupled Wave Theory for Thick Hologram Gratings, *Bell Systems Technical Journal* 48, 9, November 1969, 2909-2947, ISSN 1538-7305

Kuchel, F.M. , Tiziani, H.J. (1981). Real-time contour holography using BSO crystals, *Optics Communications* 38, 1981, 17-21, ISSN 0030-4018

Kukhtarev, N., Chen, B. S., Venkateswarlu, P. (1993). Reflection holographic gratings in [111] cut $Bi_{12}TiO_{20}$ crystal for real time interferometry, *Optics Communications* 104, December of 1993, 23-28, ISSN 0030-4018

Mallick, S., Rouède, D. (1987). Influence of the polarization direction on two beam coupling on photorefractive $Bi_{12}SiO_{20}$: Diffusion regime, *Applied Physics B: Lasers and Optics* 43, 1987 239-245, ISSN 0946-2171

Millerd, J.E., Brock, N.J. (1997) . Holographic profilometry with a rhodium-doped barium titanate crystal and a diode laser, *Applied Optics* 36, 1997, 2427-2431, ISSN 1559-128X

Pedrini, G., Fröning, P., Tiziani, H. J., Santoyo, F. M. (1999). Shape measurement of microscopic structures using digital holograms, *Optics Communications* 164 (1999) 257-268, ISSN 0030-4018

Poon, T.-C, Banerjee P. P. (2001), *Contemporary Optical Image Processing With Matlab* , Elsevier Science ISBN-10: 0080437885, New York.

Santiago Neto, R. B., de Mendonça, J. P. R. F., Lesche, B. (2004). Determination of Absolute Values of Refractive Index of Liquids Using an Interferometric Method, *Revista de Física Aplicada e Instrumentacao* 17, June 2004, 74-79, ISSN 0102-6895

Vest, C. M. (1979), *Holographic Interferometry*, Wiley, ISBN 0-471-90683-2, New York

Yamaguchi I., Ohta, S., Kato, J.-I. (2001). Surface contouring by phase-shifting digital holography, *Optics and Lasers in Engineering* 36, November 2001, 417-428, ISSN 0143-8166

5

The Composite Structure of Hologram and Optical Waveguide

Renxi Gao and Wenjun Liu
Harbin institute of technology at weihai, Weihai
China

1. Introduction

Optical holography is an excellent technology which can be applied in many fields, such as 3D-displays, information storage, products packaging, interferometer survey, optical connection and computer and so on. The conventional holography include reference wave, object wave and diffracting wave. These optical waves are spatial light and the whole hologram set up is fair size. The optical waveguides which are set up in optical routs can be integrated into a mini-chip. A new class of holography is proposed by many researchers [Suhara, et al., 1976; Putilin, et al., 1991; Singher & Shamir, 1994], in which the evanescent part of a guided wave or the guided wave in a waveguide is utilized as the reference wave or the illuminating one. This kind of hologram is defined as "waveguide hologram", and it has significant advantages in integration, wide field of view, compactness and other special functions. In this chapter, three sections are used to introduce the new development of waveguide hologram in the optical memory, optical connection and see-through planar displays fields.

2. The multilayered waveguides and holograms structure (MWH) for memory

With the development of information era, the dramatic increasing of information data needs new technologies of high density data storage. The holographic storage is regarded as a promising candidate for the next generation of optical memory. However, the conventional reconstructed holographic system includes reference wave and diffracting wave, which are spatial lights. The whole hologram system devices are large, so they are not suitable for personal use. The commercially available optical compact disc (CD) products have many advantages and play a key role in data storage area. But the data density of the optical disc can be hardly increased. We also know that the waveguide holography offers many advantages compared with conventional holograms. Waveguide holograms provide capability for the recording and reconstructing of holographic images or data information with light which propagates along optical waveguides. A higher image-to-background contrast and diffraction efficiency can be obtained because the direction of diffracted light wave is *vertical* to the illuminating light wave. Furthermore, waveguide holograms provide minimized illumination space and obstruction-free viewing. Therefore, multilayered waveguide and hologram technology is a promising candidate to solve the problems for the

next generation of optical data storage systems, many researchers have focused on this work[Imai, et al.,2003; Ishihara, et al., 2004; Mitasaki & Senda, 2006; Yagi, et al., 2007; Yagi, et al., 2008].

2.1 The principle of multilayered waveguide hologram memory

At present, there are many types of multilayered waveguide hologram structure. In this chapter, only three main types of multilayered waveguide structure for optical data storage are introduced. The first is shown in Fig. 1(a), a multi-layered structure where a layer with a high refractive index (core layer) and a layer with a low refractive index (clad layer) are alternately stacked. The hologram structures are fabricated in each waveguide core layer or on the surface of waveguide core layer in the style of pits, bumps or bubbles, as shown in Fig. 1(b). In general, they are computer-generated planar holograms. When light is introduced to this structure, the light is confined in the core layer and then propagates along waveguide core layer as reference wave. The diffracted light become spatial wave propagating in free space and containing the information from holograms, and the reconstructed images focused on the special position above of MWH, a detector such as charge-coupled device(CCD) is located at this position without any optical lenses. Because that the reference light can be confined in a target layer, and hence, the other layers do not interact with the reference light, so there are no or low crosstalk between different interlayer from the diffracted object light from one of target waveguide layer. A 100-periodic multilayered structure has been successfully fabricated and a data density of 100Gbit/inch2 can be achieved in the reference [Imai, et al., 2003], Fig. 2 shows bumps on the surface of waveguide core layer by Scanning electron microscope (SEM).

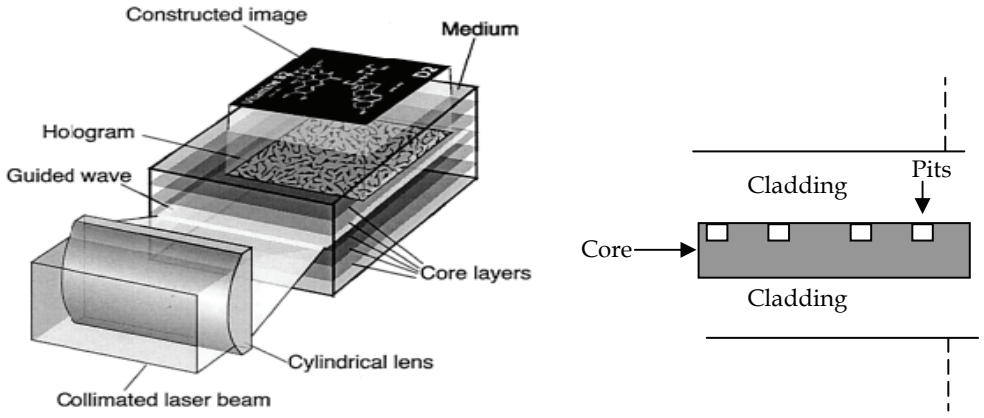


Fig. 1. (a) The firstly type structure of multilayered waveguide holograms, (b) Schematic diagram of element layer in MWH

The second type of multilayered waveguide hologram structure element is shown in Fig.3 [Mitasaki & Senda, 2006; Gao, et al., 2007], which is composite of a holographic data recording layer, a core layer(high refractive index n1)with waveguide grating, and two cladding layers with lower refractive index n_2, underneath the data recording layer. The

light was coupled into the waveguide core at the edge of MWH by cylindrical lens and then excited guide-mode light which propagated along the waveguide core. When meeting the waveguide gratings, the guide-mode light are diffracted into free space light as reference wave of holographic data recording layer. When the hologram recording layer is illuminated with these reference waves, the data information or holographic image is reconstructed by diffracted light of hologram recording layer. The direction of diffracted light of waveguide gratings can be different for each layers, thus the angle of β of light can be changed according to the structure of waveguide gratings for each layer. This type MHW also can be applied in three-dimension displays when many waveguides layers are illuminated at the same time.

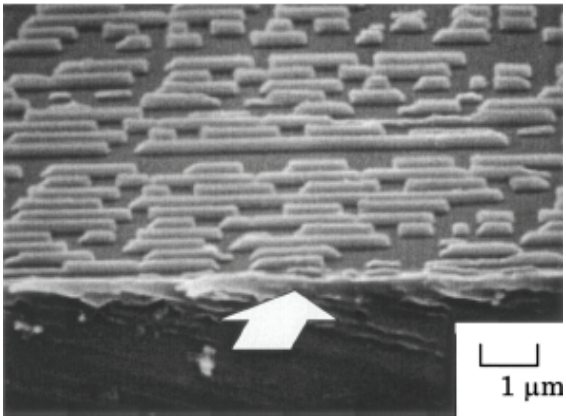


Fig. 2. SEM micrograph of a waveguide hologram (Imai, et al., 2003)

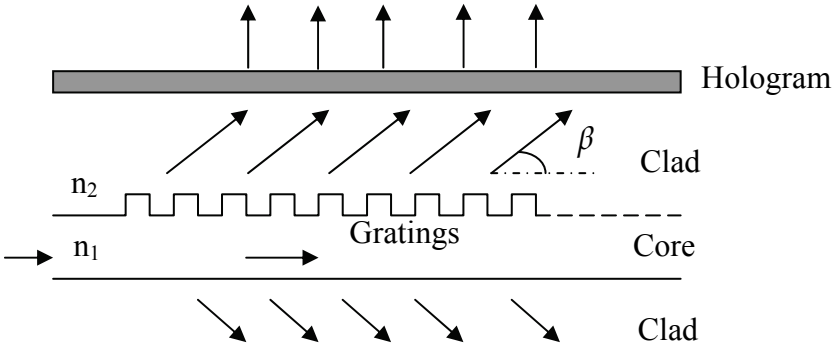


Fig. 3. Another type structure of multilayered waveguide holography element

Until now, we just discuss the type of MWH memory in which diffracted lights are coherent ly interfered. Guo et al. proposed a waveguide multilayered optical memory (WMOM) with bits stored as a refractive index change [Guo, et al., 2007], such as in Fig.4. The principle of reading data from a WMOM is based on scattering light from waveguide defects. In Fig. 4, when the incident light is coupled by the reading system L_1 into the core of the nth- *D*n layer

waveguide element of the waveguide multilayered disc (WMD), a series of guided modes and radiated modes will be excited. It is assumed that only guided modes remain and that radiated modes are attenuated to zero before meeting bits. Partial powers of guided modes confined in the core of the waveguide will be scattered from the core by bits. Hence, light scattered by the bits recorded in the core of the *D*n layer waveguide element of the WMD will form an array of bright dots against the dark background. The dot array is imaged with a confocal microscope, and only the light scattered from the bit located at the focus of the confocal microscope is collected effectively by a detector and converted to electric signals for further processing. This memory system includes a confocal microscope, so the cost of this type of waveguide multilayered memory would be higher.

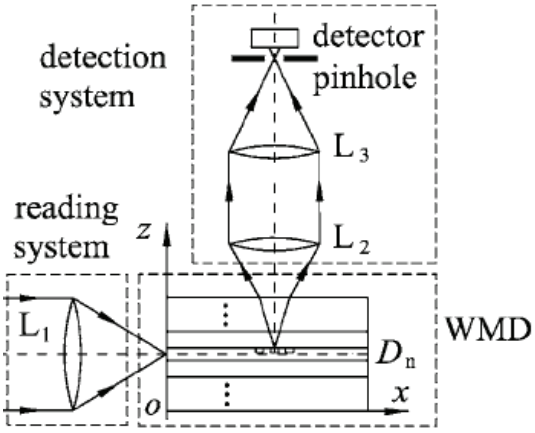


Fig. 4. Schematic diagram of a waveguide multilayered optical memory

2.2 Design methods of MWH

Generally, waveguide-hologram is based on computer generated hologram (CGH), especially for the first type MWH. Because information is implied by the style of pits or bumps on the surface of waveguide core layer or in the waveguide core layer, the distribution and shape of pits or bumps affect the quality of reconstructed image or data information. A waveguide hologram is a type of off-axis hologram where a guided wave is diffracted into free space, thus the reference light is guide-wave light propagated along the waveguide core layer and the diffracted light propagates along the direction normal to the waveguide core layer according to the distribution of bumps or pits. In a free-space CGH, the dimension of a rectangular element is typically of the order of more than ten times the optical wavelength. In order to effectively outcouple a guided wave, however, a grating period of wavelength dimension is preferred, thus the pits or bumps in rectangular element is small at the dimension of subwavelength.

In order to make diffracted light carry storage information, the bump or pit pattern needs to be designed by CGH method [Yagi, et al., 2008]. The guided-wave light disturbance in the waveguide plane can be represented by a scalar function $U(x,y)$ and the image detector plane can be represented by $V(x,y)$.When the desired image is represented by a light intensity distribution, $I(x,y)$, can be expressed as:

$$I(x,y) = |V(x,y)|^2 \tag{1}$$

Because of the image detector parallel to the waveguide, then the relations is

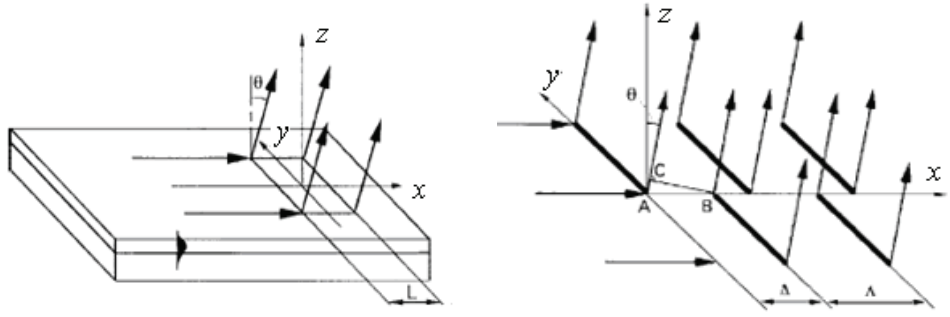
$$V(x,y) = \iint G(\xi - x, \eta - y) U(x,y) d\xi d\eta \tag{2}$$

Where G is a weight diffracted function such that

$$A \frac{\exp(jk\sqrt{x^2 + y^2 + d^2})}{x^2 + y^2 + d^2} \tag{3}$$

Here d is the distance between the plane of image detector and the waveguide storage layer. A is a constant. As V(x, y) is a convolution of U(x, y) and G(x, y), we can obtain U(x, y) for a given V(x, y) by Fourier-transform calculations. The next step involves finding a way to modulate the bump pattern and thus generate a light, represented by U(x, y), from the guided wave. The area of the bump pattern can be divided into minute square cells whose size at the level of micrometer. The above calculations are performed with sampling at this interval. The wave function of each square cell is $U_{ij} = U(x_0 + \delta i, y_0 + \delta j)$ by the magnitude a_{ij} and the phase ϕ_{ij} , where $U_{ij} = a_{ij} \exp(j\phi_{ij})$. In each cell, such as in Fig.5, the lines of bumps are perpendicular to the propagation direction of the guided wave so that the total length is proportional to a_{ij} and the locations of the lines are shifted in the propagation direction according to ϕ_{ij} .Then the drawn bump pattern works as a hologram that generates the wave represented by $U(x,y)$ as the object wave from the guided wave, which is the reference wave. Such a hologram may be regarded as a type of Lohmann's binary hologram.

We also know that the data information are implied by style of pits or bumps at the surface of waveguide core layer, and the pits or bumps are usual fabricated by using of polymers material; thus this type of hologram is of phase modulation. Each pit or bump in a unit cell can be designed by using an array of rectangular elements each containing a dislocated binary phase grating [Li, et al., 1996].



(a) (b)

Fig. 5. (a) Schematic diagram of free space light coupled from guided-wave light in a waveguide hologram. (b) Magnified picture describing the geometry of a fraction of the waveguide grating in (a). The grating period is Λ , and the two sections of the grating, separated by the x axis, have a relative dislocation of Δ in the x direction.

The phase-matching condition must be fulfilled between the guided wave and the radiation wave:

$$n_{eff}k_0 + 2\pi m / \Lambda = nk_0 \sin\theta \qquad m = 0, \pm 1, \pm 2, \cdots \tag{4}$$

Where n_{eff} is the effective refractive index of guided-wave and n is the refractive index of the material within which the diffracted light propagates, θ is the radiation angle with respect to the normal of waveguide, such as in Fig. 5(a). $k_0 = 2\pi / \lambda_0$ and λ_0 is the free-space wavelength. M is the order of diffraction. In order to increase the output efficiency of diffracted light, only the -1st diffraction order satisfies the equation (4) when m= -1.Therefore, the grating period Λ (the unit cell size) must satisfy,

$$\lambda_0 / n_{eff} \le \Lambda < \lambda_0 \tag{5}$$

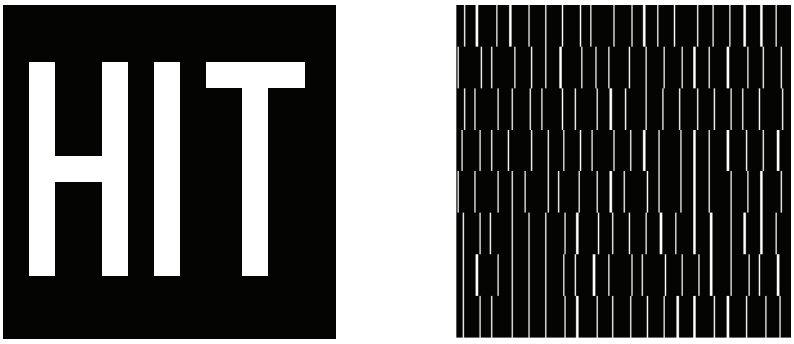
In Fig. 5(b), the unit cell is divided into two sections. Each section of the grating acts independently as an outcoupler, and the grating grooves in the two sections are relatively dislocated by a distance Δ along the guided wave propagating direction u. The grating period (Λ) satisfies (5). According to Fig. 5(b), there is an optical path difference, δL , between the radiation light outcoupled by two sections of the grating, which can be expressed as

$$\begin{aligned} \delta L &= n_{eff}\overline{AB} - n\overline{AC} \\ &= (n_{eff} - n\sin\theta)\Delta \end{aligned} \tag{6}$$

Taking into equation of (4) and noting that m= -1, the phase difference between the diffracted light, $\delta\phi$, as

$$\delta\phi = k_0\delta L = 2\pi\Delta / \Lambda \tag{7}$$

It shows that a phase shift $\delta\phi$ can be introduced by a grating dislocation Δ , and there exists a linear relation between $\delta\phi$ and Δ . Fig. 6 shows a computer-generated waveguide hologram (CGWH) structure by this method.



(a) (b)

Fig. 6. (a) The desired diffraction pattern, (b) a fraction of CGWH.

Yagi et al. discussed an optical design for multilayered waveguide holographies that utilizes an orthogonal aperture multiplexing (OAM) technique [Yagi, et al., 2007]. With OAM holography, the diffracted lights first concentrate in small isolated areas on the aperture plane (dashed line) and then expand and overlap on the imaging plane (solid line), as shown in Fig. 7. All the multiplexed images are diffracted simultaneously, as depicted by the solid and dashed lines. Any designated page (solid line) can be read without crosstalk by blocking all the diffracted lights on the aperture plane except for one selected area because one isolated area on the aperture plane corresponds to one page.
Fig. 8 shows a schematic view of the direct imaging optics. The diffracted light passes through an aperture, and then expands and constructs an image on the imaging plane, where h is the distance between the imaging and aperture planes, a is the size of the square aperture, and θ is the angle of expansion from the aperture. The aperture and imaging planes are parallel. Also, both the centres of the aperture and constructed square image form a line whose direction is normal to both planes.

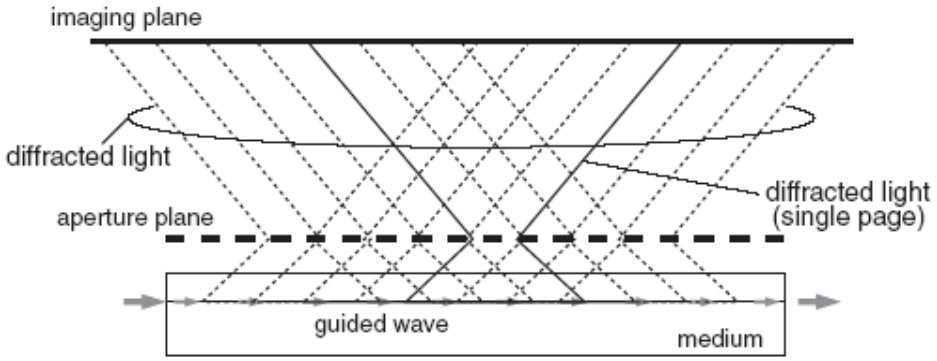


Fig. 7. Principle of orthogonal aperture multiplexing.

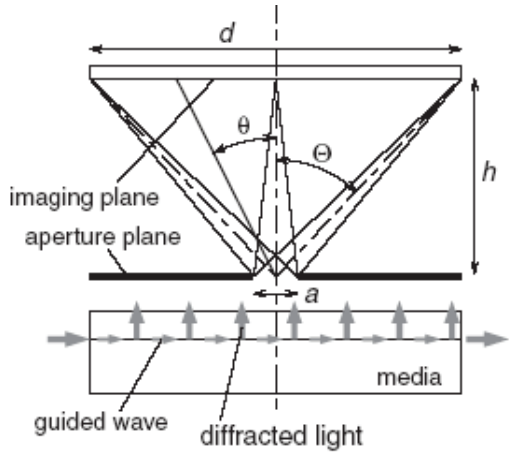


Fig. 8. Schematic view of direct imaging optics.

They discussed two types of optical design for OAM installed waveguide holograms. Both types can eliminate the need for an active focusing servo for imaging optics. The direct imaging optics offers a thinner driver, but it has a lower bit density and a shorter wavelength margin than the telecentric imaging optics. On the other hand, the telecentric imaging optics requires a thicker driver, but it offers a three fold higher bit density and a longer wavelength margin than the direct imaging optics.

2.3 Fabrication process of MWH

A multilayered waveguide hologram consists of a stack of single-mode slab waveguides with polymer as the main materials. So the fabrication methods of MWH are mainly concerned with the processing of polymer optical waveguides and the holograms. The size of pits or bumps of hologram designed by computer-generated hologram is small in dimension of subwavelength, so it is necessary to use the high precision micro-fabricating technology, such as photoresist-based patterning, direct lithographic patterning, soft lithography, embossing process and so on. In this section, we only introduce some key fabricating technologies for MWH.

- Spin-coating technology

Spin coating is an effective method to fabricate multilayered films with different thickness. We also know that the element of MHW is consist of the waveguide core layer with high refractive index of material and cladding layer with low refractive index of material. In this case, the substrate (rigid glass or flexible polyimide film) is mounted horizontally on a rotating platform. The substrate then spins very rapidly and the coating solution is dispensed onto it. The high-speed rotation throws off most of the solution, leaving behind a thin, uniform coating. Film thickness is precisely controlled by the rotational speed of the substrate. Faster rotation results in a thinner waveguide film layer. It is easy to form the film with thickness of 1~100μm by this method. However, Spin coating technology must be combined with other micro-manufacturing technology (such as photoresist patterning, reactive ion etching) to process computer-generated waveguide hologram.

- Lithographic patterning

Lithography is playing a critical role in micro- and nano-fabricated patterns for semiconductor devices and optics devices. It also can be adapted to process bits or bumps patterns on the surface of waveguide core layer in the computer-generated hologram. The techniques that can be used for patterning MWH included photoresist-based patterning, direct lithographic patterning, and soft lithography. The technology of photoresist-based patterning is seldom applied in fabrication of MWH, because that it needs cleaner room and reactive ion etching process, thus the cost of MWH is higher that other lithography patterning technology.

The direct laser lithography is a useful technique to fabricate a large planar areas with precision patterns, and it has the advantage of being maskless, allowing rapid and inexpensive prototyping in contrast to conventional mask-based photolithographic approaches. Mitasaki et al. proposed a write-once recording technique for MHW cards by laser directly writing, which can be suitable to record individual data easily in each MWH card [Mitasaki et al., 2006]. Soft lithography is a micro-fabrication technique that has been shown to generate high quality micro and nanostructures. It eliminates the use of costly and time consuming lithographic techniques and equipment. Unlike photolithography, it has

flexibility in material selection, can be applied to large planar surfaces, and provided high precise control over chemistry of patterned surfaces. Some of the diverse fabrication methods known collectively as soft lithography include: replica molding, micromolding in capillaries, microcontact printing, and microtransfer molding.

- UV embossing process technology

The ultraviolet (UV) embossing process technology [Ishihara, et al., 2004] is preferred with the advantages of mass produce and low cost. To reconstruct holographic images, bumps need to be patterned at the interface between the cladding and the core. In this process, these bumps are patterned by UV embossing using a metal stamper, which has holographic patterns. By repeating this process, a multilayered waveguide structure can be fabricated. The process flow is shown in Fig. 9. Firstly, the metal stamper is fabricated by the conventional process used for stampers of compact disks (CDs) and DVDs. This stamper has patterns which are designed to produce holographic images. Two types of UV curable resin are prepared, whose refractive indices are adjusted to form the core and the cladding, respectively.

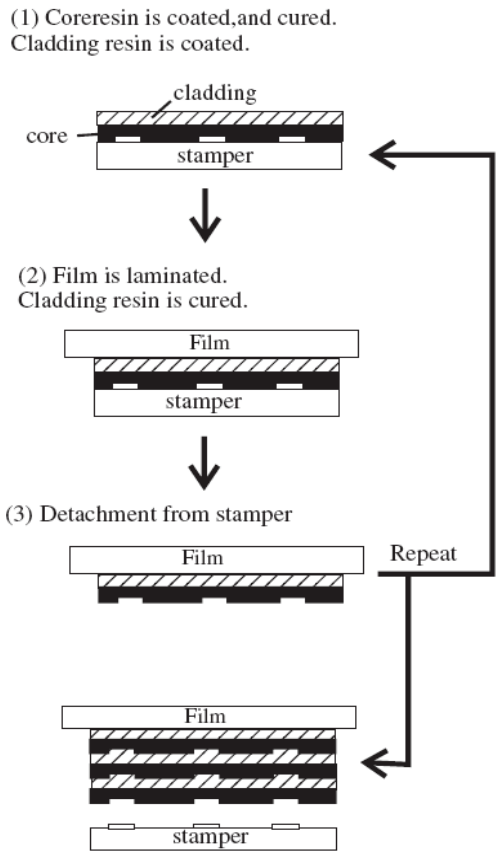


Fig. 9. Process flow of MWH

On the metal stamper, the UV curable resin for the core is spin-coated and UV-cured. Next, the UV curable resin for the cladding is spin-coated. A plastic film is laminated on it. After the resin is UV-cured, the film/cladding/core structure is detached from the metal stamper. By this process, the patterns of the metal stamper are replicated on the core layer surface. This process is repeated until the designated number of layers is fabricated. From the second cycle, the detached structure at the end of the former cycle is laminated instead of the plastic film, as shown in Fig. 9.
Using this process, Ishihara K. et al. have successfully fabricated a multilayered waveguide structure. Fig. 10 shows a picture of the 100-layer media structure. The thickness of the media is 2 mm. The thicknesses of core and cladding layers are 1.6 mm and 11 mm, respectively. 100-layer waveguides compose this media, and the holographic patterns are replicated in each layer.

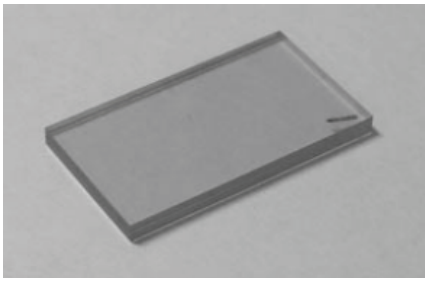


Fig. 10. 100 layer MWH , [Ishihara, et al., 2004].

2.4 Materials of MWH

Many different materials have been used in MWH in different mechanisms. In general, the materials for MWH need to have low optical absorption at the working wavelength. With the development of optical waveguide technology, many new materials with special functions have been introduced into this family. Several main optical materials are summarized in this section according to their mechanism.

- Optical polymers

Polymeric materials are particularly attractive in integrated optics because of their ability to be processed rapidly, cost-effectively, flexibility, and with high yields. Polymeric materials are allowed to form compact optical patterns by offering large refractive index contrasts and then easy to fabricate the planar waveguide structures. Some optical material, such as PMMA, PDMS, Epoxy resin and so on, now have been adopted in this field.. Polymer can be deposited by using spin coating or polymerization technique. Furthermore, the unique mechanical properties of polymers allow them to be processed by unconventional forming techniques such as molding, casting, stamping and embossing, therefore permitting rapid and low-cost shaping for waveguide formation.

- Dichromated gelatin

A normal gelatin film with less sensitive to light is sensitized through adding ammonium dichromate layer on it. Gelatin can absorb a very large amount of water and remain rigid, that means it swells. The chemical reaction happens on the interface of dichromated gelatin once it is exposed under light. The gelatin molecular chains in the exposed region have more

cross-linking. These regions swell less when immersed in water. Rapidly dehydrated by exposure to alchohol, the dichromated gelatin film shows differential strains between regions of maximum and minimum swelling. These strains modify the way of local refraction. Information can be recorded as the refraction characteristics in exposed and unexposed zones.

- Photopolymers

Some monomer molecule materials can be polymerized either through direct interaction with light or through an intermediary photosensitizer. These materials are called as photopolymer. Once exposed under light beam with information, the polymerization of such material depends on the local intensity of the recording radiation. The diffusion of surrounding monomers takes place during and after the exposure. A uniform postexposure using a fluorescent light has also been applied to increase the diffraction efficiency and to desensitize the photosensitizer. The variation in polymer concentration corresponds to the refractive index modulation.

- Photoresists

Photoresists are organic photosensitive materials, which can be classified into two types: negative and positive. The negative photoresists become insoluble in a solvent due to the polymerization. The rest unexposed area can be washed away. The positive case is just reverse when exposed on light. An image or holographic interference pattern is recorded on the surface of the photoresist layer through the insoluble and soluble area.

- Thermoplastic film

Thermoplastic film under a light beam produces a surface deformation of a transparent layer, which makes the phase of the light beam passing through the layer is modulated. The incident light pulses heat the thermoplastic layer to be molded according to the electrical field pattern.

- Two-color-absorption photopolymer

For the time being, two-color-absorption photopolymer is a new kind of recording material for MWH discs. It is attractive to be used in read-and-write MWH disks, because information can be only recorded on the target layer by illuminating the layer with a gate light, as shown in Fig. 11(b). Hirabayashi et al. of NTT photonics laboratories developed a

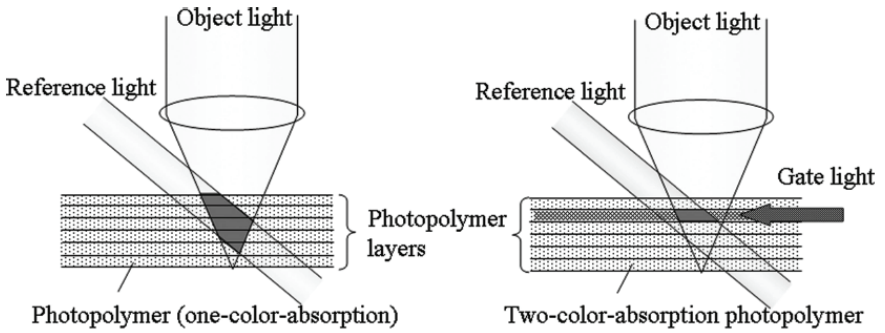


Fig. 11. (a) Conventional holographic recording in multilayer photopolymers and (b) Two-color holographic recording in multilayer photopolymers

sensitive two-color-absorption photopolymer materials doped with is(silyl)pentathiophene as a two-color-photosensitive dye and 2,2-dimethoxy-2-phenylacetophenone as a radical photopolymerization initiator and working with a 660 nm interference light and a 410 nm gate light. In 2004, Fujitsuka et al. reported that oligothiophenes show two-color absorption, and Shimizu et al. recently found that the stepwise two-color absorption of the oligothiophene in a Zeonex (Zeon Co.) matrix followed by energy transfer to an aryl azide in the matrix is efficient for two-color recording and may be applicable to holographic recording in multilayer thin films.

3. Optical interconnections based on optical waveguides holography

3.1 Introduction to optical interconnection

The term "optical interconnection" is often called simply "optical wiring" or "optical interconnect" and can be interpreted in the broadest sense of the word as "any interconnection using optical means". In term of linkages using optical means, optical fiber communication has been in practical use for more than twenty years, mainly for long distance links. In contrast, there has not been a strong need for short-reach optical links. However, as the problems of associated with hard-wired links become more apparent, this approach is gathering new focus. Optical interconnection is the most promising candidate for hard-wired circuitry. As optics become a major networking media in all communications needs, optical interconnects will also play an important role in interconnecting processors in parallel and distributed computing systems. Optical interconnect not only can provide much more connections than a traditional electronic interconnect, but also can offer much richer communication patterns for various networking applications. Such an optical interconnect can be used to serve as a cross-connect in a wide-area communication network or to provide high-speed interconnections among a group of processors in a parallel and distributed computing system.

Optical interconnections may be formed in substrates with waveguide structures and through the use of optical fiber, or in free-space either by beam spreading, or using holograms. The former is referred to as index-guided optical interconnection and the latter is referred as free-space optical interconnection. Fig. 12 depicts an example of free space interconnects [Gruber, et al., 2000]. A space between two circuit boards or a circuit board and optical interface board is purely empty. Light signals coming out from the sources propagate to designated location on the other substrate.

The purpose of the optical interconnection is to replace the electrical wire with the optical waveguide and associated optical devices. The smaller area the optical waveguide occupies on the chip area, the better integration the chip may have. All these approaches may be used to form reconfigurable interconnects with active components such as couplers for waveguides and optical fibers, modulators for beam spreading, and photo-refractive materials as dynamic holograms. Fixed interconnects are adequate for many purposes and don't require the use of active beam-steering components.

Free-space optical interconnections work in a three-dimensional volume in order to transport the signals to the desired locations. An earlier model of free-space interconnection is similar to a Fourier-plane imaging system with beam-deviating elements. The system has an array of sources in the object plane and an array of detectors in the image plane. Generally speaking, the light leaving the sources is deflected or fanned out upon passing a

holographic element and deflected or fanned out again upon traveling through the hologram in the Fourier plane. Both the object plane holographic element and the Fourier plane hologram are partitioned into independently programmed regions, each of which functions as a sub-hologram, so that light from different sources or falling on distinct regions of the Fourier plane may be directed to different detectors in the image plane.

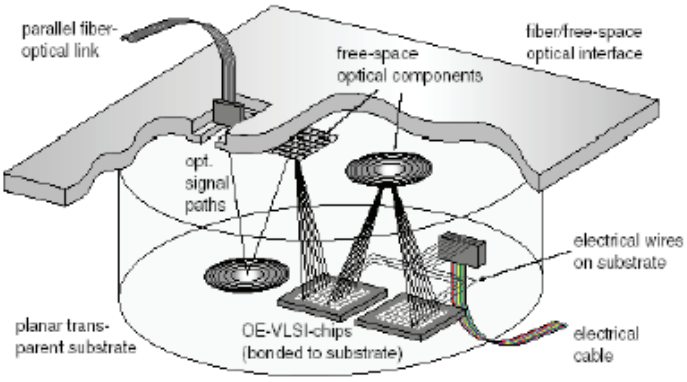


Fig. 12. Illustration of free space optical interconnect.

3.2 Optical interconnections based on optical waveguide holography

With the rapid advance of integrated optics, the importance of optical waveguides, which are the fundamental elements of optical integrated circuits, has been widely recognized. Optical waveguides are structures that confine and direct optical signals in a region of higher effective index than its surrounding media. By confining wave propagation, the waveguides provide communication between the electro-optical (E-O) transmitter and receiver, creating optical interconnects. Waveguide couplers play a key role for the realization of three-dimensional fully embedded board-level optical interconnection owing to their coupling of optical signals into and out of in-plane waveguides. A waveguide holographic grating can serve as a surface normal coupler. A classic example would be the coupling of guided-wave light to a discrete photodetector placed a few millimeters from the waveguide surface. This approach is attractive for a number of reasons. Firstly, the waveguide does not need to be cleaved and polished, as no end-faces are used in the coupling process. Secondly, making use of a slightly remote detector allows "pick-up" and place assembly components to be used in the manufacture of optics-hybrid circuits.

Optical interconnections using holographic optical elements minimize propagation delays; in addition, they reduce space requirements since several signals can propagate through the same network without mutual interference, many wiring patterns and high transmission rate. The optical signal can be divided to many output ports which are arrayed three-dimensionally and high densely by using forming function of wavefront with holograms. The holographic technology is expected to be applied for diffractive device of optical interconnection. Holographic technologies for optical interconnection devices have been studied actively[Yeh & Kostuk, 1995; Yeh & Kostuk, 1996]. Holography is the basis of many methods of optical data processing. Holograms are assigned important tasks in research on

optical interconnections. Particular interest attaches to multifocus lenses for matrix processors and optical synchronization systems, and also to focusing devices that carry out prescribed transformations of wave fronts. By using various types of holograms, interconnections can be either fixed or variable. Dynamic holograms can make these interconnections dynamic; in addition, the method of wave-front reversal is the best physical basis for the development of associative two-dimensional memory. Holograms are also indispensable in one- and two- dimensional analog-digital converters [Honma et al., 2007].

In this section, we introduce some recently results of research about waveguide Holograms (WGH). Like any other type of hologram, they have their strong and weak points, but their compatibility with integrated-optics circuits is an advantage that allows us to regard them as promising. A WGH system is shown in Fig. 13, which is based on thin substrate waveguides bearing a hologram on the surface through which light is diffracted out. A light source is optically coupled to the waveguide such that light emitted from the source is caused to propagate along the waveguide, being diffracted out at intersections with the surface of the waveguide on which the hologram is formed.

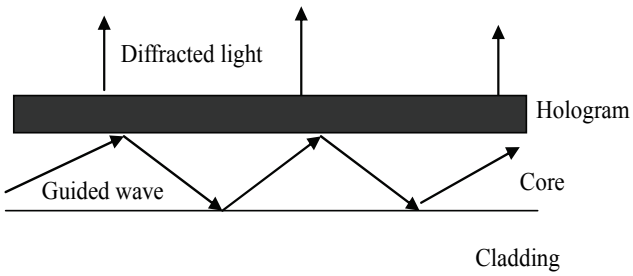


Fig. 13. Schematic of waveguide hologram system

Waveguide holograms have been used in many applications including filters, multiplexers, and DFB lasers. Since most of these applications require only one grating in the hologram, surface relief gratings become an appropriate choice. In the case of massive fanout optical holographic interconnects, many multiplexed and superimposed gratings must be recorded in the same hologram. Therefore, phase gratings are required for optical interconnect applications. Due to the nature of phase gratings, many-to-1 (fanin), 1-to-many (fanout), and many-to-many interconnects can be realized in both planar holograms and grating couplers. Another advantage of using phase gratings is that the strict phase matching condition of grating couplers can be alleviated (e.g., alignment requirements between laser diodes and grating couplers can be reduced) by either broadband processing of phase gratings" or multiple grating recording (i.e., each grating covers a particular angular acceptance).

3.2.1 Waveguide holograms of two-dimensional images and application in optical interconnects

Putilin discussed an optical interconnection employing waveguide and total internal reflection holograms as a basic element [Putilin, 1991]. It is shown that WGH have an extremely high efficiency, it forms high resolution and wide view image and permits to increase signal to noise ratio. WGH method permits to reconstruct two-dimensional images

by one-dimensional waveguide mode because of distributed interaction between them. The efficiency of WGH depends upon length of grating and perturbation of refractive index of waveguide mode caused by WGH. So the performance efficiency of WGH can be up to 100% and all waveguide mode power will be used in image. WGH is situated one by one along the direction of waveguide mode propagation and can be reconstructed simultaneously as shown in Fig.14. The general efficiency of that WGH array is defined as (diffraction efficiency of each waveguide hologram-η)

$$A = 1 - (1 - \eta)^n \tag{8}$$

Where n is a number of WGH. For uniform n the brightness of images will decrease and some information (recorded on last waveguide hologram) will be lost. For uniform brightness of images WGH must be written with increasing efficiency:

$$\eta_{i+1} = \eta_i / (1 + \eta_i) \tag{9}$$

Maximum efficiency is limited for concrete dynamics range of recording material and general efficiency will be defined as:

$$A = n\eta_{\max} / [1 + (n-1)\eta_{\max}] \tag{10}$$

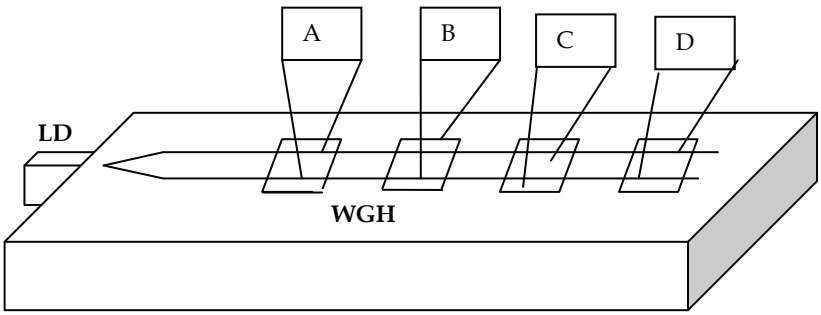


Fig. 14. Scheme of simultaneous waveguide hologram array reconstruction

The main demand of chip-to-chip interconnects is a very small distance between optical elements and image plane. Evidently requirements for chip-to-chip interconnection is similar to parameters of photolithographic equipment. WGH permit to obtain such characteristics. An application of WGH in chip-to-chip interconnection will be optimal because light can be separated from air-spaces completely. One of possible variants of that interconnection based on WGH of two-dimensional images is shown in Fig.15. The system works as explained in following steps: (1) Central WGH 1 transforms substrate waves from diode lasers to waveguide modes, (2) this light reconstructs information stored on WGH2. Each LD can reconstruct one line from WGH1 information if LD is arranged on VLSI in X direction , if LD array is arranged in Y direction then all lasers reconstruct one line from WGH1 information with individual shift in Y direction, so shape of reconstructed images from WGH2 will be changed. All variants of interconnect configurations must be recorded on WGHs. Modulation and additional reconfiguration of system can be done by integrated waveguide modulators situated between WGH1 and WGH2.

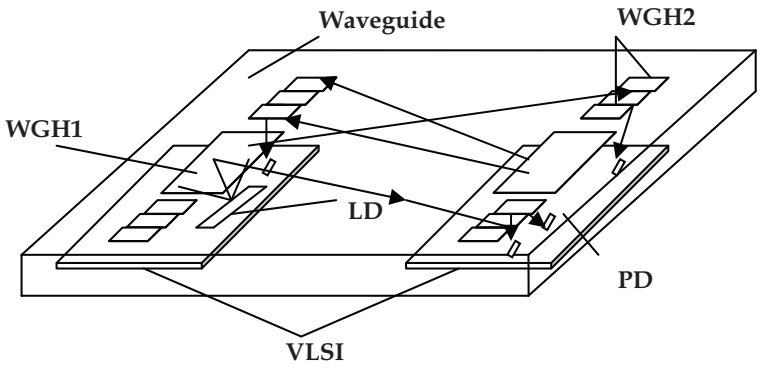


Fig. 15. The scheme of chip-to-chip interconnects based on WGH of two-dimensional images.

3.2.2 Optical multiplanar VLSI interconnects based on multiplexed waveguide holograms

Optical interconnects for very large scale integration systems based on planar waveguide holograms are analyzed in this section[Lin et al., 1990]. The combination of low loss waveguides and multiplexed waveguide holograms allows the construction of various compact planar architectures with high interconnect density and low insertion loss. The long interaction lengths possible in planar structures result in high angular and wavelength selectivity. Schematic diagram of a dual plane architecture of optical VLSI interconnects is shown in Fig. 16. One plane of the structure contains both electronic and optoelectronic elements (e.g., on a circuit board). Multiplexed holograms are fabricated in the other plane. It is also possible to achieve truly monolithic integration if a particular substrate, e.g., GaAs, is used to hold electronic, optoelectronic, and optical elements or if a mixed integration technique is used.

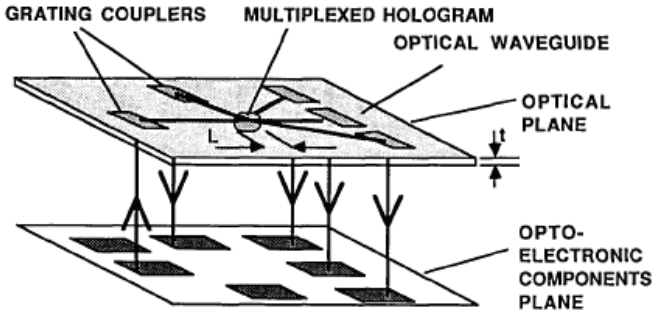


Fig. 16. Schematic diagram of a dual plane architecture of optical VLSI interconnects.

As depicted in Fig. 16, the laser diodes, driven by electronic signals, emit diverging wavefronts which can be transformed to guided waves by focusing grating couplers or a combination of microlenses and regular grating couplers. The guided waves are in turn redirected or distributed in various directions by multiplexed waveguide holograms. This configuration utilizes the advantages of holographic optical interconnects (high density, low

loss) and yet allows for compact packaging achievable with integrated optics. It also has less stringent alignment requirements compared with conventional optics. Alignment is reduced to 2-D problems in the planar waveguide, since the vertical direction alignment is provided by grating couplers. The trade-off of this approach is that coupling of laser diodes to single-mode waveguides remains a difficult problem. Solutions, such as broadband or multiplexed grating couplers, are under investigation. In addition, these 2-D alignments can be solved by a computer-generated hologram approach in which grating couplers and planar holograms are written by an e-beam gun on the same substrate. Thus alignment problems/requirements between these holographic optical elements can be significantly reduced.

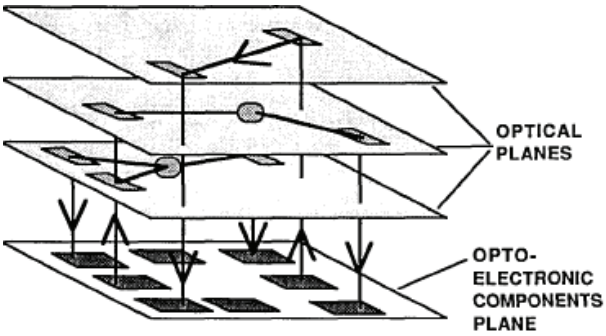


Fig. 17. Schematic diagram of a multiplanar interconnect architecture.

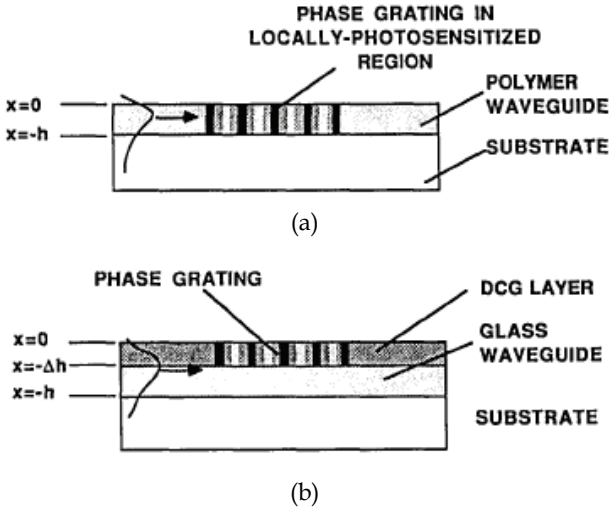


Fig. 18. (a) Planar hologram fully embedded in the waveguide; (b) Partially embedded waveguide hologram structure.

A multiplanar architecture, illustrated in Fig. 17. This multiplanar architecture preserves the advantages of a 3-D holographic optical interconnect system while simultaneously avoiding

the drawbacks of bulk system layouts. It should be emphasized that since the typical waveguide and substrate thickness can be of the order of a few hundreds of micrometers in total, the thickness of the multiplanar system can still be of the order of centimeters, even for -100 layers.

To fabricate waveguides with embedded phase gratings, two material requirements must be fulfilled. Waveguides should have low loss and be photosensitive. If a low loss holographic material can be found, the grating may be recorded in the entire guided region (the fully embedded case illustrated in Fig.18(a)). Partially embedded waveguide hologram structure, illustrated in Fig. 18(b), is composed of a low loss ion-exchanged glass planar waveguide with a layer of dichromated gelatin (DCG) as the holographic material deposited on top of the regions where holograms are to be recorded. This structure gives satisfactory results for both grating couplers and planar holograms. To obtain the desired grating vector K parallel to the waveguide surface, the two recording beams interfering in the film must be separated by an angle $2\beta_r$ determined by:

$$\sin\beta_r = \frac{\lambda_r}{2d} = \frac{n_f \lambda_r (\sin\theta_G - \sin\beta)}{2\lambda} \tag{11}$$

where β is the input beam angular deviation from normal, θ_G is the guided wave angle, λr is recording wavelength. The system can realize reconfigurable interconnects (depending on incident wavelengths) and a total of a few hundred independent interconnect paths. However, multiple planar holograms fabricated in one plane can be stacked in a multiple plane system in which thousands, even tens of thousands of interconnects are obtainable. The maximum number of interconnects in one plane is limited by the channel angular width and grating coupler and detector sizes. The upper limit of the interconnect plane size is determined by the waveguide substrate size and waveguide quality. An additional trade-off arises between interconnectivity and system bandwidth requirements due to dispersion characteristics of the structure.

3.2.3 Use of waveguide holograms for input and transmission of infromation through a optical fiber

A waveguide hologram as well as any other kind of hologram makes it possible to form a structure of a light beam that matches the optical fiber employed for input, i.e., provides focusing and input of radiation into a fiber without any additional elements. The waveguide nature of the reconstruction process is an advantage of the waveguide hologram, which provides rapid reconstruction and almost equal amplitudes of waves reconstructed from different areas of the hologram when a short reading pulse is used and the diffraction efficiency is relatively low.

Dianov et al., presented a scheme for input and transmission of information on the complex amplitude transmittance of a one-dimensional transparency through a single-mode fiber communication channel with a waveguide hologram used for input [Dianov et al., 1996]. A schematic diagram of an optical setup with specially prepared waveguide holograms for input of information on the structure of a one-dimensional light field into a fiber communication line is shown in Fig. 19. In order to form a light beam of the required structure, the waveguide hologram is illuminated by a short radiation pulse in the form of a plane wave Ep(x, y, z, t) propagating along the x axis. The one-dimensional information to

be transmitted is represented in the form of the complex amplitude transmittance t(x) of the one-dimensional transparency Tr, whose plane is practically aligned with that of the hologram. In certain cases, this information can be recorded on the hologram itself in addition to the structure that forms the radiation to be introduced into the optical fiber. To provide input of radiation modulated in accordance with the transmitted amplitude-phase information with the required spatial resolution into an optical communication line comprising a single-mode fiber, the waveguide hologram should focus radiation from each hologram element onto the fiber input, i.e., each hologram element should produce a converging spherical wave in the reconstruction process. The device interconnection level in reality is an optical fiber local area. The waveguide holograms just described may be integrated with optical fibers it will decrease number and complicacy of coupling elements.

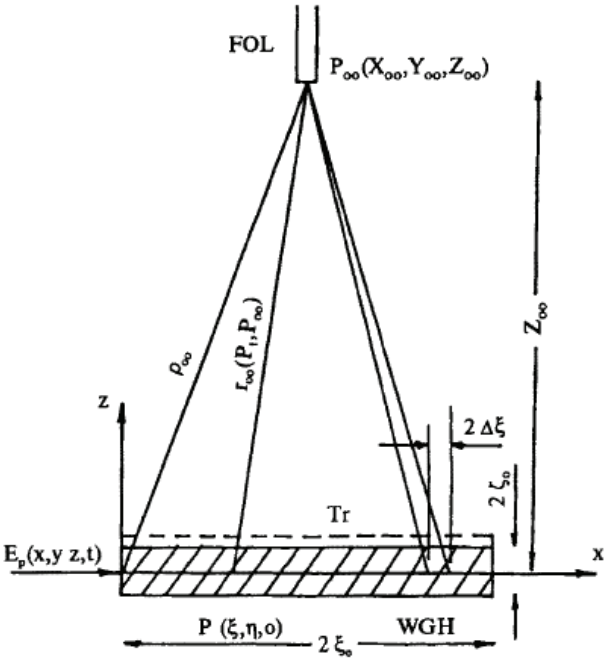


Fig. 19. Focusing of radiation into a fiber optics link by means of a waveguide hologram.

3.2.4 Waveguide coupling using holographic transmission gratings

Waveguide coupling using gratings has attracted considerable attention as an alternative to fibre butt-coupling, and a candidate for opto-hybrid integration of components. Sheard et al., studied on useing specially shaped gratings etched into the waveguide surface to couple light into the free space region above the waveguide [Sheard et al., 1997; Liao et al., 1998]. Figure 20 shows a waveguide core section with light travelling along the waveguide in the direction of the arrow on the left. The propagating light is progressively scattered by the etched grating and projects the light into the free space region above the waveguide surface. The parallelogramic shaped grating teeth are desirable, as this has been shown to be the most efficient structure for projecting the waveguide light into a single direction.

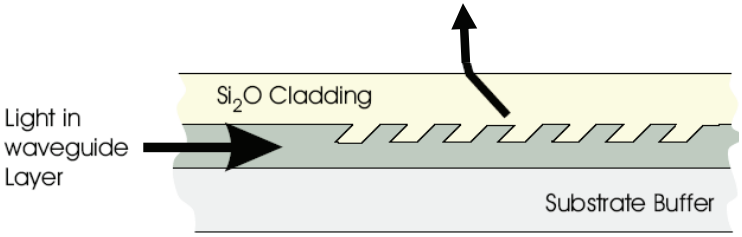
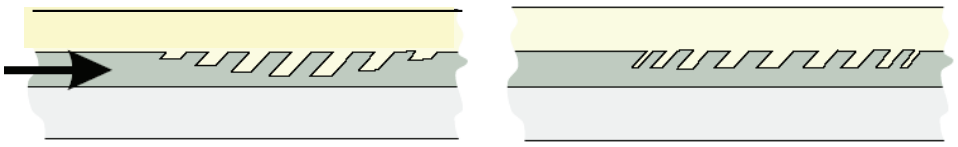


Fig. 20. Parallelogramic shaped etch grating for coupling light into the free space region above the waveguide

The intensity profile of the light coupled out of the waveguide surface can be varied by controlling the coupling efficiency. Either the grating depth can be varied along the length of the grating or the mark-to-space ratio of the grating can be varied as shown in Fig. 21 (a) and (b), respectively. Varying the mark-to-space ratio is easier using high resolution lithography. Varying the grating pitch and curving the grating lines by e-beam writing can be used to generate a focused beam from the output coupler or even an array of focussed points. A typical illustration of a focussing grating coupler is shown in Fig. 22. Here the light from a butt-coupled fibre is collected and focused by the grating a few millimetres above the waveguide surface. In most cases a diffraction limited spot size can be achieved.



(a) (b)

Fig. 21. There are two simple ways to vary the intensity profile of the light coupled out of the waveguide surface, by controlling the coupling efficiency. (a) the grating depth along the length of the grating is varied, (b) the mark-to space ratio of the grating is varied.

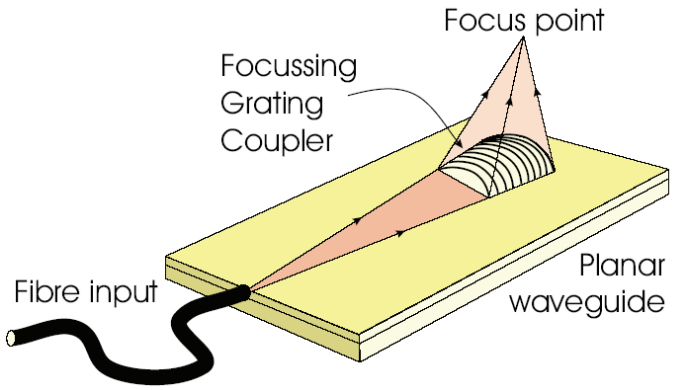


Fig. 22. Focusing grating coupler using a curved a chirped surface relief grating.

Because of its periodic nature, the grating perturbs the waveguide modes in the region underneath the grating, thus causing each one of them to have a set of spatial harmonics with z-direction propagation constants given by

$$\beta_v = \beta_0 + \frac{2v\pi}{\Lambda} \tag{12}$$

where v=0, ± 1, ± 2, . . . , and where Λ is the periodicity of the grating. The fundamental factor β_0 is approximately equal to the β_m of the particular mode in the waveguide region not covered by the grating. The principal advantage of the grating coupler is that, once fabricated, it is an integral part of the waveguide structure. Hence, its coupling efficiency remains constant and is not altered appreciably by vibration or ambient conditions.

4. The holographic waveguide for see-through planar display

The technology of see-through planar display (i.e. head-up display) is now used in the fields of military aviation, commercial aircraft, automobiles and other applications. A conventional see-through planar display system contains three primary components: a combiner, a projector unit, and a video-generating computer, so the size of the entire instrument is quite large, as shown in Fig. 23.

Because the holographic waveguide itself is transparent to the free-space light beams in the direction of perpendicular to waveguide, the observer can view the image produced by the hologram and at the same time can see through the hologram to view the scene at the opposite side of the hologram. This property is very useful for head-up display technology to eyewear display.



Fig. 23. A type of head-up display system.

In 2008, Sony Co. reported an eyeglass-shaped see-through display that can show full color video images, as shown in Fig. 24, and the holographic waveguide is the key component that realized the structure and display method in this sample. Fig. 25 illustrates this type structure of the holographic planar waveguide of the eyewear display [Mukawa et al., 2009]. The waveguide has an in-coupling and an out-coupling reflection volume hologram which have exactly the same fringe pattern and a mirror symmetrically positioned. Reflection volume holograms were employed because their diffraction bandwidths are much smaller than those of transmission holograms and could potentially enlarge the field of view of the

eyewear displays. Each of these holograms has red, green, and blue hologram layers to transmit full-color images through the waveguide.



Fig. 24. A visitor wearing the prototype of eyeglass-shaped see-through display

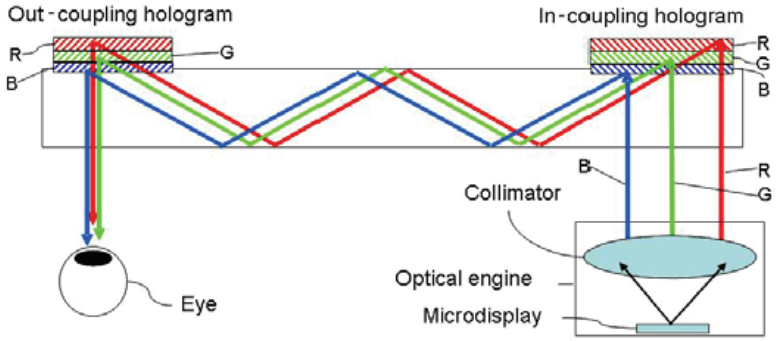


Fig. 25. The structure of holographic waveguide for eyewear display

5. Conclusion

The combination of hologram and optical waveguide is versatile technology that can be used in the optical data storage, optical connection, optical display and other applications. With the development of optoelectronic technology, we think that many new kinds of waveguide hologram devices will be invented and used widely in the future.

6. Reference

Suhara, T.; Nishihara, H. & Koyama, J. (1976). Waveguide Holograms: A New Approach to Hologram Integration. *Optics Communications,* Vol.19, No.3. 353-358.

Putilin, A.; Huang, Q. & Caulfield, H. (1991). Waveguide Holograms With White Light Illumination. *Optical Engineering,* Vol. 30, No. 10. 1615-1619.

Singher, L. & Shamir, J. (1994). Waveguide Holographic Elements Recorded by Guided Modes. *Applied Optics,* Vol. 33, No. 7. 1180-1186.

Imai, T.; Kurihara, T.; Yagi, S.; Kurokawa, Y.; Endo,M. & Tanabe, T. (2003). Orthogonal Aperture Multiplexing for Multilayered Waveguide Holographic Read-only Memories. *Optical Society of America,* Vol. 42, No. 35. 7085-7092.

Ishihara, K.; Jujiwara, T. & Esaki, A. (2004). Evaluation of Multilayered Waveguide Holographic Memory Media. *Japanese Journal of Applied Physics,* Vol. 43, No. 7B, 4915-4918.

Mitasaki, T. & Senda, M. (2006). Write-once Recording for Multilayered Optical Waveguide-type Holographic Cards. *Optical Society of America,* Vol. 23, No. 3, 659-663.

Gao, R. X.; Chen, B. X.; Qu, S. L. (2007). Design of the novel multilayered waveguide holography structures. Infrared and Laser Engineering, Vol.36, No.4, 447-449.

Yagi, S.; Nakamura, S.; Imai, N. T.; Kurihara, T; Ueno, M.; Endo, M.; Kurokawa, Y.; Yoshikawa, H. & Ohtani, Y. (2007). Optical Design for Orthogonal Aperture-Multiplexed Waveguide Holography. *Japanese Journal of Applied Physics*, Vol.46, No.6A, 3433–3437.

Yagi, S.; Imai, T.; Ueno, M.; Ohtani, Y.; Endo, M.; Kurokawa, Y.; Yoshikawa, H.; Watanabe, T. & Fukuda, M. (2008) Pit Distribution Design for Computer-Generated Waveguide Holography . *Japanese Journal of Applied Physics,* Vol. 47, No.2, 2008, 942-946.

Guo, H. M.; Zhuag, S. L.; Chen, J. B. & Liang, Z. C. (2007). Multilayered Optical Memory With Bits Stored As Refractive Index Change.I.Electromagnetic theory. *Optical Society of America,* Vol. 24, No. 6, 1776-1785.

Li, M.; Bengtsson, J.; Hagberg, M.; Larsson, A. & Suhara, T. (1996). Off-Plane Computer-Generated Waveguide Hologram. *IEEE Journal of Selected Topics In Quantum Electronics,* Vol. 2, No. 2, 226-235.

Hirabayashi, K.; Kanbara, H.; Mori, Y.; Kurihara, T.; Shinizu, M. & Hiyama, T. (2007). Multilayer Holographic Recording Using A Two-color-absorption Photopolymer. *Applied Optics,* Vol. 46, No. 35,8402-8410.

Fujitsuka, M.; Oseki, Y.; Hara, M.; Cai, X.; Sugimoto, A. & Majima, T. (2004). First direct observation of the higher triplet excited states of substituted oligothiophenes by two-color twolaser flash photolysis. *Chem. PhysChem* , Vol. , No.5, 1240–1242.

Shimizu, M.; Schelper, M.; Mochida, K.; Hiyama, T.; Adachi, M.; Sasaki, Y.; Akiyama, S.; Maeda, S.; Kanbara, H.; Mori, Y. & Kurihara, T. (2007). Novel strategy for two-photon holographic recording:stepwise two-photon absorption of a-quinquethiophene followed by energy transfer to an aryl azide. *Adv. Mater*. Vol.19, 1826–1829.

Gruber,M.; Sinzinger, S. & Jahns, J. (2000). Optoelectronic multichip module based on planar-integrated free-space optics. *Proc. SPIE, Optics in Computing*, 4089,

Yeh, J. H. & Kostuk, R. K. Substrate-mode holograms used in optical interconnects: design issues. (1995). *Applied Optics,*Vol. 34, No. 17, 3152- 3164.

Yeh, J. H. & Kostuk, R. K. (1996). Free-space holographic optical interconnects form board-to-board and chip-to-chip interconnections. *OPTICS LETTERS*, Vol. 21, No. 16, 1274-1276.

Honma, S.; Muto, S.; Okamoto, A. (2007). Shift-phase code multiplexing technique for holographic memories and optical interconnection. *Proc. of SPIE*, Vol. 6832, 68322p1-68322p11.

Putilin, A. N. (1991). Optical interconnections based on waveguide holograms. *Proc. of SPIE*, Vol. 1621, 93-101.

Lin, F; Eva M.; Strzelecki, E. M. & Jannson, T. (1990). Optical multiplanar VLSI interconnects based on multiplexed waveguide holograms. *Applied Optics,*Vol. 29, No. 8, 1126-1133.

Dianov, E. M.; Putilin, A. N. & Zubov, V. A. (1996). Use of waveguide holograms for input and transmission of information on a light field through a single-mode optical fiber. *Journal of Russian Laser Research*, Vol. 17, No. 1, 55-84.

Sheard S. J.; Liao T. D.; Yang G. G.; Prewett P. D. & Zhu J. G. (1997). Focusing waveguide grating coupler using diffractive doublet. *Applied Optics,* Vol. 36, No. 19, 4349-4353.

Liao T. D. & Sheard S. J. (1998). Radiation characteristics of waveguide diffractive doublets. *Applied Optics,* Vol.37, No.10, 1776-1783.

Mukawa, H; Akutsu, K.; Matsumura, I.; Nakano, S.; Yoshida, T.; Kuwahara, M. & Aiki, K. (2009). A full-color eyewear display using planar waveguides with reflection volume holograms. *Journal of the society for information display*, Vol. 17, No. 3, 185-193

Part 3

Holographic Techniques

6

FINCH: Fresnel Incoherent Correlation Hologram

Joseph Rosen[1], Barak Katz[1] and Gary Brooker[2]
[1]*Ben-Gurion University of the Negev*
Department of Electrical and Computer Engineering
P. O. Box 653, Beer-Sheva 84105
[2]*Johns Hopkins University*
Department of Biomedical Engineering and Microscopy Center,
Montgomery County Campus
9605 Medical Center Drive Suite 240, Rockville, MD 20850
[1]*Israel*
[2]*USA*

1. Introduction

Holographic imaging offers a reliable and fast method to capture the complete three-dimensional (3D) information of the scene from a single perspective. However, holography is not widely applied to the regime of white-light imaging, because white-light is incoherent and in general creating holograms requires a coherent interferometer system. In this chapter we describe our recently invented method of acquiring incoherent digital holograms. The term incoherent digital hologram means that incoherent light beams reflected or emitted from real existing objects interfere with each other. The resulting interferogram is recorded by a digital camera and digitally processed to yield a hologram. This hologram is reconstructed in the computer so that 3D images appear on the computer's screen. The coherent optical recording is not applicable for the incoherent optics because interference between reference and object incoherent beams cannot occur. Therefore, different holographic acquisition methods should be employed for generating an incoherent digital hologram.

The oldest methods of recording incoherent holograms have made use of the property that every incoherent object is composed of many source points, each of which is self-spatial coherent and therefore can create an interference pattern with light coming from the point's mirrored image. Under this general principle, there are various types of holograms, including Fourier (Stroke & Restrick, 1965; Breckinridge, 1974) and Fresnel holograms (Lohmann, 1965; Cochran, 1966; Peters, 1966; Marathay, 1987; Mugnier et al., 1993). The process of beam interfering demands high levels of light intensity, extreme stability of the optical setup, and relatively narrow bandwidth light source. These limitations have prevented holograms from becoming widely used for many practical applications. More recently, three groups of researchers have proposed computing holograms of 3D incoherently illuminated objects from a set of images taken from different points of view

(Li et al., 2001; Sando et al., 2003; Shaked & Rosen, 2008; Park et al., 2009). This method, although it shows promising prospects, is relatively slow since it is based on capturing tens of scene images from different view angles. Another method is called scanning holography (Poon & Korpel, 1979; Schilling et al., 1997; Poon, 2004; Rosen et al., 2006; Poon, 2007) in which a pattern of Fresnel zone plates (FZPs) scans the object such that at each and every scanning position, the light intensity is integrated by a point detector. The overall process yields a Fresnel hologram obtained as a correlation between the object and FZP patterns.

Our recently proposed incoherent digital hologram is dubbed Fresnel incoherent correlation hologram (FINCH) (Rosen & Brooker, 2007a). The FINCH is actually based on a single channel on-axis incoherent interferometer. Like any Fresnel holography, in the FINCH the object is correlated with quadratic phase functions, but the correlation is carried out without any movement and without multiplexing the image of the scene. This chapter reviews the latest developments of the FINCH in the field of color holography, microscopy and imaging with a synthetic aperture.

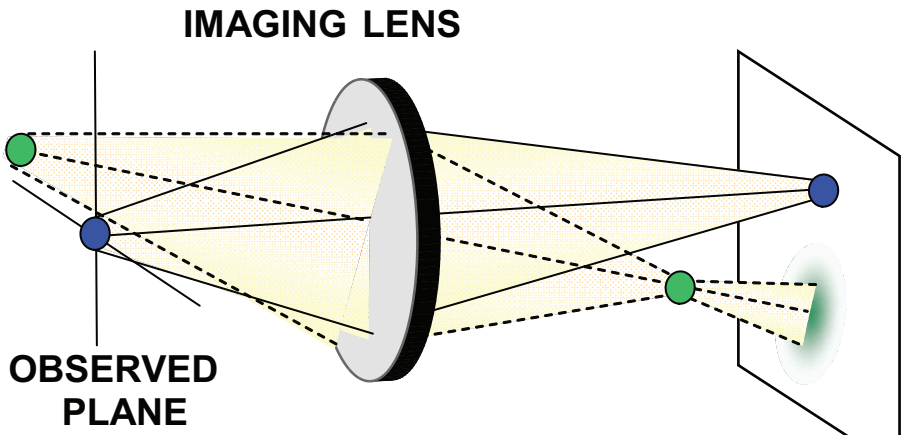
Generally, in the FINCH system, light is reflected, or emitted, from a 3D object, propagates through a spatial light modulator (SLM), and is recorded by a digital camera. The SLM is used as a beam-splitter of the single-channel incoherent interferometer, such that each spherical beam originated from each object point is split into two spherical beams with two different curve radiuses. Incoherent sum of the entire interferences between all the couples of spherical beams creates the Fresnel hologram of the observed 3D object. When this hologram is reconstructed in the computer, the 3D properties of the object are revealed.

In the following sections we describe various aspects of the FINCH, including FINCH of reflected white light (Rosen & Brooker, 2007a), FINCH of fluorescence objects (Rosen & Brooker, 2007b), a FINCH-based holographic fluorescence microscope (Rosen & Brooker, 2008), and finally a hologram recorder in a mode of a synthetic aperture (Katz & Rosen , 2010).

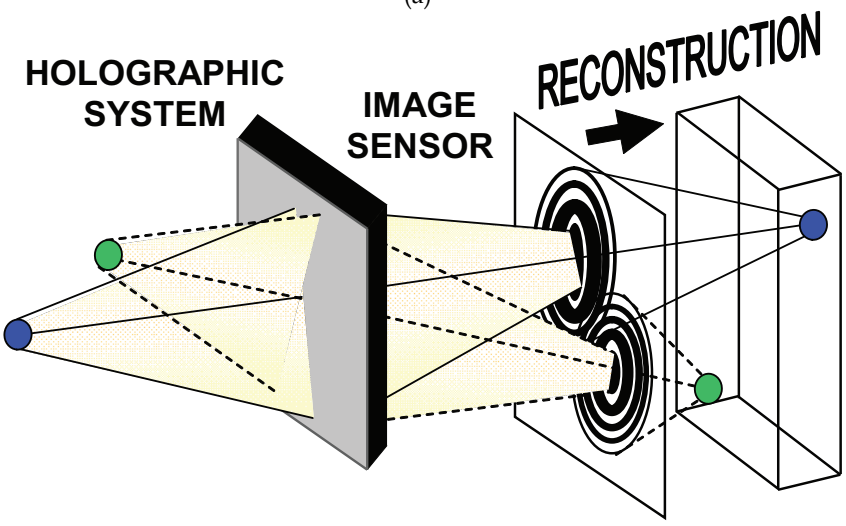
2. General properties of Fresnel holograms

The type of hologram discussed in this review is the digital Fresnel hologram, which means that a hologram of a single point has the form of the well-known FZP. The axial location of the object point is encoded by the Fresnel number of the FZP, which is the technical term for the quantity of the rings density in the FZP. To understand the operation principle of any general Fresnel hologram, let us look on the difference between regular and Fresnel-holographic imaging systems. In classical imaging, image formation of objects at different distances from the lens results in a sharp image at the image plane for objects at only one position from the lens, as shown in Fig. 1(a). The other objects at different distances from the lens are out of focus. Fresnel holographic system, on the other hand, as depicted in Fig. 1(b), projects a set of rings known as the FZP onto the plane of the image for each and every point at every plane of the object being viewed. The depth of the points is encoded by the density of the rings such that points which are closer to the system project denser rings than distant points. Because of this encoding method, the 3D information in the volume being imaged is recorded into the recording medium. Therefore, each plane in the image space reconstructed from a Fresnel hologram is in focus at a different axial distance. The encoding is accomplished by the presence of one of the holographic systems in the image path. Each holographic system, coherent or incoherent, has a different method to project the FZP on the image plane. At this point it should be noted that this graphical description of projecting

FZPs by every object's point actually expresses the mathematical two dimensional (2D) correlation (or convolution) between the object function and the FZP. In other words, the methods of creating Fresnel holograms are different from each other by the way they spatially correlate the FZP with the 3D scene. Another issue to note is that the correlation should be done with a FZP that is somehow 'sensitive' to the axial locations of the object points. Otherwise, these locations are not encoded into the hologram. The systems described in this review satisfies the condition that the FZP is dependent on the axial distance of each and every object point. This means that indeed points, which are far from the system, project FZP with fewer cycles per radial length than nearby points, and by this condition the holograms can actually image the 3D scene properly.



(a)



(b)

Fig. 1. Comparison of the Fresnel holography principle and conventional imaging. (a) Conventional imaging system. (b) Fresnel holography system.

The FZP is a sum of at least three main functions, a constant bias, a quadratic phase function and its complex conjugate. The object function is actually correlated with all these three functions. However the useful information, with which the holographic imaging is realized, is the correlation with just one of the two quadratic phase functions. The correlation with the other quadratic phase function induces the well-known twin image. This means that the detected signal in the holographic system contains three superposed correlation functions, whereas only one of them is the required correlation between the object and the quadratic phase function. Therefore, the digital processing of the detected signal should contain the ability to eliminate the two unnecessary terms.

The definition of Fresnel hologram is any hologram that contains at least, a correlation (or convolution) between an object function and a quadratic phase function. Moreover, the quadratic phase function must be parameterized according to the axial distance of the object points from the detection plane. In other words, the number of cycles per radial distance of each quadratic phase function in the correlation is dependent on the z distance of each object point. Formally, a hologram is called Fresnel hologram if its distribution function contains the following term:

$$H(u,v)=\iiint g(x,y,z)\exp\left\{i\frac{2\pi\beta}{z}\left[(u-x)^2+(v-y)^2\right]\right\}dxdydz, \tag{1}$$

where $g(x,y,z)$ is the 3D object function and β is a constant. Indeed, in Eq. (1) the phase of the exponent is dependent on z, the axial location of the object. In case the object is illuminated by coherent wave, $H(u,v)$ given by Eq. (1) is the complex amplitude of the coherent electromagnetic field directly obtained, under the paraxial approximation (Goodman, 1996), by a free space propagation from the object to the detection plane. However, we deal here with incoherent illumination, for which alternative methods to the free propagation should be applied. In fact, in this chapter we describe such method to get the desired correlation with the quadratic phase function given in Eq. (1), and this method indeed operates under incoherent illumination. Like any Fresnel holography, in the FINCH the object is correlated with a FZP, but the correlation is carried out without any movement and without multiplexing the image of the scene. Section 3 reviews the latest developments of the FINCH in the field of holography, microscopy and incoherent synthetic aperture imaging.

3. Fresnel incoherent correlation holography

Various aspects of the FINCH have been described in the literature, including FINCH for reflected white light (Rosen & Brooker, 2007a), FINCH of fluorescence objects (Rosen & Brooker, 2007b), FINCHSCOPE (Rosen & Brooker, 2008) – a holographic fluorescence microscope and finally SAFE (Katz & Rosen, 2010) – a FINCH based process of recording incoherent holograms in a synthetic aperture mode.

Generally, in the FINCH system the reflected incoherent light from a 3D object propagates through a spatial light modulator (SLM) and is recorded by a digital camera. To solve the twin image problem, three holograms are recorded sequentially, each with a different phase factor of the SLM pattern. The three holograms are superposed in the computer such that the result is a complex-valued Fresnel hologram. The 3D properties of the object are revealed by reconstructing this hologram in the computer.

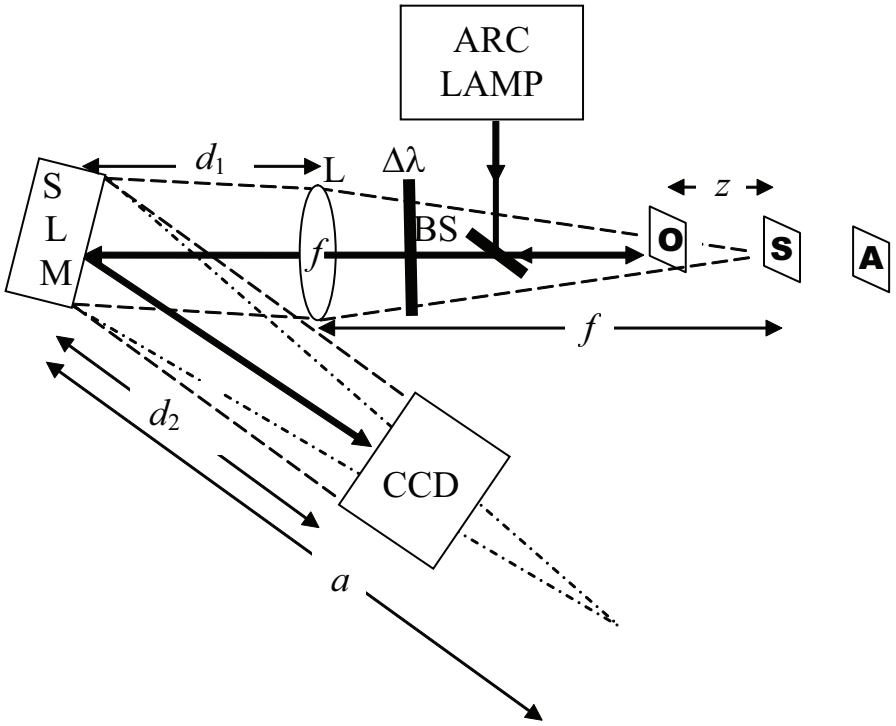


Fig. 2. Schematic of FINCH recorder. BS: beam splitter; SLM: spatial light modulator; CCD: charge-coupled device; L is a spherical lens with f=25cm focal length. Δλ indicates a chromatic filter with a bandwidth of Δλ=60nm.

3.1 FINCH for reflected white light

The first FINCH system (Rosen & Brooker, 2007a) is shown in Fig. 2. A white-light source illuminates a 3D object, and the reflected light from the object is captured by a digital camera after passing through a lens L and a SLM. In general, such a system can be analyzed as an incoherent correlator, where the SLM function is considered as a part of the system's transfer function. However, we find it easier to regard the system as an incoherent interferometer, where the grating displayed on the SLM is considered as a beam splitter. As is common in such cases, we analyze the system by following its response to an input object of a single infinitesimal point. Knowing the system's point spread function (PSF) enables one to realize the system operation for any general object. Analysis of a beam originated from a narrowband infinitesimal point source is done using Fresnel diffraction theory, (Goodman, 1996) since such a source is spatially coherent by definition.

A Fresnel hologram of a point object is obtained when the two interfering beams are, for instance, plane and spherical beams. Such a goal is achieved if the SLM's reflection function $R(x,y)$ is of the form

$$R(x,y)=\frac{1}{2}+\frac{1}{2}\exp\left[-\frac{i\pi}{\lambda a}\left(x^2+y^2\right)+i\theta\right]=\frac{1}{2}+\frac{1}{2}\mathrm{Q}\left(-\frac{1}{a}\right)\exp(i\theta). \tag{2}$$

For the sake of shortening, the quadratic phase function is designated by the function Q, such that $Q(b)=\exp[(i\pi b/\lambda)(x^2+y^2)]$. When a plane wave hits the SLM, the first constant term 1/2 in Eq. (2) represents the reflected plane wave, and the quadratic phase term is responsible for the reflected spherical wave in the paraxial approximation. The angle θ plays an important role later in the computation process to eliminate the twin image and the bias term.

A point source located at the point (x_s,y_s,z_s) a distance z_s from a spherical positive lens, with f focal length, induces on the lens plane a diverging spherical wave of the form of $Q(1/z_s)$. Right after the lens, which has a transmission function of $Q(-1/f)$, the complex amplitude is of a wave of the form $C_1(\bar{r}_s)Q(1/f_e)L(-\bar{r}_s/z_s)$ where $f_e=z_sf/(f-z_s)$ and the function L stands for a the linear phase function, such that $L(\bar{s})=\exp\left[i2\pi\lambda^{-1}\left(s_xx+s_yy\right)\right]$, $\bar{r}_s=(x_s,y_s)$ and $C_1(\bar{r}_s)$ is a complex constant dependent on the source point's location. On the digital camera plane at a distance d_2 from the SLM, the intensity of the recorded hologram is,

$$\begin{aligned} I_P(x,y) &= \left| C_1(\bar{r}_s)L\left(\frac{-\bar{r}_s}{z_s}\right)Q\left(\frac{1}{z_s}\right)Q\left(\frac{-1}{f}\right)*Q\left(\frac{1}{d_1}\right)\left[1+\exp(i\theta)Q\left(\frac{-1}{a}\right)\right]*Q\left(\frac{1}{d_2}\right)\right|^2 \\ &= \left| C_2(\bar{r}_s)L\left(\frac{-\bar{r}_sf_e}{z_s(f_e+d_1)}\right)Q\left(\frac{1}{f_e+d_1}\right)\left[1+\exp(i\theta)Q\left(\frac{-1}{a}\right)\right]*Q\left(\frac{1}{d_2}\right)\right|^2, \end{aligned} \tag{3}$$

where the asterisk denotes 2D convolution. The result of $I_P(x,y)$, after calculating the square magnitude in Eq. (3), is the PSF for any source point located at any point (x_s,y_s,z_s) on the object space of the FINCH, as follows,

$$\begin{aligned} I_P(x,y) = A_o\Bigg(& 2+\exp\left\{\frac{i\pi}{\lambda z_R}\left[\left(x-\frac{d_2f_ex_s}{z_s(f_e+d_1)}\right)^2+\left(y-\frac{d_2f_ey_s}{z_s(f_e+d_1)}\right)^2\right]+i\theta\right\} \\ & +\exp\left\{\frac{-i\pi}{\lambda z_R}\left[\left(x-\frac{d_2f_ex_s}{z_s(f_e+d_1)}\right)^2+\left(y-\frac{d_2f_ey_s}{z_s(f_e+d_1)}\right)^2\right]-i\theta\right\}\Bigg), \end{aligned} \tag{4}$$

where A_o is a constant and z_R, the reconstruction distance of the object point, is given by

$$z_R=\pm\frac{(f_1+d_2)(f_e+d_1+d_2)}{f_1-f_e-d_1}, \text{ where } f_1=\frac{a(f_e+d_1)}{a-(f_e+d_1)}. \tag{5}$$

The sign '±' in Eq. (5) indicates the possibility to reconstruct from the hologram either the virtual or the real image according to which term, second or third, is chosen from Eq. (4).

For the special object plane of the front focal plane of the lens where $z_s=f$ the expressions in Eqs. (4) and (5) become simpler because for that plane $f_e\to\infty$ and therefore $f_1=-a$ and $z_R=a-d_2$. The reconstruction distance of the point image from an equivalent optical hologram is z_R, although in the present case the hologram is of course digital, and the reconstruction is done by the computer. Note that z_R is obtained specifically in the case that one of the phase masks on the SLM is constant. This choice is used in all the FINCH experiments because practically the fill factor of the SLM is less than 100%, and therefore the constant phase modulation

inherently exists in the SLM. Consequently, choosing two diffractive lenses could cause unwanted three, instead of two, waves mixing on the hologram plane, one wave due to the constant phase and another two from the two different diffractive lenses.

Eq. (4) is the expression of the transparency function of a hologram created by an object point and recorded by a FINCH system. This hologram has several unique properties. The transverse magnification M_T is expressed as $M_T = \partial x_r / \partial x_s = d_2 / f$ for an object located on the front focal plane, and $M_T = f_e d_2 / z_s (f_e + d_1)$ for any other plane.

For a general 3D object g(x_s,y_s,z_s) illuminated by a narrowband incoherent illumination, the intensity of the recorded hologram is an integral of the entire PSF given by Eq. (4), over all the object intensity $g(x_s,y_s,z_s)$, as follows

$$H(x,y) \cong A_o \left(C + \iiint g(x_s,y_s,z_s) \exp\left\{ \frac{i\pi}{\lambda z_R} \left[\left(x - \frac{d_2 f_e x_s}{z_s (f_e + d_1)} \right)^2 + \left(y - \frac{d_2 f_e y_s}{z_s (f_e + d_1)} \right)^2 \right] + i\theta \right\} dx_s dy_s \, dz_s \right.$$
$$\left. + \iiint g(x_s,y_s,z_s) \exp\left\{ \frac{-i\pi}{\lambda z_R} \left[\left(x - \frac{d_2 f_e x_s}{z_s (f_e + d_1)} \right)^2 + \left(y - \frac{d_2 f_e y_s}{z_s (f_e + d_1)} \right)^2 \right] - i\theta \right\} dx_s dy_s dz_s . \right) \quad (6)$$

Besides a constant term *C*, Eq. (6) contains two terms of correlation between an object and a quadratic phase function, z_s-dependent via z_R, which means that the recorded hologram is indeed a Fresnel hologram. In order to remain with a single correlation term out of the three terms given in Eq. (6), we again follow the usual procedure of on-axis digital holography. Three holograms of the same object are recorded each of which with a different phase constant θ. The final hologram H_F is a superposition according to the following equation,

$$H_F(x,y) = H_1(x,y)\left[\exp(\pm i\theta_3) - \exp(\pm i\theta_2)\right] + H_2(x,y)\left[\exp(\pm i\theta_1) - \exp(\pm i\theta_3)\right]$$
$$+ H_3(x,y)\left[\exp(\pm i\theta_2) - \exp(\pm i\theta_1)\right], \quad (7)$$

where $H_i(x,y)$ is the *i*-th recorded hologram of the form of Eq. (6) and θ_i is the phase constant of the *i*-th SLM's quadratic phase used during the recording process. The choice between the signs in the exponents of Eq.(7) determines which image, virtual or real, is kept in the final hologram. A 3D image $g'(x,y,z)$ can be reconstructed from $H_F(x,y)$ by calculating the Fresnel propagation formula, as follows,

$$g(x,y,z) = H_F(x,y) * \exp\left[\frac{i\pi}{\lambda z} \left(x^2 + y^2 \right) \right], \quad (8)$$

The system shown in Fig. 2 was used to record the three holograms (Rosen & Brooker, 2007a). The SLM (Holoeye HEO 1080P) is phase-only, and as so, the desired function given by Eq. (2) cannot be directly displayed on this SLM. To overcome this obstacle, the phase function Q(-1/*a*) is displayed randomly on only half of the SLM pixels. These pixels were represented in the second term of Eq. (2), whereas the rest of the pixels representing the first constant term in Eq. (2) were modulated with a constant phase. The randomness in distributing the two phase functions has been required because organized non-random structure produces unnecessary diffraction orders, and therefore results in lower interference efficiency. The pixels were divided equally, half to each diffractive element, to

create two wavefronts with equal energy. By this method the SLM function becomes a good approximation to $R(x,y)$ of Eq. (2).

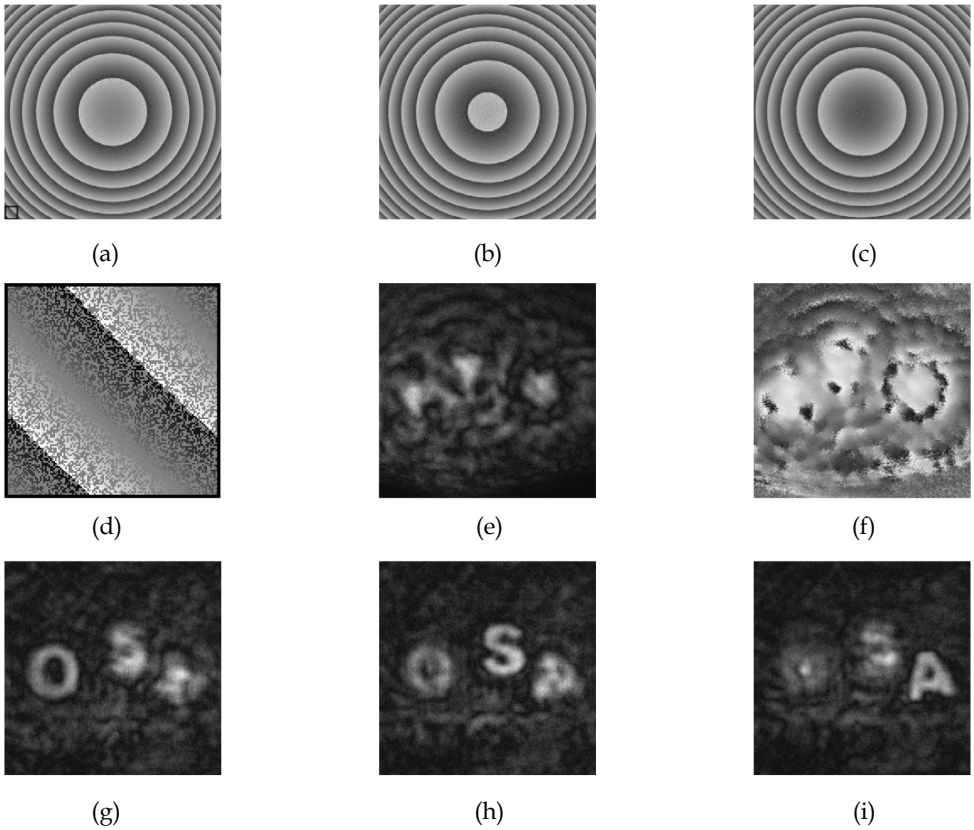


Fig. 3. (a) Phase distribution of the reflection masks displayed on the SLM, with θ=0^o, (b) θ=120^o, (c) θ=240^o. (d) Enlarged portion of (a) indicating that half (randomly chosen) of the SLM's pixels modulate light with a constant phase (e) Magnitude and (f) phase of the final on-axis digital hologram. (g)Reconstruction of the hologram of the three letters at the best focus distance of 'O'. (h) Same reconstruction at the best focus distance of 'S', and (i) of 'A'.

The SLM has 1920×1080 pixels in a display of 16.6×10.2mm, where only the central 1024×1024 pixels were used for implementing the phase mask. The phase distribution of the three reflection masks displayed on the SLM, with phase constants of 0^o, 120^o and 240^o, are shown in Figs. 3(a), (b) and (c), respectively. The other specifications of the system of Fig. 2 are: f=250 mm, a=430 mm, d_1=132 mm, d_2=260xmm.

Three white on black letters each of the size 2×2 mm were located at the vicinity of front focal point of the lens. 'O' was at z=-24 mm, 'S' was at z=-48 mm and 'A' was at z=-72 mm. These letters were illuminated by a mercury arc lamp. A filter which passed a Poisson-like power spectrum from 574 to 725 nm light with a peak wavelength of 599 nm and a bandwidth (full width at half maximum) of 60 nm was positioned between the beamsplitter

and the lens L. The three holograms, each for a different phase constant of the SLM, were recorded by a digital camera and processed by the computer. The final hologram $H_F(x,y)$ was calculated according to Eq. (7) and its magnitude and phase distributions are depicted in Figs. 3(e) and (f), respectively.

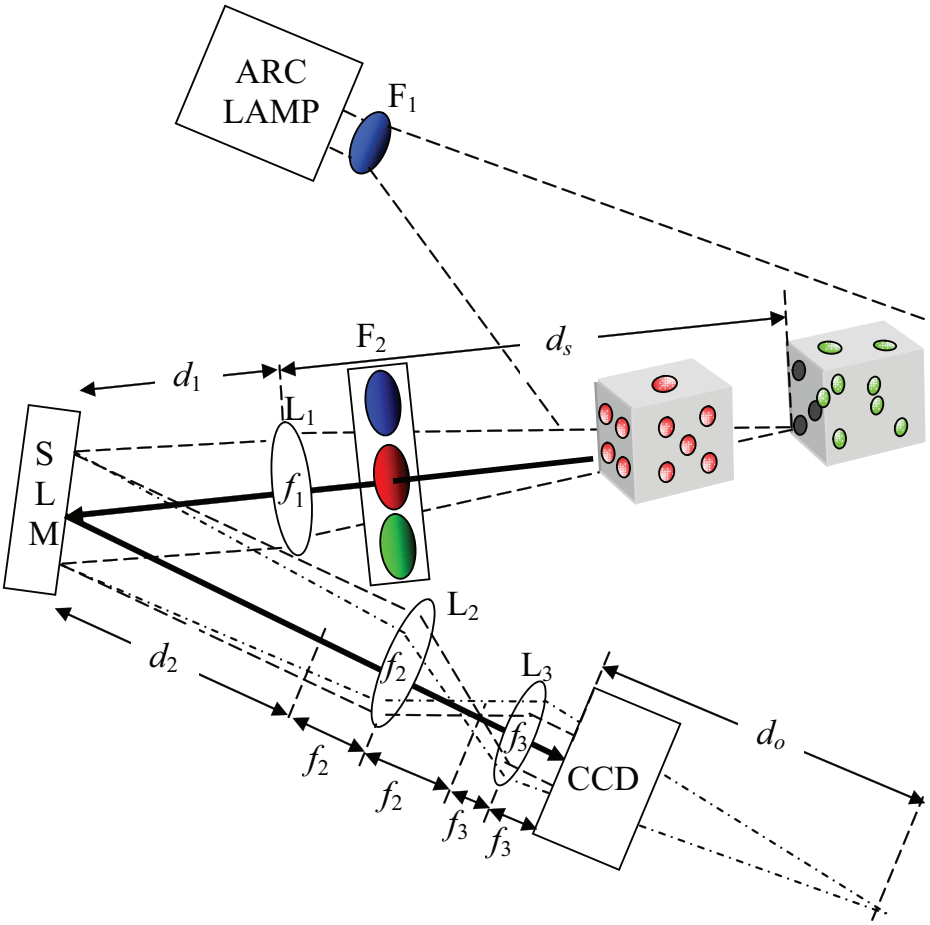


Fig. 4. Schematics of the FINCH color recorder. SLM: spatial light modulator; CCD: charge-coupled device; L_1, L_2, L_3 are spherical lenses and F_1, F_2 are chromatic filters.

The hologram $H_F(x,y)$ was reconstructed in the computer by calculating the Fresnel propagation toward various z propagation distances according to Eq. (8). Three different reconstruction planes are shown in Figs. 3(g), (h) and (i). In each plane, a different letter is in focus as is indeed expected from a holographic reconstruction of an object with a volume.

3.2. FINCH of fluorescence objects

In (Rosen & Brooker, 2007b) the FINCH has been capable to record multicolor digital holograms from objects emitting fluorescent light. The fluorescent light, specific to the

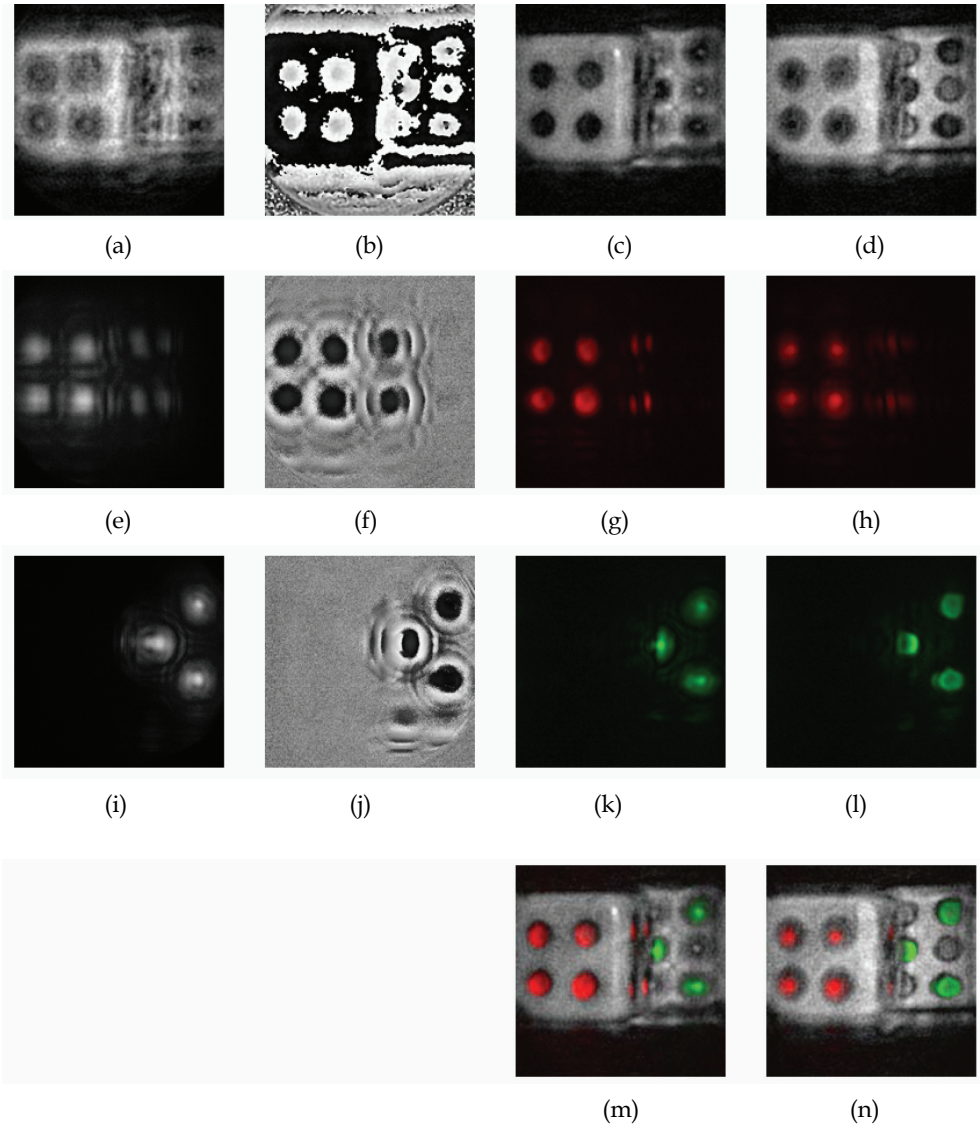


Fig. 5. (a) Magnitude and (b) phase of the complex Fresnel hologram of the dice. Digital reconstruction of the non-fluorescence hologram: (c) at the face of the red-dots on the die, and (d) at the face of the green dots on the die. (e) Magnitude and (f) phase of the complex Fresnel hologram of the red dots. Digital reconstruction of the red fluorescence hologram: (g) at the face of the red-dots on the die, and (h) at the face of the green dots on the die. (i) Magnitude and (j) phase of the complex Fresnel hologram of the green dots. Digital reconstruction of the green fluorescence hologram: (k) at the face of the red-dots on the die, and (l) at the face of the green dots on the die. Compositions of Figs 5(c), (g) and (k) and Figs 5(d), (h) and (l) are depicted in Fig. 5(m) and (n), respectively.

emission wavelength of various fluorescent dyes after excitation of 3D objects, was recorded on a digital monochrome camera after reflection from the SLM. For each wavelength of fluorescent emission, the camera sequentially records three holograms reflected from the SLM, each with a different phase factor of the SLM's function. The three holograms are again superposed in the computer to create a complex-valued Fresnel hologram of each fluorescent emission without the twin image problem. The holograms for each fluorescent color are further combined in a computer to produce a multicolored fluorescence hologram and 3D color image.

An experiment showing the recording of a color fluorescence hologram was carried out (Rosen & Brooker, 2007b) on the system shown in Fig. 4. The phase constants of $\theta_{1,2,3}=0^{o},120^{o},240^{o}$ were introduced into the three quadratic phase functions. The other specifications of the system are: f_1=250 mm, f_2=150 mm, f_3=35 mm, d_1=135 mm, d_2=206 mm. The magnitude and phase of the final complex hologram, superposed from the first three holograms, are shown in Figs. 5(a) and (b), respectively. The reconstruction from the final hologram was calculated using the Fresnel propagation formula of Eq. (8). The results are shown at the plane of the front face of the front die [Fig. 5(c)], and at the plane of the front face of the rear die [Fig. 5(d)]. Note that in each plane a different die face is in focus as is indeed expected from a holographic reconstruction of an object with a volume. The second set of three holograms was recorded via a red filter in the emission filter slider F_2 which passed 614 to 640 nm fluorescent light wavelengths with a peak wavelength of 626 nm and a bandwidth of 11 nm (FWHM). The magnitude and phase of the final complex hologram, superposed from the 'red' set, is shown in Figs. 5(e) and (f), respectively. The reconstruction results from this final hologram are shown in Figs. 5(g) and (h) at the same planes as shown in Figs. 5(c) and (d), respectively. Finally, an additional set of three holograms was recorded with a green filter in emission filter slider F_2, which passed 500 to 532nm fluorescent light wavelengths with a peak wavelength of 516nm and a bandwidth of 16nm (FWHM). The magnitude and phase of the final complex hologram, superposed from the 'green' set, is shown in Figs. 5(i) and (j), respectively. The reconstruction results from this final hologram are shown in Figs. 5(k) and (l) at the same planes as shown in Fig. 5(c) and (d), respectively. Compositions of Figs. 5(c), (g) and (k) and Figs. 5(d), (h) and (l) are depicted in Figs. 5(m) and (n), respectively. Note that all the colors in Fig. 5 are pseudo-colors. These last results yield a complete color 3D holographic image of the object including the red and green fluorescence. While the optical arrangement in this demonstration has not been optimized for maximum resolution, it is important to recognize that even with this simple optical arrangement, the resolution is good enough to image the fluorescent emissions with good fidelity and to obtain good reflected light images of the dice. Furthermore, in the reflected light images in figure 5(c) and 5(m) the system has been able to detect a specular reflection of the illumination from the edge of the front die.

3.3 FINCHSCOPE– a holographic fluorescence microscope

The next system to be reviewed here is the first demonstration of a motionless microscopy system (FINCHSCOPE) based upon the FINCH, and its use in recording high-resolution 3D fluorescent images of biological specimens (Rosen & Brooker, 2008). By using high-numerical-aperture lenses, a spatial light modulator, a charge-coupled device camera and some simple filters, FINCHSCOPE enables the acquisition of 3D microscopic images without the need for scanning.

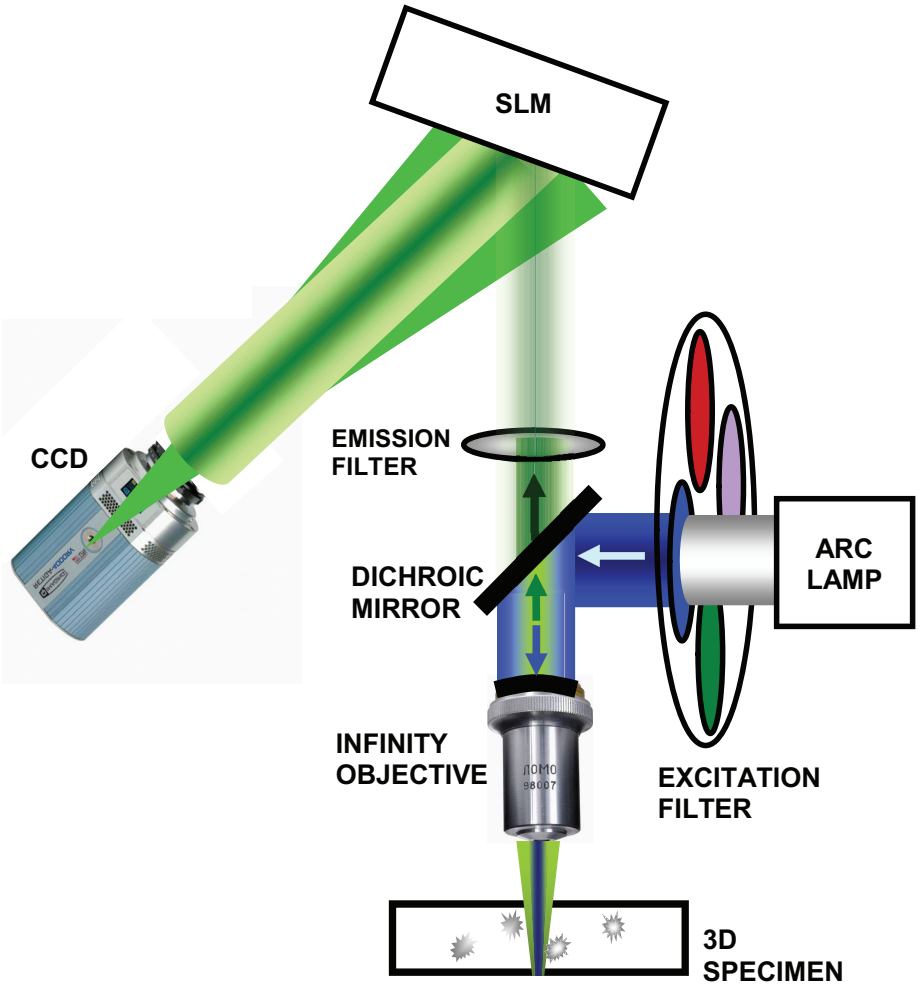


Fig. 6. FINCHSCOPE schematic in upright fluorescence microscope. The upright microscope was modified with a reflective SLM positioned at a tilt angle of 11° to reflect emission light from the objective onto the camera.

A schematic diagram of the FINCHSCOPE for an upright microscope equipped with an arc lamp source is shown in Fig. 6. The beam of light that emerges from an infinity-corrected microscope objective transforms each point of the object being viewed into a plane wave, thus satisfying the first requirement of FINCH (Rosen & Brooker, 2007a). An SLM and a digital camera replace the tube lens, reflective mirror and other transfer optics normally present in microscopes. Because no tube lens is required, infinity-corrected objectives from any manufacturer can be used. A filter wheel was used to select excitation wavelengths from a mercury arc lamp, and the dichroic mirror holder and the emission filter in the microscope were used to direct light to and from the specimen through infinity-corrected objectives.

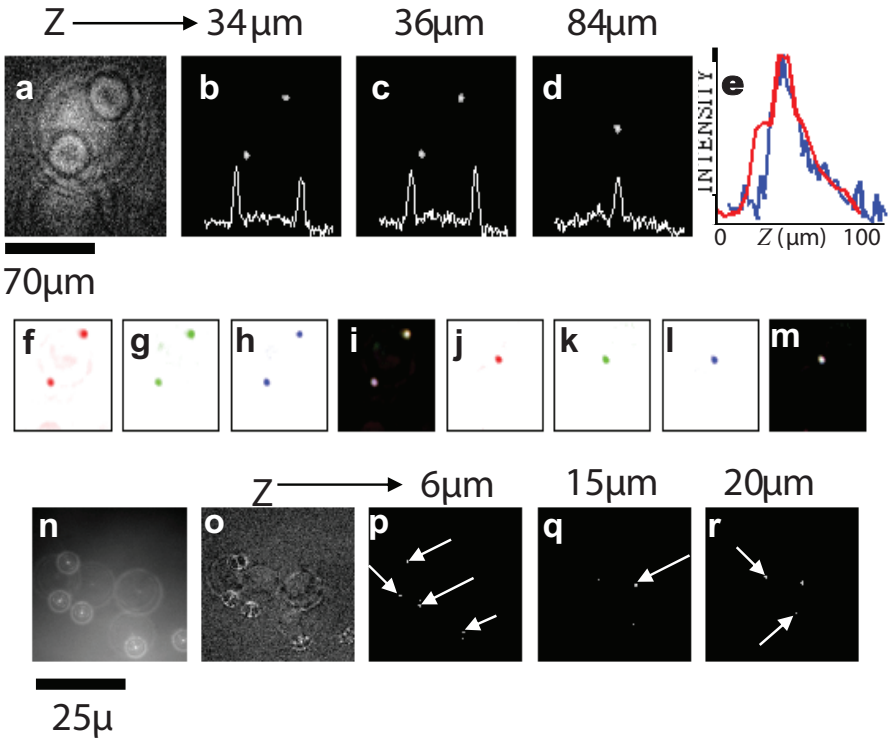


Fig. 7. FINCHSCOPE holography of polychromatic beads. (a) Magnitude of the complex hologram 6 μm beads. Images reconstructed from the hologram at z distances of (b) 34 μm, (c) 36 μm and (d) 84 μm. Line intensity profiles between the beads are shown at the bottom of panels b–d. (e) Line intensity profiles along the z axis for the lower bead from reconstructed sections of a single hologram (blue line) and from a widefield stack of the same bead (28 sections, red line). (f)–(h) Beads (6 μm) excited at 640, 555 and 488 nm with holograms reconstructed at planes b and (j–l) d. (i) and (m) are the combined RGB images for planes b and d, respectively. (n)–(r), Beads (0.5 μm) imaged with a 1.4-NA oil immersion objective: (n) holographic camera image; (o) magnitude of the complex hologram; (p–r), reconstructed image planes 6, 15 and 20 μm. Scale bars indicate image size.

The ability of the FINCHSCOPE to resolve multicolour fluorescent samples was evaluated by first imaging polychromatic fluorescent beads. A fluorescence bead slide with the beads separated on two separate planes was constructed. FocalCheck polychromatic beads (6 μm) were used to coat one side of a glass microscope slide and a glass coverslip. These two surfaces were juxtaposed and held together at a distance from one another of ~50 μm with optical cement. The beads were sequentially excited at 488, 555 and 640 nm center wavelengths (10–30 nm bandwidths) with emissions recorded at 515–535 nm, 585–615 nm and 660–720 nm, respectively. Figures 7(a)–(d) show reconstructed image planes from 6 μm beads excited at 640 nm and imaged on the FINCHSCOPE with a Zeiss PlanApo ×20, 0.75 NA objective. Figure 7(a) shows the magnitude of the complex hologram, which contains all the information about the location and intensity of each bead at every plane in the field. The Fresnel reconstruction from this hologram was selected to yield 49 planes of the image, 2 μm

apart. Two beads are shown in Fig. 7(b), with only the lower bead exactly in focus. The next image [Fig. 7(c)] is 2 μm into the field in the Z-direction, and the upper bead is now in focus, with the lower bead slightly out of focus. The focal difference is confirmed by the line profile drawn between the beads, showing an inversion of intensity for these two beads between the planes. There is another bead between these two beads, but it does not appear in Figs. 7(b) or (c) (or in the intensity profile), because it is 48 μm from the upper bead; it instead appears in Fig. 7(d) (and in the line profile), which is 24 sections away from the section in Fig. 7(c). Notice that the beads in Figs. 7(b) and (c) are no longer visible in Fig. 7(d). In the complex hologram in Fig. 7(a), the small circles encode the close beads and the larger circles encode the distant central bead. Figure 7(e) shows that the Z-resolution of the lower bead in Fig. 7(b), reconstructed from sections created from a single hologram (blue line), is at least comparable to data from a widefield stack of 28 sections (obtained by moving the microscope objective in the Z-direction) of the same field (red line). The co-localization of the fluorescence emission was confirmed at all excitation wavelengths and at extreme Z limits as shown in Figs. 7(f)–(m) for the 6 μm beads at the planes shown in Figs. 7(b) (f–i) and 7(d) (j–m). In Figs. 7(n)–(r), 0.5 μm beads (TetraSpeck, Invitrogen) imaged with a Zeiss PlanApo ×63 1.4 NA oil-immersion objective are shown. Figure 7(n) presents one of the holograms captured by the camera and Fig. 7(o) shows the magnitude of the complex hologram. Figures 7(p)–(r) show different planes (6, 15 and 20 μm, respectively) in the bead specimen after reconstruction from the complex hologram of image slices in 0.5 μm steps. Arrows show the different beads visualized in different Z image planes.

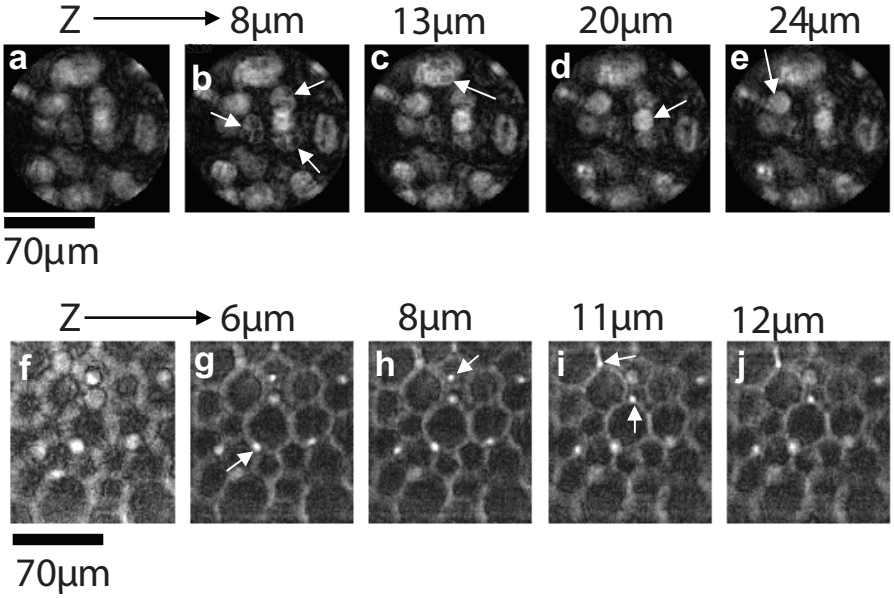


Fig. 8. FINCHSCOPE fluorescence sections of pollen grains and *Convallaria rhizome*. The arrows point to the structures in the images that are in focus at various image planes. (b)–(e), Sections reconstructed from a hologram of mixed pollen grains. (g)–(j), Sections reconstructed from a hologram of *Convallaria rhizome*. (a),(f), Magnitude of the complex holograms from which the respective image planes were reconstructed. Scale bars indicate image sizes.

The computer reconstruction along the Z-axis of a group of fluorescently labeled pollen grains (Carolina Biological slide no. 30-4264) is shown in Figs. 8(b)–(e). As is expected from a holographic reconstruction of a 3D object with volume, any number of planes can be reconstructed. In this example, a different pollen grain was in focus in each transverse plane reconstructed from the complex hologram whose magnitude is shown in Fig. 8(a). In Figs. 8(b)–(e), the values of Z are 8, 13, 20 and 24 μm, respectively. A similar experiment was performed with the autofluorescent Convallaria rhizome and the results are shown in Figs. 8(g)–(j) at planes 6, 8, 11 and 12 µm.

3.4 SAFE- Synthetic aperture with Fresnel elements

The recent development in FINCH is a new lensless incoherent holographic system operating in a synthetic aperture mode (Katz & Rosen, 2010). Synthetic aperture is a well-known super-resolution technique which extends the resolution capabilities of an imaging system beyond the theoretical Rayleigh limit dictated by the system's actual aperture. Using this technique, several patterns acquired by an aperture-limited system, from various locations, are tiled together to one large pattern which could be captured only by a virtual system equipped with a much wider synthetic aperture.

The use of optical holography for synthetic aperture is usually restricted to coherent imaging (Beck, et al., 2005; Mico, et al. 2006; Martínez-León & Javidi, 2008). Therefore, the use of this technique is limited only to those applications in which the observed targets can be illuminated by a laser. Synthetic aperture carried out by a combination of several off-axis incoherent holograms in scanning holographic microscopy has been demonstrated by (Indebetouw et al., 2007). However, this method is limited to microscopy only, and although it is a technique of recording incoherent holograms, a specimen should also be illuminated by an interference pattern between two laser beams.

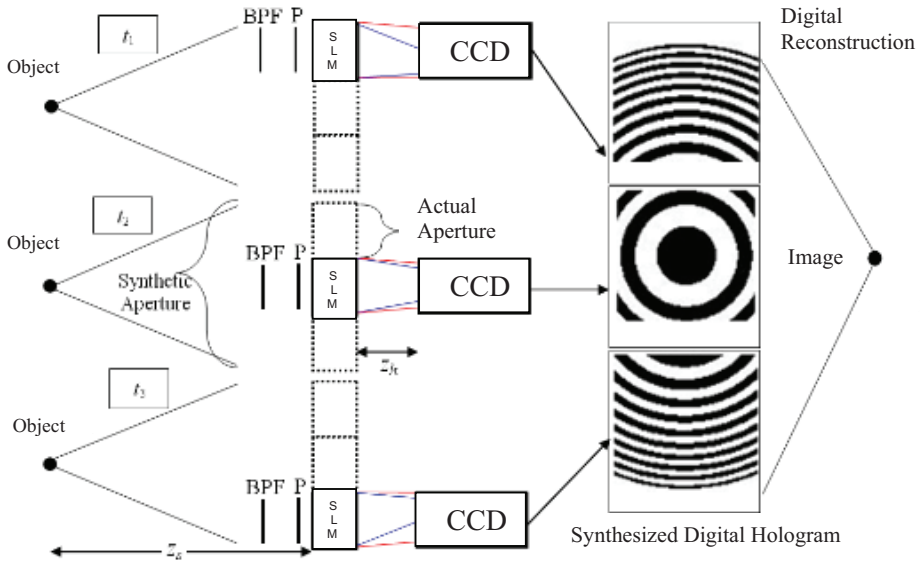


Fig. 9. Scheme of SAFE operating as synthetic aperture radar to achieve super-resolution. P indicates polarizer and BPF stands for band-pass filter.

Our new scheme of holographic imaging of incoherently illuminated objects is dubbed a synthetic aperture with Fresnel elements (SAFE). This holographic lensless system contains a band-pass filter (BPF), a polarizer, an SLM and a digital camera. SAFE has an extended synthetic aperture in order to improve the transverse and axial resolutions beyond the classic limitations. The term synthetic aperture, in the present context, means time (or space) multiplexing of several Fresnel holographic elements captured from various viewpoints by a system with a limited real aperture. The synthetic aperture is implemented by shifting the BPF-polarizer-SLM-camera set, located across the field of view, between several viewpoints. At each viewpoint a different mask is displayed on the SLM, and a single element of the Fresnel hologram is recorded (See Fig. 9). The various elements, each of which is recorded by the real aperture system during the capturing time, are tiled together so that the final mosaic hologram is effectively considered as captured from a single synthetic aperture which is much wider than the actual aperture.

An example of such system with the synthetic aperture, which is three times wider than the actual aperture, can be seen in Fig. 9. For simplicity of the demonstration, the synthetic aperture was implemented only along the horizontal axis. In principle this concept can be generalized for both axes and for any ratio of synthetic to actual apertures. Imaging with the synthetic aperture is necessary for cases where the angular spectrum of the light emitted from the observed object is wider than the numerical aperture of a given imaging system. In SAFE shown in Fig. 9, the BPF-polarizer-SLM-camera set moves in front of the object. The complete Fresnel hologram of the object, located at some distance from the SLM, is a mosaic of three holographic elements, each of which is recorded from a different position by the system with the real aperture of the size $A_x \times A_y$. The complete hologram tiled from the three holographic Fresnel elements has the synthetic aperture of the size $3 \cdot A_x \times A_y$ which is three times larger than the real aperture at the horizontal axis.

The method to eliminate the twin image and the bias term is the same as has been used before; three elemental holograms of the same object and for each point of view are recorded, each of the holograms has a different phase constant of the SLM's phase mask. The final holographic element is a specific superposition of the three recorded elements. The digital reconstruction of the final complex-valued mosaic hologram is conventionally computed by Fresnel back propagation.

SAFE has been tested in the lab by the system shown in Fig. 9. The object in this experiment is a binary grating with cycle length of 4 lines per *mm*. The distance from the object to the SLM has been 52 *cm*, and the distance between the phase-only SLM (Holoeye, PLUTO) and the digital camera (E-VISION, EVC6600SAM-GE5) has been 38.5 cm. A 100W Halogen ARC lamp has been used for objects illumination, and a BPF (with an 80 nm bandwidth surrounding 550 nm central wavelength) has been placed just in front of the SLM. The results of the experiments are summarized in Fig. 10. In the first experiment we have recorded a hologram only by the actual aperture without shifting the system, in the setup shown in Fig. 9 at the time t_2. Fig. 10(a) shows one of the three masks displayed on the SLM in this experiment. Each of the three masks has one of the three different phase factors: 0^o, 120^o or 240^o. As mentioned above, these three phase masks with different phase constants are required in order to eliminate the bias term and the twin image from the holographic reconstruction. As stated earlier, another problem with the SLM is that its fill factor is 87%, which means that part of the light is reflected from the SLM without any modulation.

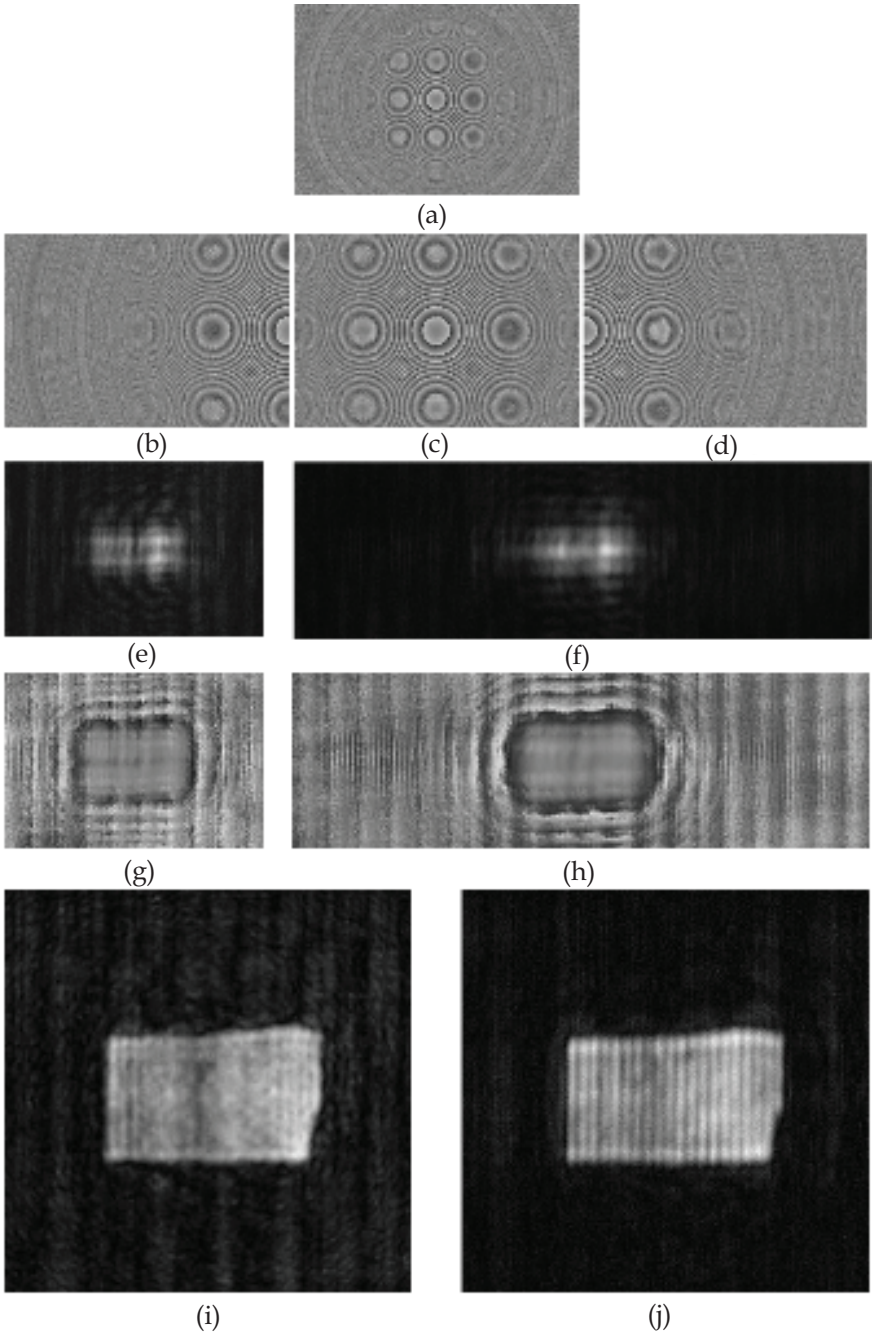


Fig. 10. Results of SAFE for the grating object with the real aperture and the synthetic aperture. See text for details.

In order to avoid the interference of three waves projected on the camera, we have chosen one of the phase elements to be constant. The other phase element has been chosen to be a negative diffractive lens with the shortest focal distance that can be achieved with the SLM having the pixel size of 8 μm. The shortest focal distance guarantees maximum resolution power for a given aperture size. In the case of the actual aperture (1500×1000 pixels) and the synthetic aperture (3000×1000 pixels), the focal distances have been -34 *cm* and -68 *cm*, respectively. The input numerical aperture is 0.0115 and 0.0231 for the real and synthetic apertures, respectively, whereas the output numerical aperture is 0.0035 and 0.0044 for the real and synthetic apertures, respectively. Note that the sum of two pure phase functions, i.e., the quadratic phase function $Q[-1/a]$ and the constant phase function in Eq. (2), is no longer a pure phase function, but a complex function with non-constant magnitude. Since the SLM is a phase-only modulator, we use the previous method of recording general complex function on a phase-only SLM. Each phase function is distributed randomly among half of the SLM pixels.

The three recorded holograms are superposed according to the same superposition equation (7). Figs. 10(e), 10(g) are the magnitude and the phase of the superposed holograms for the object. It can be seen that the resolution along the horizontal direction of the reconstructed image, computed by Fresnel back propagation, is damaged in the sense that the image is lacking the original high-frequency gratings along the horizontal direction because the aperture is too narrow to capture the entire gratings spectral content. This damaged reconstructed image is shown in Fig. 10(i).

In the SAFE experiment nine different phase masks have been displayed on the SLM, three for each location of the SLM-camera set; left, central and right. Each of the masks has an actual aperture of 1500×1000 pixels. Each of the three masks at every location has one of the three different phase factors: 0^o, 120^o or 240^o. In order to avoid edge effects on the recorded holograms there is an overlap of 750 pixels among the three actual apertures combining the synthetic aperture. For each location of the system, the three recorded holograms have been superposed as mentioned above. Figs. 10(b)-10(d) represent three masks out of nine, each of which has been displayed at different time and at a different location of the setup along the horizontal axis. The superposed complex-valued holographic element from each system's viewpoint is stored in the computer. Upon completing the system movement along the entire synthetic aperture, all three holographic elements are tiled to a single mosaic hologram. Figs. 10(f) and 10(h) represent the magnitude and the phase of the complete mosaic hologram. The reconstruction result of the mosaic hologram, computed by Fresnel back propagation, is depicted in Fig. 10(j). The binary grating on the observed objects is seen well in the reconstructed images, indicating that the synthetic aperture is wide enough to acquire most of the horizontal spectral information of the objects.

5. Discussion and conclusions

We have reviewed a new method of generating incoherent digital Fresnel holograms. The reviewed hologram, the FINCH, is actually recorded by an on-axis, single-channel, incoherent interferometer. This method inherently does not scan the object neither in the space nor in the time. Therefore, the FINCH can generate the holograms rapidly without sacrificing the system resolution. This system offers the feature of observing a complete volume from a hologram, potentially enabling objects moving quickly in three dimensions to be tracked. The FINCH technique shows great promise in rapidly recording 3D information in any scene, independently of the illumination. In addition, we have described

a rapid, non-scanning holographic fluorescence microscope that produces in-focus images at each plane in the specimen from holograms captured on a digital camera. This motionless 3D microscopy technique does not require complicated alignment or a laser. The fluorescence emission can be of relatively wide bandwidth because the optical path difference between the beams is minimal in this single-path device. Although at present each reconstructed section is not completely confocal, 3D reconstructions free of blur could be created by deconvolution of the holographic sections as is typically carried out in widefield microscopy. Time resolution is currently reduced because three holograms need to be captured sequentially. However, in the future, it will be possible to record the three holograms faster using more sensitive cameras, simultaneously capture all three holograms, or to overcome the holographic twin image problem and capture only one hologram, as any of the three holograms contain all the essential 3D information. In the present studies the image sections were obtained by a process of first capturing three holograms, computing the image z sections from the complex hologram and then, in some cases, further enhancing them by deconvolution. This process could be simplified in the future for real-time display of the holographic image, either with a holographic display system or by algorithms that create the enhanced sections and the 3D representation directly from the single hologram. There is no need for sectioning or scanning or any mechanical movement. Therefore, this system would be expected ultimately to be faster, simpler and more versatile than existing 3D microscopy techniques, which rely on pinhole imaging or deconvolution of stacks of widefield images. At present, the FINCHSCOPE is already considerably faster than conventional 3D sectioning. For example, the total image capture time for the three FINCHSCOPE images of the pollen grains in Fig. 8 was just over 1 s, compared with the 30–45 s needed to create a stack of 48 widefield or spinning disk confocal images. We have also demonstrated fluorescence holography using the high-NA objectives widely used in biological imaging. FINCHSCOPE is able to spatially resolve small beads, biological specimens and different fluorescence emission colours in x, y and z planes with perfect registration. The system provides a simple, flexible, cost-effective and powerful microscopic platform for 3D imaging.

Finally, we have demonstrated a process of recording incoherent holograms in the synthetic aperture mode. The synthetic aperture of SAFE considerably increases both the transverse and the axial resolving power. The concept of the present system can be applied to all regimes of imaging from microscopy to telescopes, and either for 2D or 3D imaging. Our demonstration of this advance in imaging, based on a new, but simple holographic principle, should open up opportunities in many life science and engineering fields, so that interesting scenes may be readily observed in three dimensions and possibly at higher resolution than with currently existing techniques.

6. References

Beck, S. M.; Buck, J. R.; Buell, W. F.; Dickinson, R. P.; Kozlowski, D. A.; Marechal, N. J. & Wright, T. J. (2005). Synthetic-aperture imaging laser radar: laboratory demonstration and signal processing. *Applied Optics* 44(35), 7621–7629

Breckinridge, J. B. (1974). Two-Dimensional White Light Coherence Interferometer. *Applied Optics* 13(12), 2760 -2762

Cochran, G. (1966). New method of making Fresnel transforms with incoherent light. *J. Optical Society of America* 56(11), 1513–1517

Goodman, J. W. (1996). *Introduction to Fourier Optics;* 2nd ed. McGraw-Hill: New York, NY.

Indebetouw, G.; Tada, Y.; Rosen, J. & Brooker, G. (2007). Scanning holographic microscopy with resolution exceeding the Rayleigh limit of the objective by superposition of off-axis holograms. *Applied Optics* 46(6), 993–1000

Katz, B. & Rosen, J. (2010). Super-resolution in incoherent optical imaging using synthetic aperture with Fresnel elements. *Optics Express* 18(2), 962–972

Li, Y.; Abookasis, D. & Rosen, J. (2001). Computer-generated holograms of three-dimensional realistic objects recorded without wave interference. *Applied Optics* 40(17), 2864-2870

Lohmann, A. W. (1965). Wavefront reconstruction for incoherent objects. *J. Optical Society of America* 55(11), 1555–1556.

Marathay, A. S. (1987). Noncoherent-object hologram: its reconstruction and optical processing. *J. Optical Society of America* A 4(10), 1861-1868

Martínez-León, L. & Javidi, B. (2008). Synthetic aperture single-exposure on-axis digital holography. *Optics Express* 16(1), 161–169

Mico, V.; Zalevsky, Z.; García-Martínez, P. & García, J. (2006) Synthetic aperture superresolution with multiple off-axis holograms. *J. Optical Society of America A* 23(12), 3162–3170

Mugnier, L. M.; Sirat, G. Y. & Charlot, D. (1993). Conoscopic holography: two-dimensional numerical reconstructions. *Optics Letters* 18(1), 66-68

Park, J.-H.; Kim, M.-S.; Baasantseren, G. & Kim, N. (2009). Fresnel and Fourier hologram generation using orthographic projection images. *Optics Express* 17(8), 6320–6334

Peters, P. (1966). Incoherent holograms with a mercury light source. *Applied Physics Letters* 8(8), 209–210

Poon, T.-C. & Korpel, A. (1979). Optical transfer function of an acousto-optic heterodyning image processor. *Optics Letters* 4(10), 317-319

Poon, T.-C. (2004). Recent Progress in Optical Scanning Holography. *J. Holography Speckle* 1(1), 6-25

Poon, T.-C. (2007). *Optical Scanning Holography with MATLAB*. Springer, New York

Rosen, J.; Indebetouw, G. & Brooker, G. (2006). Homodyne scanning holography. *Opt Express*, 14(10), 4280-4285.

Rosen, J. & Brooker, G. (2007a). Digital spatially incoherent Fresnel holography. *Optics Letters* 32(8), 912-914

Rosen, J. & Brooker, G. (2007b). Fluorescence incoherent color holography. *Optics Express* 15(5), 2244-2250

Rosen, J. & Brooker, G. (2008). Non-Scanning Motionless Fluorescence Three-Dimensional Holographic Microscopy. *Nature Photonics* 2, 190-195

Sando, Y.; Itoh, M. & Yatagai, T. (2003). Holographic three-dimensional display synthesized from three-dimensional Fourier spectra of real-existing objects. *Optics Letters* 28(24), 2518-2520

Schilling, B. W.; Poon, T.-C.; Indebetouw, G.; Storrie, B.; Shinoda, K.; Suzuki, Y. & Wu, M. H. (1997). Three-dimensional holographic fluorescence microscopy. *Optics Letters* 22(19), 1506-1508

Shaked, N. T. & Rosen, J. (2008). Multiple-viewpoint projection holograms synthesized by spatially incoherent correlation with broadband functions. *J. Optical Society of America A* 25(8), 2129-2138

Stroke, G. W. & Restrick, R. C. III, (1965). Holography with spatially noncoherent light. *Applied Physics Letters* 7(9), 229–231

Worthington, Jr. H. R. (1966). Production of holograms with incoherent illumination. *J. Optical Society of America* 56(10), 1397–1398

7

Programmable Point-source Digital In-line Holography Using Digital Micro-mirror Devices

Adekunle A. Adeyemi and Thomas E. Darcie
University of Victoria, Victoria BC
Canada

1. Introduction

Digital holographic imaging techniques allow fast retrieval of three-dimensional (3-D) amplitude and phase information of an object volume through numerical reconstruction of a two-dimensional (2-D) hologram. Digital holography consists of digital sampling of a hologram on an array of charged-coupled device detectors (CCD), and digital reconstruction of the object field through a numerical algorithm [Sch94, SJ02, Yar03]. The recording process encodes 3-D information of an object into the form of interference fringes on a two-dimensional recording screen. These fringes usually contain high spatial frequencies that represent the mixing between the scattered object field and coherent reference wave. A reconstruction process is performed on the recorded hologram to recover the object wave. This numerical acquisition method eliminates the need for any chemical processing of the hologram and mechanical refocusing of the reconstructed image that is commonly required in the traditional holographic imaging system [CMD99, DJL99, SPISSSW97]. This process has opened new frontiers in digital holography with emerging applications in research and industry [Kre05, SJ05].

Typically, common digital holography recording set-ups include off-axis and in-line configurations [SJ94]. Digital in-line holography (DIH) represents the simplest realization of the digital holography (DH), allowing for rapid acquisition of hologram images without the use of lenses. Recently, DIH with a spherical reference field has emerged as an attractive tool in 3D imaging of biological objects and micro-spheres without the use of lenses [GXJKJK06a, RGMNS06, XJMK02a, XJMK02b]. Basically, the configuration consists of a coherent light source – a pinhole – to generate a spherical reference field. An array of charge-coupled device detectors (CCD) provides digital sampling of a hologram. However, the characteristics and parameters of these components can affect the overall performance of the system.

Among factors that limit the performance of a digital in-line holographic microscope (DIHM) are the size and spatial location of the pinhole used. These affect the resolution, obtainable field of view (FOV) and object illumination angle (which determines the projection view in the reconstructed image volume.) In [GPO08], the effect of the pinhole size on the spatial coherence of the reference beam in a DIHM system was studied. The results showed that a reduction in the coherence of the light, due to increase in the size of the pinhole used, leads to broadening of the impulse response of the system. Consequently, this limits the obtainable resolution in the reconstructed image. Other resolution-limiting

factors in DIHM include the size and position of the CCD, pixel density in the hologram, location of the object between the pinhole and CCD, and wavelength of the reference wave [GXJKJK06b, JGXJK06, RPP04, KJMX01].

Another limiting effect of the size of pinhole used is the restriction of the field of view (area captured) in the reconstructed image, especially when imaging an object that extends over an area of larger dimension compared to the numerical aperture of the system. In [GXJKJK06a, GXJKJK06b], it has been shown that the illuminated area in the object plane, as well as the FOV depends on the pinhole size. Provided the CCD is large enough to capture sufficient interference fringes, the higher spatial coherence obtained from smaller pinhole leads to an increase in the illuminated area in the object plane and hence wider FOV.

For three-dimensional imaging, the spatial location of the pinhole determines the object illumination angle and projection view in the reconstructed image. Thus the use of static pinhole in the current DIHM configuration limits the object illumination angle to that provided by the pinhole location. Consequently, this also restricts the angular directional view of a reconstructed image volume to the position of the pinhole. These projections can potentially be used to extract depth information between different planes through the image volume.

The aim of this chapter is to describe one possible way to address key limitations in the present configuration of DIHM, especially in recording different projections to extract depth information by point-source translation. This involves recording of holograms at different spatial locations of the point-source. The numerically reconstructed images of these holograms can be either combined into a single image with a wider FOV or used to extract depth information in the image volume. Among other techniques of achieving a scanning point source, a DIHM system that utilized digital micro-mirror devices (DMD) to program the source of the spherical reference field (point source) has been proposed [AD09a]. The DMD is a silicon-based reflective spatial light modulator that consists of more than a million individually addressable and switchable aluminum mirror pixels on a complementary metal oxide semiconductor (CMOS) static random access memory. Early application of the DMD is in projection systems but several emerging applications have evolved and these include confocal microscopy, high dynamic range imaging, 3D metrology and holography [ABD09, BPS03, DM99]. Applications of DMD in off-axis digital holographic recording and reconstruction have been demonstrated [DDS03, KN98].

In DIHM with a programmed point-source, the scanning of the spatial location of the point source was achieved by using the DMD as the primary source of the spherical reference field that illuminates the object. Thus, translation of the addressable ON-state mirror pixel on the DMD plane leads to scanning of the point source with respect to the object position. In addition to programming the point-source, another advantage of this system is the flexibility in reconfiguring the point-source size by simply varying the number of the ON-state DMD pixels. The application of the system to enhancement of the limited FOV in DIHM has been reported [AD09a]. The results obtained through numerical reconstruction of the recorded-translated holograms of Latex micro-spheres deposited on a microscope slide show the possibility of expanding the field of view by 263%. Recently [AD09b], we have extended the application of the system, using microspheres deposited on both side of microscope slide, to record different projections of holograms. Early published results show the possibility of recording different projections of the hologram.

In this chapter, we provide an in-depth analysis on the application of the programmable point-source DIHM to extract depth information using projections of the reconstructed

image volume. A relationship between the change in illumination angle and the corresponding depth separation using a point source on two parallel planes is established. A review on the expansion of the field of view in DIHM is provided as a complete analysis and demonstration of the capability of the programmable point-source DIHM. We begin with a detailed description of the DMD-based programmable point-source DIHM system and a review of the hologram recording and reconstruction techniques.

2. DMD-based programmable point-source DIHM system

The DMD plane axis and schematic diagram of a spherical reference field digital in-line holographic microscope with a programmable point source are shown in Fig. 1 (a) and (b), respectively. A collimated laser beam (0.38 watts, 532nm wavelength and 3mm in diameter) illuminates the Texas Instruments DMD chip with 1024 x 768 pixel elements; each mirror is 13.68 μm x 13.68 μm square in size, with 14.68 μm pitch.

We mounted the DMD chip at 45 degrees to the surface of the optical table with illumination beam directed at about 20 degrees to the chip normal so that light from the ON-state mirror element is reflected perpendicularly to the chip. One of the main challenges we encountered during illumination of DMD array with coherent light was the appearance of interference fringes when a mirror element is switched ON. This was caused by the interference between the background diffraction light (resulting from the spreading of the illumination beam by the 2-D periodic pattern of the square elements on the DMD array) and light reflected from the ON-state mirror element. We noted the background diffracted orders from the array are fixed in space and any changes to the states of the mirror elements only redistribute the intensity of light among the orders.

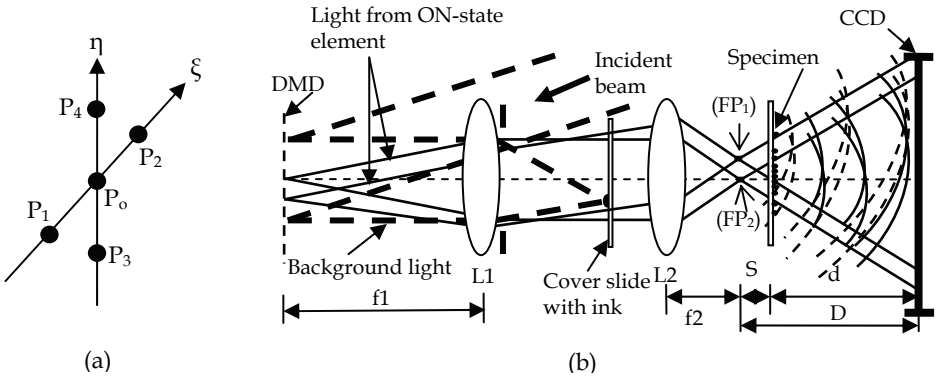


Fig. 1. Schematic diagram of the programmable pinhole DIH

To reduce this fringe effect, we employ a simple but effective method to spatially separate the ON-state reflected light from the background. This is achieved by locating lens L1 (f1=150mm and 25mm diameter) at a focal distance from the DMD elements such that the diffracted-collimated background light is focused at a certain distance behind the lens and ON-state reflected light is collimated behind the lens simultaneously, as shown in Fig. 1 (b). An ink drop on a cover slide (perpendicular to the beam path and positioned at back focal plane of L1) is used to stop the focused background diffracted light that falls directly within

the field of view, while the other background diffracted orders are blocked with an iris located behind lens L1. In order to completely block this focused background diffracted light, the diameter of the ink spot is made slightly larger than the focused spot. The light from the ON-state element, transmitted through the cover slide, is brought into focus by lens L2 (f2=25mm and 25mm diameter). This focused spot represents the point source in the system.

A typical diffraction pattern from 10x10 ON-state mirror elements (without the ink spot) captured at approximately 10mm from the focal plane is shown in Fig. 2. The figure also reveals the appearance of the interference fringes resulting from the superposition of the background diffracted light and ON-state reflected light. As the number of the "ON" elements is reduced, the size of the central main lobe (zero order) is increased. Hence, for a single DMD element, the size of the central lobe is increased beyond the diameter of the iris opening so that only light from this lobe is admitted into the aperture of L1.

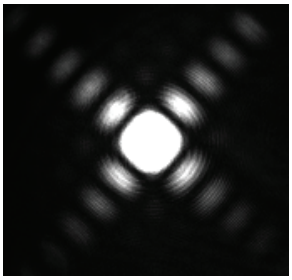


Fig. 2. Diffraction pattern from 10x10 DMD elements

The size of the point source is determined by the number of ON-state DMD elements and the demagnification between lens L1 and L2 (f1/f2). Typically, the minimum achievable size of the point source in this configuration for a single element in the "ON" state and demagnification factor of 6 (150/25) is approximately 2.2 μm. However, smaller point-source size can be obtained by using more powerful lens combination with high NA. The size of the point source can be reconfigured by simply varying the number of the elements in the "ON" state. Light from the focused spot propagates spherically towards the object plane (located at distance S from the focal plane) and generates an object scattered field that interferes with the undiffracted spherical reference field at the CCD plane (located at distance d from the object). The CCD (Qimaging Retiga 2000R) is set to a resolution of 1200 x 1200 pixels with 7.4 μm x 7.4 μm pixel size.

3. Hologram recording and reconstruction

3.1 Recording

To record holograms, one DMD mirror element is switched to ON-state at position $P_o(\xi_o, \eta_o)$ and translated to different position $P_N(\xi_o \pm n_\xi T_\xi, \eta_o \pm n_\eta T_\eta)$, where T_ξ, n_ξ and T_η, n_η denotes the number of switching step and step size respectively, on ξ – and η – axis of the DMD plane. Consequently, this process moves the focused point source to a different lateral position to create translated holograms of the object. The holograms captured for each DMD position represent the intensity of the interference field.

If the complex spherical reference and object field at the CCD plane be represented by $R(x,y)$ and $O(x,y)$ respectively, such that

$$\begin{aligned} R(x,y) &= A_R(x,y)\exp^{i\varphi_R(x,y)}, \quad \text{and} \\ O(x,y) &= A_o(x,y)\exp^{i\varphi_o(x,y)} \end{aligned} \tag{1}$$

where $A_R(x,y)$ and $A_o(x,y)$ are real amplitudes and $\varphi_R(x,y)$ and $\varphi_o(x,y)$ represent the phase of reference and object waves respectively. The mixing of the two fields at the CCD plane generates an interference field $H(x,y)$ such that,

$$\begin{aligned} H(x,y) &= R(x,y) + O(x,y) \\ &= \exp^{i\varphi_R(x,y)}\left(A_R + A_o\exp^{i(\varphi_o(x,y)-\varphi_R(x,y))}\right) . \end{aligned} \tag{2}$$

The intensity pattern $W(x,y)$ measured by the CCD is

$$\begin{aligned} W(x,y) &= \left|H(x,y)\right|^2 \\ &= \mathrm{Re}\left[{A_R}^2 + {A_o}^2 + 2A_RA_o\exp^{i(\varphi_o(x,y)-\varphi_R(x,y))}\right] . \end{aligned} \tag{3}$$

In addition to the recorded holograms, background light intensity is captured for each position of the point source. This is achieved by translating the object out of the hologram field of view such that only the spherical reference light transmitted through the clear part of the microscope glass slide is captured.

3.2 Numerical reconstruction

To reduce the effect of the background noise in the reconstructed image, the background intensity image of the reference beam (captured at every position of the point source) is subtracted from the corresponding hologram to generate a contrast hologram. The scattered wave U(X,Y,d), which carries the information of the object, is recovered from the contrast holograms using Fresnel-Kirchhoff diffraction integral with the well-known convolution approach [5].

$$U(X,Y,d) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} W(x,y)G(X-x,Y-y,\hat{d})\,dx\,dy . \tag{4}$$

In equation (4), $G(X-x,Y-y,\hat{d})$ is the impulse response function of the system with distance $\hat{d}$ representing a multiplication between distance d and fringe magnification $m = D/s$ that results from using the spherical reference field. The term $W(x,y)$ represents the contrast hologram. Using equation (4), the Fourier transforms of both $W(x,y)$ and $G(X,Y,\hat{d})$ are multiplied and an inverse Fourier transform applied to the product to obtain the object wave $U(X,Y,d)$. Depth reconstruction can be obtained by reconstructing $U(X,Y,d)$ on different planes. The intensity $I(X,Y)$ of the reconstructed wave is calculated as

$$I(X,Y) = \left|U(X,Y)\right|^2 . \tag{5}$$

4. Analysis

4.1 Translations in the DMD and hologram plane

Control of the DMD elements is achieved through control software that communicates directly with the chip via a USB port. This allowed translation of individual elements over a specified switching interval, equivalent to some number of elements. Consequently, the translation of the ON-state DMD elements generates two main effects in the object illumination. First, it leads to shifting of the point source (spherical reference beam) with respect to the object position, since the DMD serve as the primary source of light. Hence, a translated hologram of the object is generated with the interference fringes moving in the opposite direction of the element switching. Second, for objects with some depth features, the translation also changes the illumination angle of the depth features thereby encoding information about different projections through the object volume in the translated hologram.

If T_{dmd} and T_{holo} represents the linear translations distance in the $\xi - \eta$ plane of the DMD and hologram plane respectively, the relationship can be expressed as follows.

$$\frac{T_{dmd}}{T_{holo}} = M_L \,, \tag{6}$$

where $M_L = \frac{f_1}{f_2}$ is the magnification between lens L1 and L2. Thus for a given translation on the DMD plane, the corresponding translation distance in the hologram plane is determined by the magnification (M_L) in the system. A higher translation distance in the hologram plane can be obtained by decreasing M_L.

4.2 Focused spot size

For light admitted by L1 from the ON-state DMD elements, the combination of lens L1 and L2 acts as a de-magnifier depending on the ratio between f1 and f2. This process recreates the image of the ON-state DMD elements at the focal plane of L2 (FP1). Thus the focused spot size (W_o) is determined by the achievable de-magnification (M_d) between lens L1 and L2 such that

$$W_o = \Delta_{dmd} M_d \tag{7}$$

where $M_d = 1/M_L$ and Δ_{dmd} denotes the DMD mirror element size. For the configuration shown in Fig. 1, the dimension of the ON-state DMD elements is reduced by a factor of 6 (i.e. $M_d = 6$). Thus, the minimum focal spot is 2.2 μm and this corresponds to a single element (13.68um x 13.68um) in On-state (i.e. in position P_1).

4.3 Field of view and reconstructed object magnification

Factors that limit the achievable resolution in DIHM have been treated extensively with developed criteria for achieving lateral as well as depth resolution [GXJKJK06a, GXJKJK06b, JGXJK06]. The FOV in DIHM is defined as the area in the object plane magnified by the incident spherical reference field and captured in the reconstructed hologram. Since the resolution and effective number of interference fringes in the recorded hologram are limited significantly by the size and position of the CCD, it is difficult to achieve the desired

resolution (with adequate magnification) and wide field of view simultaneously in the reconstructed images. This is applicable to imaging an object that extends over an area of larger dimension compared to the numerical aperture of the system. Increasing the object area (FOV) captured in a hologram, for a given resolution, leads to reduction in the magnification (M) of the reconstructed object features.

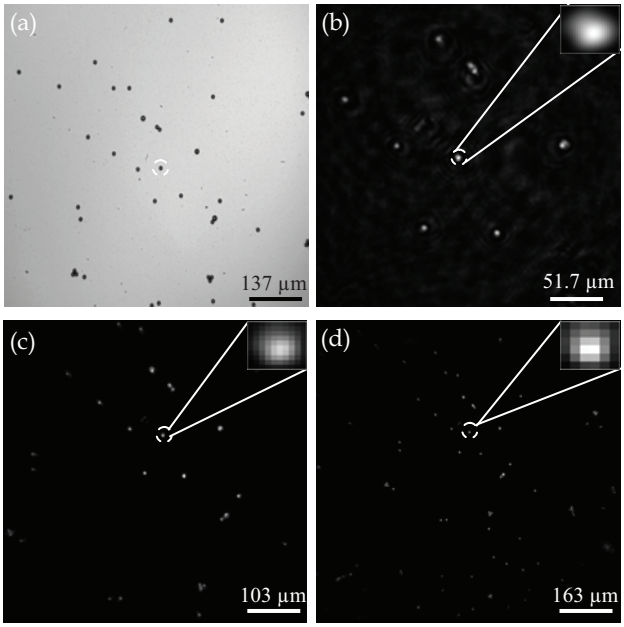


Fig. 3. Bright-field image and reconstructions of 9-µm spheres deposited on a microscope glass slide: (a) Image obtained from compound bright-field microscope. (b) - (d) Reconstructed images at distance S = 1.875 mm, 4.7 mm and 7.3 mm from the point source respectively. Green laser; 2.2-µm point source; D = 4.3 cm; NA = 0.103.

To illustrate this, Fig 3(a)-(d) shows the holographic reconstructed images of 9 µm spheres (D=4.3 cm) at distances S = 1.875 mm, 4.7 mm and 7.3 mm from the point source respectively. The corresponding estimated FOV in Fig. 3(b)-(d) are approximately 0.3 mm, 0.53mm and 0.95 mm respectively. From the images, we observe a reduction in the magnification of the objects as the FOV is increased i.e. as more object areas are captured in the recorded hologram. Thus, for a given resolution, determined by the DIHM configuration parameters, the obtained magnification in DIHM is compromised as the FOV in the reconstructed images increases. However, since one of the primary tasks of any microscope is to produce a magnified and resolved image of a given specimen with high contrast, a system that improves the range of the FOV for a given resolution and magnification at certain CCD position is necessary.

4.4 Depth information extraction

The spatial location of the point source in the programmable point-source DIHM, with respect to the object position, determines the object illumination angle. For a 2-D object

occupying a lateral plane and located at some distance from the point source (Fig. 1(b)), a change in the position of the point source leads to lateral translation of the object in the reconstruction plane. However, for 3-D object with features occupying different planes in depth, the translation of the point source leads to a change in the illumination angle. Figure 4 shows an illustration with two point objects A and B separated by axial distance t and lying on optical axis G_0G (perpendicular to the hologram plane).

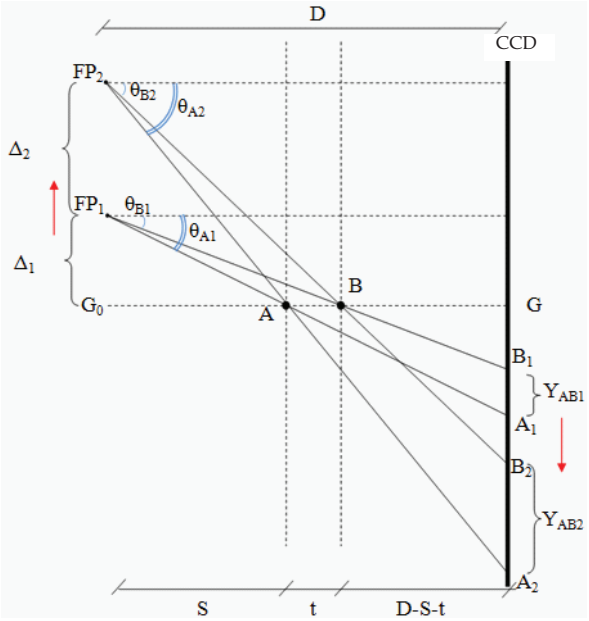


Fig. 4. Schematic illustration of change in object illumination angle resulting from point-source translation

When these objects are illuminated from point source located at G_0, the illumination angle (θ_{A0} and θ_{B0}, not shown in the illustration) corresponding to these objects are equal to zero. Consequently, the corresponding hologram of these objects will superimpose with the center of the interference pattern at G. The reconstructed image of the objects (when the reconstructed planes are combined in depth) will overlap when viewed from the direction indicated by arrow V (from the top of the aligned images) as shown in Fig. 5(a). However, shifting the point source from G_0 to position FP_1 through distance Δ_1 changes the illumination angle such that $\theta_{B1} < \theta_{A1}$. The difference between these angles i.e. ($\theta_{A1} - \theta_{B1}$) transform to a separation ($Y_{AB1} = A_1 - B_1$) between the center of hologram of A and B on the CCD plane. An illustration of the separation between the reconstructed image of A and B is shown in Fig. 5(b). Similarly, when the point source is translated from FP_1 to FP_2 through a distance Δ_2, the illumination angles θ_{A1} and θ_{B1} are increased to θ_{A2} and θ_{B2} respectively. This consequently leads to an increment in the separation between the holograms of A and B such that $Y_{AB2} > Y_{AB1}$.

The illustration in Fig. 4 is based on an initial condition which assumes the objects (A and B) and point source are perfectly located on the optical axis G_0G (i.e. $\theta_{A0} = \theta_{B0} = 0$). In practice,

however, it is difficult to satisfy this condition. This is due to some offset introduced by the error in alignment of the central element in the DMD array with the optical axis. Hence, the initial illumination angle θ_{A0} and θ_{B0} are not equal to zero. In this circumstance, FP_1 may be used to represent the initial spatial location of the point source with Δ_1 representing the offset from the optical axis. Based on the geometry in Fig. 4, Δ_1 can be expressed as follows.

$$\Delta_1 = \frac{\frac{S^2 Y_{AB1}}{t} + Y_{AB1} S}{D} \tag{8}$$

Where D and S denotes the point-source-to-CCD and point-source-to-A distance respectively. The separation between the objects A and B, denoted by t, can be expressed as follows.

$$t = \frac{S^2 (Y_{AB2} - Y_{AB1})}{S(Y_{AB1} - Y_{AB2}) + D\Delta_2} \tag{9}$$

Thus, the depth information of a 3-D object can be extracted by translating the point source of the spherical reference field and using the corresponding information (Y_{AB1} and Y_{AB2}) obtained from the recorded holograms.

In addition to the possibility of extracting the depth information, the change in the illumination angle also allows the different projection view of the 3-D object to be captured. If the ON-state mirror element is translated across the DMD plane as shown in Fig. 1(b), the reconstruction of the recorded holograms will show different view of the 3-D features with respect to the spatial location of the point source.

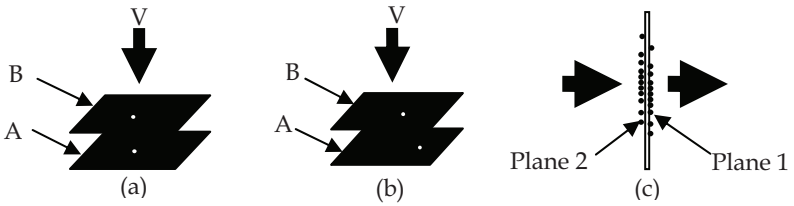


Fig. 5. Illustrations of the effect of change in illumination angle on 3D reconstruction

While this analysis applies only for point sources separated into 2 planes, it does demonstrate that the ability to translate the point source does provide information by which axial information can be extracted. We anticipate that more sophisticated approaches could be devised to convert holograms obtained for more complex samples into reconstructed images that include axial information. Such an undertaking is beyond the scope of this work.

5. Experimental demonstrations and results

5.1 Demonstration of Enhanced FOV

To demonstrate an expansion of the field of view using the programmable point-source DIHM with 2.2 μm, 20 holograms (10 holograms along each axis) and their corresponding background intensity are recorded along $\xi - \eta$ axis. The object, consisting of polystyrene

microspheres (9um diameter) deposited on a microscope glass slide, and CCD are located at a distance S = 1.875 mm and D = 43 mm from the point source, respectively. This corresponds to a magnification m = 23. We reconstructed all the hologram captured to be combined to generate a single reconstructed image with a wider field of view.

To summarize the hologram recording process described in Section 3.1, we switch "ON" a DMD element at position $Po(\xi_o, \eta_o)$ close to the optical axis, capture the corresponding hologram and the background intensity, then switch the element to a different position $P_N(\xi_o \pm n_\xi T_\xi, \eta_o \pm n_\eta T_\eta)$ to capture the corresponding hologram and background intensity. T_ξ and T_η represent the switching steps on $\xi-$ and $\eta-$ axis of the DMD plane respectively. The step size (number of DMD element in each switching step) along $\xi-$ and $\eta-$ axis is denoted by n_ξ and n_η. It should be noted that the positions corresponding to $T_\xi = T_\eta = \pm 10$ set the boundary of the mirror elements admitted into the lens system and constitute the limits of the FOV expansion in this demonstration.

Figure 6 shows the captured holograms and reconstructed images for 6 positions $P_o(\xi_o, \eta_o)$, $P_1(\xi_o - 5T_\xi, \eta_o)$, $P_2(\xi_o + 5T_\xi, \eta_o)$, $P_3(\xi_o, \eta_o - 5T_\eta)$, $P_4(\xi_o, \eta_o + 5T_\eta)$ and $P_5(\xi_o, \eta_o + 10T_\eta)$ with camera exposures 12 milliseconds (ms) at P_o, 18 ms at P_1, P_2, P_3, and P_4, and 29 ms at P_5 mirror element positions. Lower exposure values are possible by increasing the laser power (we use 0.3W, but the maximum power available from our laser is 4W). The dark circle in the holograms represents the ink drop used to block the background diffracted light as described in Section 2.

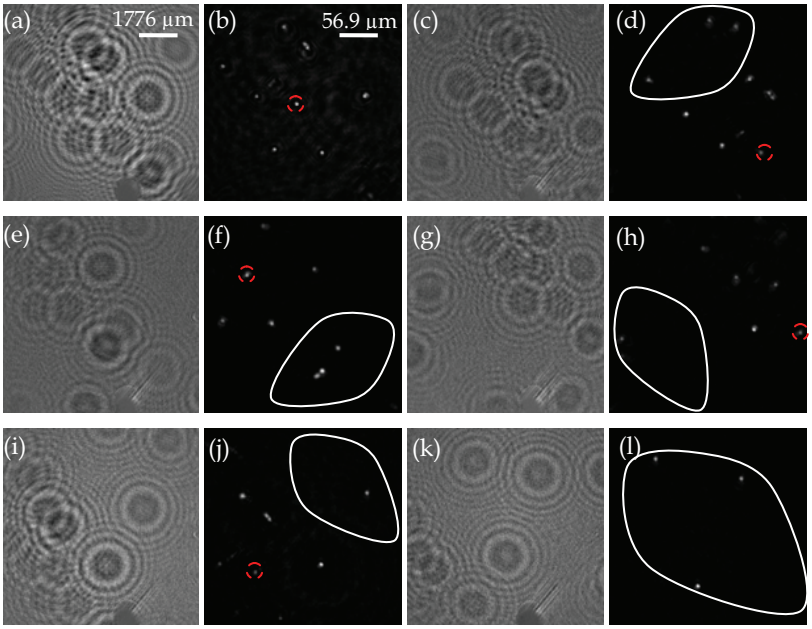


Fig. 6. Holograms and reconstructed images for "ON" DMD mirror element positions $P_0(\xi_o, \eta_o)$, $P_1(\xi_o - 5T_\xi, \eta_o)$, $P_2(\xi_o + 5T_\xi, \eta_o)$, $P_3(\xi_o, \eta_o - 5T_\eta)$, $P_4(\xi_o, \eta_o + 5T_\eta)$ and $P_5(\xi_o, \eta_o + 10T_\eta)$. S = 1.875 mm; Laser wavelength=532nm; 2.2-µm point source; D = 4.3 cm.

The reconstructed image of the hologram at position P_0 (in Fig. 6(b)) marks the initial FOV (approximately 0.3 mm) to be expanded. The area corresponding to this FOV in the bright-field microscope image of the spheres is marked with a broken red circle in Fig. 7(b). Regions in the reconstructed images (Fig. 6(d), (f), (h), (j), and (l)) corresponding to new objects translated into the initial FOV are indicated with a closed white line.

To expand the FOV, a process that combines regions with new objects in all the reconstructed images into a single image is necessary. This is achieved through an algorithm that utilizes edge detection to obtain the location of objects at the peripheral of the field of view. Using these locations and the recording parameters (T_ξ, T_η, n_ξ and n_η) of the recorded holograms, a single image with wider field of view is generated as shown in Fig. 7(a). The corresponding field of view is calculated to be approximately 0.79 mm (region within the red rectangle). We obtained this value by multiplying the size of the array with the pixel size in the reconstructed image. This value gives an expansion of the initial FOV by approximately 263%.

It should be noted that the results obtained are for a 2.2 µm point-source size as achieved in our current configuration. However, a wider FOV can be achieved with a smaller point-source size since more area is illuminated in the object plane than using larger point-source.

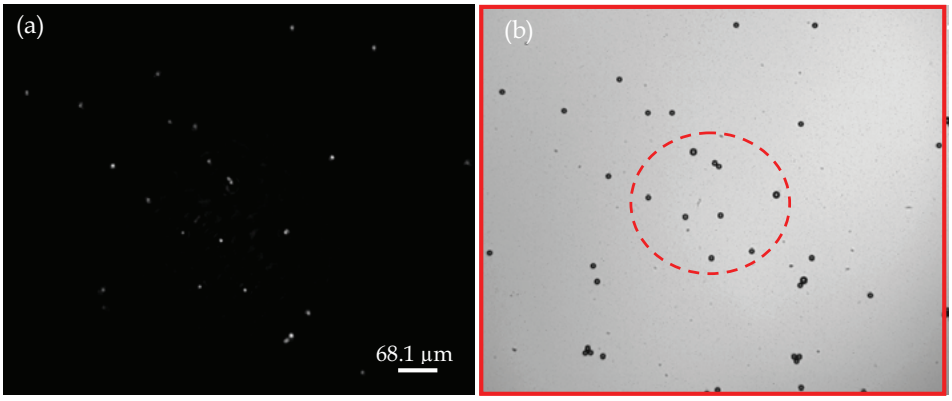


Fig. 7. Single image with wider FOV generated from combination of all reconstructed images: (a) Single image with wider FOV generated from combination of all reconstructed images with new objects in the initial FOV using 2.2-µm point source (b) bright-field compound microscope of the microspheres showing initial FOV ≈ 0.3 mm (in red circle) and enhanced FOV ≈ 0.79 mm (region enclosed in red rectangle).

5.2 Demonstration of depth feature extraction

To demonstrate the extraction of depth information, a sample with depth information was obtained by depositing drops of polystyrene microspheres (9 µm diameter) from suspension in distilled water on both sides of a microscope glass slide of 1 mm thickness (see illustration in Fig. 5(c)). The slide was located at distance S = 4.1 mm from the point source and illuminated in the specified direction of the arrow. Figure 8 shows the brightfield images of plane 1 and plane 2 of the slide captured with 10x magnification objective lens.

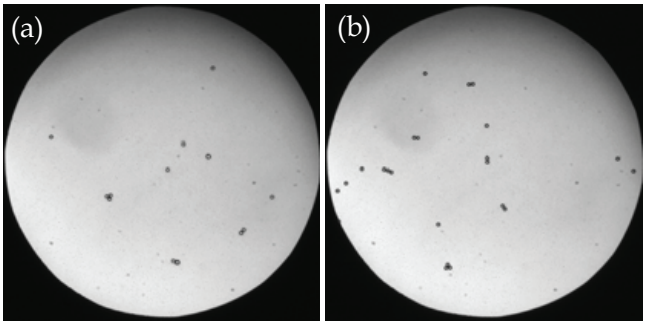


Fig. 8. Brightfield image of 9 μm spheres deposited on both sides of a glass slide as illustrated in Fig. 4.15 (c): (a) plane 1, and (b) plane 2.

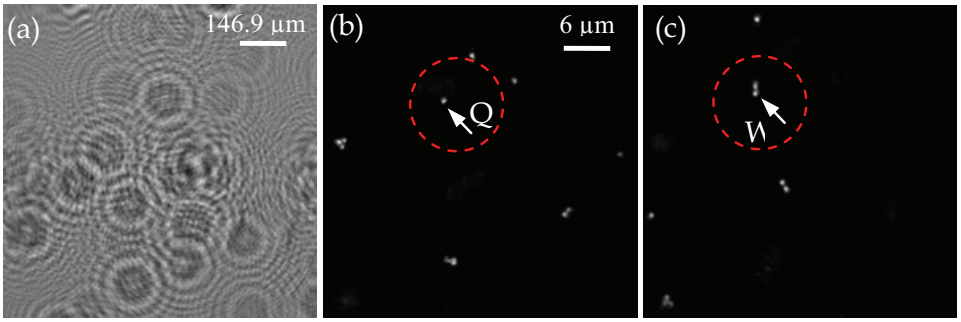


Fig. 9. Depth Reconstruction images: (a) Hologram contrast image, (b) Reconstructed image at plane 1, and (c) Reconstructed image at plane 2, Laser wavelength=532nm; S=4.1 mm, D=50 mm; 2.2-μm point source.

Figure 9(a) shows the contrast image of the captured hologram at position $P_o(\xi_o, \eta_o)$ - the nearest position to the optical axis. Reconstructions at plane 1 and plane 2 of the hologram are shown in Fig. 9(b) and (c) respectively. These reconstructions at different planes are obtained by changing the distance d in the diffraction integral (expression 4). We mark specific spheres Q and W on plane 1 and 2 respectively.

Figure 10(a) shows the rendering of both reconstructed planes by simply combining them into a single 2-D image. This rendering method introduces blurry artifacts due to the background intensity in the reconstructed images of on both planes. The combined image is similar to the illustration in Fig. 5(b) when viewed from the direction of the arrow V.

To demonstrate the capability of the PP-DIHM system to capture different projections of features Q and W on plane 1 and 2 respectively, holograms corresponding to point-source positions $P_1(\xi_o - 5T_\xi, \eta_o)$, $P_2(\xi_o + 5T_\xi, \eta_o)$, $P_3(\xi_o, \eta_o - 5T_\eta)$ and $P_4(\xi_o, \eta_o + 5T_\eta)$ are captured as shown in Fig. 10(b)–(e). The reconstructed images of these holograms are shown in Fig. 11(a)-(d), respectively. It can be observed that the reconstructed features on both planes (i.e. Q and W) are translated with respect to the position of the point source (illustrated in Fig. 4). Thus for features located longitudinally on opposite sides of the slide, the illumination angle

θ_B is greater than θ_A. This leads to a change in the projection view of the reconstructed images and lateral shifts between features on both planes.

To extract the depth information using expression (9), only two reconstructed holograms are necessary. For convenience, the images in Fig. 10(a) and Fig. 11(a) are shown in Fig. 12. We estimate the distances Y_{AB1} and Y_{AB2} (referred to the hologram plane). These distances are calculated by multiplying the number of image pixels separating Q and W by the CCD (7.4 μm) pixel size. Thus Y_{AB1} and Y_{AB2} are estimated to be 162.3 μm and 418.6 μm respectively.

The translated distance (Δ_2) of the point source between positions corresponding to ON-state DMD element at $P_o(\xi_o, \eta_o)$ and $P_1(\xi_o - 5T_\xi, \eta_o)$ is equivalent to

$$\Delta_2 = \frac{N_{dmd} * \Delta_{dmd}}{M_L} = 114\mu m. \tag{10}$$

where $N_{dmd} = 50$ is the number of DMD elements in the translation interval on the DMD plane, $\Delta_{dmd} = 13.68\mu m$ is the DMD pixel size and $M_L = 6$ is the magnification between lens L1 and L2 in Fig. 1. Thus, using expression (9) and given the distance D (between the point source and CCD plane), and S (between the point source and plane 2 of the slide) as 50mm and 4.1mm respectively, the depth information (t) separating the two planes can be obtained as approximately 0.93mm. This value and the known separation (1 mm thickness of the slide) are quite close.

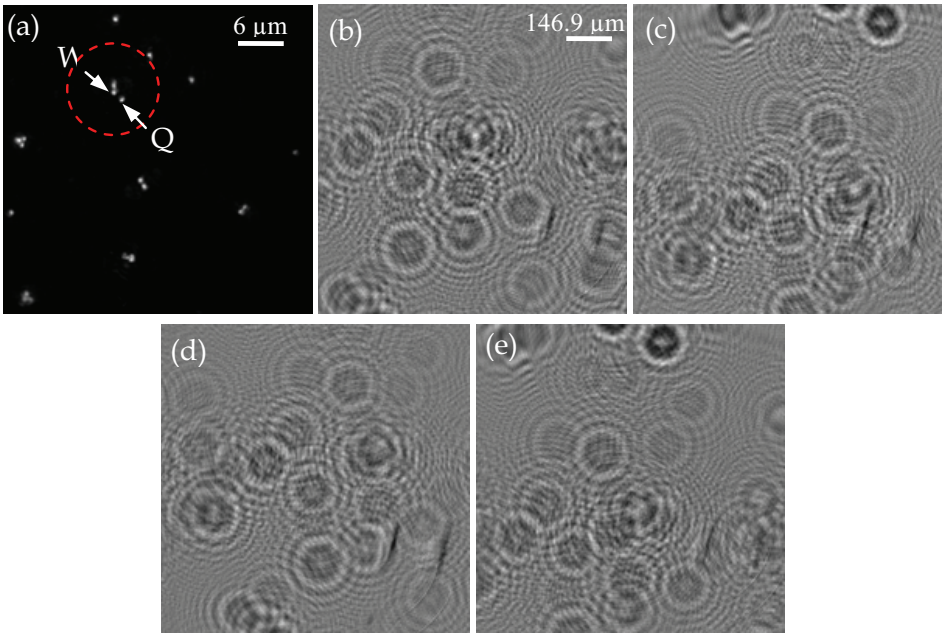


Fig. 10. (a) Display of both reconstructed plane in Fig. 9; (b)-(e) Hologram at position $P_1(\xi_o - 5T_\xi, \eta_o)$, $P_3(\xi_o, \eta_o + 5T_\eta)$, $P_2(\xi_o, \eta_o - 5T_\eta)$, and $P_4(\xi_o + 5T_\xi, \eta_o)$ respectively. Laser wavelength=532nm; S=4.1 mm, D=50 mm; 2.2-μm point source.

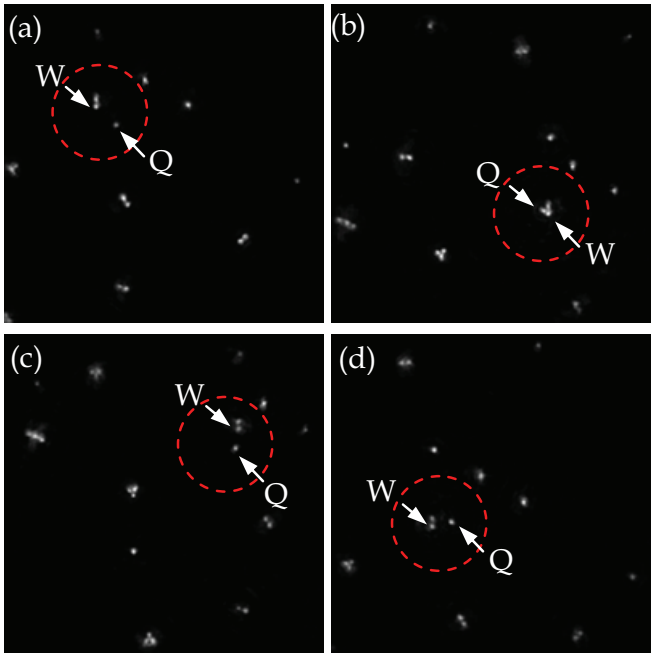


Fig. 11. (a) – (d). 3D reconstructed images of holograms in Fig. 10 (b) – (e) respectively. Laser wavelength=532nm; S=4.1 mm, D=50 mm; 2.2-μm point source.

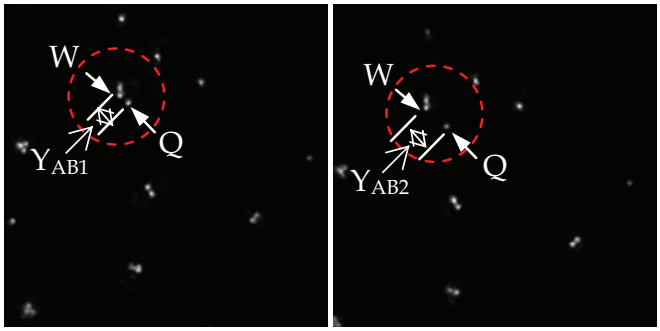


Fig. 12. Reconstructed images to estimate the depth separation

The difference between these two values can be attributed to some factors. One reason may be the inhomogeneities of the glass slide as the scattered wave from plane 2 is transmitted through the slide before reaching the CCD plane. Also, errors due to imperfection of the spherical reference field in our system may contribute to this difference. The reconstruction process based on the Fresnel-Kirchhoff diffraction integral in equation 4 assumed a perfect spherical reference field. However, our method of generating the spherical reference field by focusing light from the ON-state DMD element may introduce some error into the reconstruction process. Another factor that may also contribute to the error is inaccuracy in the measurement of the dimensions used.

6. Discussions

Despite many potential strengths, our DMD-based programmable point-source DIHM system does have challenges. Basically, the main challenges include the new challenges related to the use of the DMD, and more familiar challenges associate with the use of a CCD in holography.

Problems introduced by the DMD include background diffraction orders resulting from the 2-D periodic structure of the array and falling within the field of view of the system. These lead to unwanted interference fringes when light from the ON-state element is superimposed, as shown in Fig. 2. For demonstration purposes, our method of reducing the background light in the field of view by locating a cover slide with an ink drop at the back focal plane of lens L1 is quite effective, but introduces complexities in the system. Care is required to locate the cover slide perpendicularly along the beam path and to obtain the required size of the ink drop in comparison to the focused spot size of the background diffracted light. This method at least demonstrates a path to more elegant remediation of this problem through clever spatial filtering.

Also, the location of iris P, which blocks some of the diffracted background light, reduces the effective NA of lens L1 in capturing the light beam from the ON-state DMD element. This consequently limits the minimum achievable size of the point source at the focal plane of lens L2. The key to achieving larger NA and removing the complexity introduced by the spatial filter (e.g. ink drop on a cover slide) is to reduce the diffraction effect from the DMD. Given the trends in technological advancement and the general desire to increase contrast, future generations of DMD chips may have less significant diffraction effects. Light scattering from the DMD chip has been improving steadily by advances such as reducing the spacing between the mirror elements and depositing a dark layer on the backplane of the DMD mirror elements.

The Gaussian profile of the collimated input beam and its spreading over many DMD elements has potential limiting effects. The spreading of the incident beam leads to low brightness obtained from the ON-state mirror elements. Camera exposure time can be used to compensate for the non-uniform profile of the input beam as the ON-state mirror element is translated away from the origin. However, to achieve high intensity of the light from a single ON-state element, a high-power input laser beam is required. Fortunately, affordable and powerful diode-pumped lasers are available.

The challenges associated with the use of CCD in holography are well documented in the literature. In summary, the extent to which a CCD can be located close to the object is limited by the pixel size (affects the sampling of interference pattern) and CCD array size (affects the numerical aperture). Consequently, this also limits the obtainable resolution and restricts the field of view in the reconstructed images. Also, due to the use of a static pinhole in the traditional setup, the obtainable projection view of an object is restricted to only the location of the pinhole.

As for strengths of the technique, in addition to enhancing the field of view in the reconstructed image and extracting the axial 3-D information of an object through acquisition of different projections of the hologram, the DMD-based programmable point-source scanning technique provides the flexibility in reconfiguring the size of the point source. This removes the physical limitation imposed by the definite pinhole size in the prior spherical reference-field DIHM configurations. Our approach also eliminates the need for a mechanical scanning procedure in translating the sample or the CCD camera across the beam path.

7. Future work

In order to improve the performance of the system under the current configuration and further enhance its application, the following are some research directions to be taken.
The current set up takes considerable time to generate the result. This is due to time required to communicate between the various software components to execute certain task. Future work will improve this by speeding up the algorithms implementations and integrating all associated software. Also, an estimation of the background intensity based on a numerical method will potentially improve the speed and eliminate the need to capture background intensity for every point-source position.
Another area that needs improvement is the minimum achievable size of the point source. One possibility of reducing the limited size of the point source is to increase the magnification between lens L1 and L2 through the use of more powerful lenses with high NA. For a smaller point source, compared to 2.2 μm in this demonstration, the radial intensity variation in the spherical reference beam profile across the captured hologram field of view is reduced significantly. However, the minimum point source size that can be achieved will also depend on the diffraction limits of the system.
Some potentially interesting applications of the programmable point-source DIHM include particle tracking and exploring the advantage of multiple point-source illumination. For the former, the different fields of view in the holograms captured at different scanning positions of the point source can be used to obtain the path taken by an object in the field of view.
In principle, multiple point sources can improve depth resolution in DIHM since information about the object from different projections is captured in a single hologram. However, there is need for a smaller point-source size and higher numerical aperture compared to our present configuration. We also envisage a complex reconstruction algorithm will be required to recover the object field due to the multiple point sources of the reference field

8. Conclusion

This chapter has described the technique of programming the point source in digital in-line holographic microscopy using Digital Micro-mirror Devices. By individually addressing the DMD pixels, the location of an ON-state pixel is translated to different locations in the DMD plane. This leads to translation of the source of the spherical reference beam with respect to the object location. Results from the demonstrations showed the system can potentially be used to extract depth information in an object volume and enhance the limited field of view.

9. References

[AD09a] A.A. Adeyemi and T.E. Darcie, *Expansion of field of view in digital in-line holography with a programmable point source*, Appl. Opt. vol. 48, pp. 3291-3301, 2009.

[AD09b] A.A. Adeyemi and T.E. Darcie, *Extraction of 3D Axial Features in Programmable Point-Source Digital In-line Holographic Microscope with Spherical Reference Field, IEEE* Pacific Rim (PACRIM) Conference on Communications, Computers and Signal Processing, pp. 736 – 739, Victoria BC, 2009.

[ABD09] A. A. Adeyemi, N. Barakat, and T. E. Darcie, *Applications of digital micro-mirror devices to digital optical microscope dynamic range enhancement*, Opt. Express, vol. 17, Iss. 3, pp. 1831-1843, 2009.

[BPS03] V. Bansal, S. Patel, P. Saggau, *A High speed confocal laser-scanning microscope based on acousto-optic deflectors and a digital micromirror device, Proceedings of the IEEE conference on engineering in medicine and biology society* (Institute of Electrical and Electronics Engineers, TX), pp. 17-21, 2003.

[CMD99] E. Cuche, P. Marquet, and Ch. Depeursinge, *Simultaneous amplitude-contrast and quantitative phasecontrast microscopy by numerical reconstruction of Fresnel off-axis holograms*, Appl. Opt. vol. 38, pp. 6994-7001, 1999.

[DDS03] D. Dudley, W.M. Duncan, J. Slaughter, *Emerging digital micromirror device (DMD) applications, Proceedings of SPIE*, vol. 4985, pp. 14-25, 2003.

[DJL99] F. Dubois, L. Joannes, and J.-C. Legros, *Improved three-dimensional imaging with a digital holographic microscope with a source of partial spatial coherence*, Appl. Opt. vol. 38, pp. 7085-7094, 1999.

[DM99] A.L.P. Dlugan and C.E. MacAulay, *Update on the use of digital micromirror devices in quantitative microscopy, Proceeding of SPIE*, vol. 3604, pp. 253-262, 1999.

[GPO08] U. Gopinathan, G. Pedrini, and W. Osten, *Coherence effects in digital in-line holographic microscopy*, J. Opt. Soc. Am. vol. 25, pp. 2459-2466, 2008.

[GXJKJK06a] J. Garcia-Sucerquia, W. Xu, S.K. Jericho, P. Klages, M.H. Jericho, H.J. Kreuzer, *Digital in-line holographic microscopy*, Appl. Opt. vol. 45, pp. 836-850, 2006.

[GXJKJK06b] J. Garcia-Sucerquia, W. Xu, S. Jericho, M. H. Jericho, P. Klages and H.J. Kreuzer, *Resolution power in digital in-line holography, Proceedings of SPIE* vol. 6027, pp. 637-644, 2006.

[JGXJK06] S.K. Jericho, J. Garcia-Sucerquia, W. Xu, M.H. Jericho, H.J. Kreuzer, *Submersible Digital In-line Holographic Microscope*, Rev Sci Instrum, vol. 77, 043706, 2006.

[KJMX01] H. J. Kreuzer, M. H. Jerico, I. A. Meinertzhagen, and W. Xu, *Digital in-line holography with photons and electrons*, J. Phys.: Condens. Matter vol. 13, pp. 10729–10741, 2001.

[KN98] K. J. Kearney and Z. Ninkov, *Characterization of a digital micromirror device for use as an optical mask in imaging and spectroscopy, Proceedings of SPIE*, vol. 3292, pp. 81-92, 1998.

[Kre05] T. Kreis, *Hankbook of holographic interferometry: optical and digital methods*, Wiley-VCH, 2005.

[RGMNS06] J. P. Ryle, U. Gopinathan, S. McDonnell, T. J. Naughton, and J. T. Sheridan, *Digital in-line holography of biological specimens*, *Proceedings of SPIE* 6311, 63110C, 2006.

[RPP04] L. Repetto, E. Piano, and C. Pontiggia, *Lensless digital holographic microscope with light-emitting diode illumination*, Opt. Lett. vol. 29, pp. 1132-1134, 2004.

[Sch94] U. Schnars, *Direct phase determination in hologram interferometry with use of digitally recorded holograms*, J. Opt. Soc. Am. vol. 11, pp. 2011–2015, 1994.

[SJ02] U. Schnars and P. O. Jüptner, *Digital recording and numerical reconstruction of holograms*, Meas. Sci. Technol. 13, R85-R101, 2002.

[SJ05] U. Schnars and W. P. Jueptner, *Digital Holography*, Springer, 2005.

[SPISSSW97] B. W. Schilling, T.-Ch. Poon, G. Indebetouw, B. Storrie, K. Shinoda, Y. Suzuki, and M. H. Wu, *Three-dimensional holographic fluorescence microscopy*, Opt. Lett. vol. 22, pp. 1506-1508, 1997.

[XJMK02a] W. Xu, M. H. Jericho, I. A. Meinertzhagen, and H. J. Kreuzer, *Digital In-Line Holography of Microspheres*, Appl. Opt. vol. 41, pp. 5367-5375 2002.

[XJMK02b] W. Xu, M. H. Jerico, I. A. Meinertzhagen, and H. J. Kreuzer, *Digital in-line holography for biological applications*, Proc. Natl. Acad. Sci. USA vol. 98, pp. 11301–11305, 2002.

[Yar03] L. P. Yaroslavsky, *Digital Holography and Digital Image Processing: Principles, Methods, Algorithms*, Kluwer, 2003.

8

Pulsed Full-Color Digital Holography with a Raman Shifter

Percival Almoro, Wilson Garcia and Caesar Saloma
National Institute of Physics, University of the Philippines
Diliman, Quezon City,
Philippines

1. Introduction

Holography deals with processes involving the transformation of waves by interference structures that are formed when coherent electromagnetic waves interact with matter (Andreeva, 2002). As an optical technique it was introduced in the seminal works of Gabor, Leith, Denisyuk and others in the 1960's (Benton, 2005). By succeeding to reconstruct the complete wave (full amplitude and phase information), holography yields depth information of a volume (three-dimensional) scene.

The main goal of holography in imaging is to reconstruct wavefronts from 3-D colored and moving objects in real time. Sustained success however, is still hampered by a number of technical hurdles that need to be overcome. Full-color reconstruction requires a multi-wavelength light source and a wide bandwidth light-sensitive recording medium. Conventional color holography utilizes three separate continuous-wave (CW) lasers as illuminator which is more expensive and difficult to operate and maintain. The narrow temporal bandwidth of CW lasers also restrict their applications only to holographic imaging investigations involving stationary objects.

Here we discuss the use of the hydrogen Raman shifter as a pulsed color holographic light source. The Raman shifter is an attractive holographic light source because it is pulsed, compact, inexpensive, and multi-wavelength (Almoro et al., 2004; Almoro et al., 2007). Pulsed color light sources extend the range of possible applications of color holography to fast moving objects and rapid events. Due to significant advances in photodetector and computer technology, holograms can now be recorded with high-resolution digital cameras and reconstruction could be carried out numerically in a fast manner (Schnars & Jüptner, 2002).

Digital holography is suitable for industrial applications since it does not involve the cumbersome development of photographic films. In full-color digital holography, a minimum set of three holograms are captured corresponding to the primary color channels (i.e., red, green and blue color channels). For a proper fit of the superposition of the reconstructions, it is important to characterize and control the wavelength dependences of the image size, lateral resolution and depth of focus. These image variations result in the dispersion of the color hologram reconstructions making the final image blurred and unacceptable. Here we discuss the experimental evidence of the said effects and describe a technique to control the chromatic dispersion of full-color holograms.

2. Basic principles

2.1 Pulsed full-color digital holography

2.1.1 Digital holography

In conventional digital holography, a test object is illuminated using a CW monochromatic light and the recorded hologram is reconstructed in a computer (Schnars & Jüptner, 2002). Numerical reconstruction is achieved by calculating the diffraction of the monochromatic light incident on the hologram through a diffraction formula. The derived complex amplitude is then used to plot the intensity or phase distributions at any given plane. A three-dimensional (3D) image of a volume scene is generated from a stack of 2D intensity distributions. An unwrapped phase distribution may also be used to depict the surface shape of an object or the internal variations in the refractive index. In addition, phase distributions corresponding to different states of the test object may be compared and used to study changes in the physical state of the object or sample (Schnars & Jüptner, 2002).

Applications of digital holography have found their way in optical metrology, 3D microscopy and object recognition. In optical metrology, the reconstructed phase is used to detect material defects (Schulze et al., 2003) and to determine material properties such as the Poisson ratio, Young's modulus and thermal expansion coefficient (Seebacher, 2001). Digital holographic microscopy has been applied in the imaging of biological specimens (Zhang & Yamaguchi, 1998) and micro-optical elements (Colomb, 2010). Direct access of the complex amplitude of the reconstructed object beam allows for a more versatile discrimination of dissimilar 3D objects (Javidi & Tajahuerce, 2000).

The maximum resolution of the reconstructed image depends on the CCD pixel pitch. High-end CCD cameras that are now in the market have a pixel pitch of about 3.5 microns. The applications of digital holography are limited by the properties of the available light source. For the investigations of colored objects, a multi-wavelength light source with its wide spectral bandwidth is desirable since it enables us to sample the color information of the object in a single setting. For the investigations of dynamic objects and events, a pulsed light source is necessary.

2.1.2 Pulsed holography

Pulsed lasers extend the applications of holography to fast moving objects and transient events (Mallick, 1975). Due to the short illumination time of a pulsed laser (typically 10^{-9} second), object movements are literally frozen and the measurement resolves the rapid deformations of samples undergoing rapid kinetics. Ruby lasers (wavelength, λ = 694 nm) (Wedendal & Bjelkhagen, 1974) and Nd:YAG (λ = 532 nm) (Bates,1973) lasers, both noted for their long coherence lengths, are the commonly used pulsed lasers. Primary consideration for their popular use is the spectral sensitivity of the available holographic films which overlap with their emission wavelengths.

The advantage of pulsed holography is clearly demonstrated in applications where it is impossible for the test object to be stationary. Pulsed digital holography has been used to carry out measurements of micro-deformation occurring on the object undergoing unsteady vibrations (Pedrini, et al., 2002). Another practical application of pulse digital holography has been in the investigation of the dynamic phase change in an arc discharge (Liu, et al., 2002). Pulsed holography has also been combined with endoscopy, making the technique more compact and adaptable to industrial setting (Pedrini, et al., 2003).

The experimental issues against pulsed lasers are the aberrations and poor spatial distribution of the beams (Leith et al., 1991; Gustafsson, 2000). Degradation in the images during optical reconstructions in pulsed film holography is addressed by the use of anamorphic optics and an inter-channel coupling filter (Leith et al., 1991). For pulsed digital holography, the time between successive pulses limits the temporal resolution for hologram recording.

2.1.3 Full-color holography

Full-color holography involves the use of a multi-wavelength light source in the recording of color holograms and the subsequent superposition of the colored reconstructions. Any color in the visible spectrum is represented as a weighted linear combination of the three channel outputs that represent the primary-color channels of red (*R*), green (*G*) and blue (*B*). The red color aspects or portions of the object reflect or transmit only the red component of illumination beam. Other object colors reflect certain proportions of *R*, *G* and *B*. For example, yellow reflects equal proportions of *R* and *G* components of light, demonstrating the color addition during recording. Combined with the 3D effect, full-color holography adds realism to the reconstructions especially in imaging applications.

Full-color film holography was utilized in constructing replicas of full-color art works such as oil paintings (Bjelkhagen & Vukicevic, 2002). The quality of the reconstructions in full-color holography is improved by reducing the undesirable speckle artefacts that remain even with the use of a multi-wavelength light source (Harthong, et al., 1997). In optical metrology, a multi-wavelength light source enables measurements with variable resolution. During reconstruction in holographic interferometry, the zeroth-order fringe which indicates the reference undeformed region can be identified because of its white color that results from superposition of the primary colors used (Desse, et al., 2002; Jeong, 1999; Demoli, et al., 2003). Implementation of full-color holography has been constrained by several experimental issues. During recording, the first issue concerns with finding a suitable light source. The factors that should be considered are the range of colors, coherence lengths and relative intensities of the laser lines used. The range of colors of the reconstructions can be assessed based on the *CIE* chromaticity diagram for a given set of wavelengths (Bjelkhagen, 2002; Kubota, 2001; Peercy & Hesselink, 1994).

Four or more wavelengths provide the best color reproduction (Kubota, 2001; Peercy, 1994). One possible solution to the problem of limited coherence is the introduction of an intra-cavity Fabry-Perot etalon in the laser light source (Lin & LoBianco,1967). Previous investigations on the sampling characteristics of the holographic process have suggested using sets of wavelengths and relative intensities for color holography to improve color reproduction (Peercy & Hesselink, 1994).

The second issue in film holography recording is the spectral sensitivity of the holographic emulsion. Recording of color holograms could be achieved with two composite films (Kubota & Ose, 1979; Kubota, 1986) with sensitivities in red and blue-green or by using a single panchromatic broadband film (Bjelkhagen, 2002).

In the reconstruction in film holography the critical issues are the color shift due to emulsion shrinkage, wavelength dependence of magnification and location of the image, and the "cross-talk" or spatial overlap of the various color reconstructions resulting in a blurred image. Color shift may be addressed by chemical treatment of the holographic films (Lin & LoBianco, 1967). To achieve acceptable fit of color reconstructions, the magnifications and

locations could be adjusted using various geometry and laser sources during optical reconstruction (Olivares-Perez, 1989). The Lippman configuration has been able to address sufficienty unwanted crosstalk in film holography (Kubota & Ose, 1979). It utilizes the strict Bragg wavelength selectivity in volume reflection holograms.

The first demonstration of full-color digital holography (FCDH) was achieved using CW Helium Cadmium white light laser with outputs in the 442, 538, and 636 nm (Yamaguchi et al., 2002). The three hologram channels that were produced by the *R*, *G* and *B* channels contain all the essential information about the color of the object. To reconstruct the image of the colored object, one applies the Fresnel method to each of the holograms using the respective recording wavelengths. The grayscale images obtained from the intensity plots of the three-color channels were rendered with the corresponding color equivalents based on the proportions of the intensities of the illumination beams and the spectral sensitivity of the CCD camera (Yamaguchi et al., 2002). Finally, the *R*, *G* and *B* reconstructions were superimposed to produce a white-balanced full color image as viewed in the computer monitor.

Unlike full-color film holography, FCDH is not susceptible to the unwanted effects of emulsion shrinkage, cross talk, and limited spectral sensitivity. Color image dispersion remains an attendant issue during reconstruction that arises from the wavelength dependence of the image size, lateral resolution and depth of focus resulting in a smeared image.

2.1.4 Control of chromatic dispersion

The Fresnel method of digital hologram reconstruction which is applicable for large recording distances could be implemented via a discrete fast Fourier transform algorithm (Garcia et al., 1996). The scaling factor between the pixel sizes Δx in the hologram plane (considered as spatial domain) and $\Delta \tilde{x}$ in the image plane (or the spatial frequency domain) depends on the wavelength λ of the light source and the image distance z (Garcia et al., 1996):

$$\Delta \tilde{x} = \frac{\lambda z}{N \Delta x} \tag{2.1}$$

where $N\Delta x$ is the hologram size or the dimension of the CCD chip. The image size of the reconstruction varies for holograms recorded at different wavelengths and focused at different image distances. Longer wavelengths and a farther image distance result in a large value for $\Delta \tilde{x}$.

Because $\Delta \tilde{x}$ represents a range of spatial frequencies, the entire spatial frequency spectrum of the hologram then can be represented by a fewer number of pixels. Fewer pixels implies a smaller reconstructed image for holograms recorded using longer λ values. The lateral resolution of the reconstructed image varies with λ. Since a reconstructed image from holograms that are recorded at shorter λ's (and closer image distances) are bigger and more details are displayed, it will feature a higher lateral resolution. The depth of focus or longitudinal (axial) resolution of the reconstructions is proportional to λ and is inversely proportional to the square of the effective numerical aperture. Hence, reconstructions that are obtained with shorter λ's have shallower depths of focus. Variances in the image size, lateral resolution and depth of focus of the primary color reconstructions result in chromatic

dispersion. Compensation for chromatic dispersion is important for proper full-color overlay - failure to compensate results in a blurred image. A summary of the dependence of holographic image size, resolution and depth of focus on wavelength and distance is shown in Table 2.1.

Image parameters	Wavelength (λ)	Recording/Reconstruction Distance
Size	Longer λ, smaller size	Farther distance, smaller size
Resolution	Longer λ, lower resolution	Farther distance, lower resolution
Depth of focus	Longer λ, larger DoF	Farther distance, larger DoF

Table 2.1. Dependence of image parameters on wavelength and distance

A method for controlling chromatic dispersion in FCDH was demonstrated using two CW lasers of different wavelengths (Ferraro, 2004). The effects of varying sizes of the reconstructions from the color holograms were compensated by resizing the holograms based on the ratio of wavelengths for a fixed image distance using the relation (Ferraro, 2004):

$$N_2 = \left(\frac{\lambda_2}{\lambda_1} \right) N_1 \tag{2.2}$$

where N_1 and N_2 are the original and resized hologram dimensions, respectively. λ_2 and λ_1 are the longer and shorter wavelengths, respectively.

For example, consider two holograms with dimensions 480 x 640 pixels, one is recorded in red (λ_2 = 636 nm) and the other in green (λ_1 = 503 nm). To equalize the image pixel size for the red and green reconstructions, augment the red hologram by padding the matrix with zeros in both horizontal and vertical directions. Hence, the new dimensions for the red become 734 x 979 pixels. Upon application of the Fresnel method on the resized red hologram at the same image distance z as in the green hologram, the image sizes of the reconstructions from both red and green holograms are equalized. We employ similar procedure for image resizing with one major difference – the image distance is not fixed. A summary of the imaging equations for single-color holography is provided by Goodman (1996).

The following imaging condition is pertinent in FCDH: if the recording and reconstruction wavelengths are the same then the coordinates of the image depend only on the coordinates of the object position, location of the point reference beam and location of the point reconstruction beam. Thus, images of the same object that are reconstructed from primary color holograms are located at the same coordinates if and only if the coordinates of the point reference and reconstruction beams are the same.

Upatnieks presented similar imaging relations and stressed the effects of light source on the quality of reconstructions (Upatnieks, 1979). Exact image formation without size change and aberration requires that the wavelength for recording and reconstruction be the same and the curvature of the reconstruction wave is a replica of the original reference beam or an exact conjugate of it. During computer reconstructions of digital holograms, it is convenient to approximate the reconstruction beam as a plane wave at normal incidence to the

hologram and with the same wavelength as the recording beam. Such a numerical reconstruction beam, however, does not always satisfy the curvature requirement for the reconstruction wave.

The image position of the reconstruction usually differs from object position. Although not a serious issue in single-color holography it becomes critical in FCDH. The wavelength dependence of the beam divergence and chromatic aberrations alter the positions of the reference and illumination beams leading to wavelength-dependent variance between image and object distances. Consequently, the reconstructions from the primary color holograms of the same object are focused at different image distances. The exact amounts of displacements between the color reconstructions are normally difficult to determine since they are influenced by the specific optical elements that are used in the FCDH system. Here, we describe a technique to minimize the chromatic dispersion using the best-focused image distance during the reconstruction of the color hologram channels.

2.2 Stimulated Raman Scattering

Stimulated Raman scattering (SRS) is an inelastic scattering effect whereby an incident photon with frequency ω_p is absorbed by a molecule in the ground state. This causes the molecule to be excited to a virtual state. A second incoming photon can stimulate this molecule to relax to the ground state or to the first excited (vibrational, rotational, or vibrational-rotational) state (Garcia et al., 2002). If the Raman transition involves vibrational energy levels only, the process is known as stimulated vibrational Raman scattering (SVRS). SVRS generates photons shifted by the vibrational Raman frequency ω_{vib}. For H_2, ω_{vib} = 4155 cm^{-1}. By using a linearly polarized 532 nm pump, SVRS can generate the following vibrational Stokes lines: S_{10} = 683 nm, S_{20} = 953.6 nm, S_{30} = 1579.5 nm, …, and the following vibrational anti-Stokes lines: AS_{10} = 435.7 nm, AS_{20} = 368.9 nm, AS_{30} = 319.9 nm, … . If the transition occurs between the rotational energy levels at a single vibrational energy level, the process is known as stimulated rotational Raman scattering (SRRS).

In SRRS the photons generated are shifted by the rotational Raman frequency ω_{rot}. For H_2, ω_{rot} = 587 cm^{-1}. An elliptically polarized 532 nm pump can generate, via SRRS in H_2, the following rotational Stokes lines: S_{01} = 549.1 nm, S_{02} = 567.4 nm, S_{03} = 587.0 nm, … and the following rotational anti-Stokes lines: AS_{01} = 515.9 nm, AS_{02} = 500.7 nm, AS_{03} = 486.4 nm, …, . If the Raman medium has both vibrational and rotational excited states, as in the case of H_2, a photon shifted by both ω_{vib} and ω_{rot} can be generated via the vibrational-rotational Raman effect. This process can generate the following vibrational-rotational Stokes lines: S_{10} – AS_{02} = 632.3 nm, S_{10} – AS_{01} = 656.7 nm, S_{10} – S_{01} = 711.5 nm, S_{10} – S_{02} = 742.5 nm, … and the following vibrational-rotational anti-Stokes lines: AS_{10}- S_{02} = 459.2 nm, AS_{10}- S_{01} = 447.1 nm, AS_{10}- AS_{01} = 424.8 nm, AS_{10}- AS_{02} = 414.5 nm, AS_{20}- S_{01} = 377.1 nm, AS_{10}- AS_{01} = 361.1 nm, ….

3. Experiments in full-color digital holography

3.1 The light source

Figure 1 presents the schematic diagram of the experimental setup for pulsed FCDH which consists of three main sections: 1) Pulsed multi-wavelength light source; 2) Beam conditioning optics; and, 3) Recording and reconstruction setup. The light source is a H_2 Raman shifter that is pumped by the 355 nm output beam of a pulsed Nd:YAG laser (Spectra Physics GCR-230-10) with a pulse duration of about 5 ns. The plane mirrors direct the pump beam into the H_2 Raman shifter. Lens *L1* (focal length, 50 cm) focuses the pump

laser beam into a Raman cell (length, 58 cm) that is filled with hydrogen gas (99.9999% purity). At the exit port of the cell is lens *L2* (focal length, 50 cm), which collimates the Raman output beams.

The Raman output beams are directed towards the beam conditioning section. The output beams of the Raman shifter consist of the following spectral components: Rayleigh (355 nm), Stokes (S_1, S_2, …) and anti-Stokes (aS_1, aS_2, …). The output beam energies depend on the hydrogen gas pressure in the Raman cell and on the pump energy (Garcia, W. et al., 2002). At a hydrogen gas pressure of 1.38 MPa and a pump energy of 6.5 mJ, the following Raman output lines (in nm) are obtained: 415.9 (S_1; blue), 502.9 (S_2; green), 635.9 (S_3; red), 864.5 (S_4), 309 (aS_1), 273.8 (aS_2), 245.8 (aS_3), and 222.9 (aS_4). The beam energy decreases with the Stokes (or anti-Stokes) number. The S_1 beam energy is 2.12 mJ which is 1.43 times greater than that of S_2 and 3.3 times higher than that of S_3. The Stokes beam energies are also larger than their anti-Stokes counterparts. For pulsed full-color digital holography, the first three Stokes output beams S_1 (blue), S_2 (green) and S_3 (red) are utilized as the primary additive color channels.

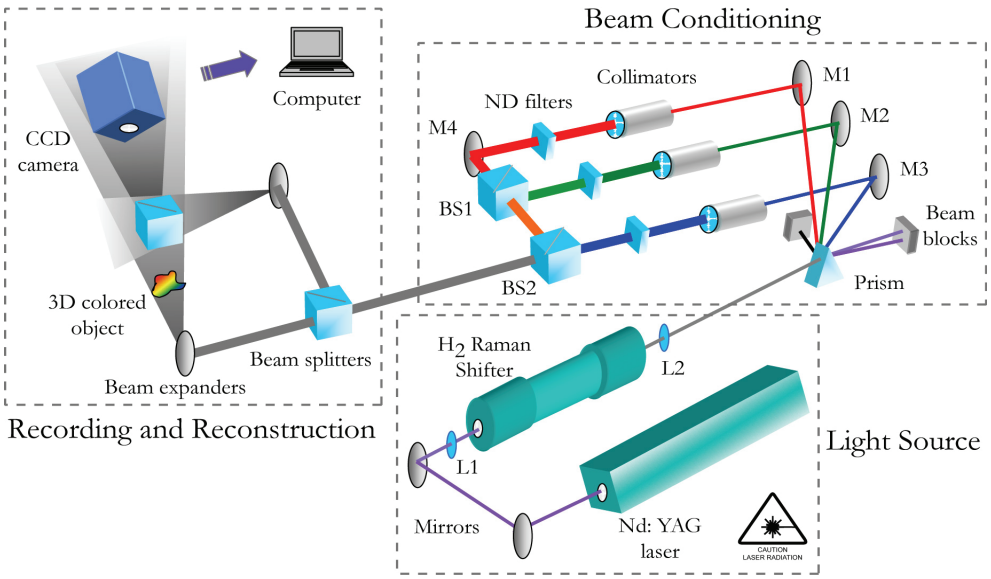


Fig. 1. Experimental setup for pulsed full-color digital holography.

3.2 Beam conditioning

To make the Raman shifter suitable for pulsed FCDH, we need to address the unequal relative energies and poor transverse intensity distributions of the S_1, S_2 and S_3 beams. The transverse intensity distribution of a Raman output beam is non-Gaussian. Uniform beam profiles and equal beam intensities are important for achieving uniform object illumination and better control of the beam ratio. Conditioning is required for the Stokes beams before they can be fully utilized as a holographic light source.

The beam-conditioning section consists of a Pellin-Broca prism, beam blocks, deflecting mirrors (*M1, M2, M3,* and *M4*), three sets of beam collimators, neutral density filters and

beam splitters (*BS1*, *BS2*). The prism disperses the output beams of the Raman shifter such that only the S_1, S_2 and S_3 beams are able to proceed to the collimators. The components of the collimators are not shown in Fig 1 for simplicity. Each collimator is composed of a first lens (focal length f = 11 cm, diameter d = 5.08 cm) that individually focuses the Stokes beams to spatial filters (d = 0.15 cm). The light that is transmitted by each filter is collimated by a second lens (f = 25 cm, d = 4.13 cm). The relative intensities of the collimated S_1, S_2 and S_3 beams are equalized with a set of neutral density filters (*ND* filters). The equalized Stokes beams are then recombined via mirror *M4* and beam splitters *BS1* and *BS2*. After *BS2*, the intensity of each Stokes beam has been reduced to 50 μW which is safe for the operation of the CCD camera (SONY XC-75, 640 x 480 pixels, monochrome). The recombined beams form the new holographic pulsed white light source.

Figure 2 demonstrates the benefits of beam conditioning on the beam profiles and relative intensities of the primary color channels from S_1, S_2 and S_3 Raman outputs. In Fig. 2(a) is an image of the selected Raman outputs before performing beam conditioning. The beam is characterized by the presence of intense high-frequency spatial structures and scattered color components. The beam intensity also exhibits temporal and spatial variations between pulses that are due to non-uniform gas heating in the Raman cell. These deleterious effects are minimized by the constructed beam conditioning section.

Figures 2(b) - 2(d) illustrate the improvement of the beam profiles and relative intensities which is 50 μW for each of the laser lines. Figure 2(e) presents the white light beam after superimposing the color channels (beam width ~ 2.54 cm). A slight chromatic dispersion remains in the superposition (manifested as halo) due to the wavelength dependence of beam divergence and chromatic aberration in the optics elements. We utilized the white light beam to record the RGB holograms of various test samples.

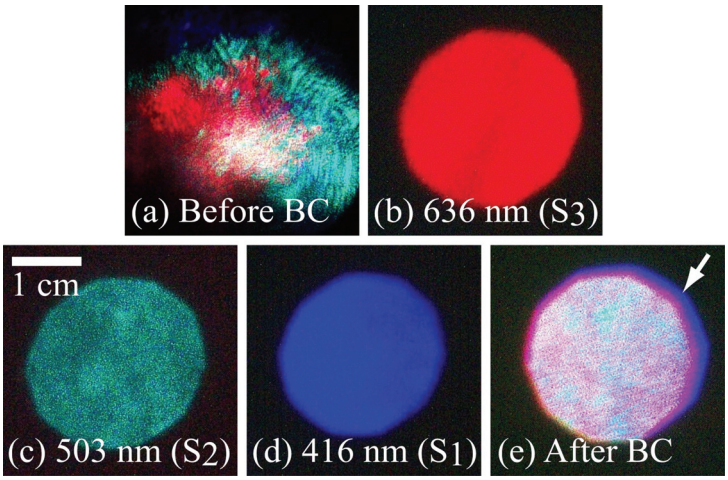


Fig. 2. Effects of beam conditioning (BC) on beam quality. (a) Before BC: overlapped S_1, S_2 and S_3 output beams of the Raman shifter. Uniform profiles and equalized intensities after BC for (b) S3; (c) S2; (d) S1 beams; and, (e) superposition of the beams. Arrow indicates the presence of a halo.

3.3 Recording and reconstruction

Hologram recording and reconstruction are performed with a typical in-line digital holographic setup. In the in-line configuration, the object and reference beams are approximately co-linear which result in widely-spaced interference fringes. Large fringes are ideal for enhanced CCD sampling. We adjusted the positions of the beam expanders (concave mirrors) for back illumination in the case of a transmitting object and for front illumination in the case of a reflecting opaque object. During hologram recording, the test object is illuminated, one beam at a time, by blocking the other two beams in the beam conditioning section. We used a negative resolution target as a transmitting test object and captured the interference of the object and reference beams with the CCD camera. The recorded *R*, *G* and *B* holograms produced by the S_3, S_2 and S_1 beams respectively, are resized, reconstructed, rendered with color and, finally, unified in the computer.

3.4 Resizing the holograms

Figure 3 depicts the resizing of the holograms based on the wavelength ratios. The original size of the *R*, *G* and *B* holograms is 480 x 640 pixels [Fig. 3(a) – 3(c)]. Figure 3(d) is the resized hologram for the *R* channel with new dimensions 734 x 979 pixels that is obtained after embedding the hologram in Fig 3(a) in a matrix with zero amplitude. We obtained the new dimensions by applying Eqn. 2.2 in both the vertical and horizontal directions with λ_2 = 636 nm and λ_1 = 416 nm.

Figure 3(e) is the resized hologram for the green channel with new dimensions 581 x 774 pixels. The new dimensions are obtained by applying Eqn. 2.2 in the vertical as well as in the horizontal direction with λ_2 = 503 nm and λ_1 = 416 nm. The blue hologram retains its original dimensions.

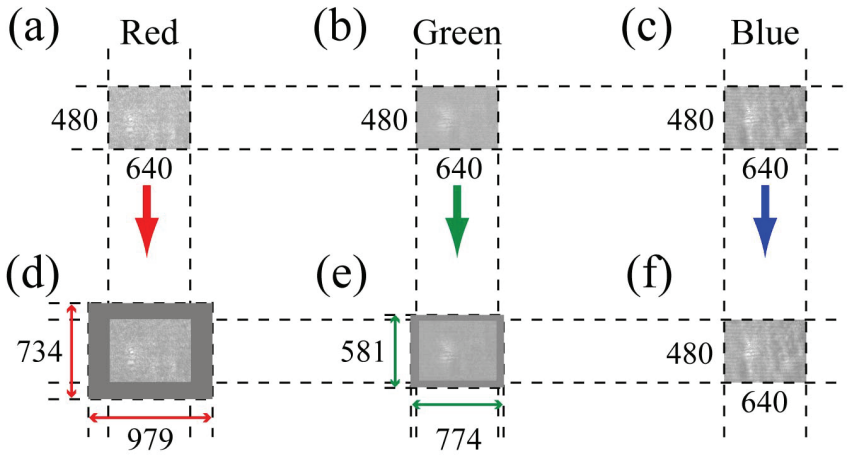


Fig. 3. Resizing of the color holograms. Original size of the (a) red, (b) green and (c) blue holograms with dimensions 480 pixels (V) x 640 pixels (H). (d) and (e) are the resized red and green holograms, respectively, with new dimensions indicated. Blue hologram retains its original dimensions.

3.5 Reconstruction of holograms

The *B* hologram and the resized *R* and *G* holograms are reconstructed at distances where the images are best focused. The distance of the best-focused image is established by numerically focusing the hologram at various distances and then selecting the sharpest image.
For the purposes of comparison, reconstructions from resized holograms are also plotted at the same reconstruction distance (Ferraro et al., 2004). The fixed distance condition corresponds to the reconstruction distance of the blue channel. The *RGB* reconstructions are then compared in terms of size and resolution.

4. Control of chromatic dispersion

4.1 Wavelength dependence of image parameters

Figure 4 shows the recorded holograms and the test object used. In Figs. 4(a) to 4(c) are the recorded *R*, *G* and *B* holograms, respectively. The test object is an element number "5" of the negative resolution target with a height of 4 mm. Figure 4(d) is an image of the test object that is illuminated by the Raman white light source. The *RGB* holograms are then evaluated using the Fresnel method using the respective recording wavelengths. The lens-like nature of holograms results in the focusing of the incident reconstruction beam forming an image. The image size, lateral resolution and longitudinal resolution are investigated for their wavelength dependences.

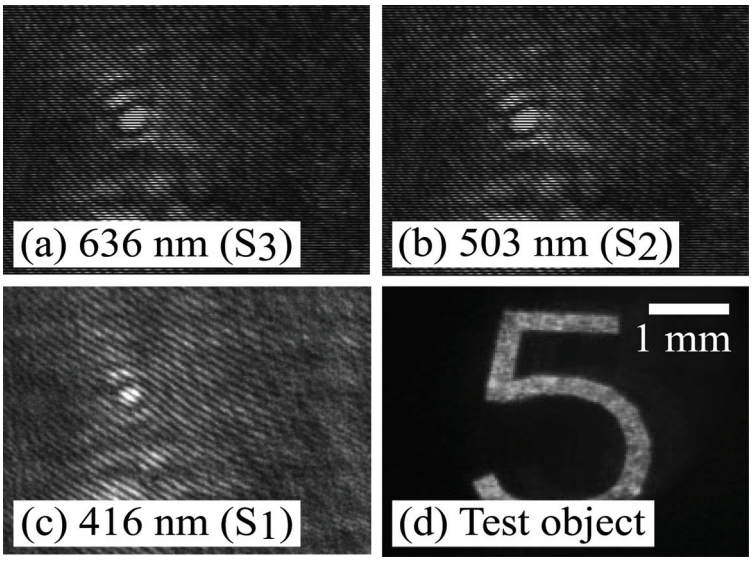


Fig. 4. Recorded holograms. (a) Red, (b) green and (c) blue holograms of the transmitting test object shown in (d).

Figure 5 presents a matrix of the color reconstructions that reveal the effects of wavelength and object-to-CCD recording distance (*z*) on the image size. For a fixed distance, down a column, the image size increases with decreasing wavelength. For a constant wavelength, moving from left to right, the image size decreases as distance increases. The results demonstrate that the image size increases with decreasing wavelength.

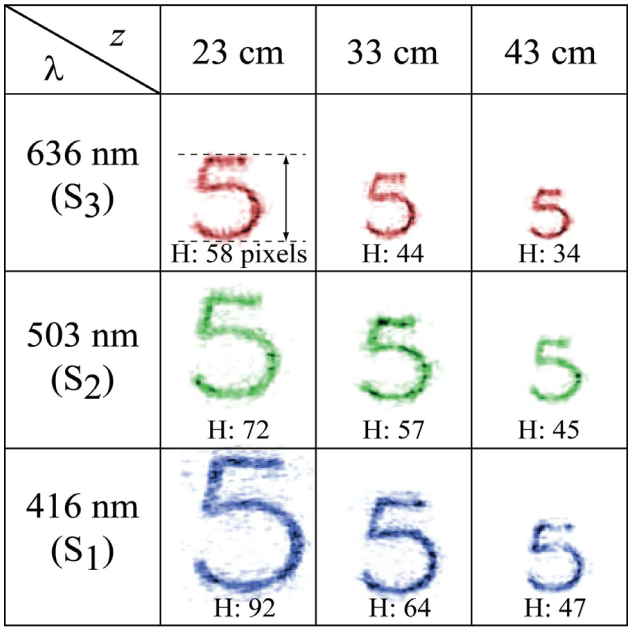


Fig. 5. Dependences of the image size on illumination wavelength and object-to-CCD distance (z) during recording. Evaluation of the color reconstructions is based on image height (H, in number of pixels).

Figure 6 illustrates the effects of the light source wavelength on the longitudinal (axial) resolution of the reconstructions. At a particular wavelength, an intensity line scan is obtained across an array of reconstructions plotted at equally-spaced axial positions. Figure 6(a) shows the arrays of *RGB* reconstructions from holograms recorded at the same object-to-CCD distance. The intensity line scans are obtained through the middle portions of the number "5" and then are plotted at the different axial distances.

Figures 6(b) - 6(d) show the intensity scans for the *R*, *G* and *B* arrays, respectively. To estimate how fast a focused image blurs with axial distance, here we considered the distance range (Δz) where the intensity decreases to half its maximum value. The intensity scan for the red reconstructions [Fig. 6(b)] exhibits slower change across the axial direction (Δz = 15 cm). While the blue reconstructions [Fig 6(d)] exhibit faster change in intensity (Δz = 7.5 cm). The results demonstrate that the longitudinal resolution increases with shorter wavelengths. The reconstructions from the blue hologram is said to have a narrow depth of focus while those from the red hologram have a larger depth of focus.

Lateral resolution also improves at shorter wavelengths as illustrated in Fig. 5 where the images are larger for the case of blue reconstructions. A larger reconstructed image of an object would reveal more details – an observation that is consistent with the wavelength dependence of the pixel size in the image plane. Since each image pixel represents a range of spatial frequencies (by virtue of the Fourier transformation) a bigger reconstruction can also be interpreted as a representation of higher spatial frequencies that contain the finer details of the object.

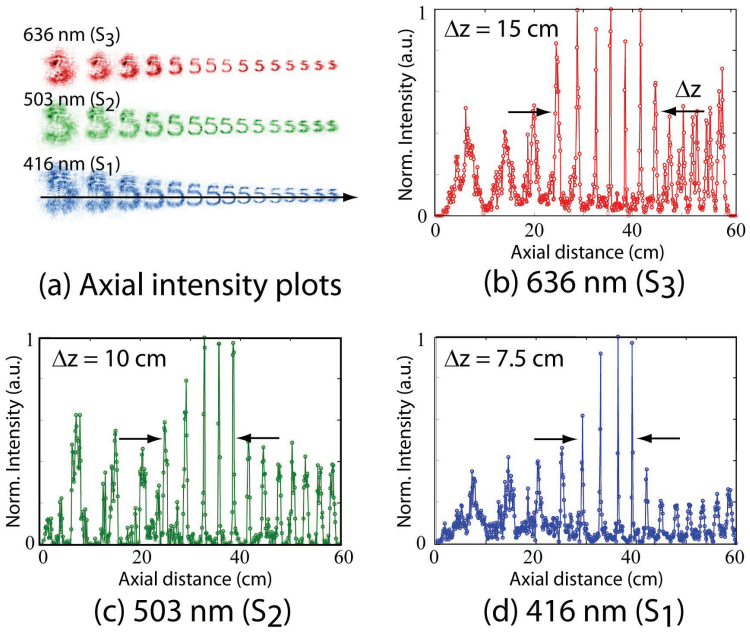


Fig. 6. Dependence of longitudinal (axial) resolution of the reconstructions on illumination wavelength. (a) Arrays of *RGB* reconstructions plotted at equally-spaced axial positions. The arrowed line indicates the scanned portions across an array. (b), (c) and (d) are the axial intensity line scans for the *R*, *G* and *B* reconstructions, respectively. Based on the range (Δz) where intensity decreases to half its maximum, the narrowest depth of focus (highest axial resolution) is observed in the reconstructions at the shortest wavelength (416 nm, S_1) (d).

4.2 Proper fit of the color reconstructions

Chromatic dispersion of the superimposed reconstructions can be controlled by resizing the holograms and reconstructing at best-focused distances. Figure 7 demonstrates the effects of hologram resizing on the size and resolution of the superimposed *RGB* reconstructions.

Figure 7 sub-panels (a) - (c) are the *RGB* reconstructions from the original size holograms. The heights of the reconstructions are 53 pixels, 65 pixels and 76 pixels, for *R*, *G*, and *B* channels, respectively. The superimposed image reconstruction [Fig. 7(d)] evidently demonstrates the wavelength dependences of the image size. The reconstruction distances where the images are best focused are 22.0 cm, 21.0 cm and 20.5 cm for *R*, *G*, and *B* reconstructions, respectively. The variance of the image distance can be attributed to chromatic aberrations in the optics used.

Figure 7 sub-panels (e) – (g) are the reconstructions from the resized holograms and plotted using a fixed image distance of 20.5 cm (chosen from image distance of the blue color channel). The images are successfully resized but the heights are not exactly equalized as depicted in the superimposed image [Fig. 7(h)]. The heights of the reconstructions are 81 pixels, 73 pixels and 76 pixels for *R*, *G*, and *B* channels, respectively. Such an unsatisfactory result could be attributed to the assumption that the location of the image does not vary significantly with wavelength.

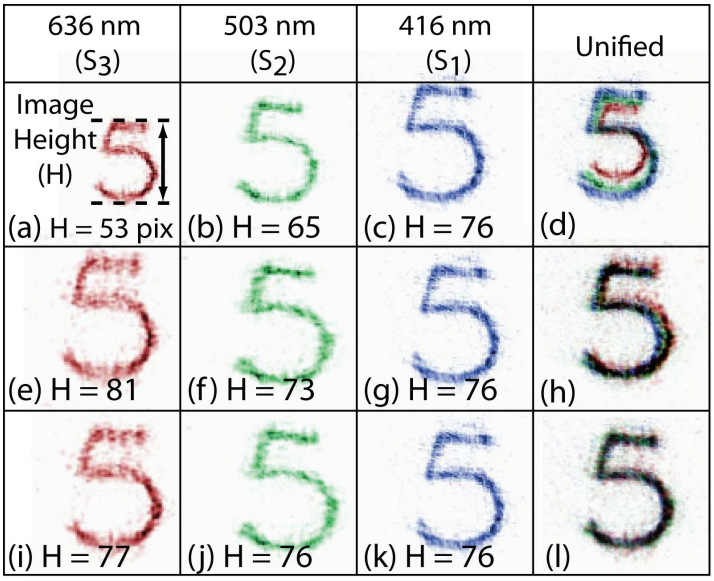


Fig. 7. Control of chromatic dispersion. (a) – (d) *RGB* reconstructions and unified image using holograms at original size. (e) – (h) Reconstructions from resized holograms and plotted at a fixed image distance. (i) – (l) Reconstructions from resized holograms and plotted at best-focused image distances.

Figures 7 sub-panels (i) – (k) are the reconstructions obtained from resized holograms using image distances of 22.0 cm, 21.0 cm and 20.5 cm for the *R*, *G*, and *B* channels, respectively. The heights of the reconstructions are 77 pixels, 76 pixels and 76 pixels for *R*, *G*, and *B* channels respectively. The heights are more uniform when the image distances that are utilized correspond to those of the best focused images. The small variance that has remained in the heights of the reconstructions is attributed to the depth of focus where the image is acceptably focused. Proper fit of the RGB reconstructions is achieved after resizing the holograms based on wavelength ratios and reconstructing at best-focused image distances [Fig. 7(l)].

Figure 8 demonstrates the improvement of the fit of the full-color reunified image after application of the resizing technique. Figure 8(a) presents the intensity line scan across a portion of the superposed reconstructions obtained without resizing the holograms. At this stage, the RGB channels are unsuitable for unification as evident from the split lines corresponding to a solid portion of the number "5". Figure 8(b) shows the line scan for the unified image after resizing the holograms at a fixed image distance. The presence of bifurcation in the intensity scan indicates a poor fit of the *RGB* channels after the reconstructing at a fixed distance. Figure 8(c) presents the line scan of the unified image after hologram resizing and plotting at the best-focused image distances. The smooth intensity scan demonstrates proper fit of the *RGB* channels when the holograms are resized based on wavelength ratios and reconstructed at best-focused image distances.

Figure 9 conveys the importance of hologram resizing on the longitudinal resolution. Figure 9(a) presents a horizontal array of full-color reconstructions in the absence of hologram resizing. The reconstructions are plotted at equally-spaced axial planes in the foreground and background of the best focused image planes. We used the contrast of the intensity profile as a measure of the axial resolution. High contrast in the axial intensity plot within a short axial range would mean enhanced localization of the colored test object.

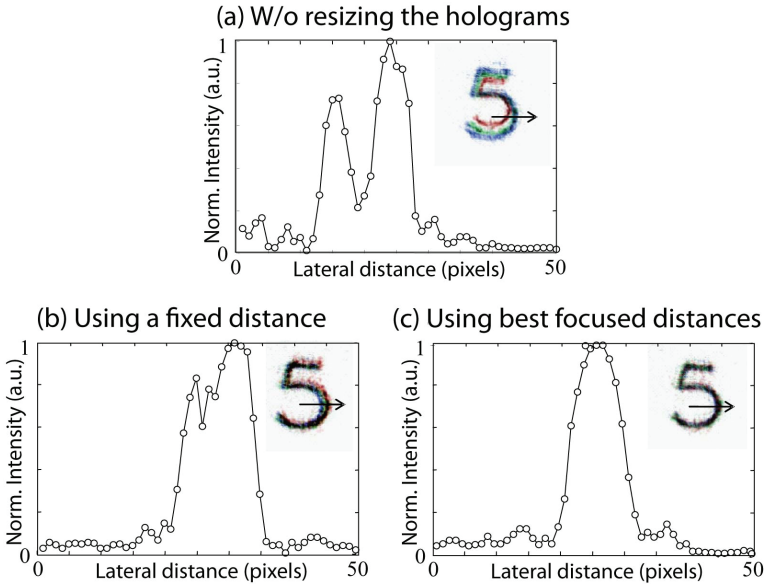


Fig. 8. Intensity line scans to demonstrate the effects of hologram resizing on the proper fit of the *RGB* reconstructions. (a) Without hologram resizing. (b) With hologram resizing and reconstruction at a fixed distance. (c) With hologram resizing and reconstruction at best-focused image distances.

Figure 9(b) shows the horizontal intensity line scan across the array in Fig. 9(a). The line scan passes through the mid portion of the reconstructed images of "5" as indicated by the arrowed line in Fig. 9(a). The intensity profile has a low contrast (difference between maximum and minimum intensity values as depicted by the circles) and exhibits a gradual change across the array. Figure 9(c) shows an array of full-color reconstructions after resizing the holograms. The high contrast in the intensity profile [Fig. 9(d)] indicates an improvement of the longitudinal or axial resolution as a result of the hologram resizing. Enhanced axial resolution is important in optical sectioning where a specific plane of a three dimensional scene has to be accessed. The full-color reconstruction must exhibit high longitudinal resolution or narrow depth of focus in order to resolve the various axial planes.

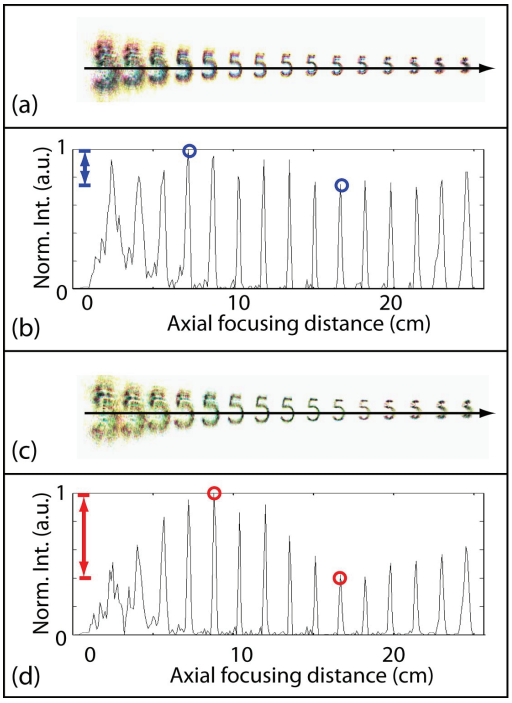


Fig. 9. Effects of hologram resizing on longitudinal resolution. (a) Array of full-color reconstructions without hologram resizing. (b) Intensity scan across the array in (a) shows low contrast and gradual change. (c) Full-color reconstructions from resized holograms. (d) Intensity scan across the array in (c) shows high contrast and rapid change demonstrating enhanced axial resolution.

5. Summary and conclusions

We have discussed pulsed FCDH with a hydrogen Raman shifter as the single source of pulsed, highly-directional multi-wavelength light. The first three Stokes beam outputs (λ_{S1} = 415.9 nm, λ_{S2} = 502.9 nm, λ_{S3} = 635.9 nm) of the Raman shifter are utilized as the three primary color channel for carrying all the color information from the object. A simple beam conditioning procedure for improving the beam quality of the Stokes beams and for equalizing their beam was developed and implemented and holographic recording and reconstruction were both successfully demonstrated. The Raman shifter is a promising light source for pulsed FCDH because it is inexpensive to construct and maintain.

The usefulness of full color holograms depends on the quality and resolution of the reconstructed images. Higher resolution and narrower depth of focus are gained for reconstructions from holograms that are recorded at shorter wavelengths. The wavelength dependence of the image size and resolution spatially disperses the reconstructions from the primary color channels resulting in blurred unacceptable image. Compensation for color dispersion could be done by resizing the holograms using the ratio of wavelengths and reconstructing at a fixed image distance. However, the wavelength dependence of the beam

divergence and optical aberrations result in the spatial displacement of the point reference beam that directly changes the direction and shape of the reference beam and therefore, the image distance. Knowledge of the variance in the image distance for the primary color channels, has allowed us to rescale using the image distance that yielded the best focused image thereby assuring better compensation of size, resolution and depth of focus.

6. Acknowledgments

We acknowledge the financial support of PCASTRD-DOST and OVCRD-UP Diliman. We also thank M. Cadatal, F. Daquiado, S. Ledesma, M. Francisco, I. Quiatchon, R. Ibarreta, J. Palero and C.A. Alonzo of NIP for their valuable technical assistance in the various stages of the FCDH work.

7. References

[1] Almoro, P., Cadatal, M., Garcia, W. & Saloma, C. (2004). Pulsed multicolor digital holography with hydrogen Raman shifter. Appl Opt. 43, 2267-2271.

[2] Almoro, P., Garcia, W. & Saloma, C. (2007). Colored object recognition by digital holography and a hydrogen Raman shifter. Opt Express 15, 7176-7181.

[3] Andreeva, O.V. (2002). Proposals for a terminological dictionary of optics. Holography, J. Opt. Technol. 69 , 367-374.

[4] Bates, H. (1973). Burst-Mode Frequency-Doubled YAG:Nd3+ Laser for Time-Sequenced High-Speed Photography and Holography. Appl. Opt. 12, 1172-1178.

[5] Benton, S.A. (2002). Holography reinvented, Proc. SPIE 4737, 23-26.

[6] Bjelkhagen, H. I. & Vukicevic, D. (2002). Color holography: a new technique for reproduction of paintings. Proc. SPIE 4659, 83-90.

[7] Bjelkhagen, H. I. (2002). Super-realistic imaging based on color holography and Lippmann photography. Proc. SPIE 4737, 131-141.

[8] Colomb, T., Pavillon, N., Kühn, J., Cuche, E., Depeursinge, C., & Emery, Y. (2010). Extended depth-of-focus by digital holographic microscopy. Opt. Lett. 35, 1840-1842.

[9] Desse, J., Albe, F. & Tribillon, J. (2002). Real-time color holographic interferometry. Appl. Opt. 41, 5326-5333.

[10] Demoli, N., Vukicevic, D. & Torzynski, M. (2003). Dynamic digital holographic interferometry with three wavelengths. Opt. Express 11, 767-774.

[11] Ferraro, P., De Nicola, S., Coppola, G., Finizio, A., Alfieri, D. & Pierattini, G. (2004). Controlling image size as a function of distance and wavelength in Fresnel-transform reconstruction of digital holograms. Opt. Lett. 29, 854-856.

[12] Garcia, J., Mas, D. & Dorsch, R. (1996). Fractional-Fourier-transform calculation through the fast-Fourier-transform algorithm. Appl. Opt. 35, 7013-7018.

[13] Garcia, W., Palero, J. & Saloma, C. (2001). Temporal coherence control of Nd:YAG pumped Raman shifter. Opt. Commun. 97, 109-114.

[14] Goodman, J. (1996). Introduction to Fourier Optics. 2nd ed., McGraw-Hill, Singapore.

[15] Gustafsson, J. (2000). Finding new wavelengths for pulsed holography. Proc. SPIE 4149, 353-358.

[16] Harthong, J., Sadi, J., Torzynski, M. & Vukicevic, D. (1997). Speckle phase averaging in high-resolution color holography. J. Opt. Soc. Am. A 14, 405-410.
[17] Javidi B., & Tajahuerce, E. (2000). Three-dimensional object recognition by use of digital holography. Opt. Lett. 25, 610-612.
[18] Jeong, T. H., Bjelkhagen, H. I. & Spoto, L. (1999). Holographic interferometry with multiple wavelengths. Appl. Opt. 36, 3686-3688.
[19] Kreis, T. (2002). Frequency analysis of digital holography with reconstruction by convolution. Opt. Eng. 41, 1829-1839.
[20] Kubota, T., Takabayashi, E., Kashiwagi, T., Watanabe, M., & Ueda, K. (2001). Color reflection holography using four recording wavelengths. Proc. SPIE 4296, 126-133.
[21] Kubota, T. & Ose, T. (1979). Lippmann color holograms recorded in methylene-bluesensitized dichromated gelatine. Opt. Lett. 4, 289-291.
[22] Kubota, T. (1986). Recording of high quality color holograms. Appl. Opt. 25, 4141-4145.
[23] Leith, E. N., Lyon, P., & Chen, H., (1991). Imaging problems with femtosecond-pulse holography. J. Opt. Soc. Am. A 8, 1014-1018.
[24] Lin, L., & LoBianco, C. (1967). Experimental Techniques in Making Multicolor White Light Reconstructed Holograms. Appl. Opt. 6, 1255-1258.
[25] Liu, Z., Centurion, M., Panotopoulos, G., Hong, J., & Psaltis, D. (2002). Holographic recording of fast events on a CCD camera. Opt. Lett. 27, 22-24.
[26] Mallick, S. (1975). Pulse Holography of Uniformly Moving Objects. Appl. Opt. 14, 602-605.
[27] Olivares-Perez, A., Berriel-Valdos, L. & Morales, A. (1989). Magnification effect and color blur behavior in holography. Appl. Opt. 28, 4366-4369.
[28] Pedrini, G., Gusev, M., Schedin, S., & Tiziani, H. J. (2003). Pulsed digital holographic interferometry by using a flexible fiber endoscope. Opt. Lasers Eng. 40, 487-499.
[29] Pedrini, G., Schedin, S., & Tiziani, H. J. (2002). Pulsed digital holography combined with laser vibrometry for 3D measurements of vibrating objects. Opt. Lasers Eng. 38, 117-129.
[30] Peercy, M., & Hesselink, L. (1994). Wavelength selection for true-color holography. Appl. Opt. 33, 6811-6817.
[31] Schnars U. & Jüptner W. (2002). Digital recording and numerical reconstruction of holograms, Meas. Sci. Technol. 13, R85-R101.
[32] Schulze, M. A., Hunt, M., Voelkl, E., Hickson, J., Usry, W., Smith, R., Bryant R. & Thomas Jr., C. (2003). Semiconductor wafer defect detection using digital holography, Proc. SPIE 5041, 183-`193.
[33] Seebacher, S., Osten, W., Baumbach, T., & Jüptner, W. (2001). The determination of material parameters of micro-components using digital holography. Opt. Lasers Eng. 36, 103-126.
[34] Upatnieks, J., (1979) Image Formation. Handbook of Optical Holography. Caulfield, H. (Ed.), New York.
[35] Wedendal, P. R. & Bjelkhagen, H. I. (1974). Dynamics of Human Teeth in Function by Means of Double Pulsed Holography; an Experimental Investigation. Appl. Opt. 13, 2481-2485.

[36] Yamaguchi, I., Matsumura, T. & Kato, J. (2002). Phase-shifting color digital holography. Opt. Lett. 27, 1108-1110.
[37] Zhang, T. & Yamaguchi, I., (1998). Three-dimensional microscopy with phase-shifting digital holography. Opt. Lett. 23, 1221-1223.

9

Optical Holography Reconstruction of Nano-objects

Cesar A. Sciammarella[1,2], Luciano Lamberti[2] and
Federico M. Sciammarella[1]
[1]*College of Engineering & Engineering Technology, Northern Illinois University,*
[2]*Dipartimento di Ingegneria Meccanica e Gestionale, Politecnico di Bari,*
[1]*USA*
[2]*Italy*

1. Introduction

The continuous growth of many fields in nanoscience and nanotecnology puts the demand for observations at the sub-micron level. Electron microscopy and X-rays can provide the necessary short wavelengths to gather information at the nanometer and sub-nanometer range but, in their current form, are not well suited to perform observations in many problems of scientific and technical interest. Furthermore, the environment required for the observation via X-rays or electron microscopy is not suitable for some type of specimens that it is necessary to study. Another concern is the changes that may be induced in the specimen's structure by the utilized radiation. These issues have led to the return to optics and to the analysis of the optical problem of "super-resolution", that is the capacity of producing optical images beyond the classical diffraction limit.

Classical optics has limitations on the resolution that can be achieved utilizing optical microscopy in the observation of events taking place at the sub-micron level, i.e. to a few hundreds of nanometers. To overcome this limitation and achieve super-resolution, non conventional methods of illumination such as evanescent waves are utilized. The initial approach to the utilization of evanescent field properties was the creation of near-field techniques: a probe with dimensions in the nano-range detects the local evanescent field generated in the vicinity of the objects that are observed. A perspective review on super-resolution can be found in Sciammarella (2008).

New methods recently developed by C.A. Sciammarella and his collaborators (see, for example, Sciammarella, 2008; Sciammarella *et al*., 2009) rely on the emission of coherent light by the objects that are under analysis. This is done through the phenomenon of light generation produced by electromagnetic resonance. Object self-luminosity is the consequence of electromagnetic resonance. Why self-luminosity may help to increase resolution? The light generated in this way has particular properties that are not present in the light sent by an object that results from external illumination. The produced wave fronts can travel long distances or go through an optical system without the diffraction changes experienced by ordinary wave fronts (see, for example, Durnin *et al*., 1987; Buchal, 2003; Gutierrez-Vega *et al*., 2001; Hernandez-Aranda *et al*., 2006).

This paper will discuss two alternative approaches to near field techniques: (i) Using diffraction through the equivalent of a diffraction grating to generate an ample spectrum of wave vectors; (ii) Exciting the objects to be observed with the evanescent fields so that the objects become self-luminous. Examples of optical holography reconstruction of nano-objects are presented in the chapter.
The paper is structured as follows. After the Introduction section, evanescent waves and super-resolution are briefly recalled in Section 2. Section 3 describes the experimental setup. Section 4 analyzes the diffraction pattern of a 6 μm diameter polystyrene microsphere immersed in a NaCl solution and illuminated by evanescent light. Section 5 analyzes the system of rectilinear fringes observed in the image. Section 6 describes the process of formation of holograms at the nano-scale. Sections 7 and 8 respectively present the results of optical reconstruction of NaCl nanocrystals and polystyrene nanospheres contained in the saline solution. Finally, Section 9 summarizes the most important findings of this study.

2. Theoretical background

2.1 Properties of evanescent waves

The self generation of light is achieved through the use of total internal reflection (TIR). At the interface between two media such that the index of refraction of medium 1 is larger than the index of refraction of medium 2 (i.e. $n_1>n_2$), if a light beam is incident with an angle $\theta_i>\theta_{i,crit}$, a total reflection of the beam, in which essentially all of the light is reflected back into the first medium, takes place (Fig. 1a). Even though the light no longer propagates into the second medium, there is a small amount of penetration of the electromagnetic field across the interface between the two media. In the vector form solution of Maxwell equations, it can be shown that, under the condition of total reflection, the electromagnetic field does not disappear in the second medium (Born and Wolf, 1999). However, there is no energy exchange with the second medium. The components of the electromagnetic field transmitted in the second medium, vectors **E** and **H**, depend on (Born and Wolf, 1999):

$$\exp\left[-i\frac{2\pi}{\lambda}z\sqrt{\frac{\sin^2\theta_i}{{n_{12}}^2}-1}\right] \tag{1}$$

where the relative index of refraction is defined as $n_{12}=n_2/n_1$.
The intensity of the field decays by 1/e at the distance z from the interface equal to:

$$z=\frac{\lambda}{2\pi\sqrt{\frac{\sin^2\theta_i}{{n_{12}}^2}-1}} \tag{2}$$

Under this condition a particular type of waves are produced in the interface (Fig. 1a). These waves are called *evanescent waves* and travel at the interface decaying exponentially in the second medium at a depth that is a fraction of the wavelength of the utilized light as shown by Eq. (2). The decay of electromagnetic field is sketched in Fig. 1a.
The electromagnetic field of evanescent waves hence does not propagate light in the second medium. However, if a dielectric medium or a conducting medium comes in contact with the evanescent field, light is emitted by the medium itself. This interaction depends on the properties, size and geometry of the medium. In the case of a dielectric medium, through

Rayleigh molecular scattering, light is emitted in all directions as illustrated in Fig. 1b. The light emission is a function of the electronic configuration of the medium. If the medium is a metal, the Fermi's layer electrons produce resonances called plasmons. The effect is reversible, photons can generate plasmons, and the decaying plasmons generate photons. In the case of very small dielectric objects, the resonance takes place at the level of the bound electrons. The actual dimensions of the object determine the different resonance modes and light at frequencies different from the frequency of illuminating light is generated.

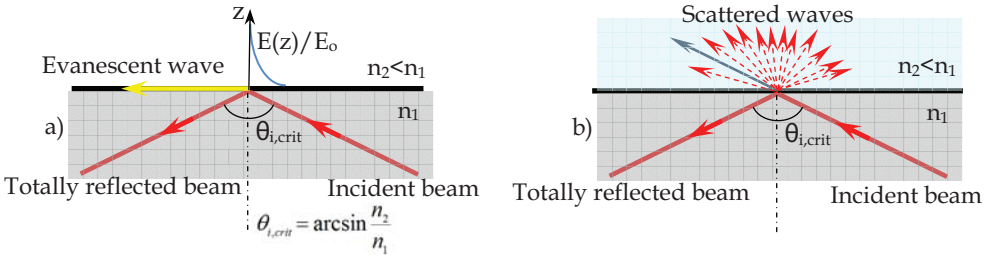


Fig. 1. a) Formation and characteristics of evanescent waves; b) Propagation in a dielectric medium.

According to the Quantum Mechanics principle of preservation of momentum for the photons, some of the energy of the incident beam continues in the second medium (the direction indicated by the blue arrow in Fig. 1b).

2.2 Super-resolution

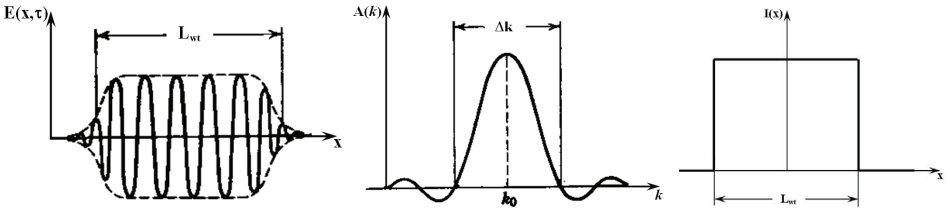
The possibility of getting higher resolutions depends on the energy available and detector spatial frequency capability. This was foreseen by Toraldo di Francia (Toraldo di Francia, 1952) when he postulated that the resolution of an optical system could be increased almost continuously beyond the classical Rayleigh's diffraction limit, provided that the necessary energy to achieve these results is available. When Toraldo di Francia presented this original work (Toraldo di Francia, 1958) it was argued that his proposed super-resolution approach violated the Heisenberg uncertainty principle. However, the arguments using the Heisenberg principle, those arguments are included in many text books of Optics, can be easily dismissed because they are based on the wave function of a single photon while one is dealing with the wave function of millions of photons. The question can be summarized according to Yu (2000). If ΔE is the amount of energy invested in an observation and Δx is a distance to be measured, the Heisenberg principle can be stated as follows:

$$\Delta E \, \Delta x \geq \frac{hc}{2} \tag{3}$$

where h is the Planck's constant, c is the speed of light. By increasing the energy there is no limit to the smallness of distance that can be measured. Therefore, increases in smallness of the distance that one wants to measure will require enormous increases in the energy that must be invested. This is because in Eq. (3) there is a large factor: the speed of the light c. There is another important application of the Heisenberg principle that is more relevant to the topic of this chapter and was pointed out by Vigoureux (2003). This relationship is:

$$\Delta x \, \Delta k_x > 2\pi \tag{4}$$

The wave propagation vector has two components $\overrightarrow{k_x}$ and $\overrightarrow{k_y}$. In Eq. (4), $\overrightarrow{k_x}$ is the wave vector component in the x-direction. Vigoureux showed that for waves propagating in the vacuum Eq. (4) leads to the Rayleigh limit of $\lambda/2$. In order to go beyond the $\lambda/2$ limit, one must have values of $\overrightarrow{k_x}$ falling in the field of evanescent waves. Therefore, to capture evanescent waves is an effective approach to getting super-resolution. This conclusion is in agreement with the conjecture made by Toraldo di Francia in 1952.



a) b) c)

Fig. 2. a) Harmonic wave train of finite extent L_{wt}; b) Corresponding Fourier spectrum in wave numbers *k*; c) Representation of a spatial pulse of light whose amplitude is described by the rect(x) function.

A further mathematical argument may be based on the Fourier expansion of the Maxwell equation solution. In the classical optics scheme of plane-wave solutions of the Maxwell equations, monochromatic waves of definite frequencies and wave numbers are considered. This idealized condition does not apply in the present case. One can start from the Fourier solution of the Maxwell equations in the vector field (time-space):

$$E(x,\tau) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} A(k) e^{ikx - i\omega(k)\tau} \, dk \tag{5}$$

where: $E(x,\tau)$ is the scalar representation of the propagating electromagnetic field, x is the direction of propagation of the field, τ is the time, $A(k)$ is the amplitude of the field, *k* is the wave number $2\pi/\lambda$, $\omega(k)$ is the angular frequency. $A(k)$ provides the linear superposition of the different waves that propagate and can be expressed as:

$$A(k) = \sqrt{2\pi} \;\; \delta(k\text{-}k_o) \tag{6}$$

where $\delta(k\text{-}k_o)$ is the Dirac's delta function. This amplitude corresponds to a monochromatic wave $E(x,\tau) = e^{ikx - i\omega(k)\tau}$. If one considers a spatial pulse of finite length (Fig. 2a), at the time $\tau=0$, $E(\xi,0)$ is (see Fig. 2c) a finite wave-train of length L_{wt} where $A(k)$ is a function spreading a certain length Δk (Fig. 2b). The dimension of L_{wt} depends on the analyzed object size. In the present case, objects are smaller than the wavelength of the light.

Since L_{wt} and Δk are defined as the RMS deviations from the average values of L_{wt} and Δk evaluated in terms of the intensities $|\mathbf{E}(x,0)|^2$ and $|\mathbf{A}(\mathbf{k})|^2$ (Jackson, 2001), it follows:

$$L_{wt} \Delta k \geq \frac{1}{2} \tag{7}$$

Since L_{wt} is very small, the spread of wave numbers of monochromatic waves must be large. Hence there is a quite different scenario with respect to the classical context in which the length L_{wt} is large when compared to the wavelength of light.
In order to simplify the notation, one can reason in one dimension without loss of generality. The spatial pulse of light represented in Fig. 2c is defined as follows:

$$A(x) = A_o \, rect(x) \tag{8}$$

where: rect(x)=1 for $|x|<1/2$; rect(x)=1/2 for $|x|=1/2$; rect(x)=0 elsewhere.
The Fourier transform of the light intensity $[A(x)]^2$ is $[sinc(x)]^2$. To the order 0 it is necessary to add the shifted orders ±1. The function $A(x\pm\Delta x)$ can be represented through the convolution relationship:

$$A(x \pm \Delta x) = \int_{-\infty}^{+\infty} A(x') \cdot \delta(x \pm \Delta x) \cdot dx' \tag{9}$$

where $x'=x\pm\Delta x$. The Fourier transform of the function $A(x\pm\Delta x)$ will be:

$$FT\,[A(x \pm \Delta x)] = A(f_x) \cdot e^{i[2\pi f_x(\mp \Delta x/2)]x} \tag{10}$$

where f_x is the spatial frequency. The real part of Eq. (10) is:

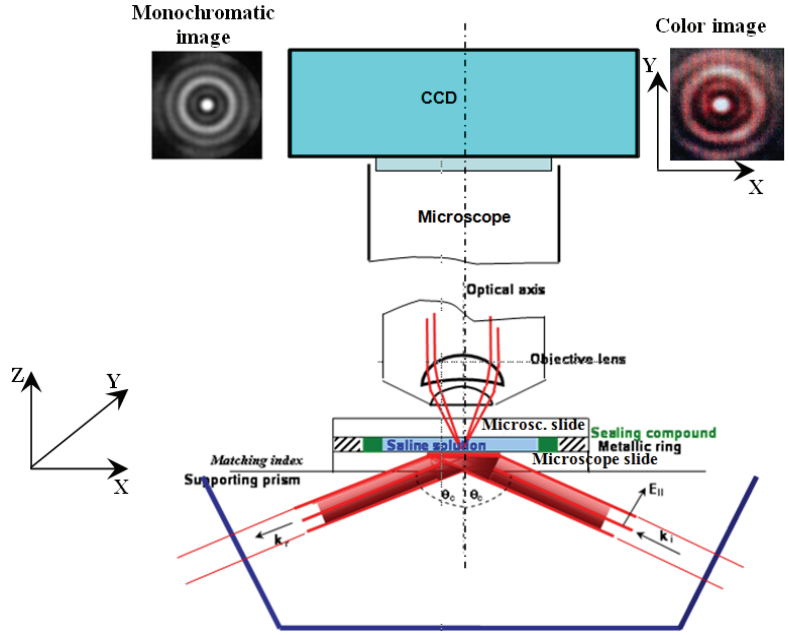
$$\mathrm{Re}\left\{FT[A(x \pm \Delta x]\right\} = A(f_x) \cdot \cos\left\{[2\pi f_x(\mp \Delta x / 2)]x\right\} \tag{11}$$

By taking the Fourier transform of Eq. (11), one can return back to Eq. (9). For the sake of simplicity, the above derivations are in one dimension but can be extended to the 3D case. Brillouin (1930) showed that for a cubic crystal the electromagnetic field can be represented as the summation of plane wave fronts with constant amplitude as assumed in Eq. (5). In such circumstances the above derivation can be extended to 3-D and can be applied to the components of the field in the different coordinates.

3. Experimental setup

Figure 3 shows the schematic representation of the experimental setup. Following the classical arrangement of total internal reflection (TIR), a helium-neon (He-Ne) laser beam with nominal wavelength λ=632.8 nm impinges normally to the face of a prism designed to produce limit angle illumination on the interface between a microscope slide (supported by the prism itself) and a saline solution of sodium-chloride contained in a small cell supported by this slide. Consequently, evanescent light is generated inside the saline solution.
The objects observed with the microscope are supported by the upper face of the microscope slide. Inside the cell filled with the NaCl solution there is a polystyrene microsphere of 6 μm diameter. The microsphere is fixed to the face of the slide through chemical treatment of the contact surface in order to avoid Brownian motions. The polystyrene sphere acts as a relay lens which collects the light wave fronts generated by nano-sized crystals of NaCl resting on the microscope slide. Polystyrene nanospheres also are injected in the solution. More details on the polystyrene sphere and the saline solution are given in the table included in Fig. 3.
The observed image is focused by a microscope with NA=0.95 and registered by a monochromatic CCD attached to the microscope. At a second port, a color camera records color images. The CCD is a square pixel camera with 1600x1152 pixels. The analysis of the

images recorded in the experiment is performed with the Holo Moiré Strain Analyzer software (General Stress Optics, 2008).



Parameter	Value	Note
Polystyrene microsphere diameter D_{sph}	6 ± 0.042 μm	Tolerance specified by manufacturer
Refraction index of polystyrene sphere n_p	1.57 ± 0.01	Value specified by manufacturer
Refraction index of saline solution n_s	1.36	From NaCl concentration at λ=590 nm

Fig. 3. Experimental set up to image nano-size objects using evanescent illumination. Two CCD cameras are attached to a microscope to record images: monochromatic, color.

Three features can be clearly distinguished in the images recorded by the CCD: the diffraction pattern of the microsphere, a system of rectilinear fringes independent of the microsphere, luminous spots well above the speckle average intensity. These features will be analyzed in detail in the chapter explaining how to extract information from the image.

4. Diffraction pattern of the polystyrene microsphere

Figure 4 shows the image of the diffraction pattern of the 6 μm diameter microsphere (Fig. 4a), the FFT of the pattern (Fig. 4b), and an expanded scale of the central part of the FFT pattern (Fig. 4c). Below are the corresponding images (Figs. 4d-f) of the diffraction pattern of a circular aperture of nominal diameter 6 μm (the classical Airy's pattern). Since the two patterns are different, the wave front of light diffracted by the microsphere is not planar. For this reason, the diffraction pattern of the particle has been simulated with an algorithm based on the multi-corona analysis originally developed by Toraldo di Francia (1952 & 1958) and later reprised by Mugnai *et al.* (2001 & 2004).

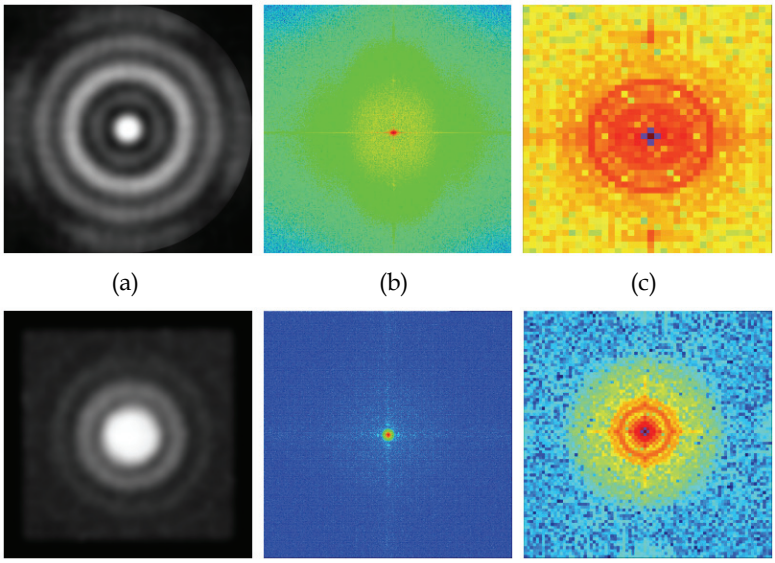


Fig. 4. Diffraction pattern of the 6 μm particle and of the 6 μm pinhole: a) Diffraction pattern (6 μm sphere); b) FFT of the diffraction pattern; c) Enlarged view of Fig. 4b); d) Diffraction pattern (6 μm pinhole); e) FFT of the diffraction pattern; f) Enlarged view of Fig. 4e.

In the multi-corona analysis, the 6 μm diameter particle is modeled as a circular pupil of diameter equal to the particle diameter. Let us assume that the pupil of diameter D is divided in N circular coronae by N+1 concentric circumferences. The i^{th} portion of the pupil is thus a circular corona limited by the diameters D_{i-1} and D_i where the generic diameter D_i is defined as $\alpha_i D$. The α_i parameter ranges between 0 and 1 and increases as the corona goes far from the center of the pupil. If all coronae have the same thickness, parameters α_i are in arithmetic progression: 0, 1/N, 2/N,…, i/N,…, 1 (i=1,…,N).

The complex amplitude – expressed as a function of the observation angle θ limited by the position vector of the observation point and the axis perpendicular to the plane of the corona – diffracted by a circular corona limited by D_{int} and D_{ext} is:

$$A(\theta) = \frac{A}{2\lambda \sin\theta}\left[D_{ext} J_1\left(\pi \frac{D_{ext}}{\lambda} \sin\theta \right) - D_{\text{int}} J_1\left(\pi \frac{D_{\text{int}}}{\lambda} \sin\theta \right) \right] \tag{12}$$

where A is the uniform complex amplitude illuminating the corona, λ is the wavelength of the light, and J_1 is the Bessel function of the first order.

Introducing the dimensionless variable $X=\pi D \sin\theta/\lambda$ and summing over the complex amplitudes diffracted by each corona, the total complex amplitude diffracted by the complete pupil comprised of N circular coronae can be expressed as:

$$A(x) = \sum_{i=0}^{N-1} \frac{k_{i+1}}{x} \left[\alpha_{i+1} J_1(\alpha_{i+1} x) - \alpha_i J_1(\alpha_i x) \right] \tag{13}$$

The k_i coefficients were determined by minimizing the difference between the normalized intensity distribution of the diffraction pattern recorded experimentally and its counterpart predicted by the theoretical model (13). For that purpose, the SQP-MATLAB® optimization routine (The MathWorks, 2006) was utilized. Figure 5 compares the experimental results with the predictions of the multi-corona analysis that was extended up to 7 coronae in order to verify the consistency of the algorithm with the experimental data outside of the central region of the pattern. In the central region of the pattern taking three coronae is enough to get an excellent agreement. By increasing the number of coronae the Bessel function expansion (13) captures very well the positions of all the minima and maxima of the experimental curve but causes secondary maxima to reduce in magnitude.

The present results were obtained in the range of visible optics and agree well with the experimental results found by Mugnai *et al*. (Mugnai *et al*, 2001 & 2004) in the microwave range; both phenomena are described very well by the three-coronae theoretical model. Experimental data are consistent also with the results of a theoretical study recently carried out by Ayyagari and Nair (Ayyagari and Nair, 2009) on the scattering of *p*-polarized evanescent waves by a dielectric spherical particle.

It can be seen from Fig. 5 that the central spot of the diffraction pattern observed experimentally is almost two times more narrow in size than its counterpart in the classical Airy's pattern: the position of the first minimum of intensity in the experimental pattern is in fact X=2.05 vs. the classical value of X=3.83 for the Airy's pattern. Therefore, the diffraction pattern of the polystyrene microsphere is equivalent to that produced by a super-resolving pupil.

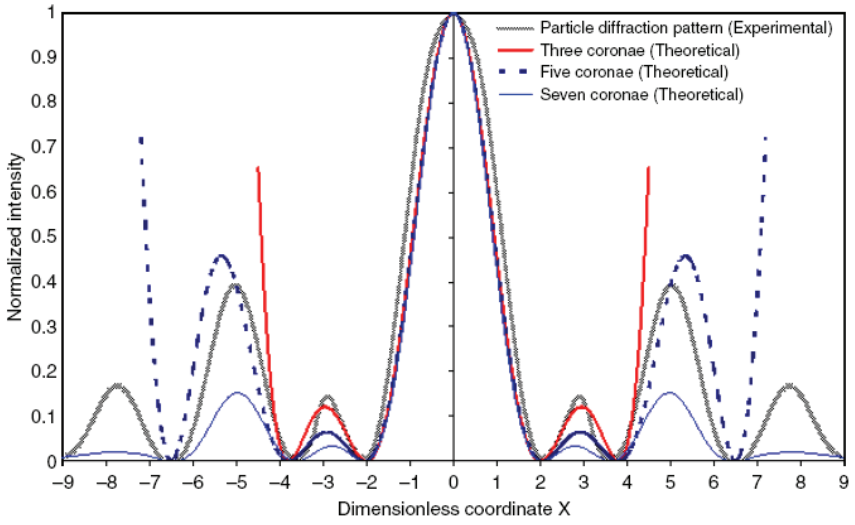


Fig. 5. Comparison between the multi-corona theoretical model and experimental results.

The first step to understand the process of formation of the diffraction pattern observed experimentally is to consider the microsphere as a lens (Fig. 6). The focal distance *f* of this lens can be computed using the equations of geometric optics. The source of the illumination generating the diffraction pattern of the microsphere and the fringes covering the image of

the particle is the resonant electro-magnetic oscillations of interaction of the microsphere with the supporting microscope slide. The whole region of contact between the particle and the microscope slide becomes the equivalent of a micro-laser. The light is generated in the sphere around the first 100 nm in depth and focused in the focal plane of the sphere. The microscope is focused to the plane of best contrast of the diffraction pattern, which must be very close to the microsphere focal plane. The diffraction pattern is not formed by a plane wave front illuminating the microsphere, as is the case with the classical Mie's solution for a diffracting sphere. In the present case, a large number of evanescent wave fronts come in contact with the microsphere and then generate propagating wave fronts that are seen as interference fringes. This mechanism will be clarified in the next section.

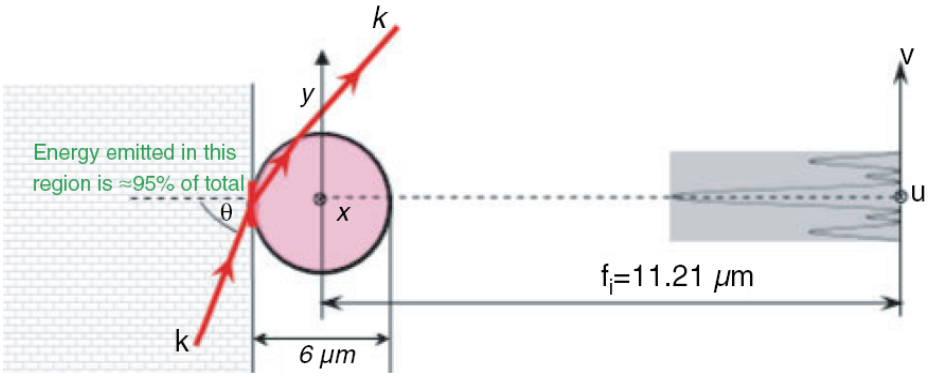


Fig. 6. Polystyrene microsphere modeled as a small spherical lens.

5. System of fringes contained in the recorded image: multi-*k* vector fields

As is mentioned above, besides the diffraction pattern of the polystyrene microsphere, the image recorded by the optical microscope contains also a system of rectilinear fringes. Figure 7 illustrates the process of generation of the evanescent beams that provide the energy required for the formation of the images. The optical setup providing the illumination is similar to the setup originally developed by Toraldo di Francia to prove the existence of evanescent waves (Toraldo di Francia, 1958). A grating is illuminated by a light beam at the limit angle of incidence θ_c. A matching index layer is interposed between the illuminated surface and a prism which is used for observing the propagating beams originated on the prism face by the evanescent waves. In the original experiment conducted by Toraldo di Francia the diffraction orders of the grating produced multiple beams that by interference generated the fringes observed with a telescope. In the present case, the interposed layer corresponds to the microscope slide. Although the microscope slide does not have exactly the same index of refraction of the prism, it is close enough to fulfill its role. The diffraction effect is produced by the residual stresses developed in the outer layers of prism. The multiple illumination beams are the result of residual stresses in the outer layers of the prism. Guillemet (1970) analyzed in detail the process of formation of interference fringes originated by evanescent illumination in presence of residual stresses on glass surfaces. The glass in the neighborhood of surface can be treated as a layered medium and the fringe orders depend on the gradients of index of refraction. A more extensive analysis

of the role played by birefringence in the present example was carried out by Sciammarella and his collaborators (Sciammarella and Lamberti, 2007; Sciammarella *et al.*, 2010).

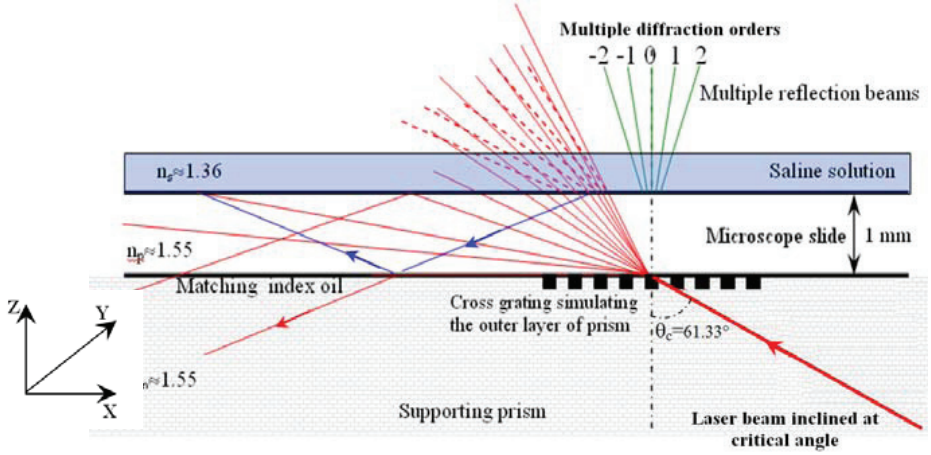


Fig. 7. Model of the interface between the supporting prism and the microscope slide as a diffraction grating causing the impinging laser beam to split into different diffraction orders.

Figure 7 provides a schematic representation of the process of illumination of the observed nano-objects. The laser beam, after entering the prism, impinges on the prism-microscope slide interface, symbolically represented by a grating, where it experiences diffraction. The different diffraction orders enter the microscope slide and continue approximately along the same trajectories determined by the diffraction process. The slight change in trajectory is due to the fact that the indices of refraction of the prism and microscope slide are slightly different. As the different orders reach the interface between the microscope slide and saline solution, total reflection takes place and evanescent wave fronts emerge into the solution in a limited depth that is a function of the wavelength of the light as shown in Eq. (2).

Since the wave fronts are originated by artificial birefringence, for each order of diffraction there are two wave fronts: the *p*-polarized wave front and the *s*-polarized wave front. Upon entering the saline solution these wave fronts originate propagating wave fronts that produce interference fringes. Figure 8 shows the diffraction orders corresponding to one family of fringes extracted from the FT of the image captured by the optical system and contains the diffraction pattern of the microsphere that acts as a relay lens. Since these wave fronts come from the evanescent wave fronts, their sine is a complex number taking values greater than 1. Figure 8 includes also the image extracted from the FT of the central region of the diffraction pattern of the microsphere (the first dark fringe of the microsphere diffraction pattern is shown in the figure) and shows the presence of the families of fringes mentioned above. The fringes form moiré patterns that are modulated in amplitude in correspondence of the loci of the interference fringes of the microsphere. A total of 120 orders can be detected for this particular family (Sciammarella *et al.*, 2010). These wave fronts play a role in the observation of the NaCl nanocrystals contained in the saline solution.

The detail view of Fig. 9 shows the ordinary beam and the extraordinary beam originated by the artificial birefringence of the prism outer layers: birefringence is caused by residual stresses at the boundary between glass slide and prism. The angles of inclination of the

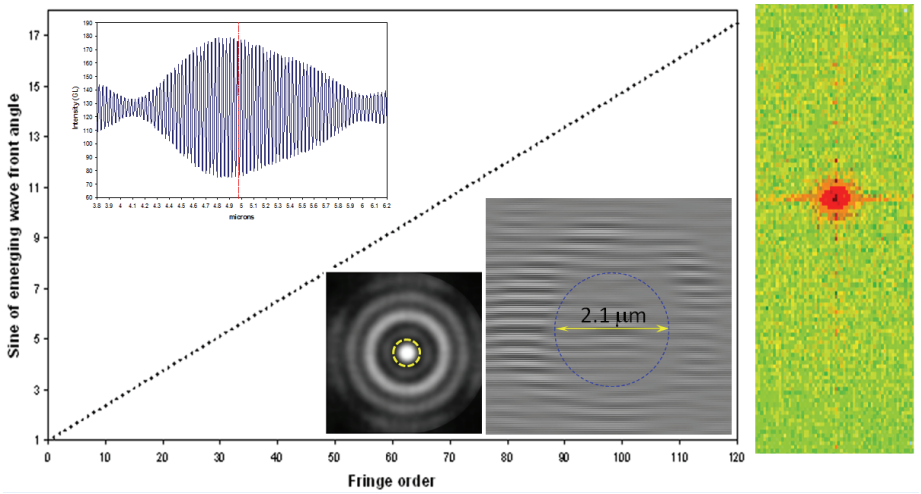


Fig. 8. System of fringes observed in the image of the 6 µm diameter polystyrene sphere. The dotted circle represents the first dark ring in the particle diffraction pattern

ordinary and the extraordinary beams with respect to the normal to the glass slide surface are respectively ϕ_{pp} and ϕ_{sp} where the first subscript "p" and "s" indicates the type of polarization while the second subscript "p" indicates that these wave fronts come from the prism. The *p*-wave front order emerges at the angle ϕ_{pm} in the saline solution while the *s*-wave front emerges at the angle ϕ_{sm} in the saline solution. Those angles can be determined experimentally by utilizing the following relationships:

$$\begin{cases} \sin\phi_{pm}{}^{sper} = \dfrac{\lambda}{p_{pm}} \\ \sin\phi_{sm}{}^{sper} = \dfrac{\lambda}{p_{sm}} \end{cases} \tag{14}$$

where p_{pm} and p_{sm}, respectively, correspond to the spatial frequencies of the two families of fringes formed by the *p* and *s*-waves traveling through the thickness of the glass slide. The pitches p_{pm} and p_{sm} can be measured experimentally by retrieving the sequence of orders produced by the two beams from the FT of the formed images. Since the fundamental pitches of *p* and *s*-wave fringes are, respectively, p_{pm}=3.356 µm and p_{sm}=2.258 µm, the corresponding angles are $\phi_{pm,sper}$=10.860° and $\phi_{sm,sper}$=16.425°.

If the birefringence effect is neglected in the glass slide, the theoretical values of the angles of the *p* and *s*-wave fronts emerging in the saline solution can be expressed as follows (the full derivation is given in Sciammarella *et al*., 2010):

$$\begin{cases} \sin\phi_{pm} = \dfrac{\sqrt{n_{pp}{}^2 - n_g{}^2}}{n_{sol}} \\ \sin\phi_{sm} = \dfrac{\sqrt{n_{sp}{}^2 - n_g{}^2}}{n_{sol}} \end{cases} \tag{15}$$

where n_{pp} and n_{sp} are the indices of refraction of the *p* and *s*-waves traveling in the prism; n_g is the refraction index in the glass slide; n_{sol} is the refraction index of the saline solution.
Theoretical values provided by Eq. (15) can be compared with the corresponding values measured experimentally. However, there are two unknowns in Eq. (15): the indices of refraction n_{pp} and n_{sp}. The index of refraction of the glass slide n_g is provided by the manufacturer: 1.5234. The index of refraction of the saline solution can be determined from the solution concentration: 1.36.
The Maxwell-Neumann equations written for the *p* and *s*-wave fronts traveling through the stressed region of the prism are:

$$n_{pp} - n_o = A\sigma_1 + B\sigma_2 \tag{16}$$

$$n_{sp} - n_o = B\sigma_1 + A\sigma_2 \tag{17}$$

where n_o is the index of refraction of the prism in the unstressed state while n_{pp} and n_{sp} correspond to the indices of refraction of the extraordinary wave fronts in the stressed prism. Equations (16) and (17) contain also the constants A and B and the residual compressive stresses σ_1 and σ_2 generated in the prism by the manufacturing process.
By comparing the expressions of the index of refraction n_o of the ordinary beam it follows:

$$n_o = n_{pp} - \left(A\sigma_1 + B\sigma_2\right) \;\Leftrightarrow\; n_o = n_{sp} - \left(B\sigma_1 + A\sigma_2\right) \tag{18}$$

where the two members of equation (18) must coincide.

$$\begin{cases} Min\left[\Psi(n_{pp}, n_{sp}, A, B, \sigma_1, \sigma_2) = \sqrt{\frac{1}{2}\left[\left(\frac{\sqrt{{n_{pp}}^2 - {n_g}^2}}{n_{sol}} - \sin\phi_{pm}^{\;sper}}{\sin\phi_{pm}^{\;sper}}\right)^2 + \left(\frac{\frac{\sqrt{{n_{sp}}^2 - {n_g}^2}}{n_{sol}} - \sin\phi_{sm}^{\;sper}}{\sin\phi_{sm}^{\;sper}}\right)^2\right]}\right] \\ 0.98 \le \left|\frac{n_{pp} - (A\sigma_1 + B\sigma_2)}{n_{sp} - (B\sigma_1 + A\sigma_2)}\right| \le 1.02 \\ 1.5235 \le n_{pp} \le 1.575 \\ 1.5235 \le n_{sp} \le 1.575 \\ -0.8 \cdot 10^{-11} \le A \le -0.3 \cdot 10^{-11} \\ -3.5 \cdot 10^{-11} \le B \le -2.5 \cdot 10^{-11} \\ -180 \cdot 10^6 \le \sigma_1 \le -120 \cdot 10^6 \\ -180 \cdot 10^6 \le \sigma_2 \le -120 \cdot 10^6 \end{cases} \tag{19}$$

The optimization problem (19) was formulated where the objective function Ψ represents the average difference between the theoretical angles at which the *p* and *s*-wave fronts emerge in the saline solution and their counterpart measured experimentally. The Maxwell-Neumann equations were included as constraints in the optimization process: the equality constraint on the ordinary wave, Eq. (18), must be satisfied within 2% tolerance. There are six optimization variables: the two indices of refraction n_{pp} and n_{sp} in the error functional; the two constants *A* and *B* and the residual stresses σ_1 and σ_2 in the constraint equation. The last unknown of the optimization problem, the refraction index n_o corresponding to the unstressed prism, can be removed from the vector of design variables by combing the Maxwell-Neumann relationships into the constraint equation (18).

Three different optimization runs were carried out starting from: (i) lower bounds of design variables (Run A); (ii) mean values of design variables (Run B); (iii) upper bounds of design variables (Run C). The optimization problem (19) was solved by combining response surface approximation and line search (Vanderplaats, 2001). SQP-MATLAB performed poorly as design variables range over very different scales.

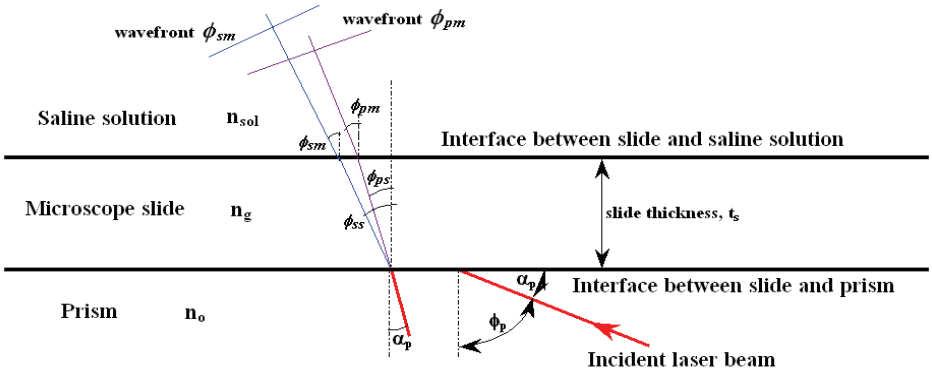


Fig. 9. Schematic of the optical path of *p* and *s*-polarized beams generated by the artificial birefringence of prism.

Parameters	Run A	Run B	Run C
n_{pp}	1.5447	1.5447	1.5468
n_{sp}	1.5716	1.5709	1.5699
n_o	1.5399	1.5399	1.5405
A (m^2/N)	$-0.5167\cdot10^{-11}$	$-0.3017\cdot10^{-11}$	$-0.6994\cdot10^{-11}$
B (m^2/N)	$-2.8096\cdot10^{-11}$	$-2.9645\cdot10^{-11}$	$-3.1390\cdot10^{-11}$
C=A–B (m^2/N)	$2.2929\cdot10^{-11}$	$2.6628\cdot10^{-11}$	$2.4396\cdot10^{-11}$
σ_1 (MPa)	–135.30	–142.50	–161.85
σ_2 (MPa)	–145.01	–145.83	–164.90
τ (MPa)	4.8549	1.6657	1.5279
Residual error Ψ (%)	0.3872	0.2732	0.3394
Error on n_o (%)	1.730	1.673	1.485

Table 1. Results of optimization runs to analyze the effect of artificial birefringence of prism

Optimization results are listed in Table 1. Besides the residual error on the optimized value of Ψ and the constraint margin (18), the table shows also the resulting value of the index of refraction of the ordinary beam n_o, the photoelastic constant $C=A-B$ and the shear stress τ in the prism determined as $|(\sigma_1-\sigma_2)/2|$. Remarkably, the residual error on emerging wave angles is always smaller than 0.39%. The error on refraction index of the ordinary beam is less than 1.75%. The optimization process was completed within 20 iterations.

6. Formation of the holograms at the nano-scale

The wave fronts emerging from the microscope slide play a role in the observation of the prismatic NaCl nanocrystals contained in the saline solution. The observed images are not images in the classical sense but they are lens holograms (theory of Fourier holography is explained in the classical textbook written by Stroke (Stroke, 1969)) generated by the self-luminous nanocrystals. Figure 10 shows the schematic representation of the optical circuit bringing the images to the CCD detector. It can be seen that different diffraction orders emerge from the interfaces of the crystals and the supporting microscope slide. These emerging wave fronts act as multiplexers creating successive shifted images of the object. Let us assume to look at a prism (Fig. 10, Part 1) approximately parallel to the image plane of the CCD. Successive shifted luminous images of the prism are then recorded.
The largest fraction of energy is concentrated in the zero order and the first order (Sciammarella *et al.*, 2009). These orders overlap in an area that depends on the process of formation of the image (see Fig. 10). The order 0 produces an image on the image plane of the optical system, that is centered at a value x of the horizontal coordinate. Let us call S(x) this image. The order +1 will create a shifted image of the particle, S(x-Δx). The shift implies a change of the optical path between corresponding points of the surface. In the present case, the trajectories of the beams inside the prismatic crystals are straight lines and the resulting phase changes are proportional to the observed image shifts. The phase change is:

$$\Delta\phi(x,\Delta x) = K_p\left[S(x) - S(x-\Delta x)\right] \tag{20}$$

where K_p is a coefficient of proportionality. Equation (20) corresponds to a shift of the image of the amount Δx. If the FT of the image is computed numerically, one can apply the shift theorem of the Fourier transform. For a function f(x) shifted by the amount Δx, the Fourier spectrum remains the same but the linear term $\omega_{sp}\Delta x$ is added to the phase: ω_{sp} is the angular frequency of the FT. It is necessary to evaluate this phase change. The shift can be measured on the image by determining the number of pixels representing the displacement between corresponding points of the image (see Fig. 10, Part 4). Through this analysis and using the Fourier Transform it is possible to compute the thickness t of the prism in an alternative way to the procedure that will be described in the following. These developments are a verification of the mechanism of the formation of the images as well as of the methods to determine prism thickness (Sciammarella *et al.*, 2009).
Let us now consider the quasi-monochromatic coherent wave emitted by a nano-sized prismatic crystal. The actual formation of the image is similar to a typical lens hologram of a phase object illuminated by a phase grating (Tanner, 1974). The Fourier Transform of the image of the nanocrystal extended to the complex plane is an analytical function. If the FT is known in a region, then, by analytic continuation, $F(\omega_{sp})$ can be extended to the entire domain. The resolution obtained in this process is determined by the frequency ω_{sp} captured

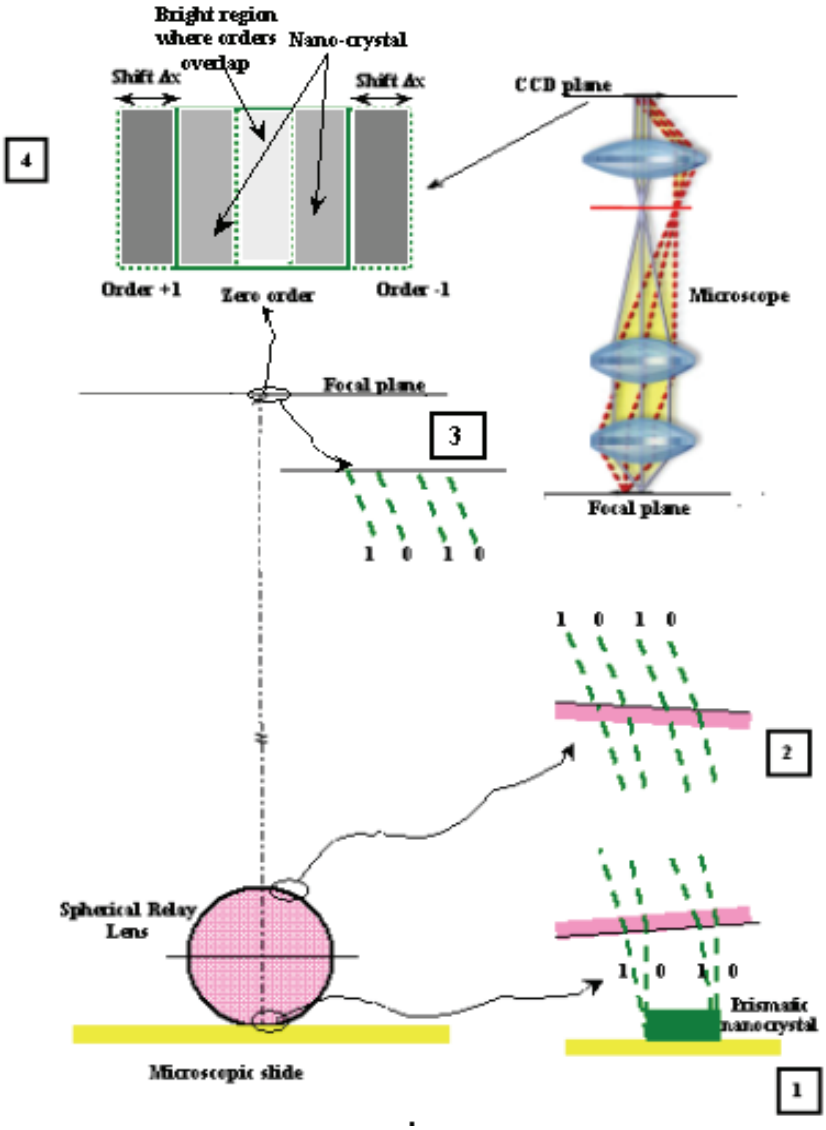


Fig. 10. Schematic representation of the optical system leading to the formation of lens hologram: 1) Prismatic nanocrystal; 2) Wave fronts entering and emerging from the polystyrene micro-sphere acting as a relay lens; 3) Wave fronts arriving at the focal plane of the spherical lens; 4) Wave fronts arriving at the image plane of the CCD. The simulation of the overlapping of orders 0, +1 and -1 in the image plane of the CCD is also shown.

in the image. The image can be reconstructed by a combination of phase retrieval and suitable algorithms. The image can be reconstructed from a $F(\omega_{sp})$ such that $\omega_{sp}<\omega_{sp,max}$, where $\omega_{sp,max}$ is determined by the wave fronts captured by the sensor.

The fringes generated by different diffraction orders experience phase changes that provide depth information. These fringes are carrier fringes that can be utilized to extract optical path changes. This type of setup to observe phase objects was used in phase hologram interferometry as a variant of the original setup proposed by Burch and Gates(Burch et al., 1966; Spencer and Anthony, 1968). When the index of refraction in the medium is constant, the rays going through the object are straight lines. If a prismatic object is illuminated with a beam normal to its surface, the optical path s_{op} through the object is given by the integral:

$$s_{op}(x,y) = \int n_i(x,y,z)dz \tag{21}$$

where the direction of propagation of the illuminating beam is the z-coordinate and the analyzed plane wave front is the plane x-y; $n_i(x,y,z)$ is the index of refraction of the object through which light propagates.
The change experienced by the optical path is:

$$\Delta s_{op}(x,y) = \int_0^t \left[n_i(x,y,z) - n_o \right] dz \tag{22}$$

where t is the thickness of the medium. By assuming that $n_i(x,y,z)=n_c$ where n_c is the index of refraction of the observed nanocrystals, Eq. (22) then becomes:

$$\Delta s_{op}(x,y) = (n_c - n_o)t \tag{23}$$

By transforming Eq. (23) into phase differences and making $n_o=n_s$, where n_s is the index of refraction of the saline solution containing the nanocrystals, one can write:

$$\Delta\phi = \frac{2\pi}{p}(n_c - n_s)t \tag{24}$$

where p is the pitch of the fringes present in the image and modulated by the thickness t of the specimen. In general, the change of optical path is small and no fringes can be observed. In order to solve this problem, carrier fringes can be added. An alternative procedure is the introduction of a grating in the illumination path (Tanner, 1974). In the present case of the nanocrystals, carrier fringes can be obtained from the FT of the lens hologram of analyzed crystals. In the holography of transparent objects, one can start with recording an image without the transparent object of interest. In a second stage, one can add the object and then superimpose both holograms in order to detect the phase changes introduced by the object of interest. In the present experiment, reference fringes can be obtained from the background field away from the observed objects. This procedure presupposes that the systems of carrier fringes are present in the field independently of the self-luminous objects. This assumption is verified in the present case since one can observe fringes that are in the background and enter the nanocrystals experiencing a shift.

7. Observation of prismatic sodium-chloride nanocrystals

The prismatic nanocrystals of sodium-chloride grow on the surface of the microscope slide by precipitation from the saline solution. Different isomers of NaCl nanocrystals are formed and they are the object of dimensional analysis. There is a variety of nano-isomers that can form at room temperature (Hudgins *et al*., 1997). In this study, only three isomers types and four crystals were analyzed out of a much larger population of possible configurations.

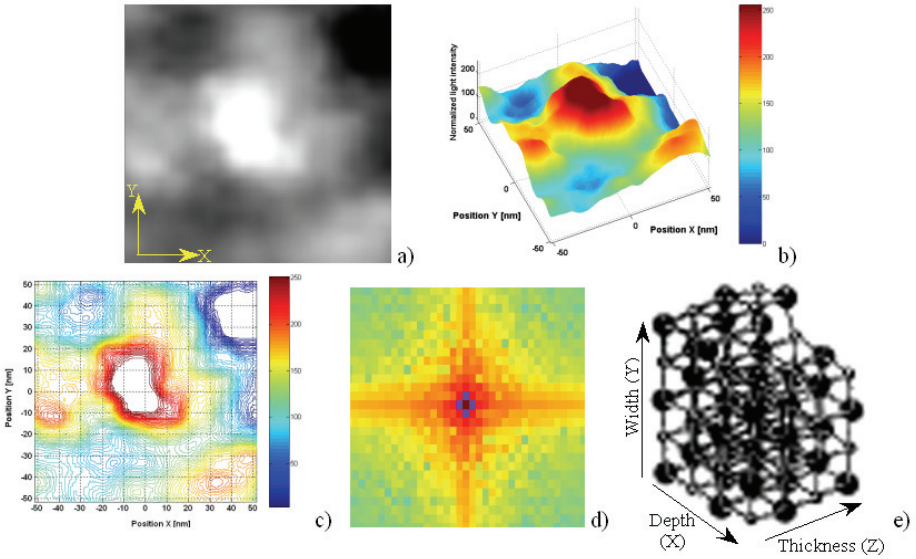


Fig. 11. NaCl nanocrystal of length 54 nm: a) gray-level image (512 x 512 pixels); b) 3-D distribution of light intensity; c) isophote lines; d) FT pattern; e) theoretical crystal structure.

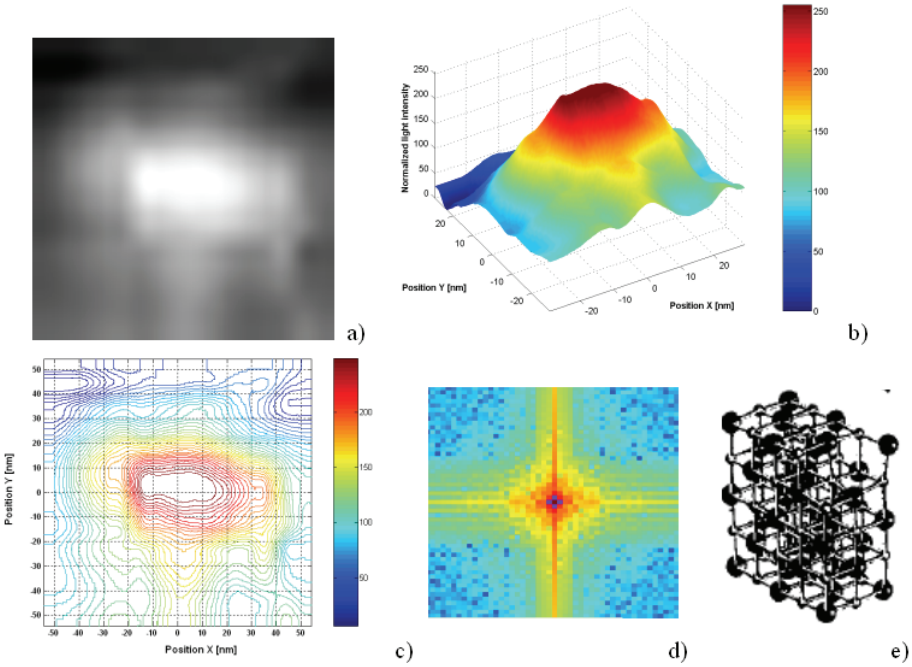


Fig. 12. NaCl nanocrystal of length 55 nm: a) gray level image (512 x 512 pixels); b) 3-D distribution of light intensity; c) isophote lines; d) FT pattern; e) theoretical crystal structure.

The sodium-chloride nanocrystals are located in the recorded images by observing the image of the lens hologram with increasing numerical zooming. A square region around a selected particle is cropped and the image is digitally re-pixelated to either 512×512 pixels or 1024×1024 pixels to increase the numerical accuracy of the FT of the crystals images. Image intensities are also normalized from 0 to 255. Figure 11 shows the nanocrystal referred to as crystal L=54 nm. The isophotes (lines of equal light intensity) of the crystal along with the corresponding isophote level lines and the FT of the image pattern are presented in the figure; the theoretical structure of the nanocrystal is also shown.
A similar representation is given in Fig. 12 for another prismatic crystal with about the same length, that will be referred as crystal L=55 nm.
Figure 13 illustrates the case of a square cross-section crystal that will be called crystal L=86 nm. Figure 13b is an image of the crystal recorded by a color camera; the crystal has a light green tone. The monochromatic image and the color image have different pixel structures, but, by using features of the images clearly identifiable, a correspondence between the images could be established and the image of the nanocrystal located. The picture indicates that the image color is the result of an electromagnetic resonance and not an emission of light at the same wavelength as the impinging light wavelength.

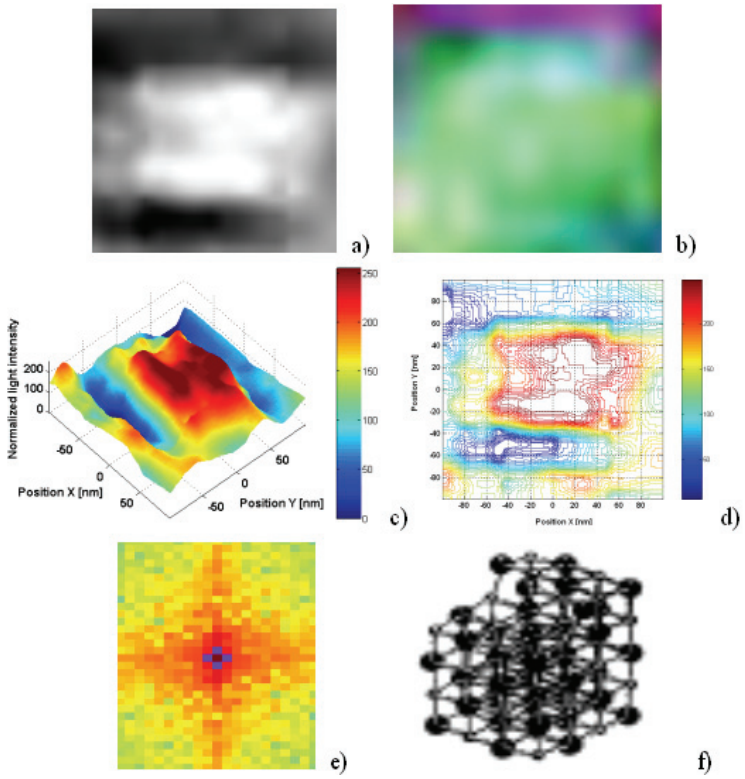


Fig. 13. NaCl nanocrystal of length 86 nm: a) gray-level image (1024 x 1024 pixels); b) image of the crystal captured by a color camera; c) 3-D distribution of light intensity; d) isophote lines; e) FT pattern; f) theoretical crystal structure.

Finally, Fig. 14 is relative to the nanocrystal referred to as crystal L=120 nm.

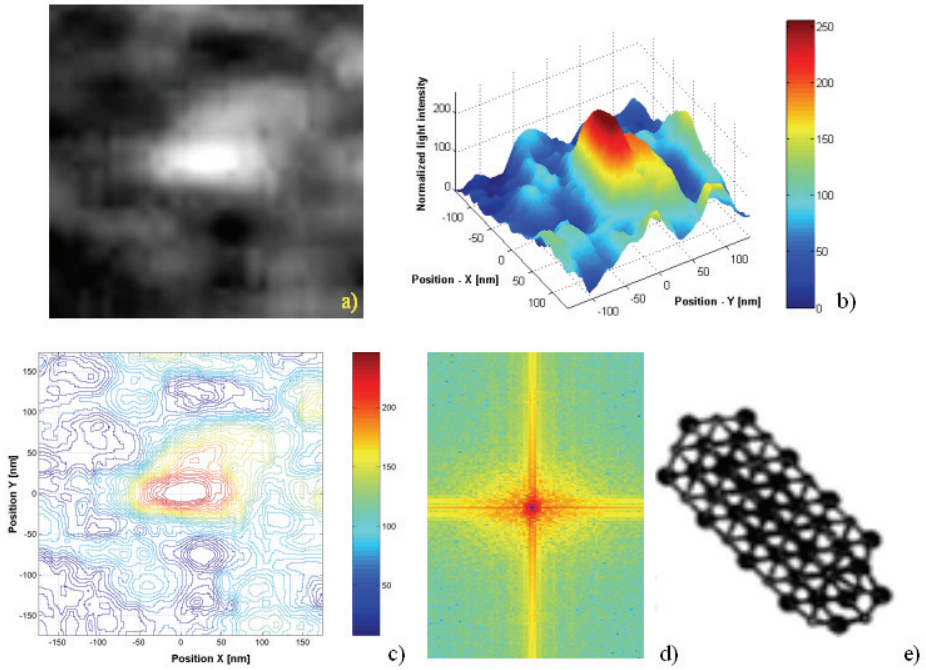


Fig. 14. NaCl nanocrystal of length 120 nm: a) gray-level image (512 x 512 pixels); b) 3-D distribution of light intensity; c) isophote lines; d) FT pattern; e) theoretical crystal structure.

Since the change of phase is very small, fractional orders must be used for measuring the change of optical path as the light travels through nanocrystals. The procedure to obtain depth information from the recorded images is now described. For each analyzed crystal a particular frequency is selected. This frequency must be present in the FT of the image and should be such that the necessary operations for frequency separation are feasible. This means that frequencies that do not depend on the thickness of the nanocrystal must not be near the selected frequency. The frequency of interest is individualized in the background of the observed object. Then, the selected frequency is located in the FT of the object. A proper filter size is then selected to pass a number of harmonics around the chosen frequency. Those additional frequencies carry the information of the change in phase produced by the change of optical path. In the next step, the phase of the modulated carrier and the phase of the unmodulated carrier are computed. The change of phase is introduced in Eq. (24) and the value of t is computed. Since the phase difference is not constant throughout the prismatic crystal face as it is unlikely that the crystal face is parallel to the image plane of camera, an average thickness is computed. The process was repeated for all prismatic crystals analyzed. For the nanocrystal L=86 nm, the face under observation is not a plane but has a step; furthermore, the face is also inclined with respect to the CCD image plane. Hence, depth was computed as an average of the depth coordinates of points of the face.

To completely describe the prisms, one needs to obtain the in-plane dimensions as well. For that purpose, an edge detection technique based on the Sobel filter was utilized in this

study. The edge detection is a difficult procedure since the Goos-Hänchen effect may cause distortions of the wave fronts at the edges. However, a careful procedure allowed the edge detection to be used within a margin of error of a few nanometers. It has to be expected that the prismatic crystals faces are not exactly parallel to the image plane of the camera. Furthermore, due to the integration effect of the pixels on the arriving wave fronts, not all the nanocrystals images can provide meaningful results.

The values of thickness determined for the different nanocrystals were in good agreement with the results provided by the image-shift procedure – Eq. (20) – outlined in Section 6: the standard deviation is within ±5 nm.

Table 2 shows the experimentally measured aspect ratios of nanocrystal dimensions compared with the corresponding theoretical values: the average error on aspect ratios is 4.59% while the corresponding standard deviation is ±6.57%.

Nanocrystal length (nm)	**Experimental dimensions** (nm)	**Experimental aspect ratio**	**Theoretical aspect ratio**
54	72 x 54 x 53	5.43 x 4.08 x 4	5 x 4 x 4
55	55 x 45 x 33.5	4.93 x 4.03 x 3	5 x 5 x 3
86	104 x 86 x 86	4.84 x 4 x 4	5 x 4 x 4
120	120 x 46 x 46	7.83 x 3 x 3	8 x 3 x 3

Table 2. Aspect ratio of the observed nanocrystals: experiments vs. theory

Table 3, on the basis of aspect ratios, shows the theoretical dimensions of the sides of nanocrystals and compares them with the measured values: the average absolute error is 3.06 nm, the mean error is –1.39 nm, the standard deviation of absolute errors is ±3.69 nm. A conservative assumption to estimate the accuracy of measurements is to adopt the smallest dimensions of the crystals as given quantities from which the other dimensions are then estimated. The smallest dimensions are the ones that have the largest absolute errors.

Nanocrystal length (nm)	**Dimensions Measured**	**Dimensions Theoretical**	**Difference** (nm)
	72	66.3	+5.7
54	54	53	+1
	53	53	-----
	55	55.8	-0.8
55	45	55.8	-10.8
	33.5	33.5	-----
	104	107.5	-3.5
86	86	86	0
	86	86	-----
	120	122.7	-2.7
120	46	46	0
	46	46	-----

Table 3. Main dimensions of the observed nanocrystals: experiments vs. theory.

The theoretical structure 5×4×4 corresponding to the crystal L=86 nm has one step in the depth dimension. The numerical reconstruction of this crystal is consistent with the theoretical structure (Fig. 15a); Fig. 15c shows the level lines of the top face; Fig. 15d shows a cross section where each horizontal line corresponds to five elementary cells of NaCl. The crystal shows an inclination with respect to the camera plane that was corrected by means of an infinitesimal rotation. This allowed the actual thickness jump in the upper face of the crystal (see the theoretical structure in Fig. 15b) to be obtained. The jump in thickness is 26 nm out of a side length of 86 nm: this corresponds to a ratio of 0.313 which is very close to theory. In fact, the theoretical structure predicted a vertical jump of one atomic distance vs. three atomic distances in the transverse direction: that is, a ratio of 0.333.

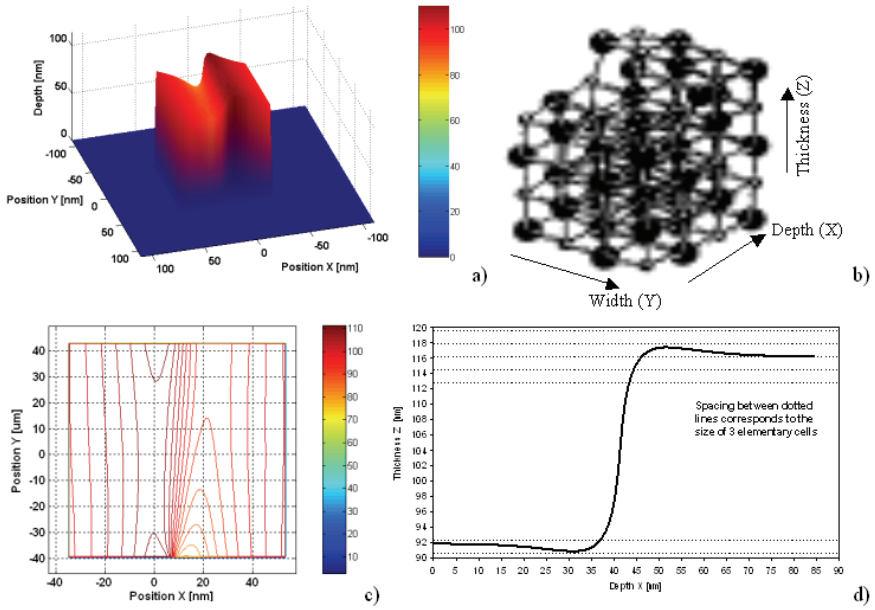


Fig. 15. a) Numerical reconstruction of the NaCl nanocrystal of length 86 nm; b) theoretical structure; c) level lines; d) rotated cross section of the upper face of the nanocrystal: the spacing between dotted lines corresponds to the size of three elementary cells.

8. Observation of the polystyrene nanospheres

Microspheres and nanospheres made of transparent dielectric media are excellent optical resonators. Unlike the NaCl nanocrystals whose resonant modes have not been previously studied in the literature, both theoretical and experimental studies on the resonant modes of microspheres and nanospheres can be found in the literature. Of particular interest are the modes localized at the surface, along a thin equatorial ring. These modes are called *whispering-gallery modes* (WGM)**.** WGM result from light confinement due to total internal reflection inside a high index spherical surface within a lower index medium and from resonance as the light travels a round trip within the cavity with phase matching (Johnson, 1993). The WG modes are within the Mie's family of solutions for resonant modes in light scattering by dielectric spheres. The WG modes can also be derived from Maxwell equations

by imposing adequate boundary conditions (Bohren and Huffman, 1998). They can also be obtained as solutions of the Quantum Mechanics Schrodinger-like equation describing the evolution of a complex angular-momentum of a particle in a potential well.
Figure 16a shows the image of a spherical particle of diameter 150 nm. This image presents the typical WG intensity distribution. Waves are propagating around the diameter in opposite directions thus producing a standing wave with 7 nodes and 6 maxima. The light is trapped inside the particle and there is basically a surface wave that only penetrates a small amount into the radial direction. The signal recorded for this particle is noisier compared with the signal recorded for the prismatic crystals. The noise increase is probably due to the Brownian motion of the spherical particles. While NaCl nanocrystals seem to grow attached to the supporting surface, nanospheres are not in the same condition. Of all resonant geometries a sphere has the capability of storing and confining energy in a small volume.

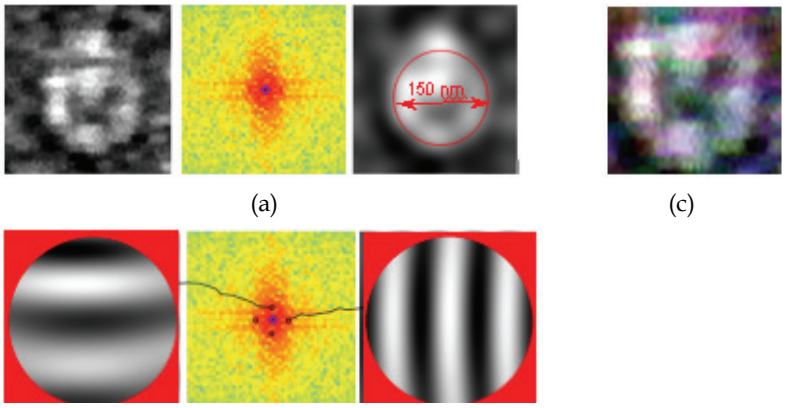


Fig. 16. Spherical nanoparticle of estimated diameter 150 nm: a) FT and zero order filtered pattern; b) systems of fringes modulated by the particle; c) color image of the particle.

The method of depth determination utilized for the nanocrystals can be applied also to the nanospheres. While in prismatic bodies made out of plane surfaces the pattern interpretation is straightforward, in the case of curved surfaces the analysis of the patterns is more complex since light beams experience changes in trajectories determined by the laws of refraction. In the case of a sphere the analysis of the patterns can be performed in a way similar to what is done in the analysis of the Ronchi test for lens aberrations. Figure 16 shows the distortion of a grating of pitch p=83.4 nm as it goes through the nanosphere. The appearance of the observed fringes is similar to that observed in a Ronchi test. The detailed description of this process is not included in this chapter for the sake of brevity.
Figure 17a shows a spherical particle of estimated diameter 187 nm while Fig. 17b shows the average intensity. Figure 17c is taken from Pack (2001) and shows the numerical solution for the WGM of a polystyrene sphere of diameter 1.4 μm while Fig. 17d shows the average intensity. There is good correspondence between experimental results and numerical simulation. The electromagnetic resonance occurs at the wavelength λ=386 nm which corresponds to UV radiation. The color camera is sensitive to this frequency and Fig. 16c shows the color picture of the D=150 nm nanosphere: the observed color corresponds approximately to the above mentioned resonance wavelength.

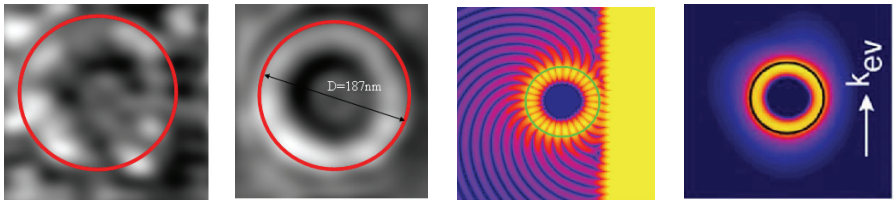


Fig. 17. Spherical nano-particle of estimated diameter 187 nm: a) Original image; b) Average intensity image; c) Numerical simulation of WG modes; d) Average intensity of numerical simulation.

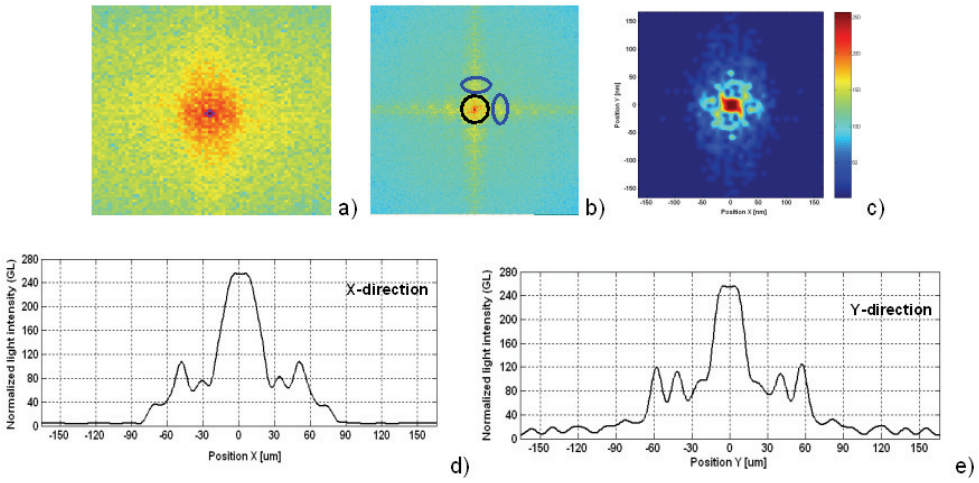
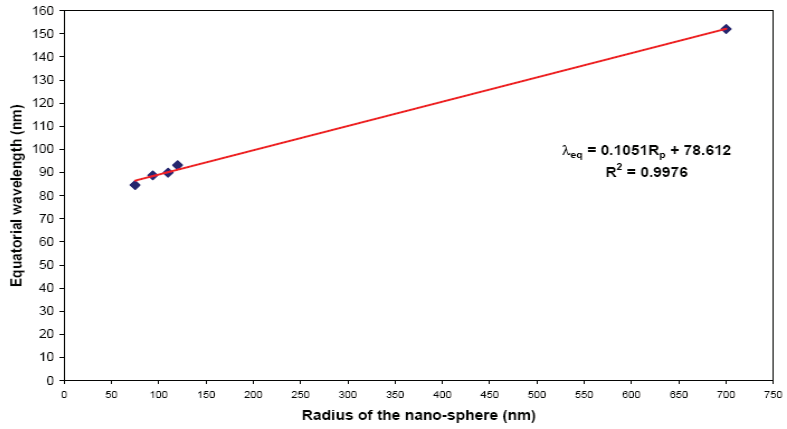


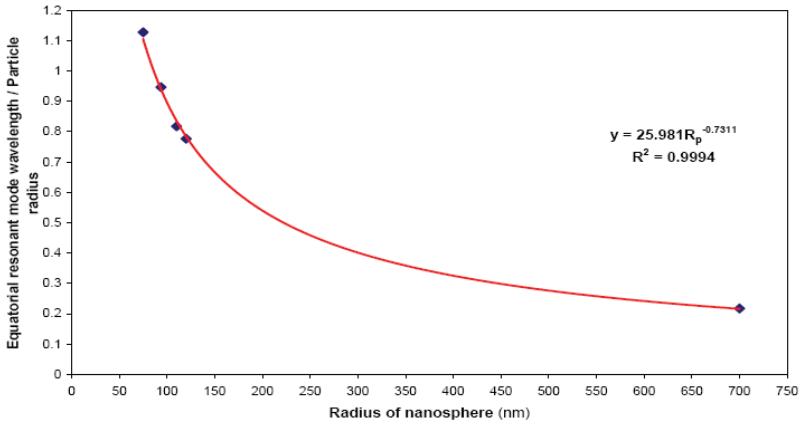
Fig. 18. a) FT pattern of the image of the 150 nm polystyrene nanosphere; b) Fourier transform of the real part of the FT of the nanosphere image; c) 2D view of the zero order extracted from the real part of the FT pattern of the nanosphere image; d) cross-section of the zero order in the X-direction; e) cross-section of the zero order in the Y-direction.

Four particles with radius ranging between 150 and 228 nm were analyzed. For example, Fig. 18 illustrates the different stages involved by the determination of the D=150 nm nanoparticle diameter. The Fourier transform of the image is shown in Fig. 18a while Fig. 18b shows the FT of the real part of the FT of the image. Figure 18c shows the intensity distribution of order 0. Different peaks corresponding to fringe systems present in the image can be observed. Figures 18d-e show the cross sections of the zero order with the corresponding peaks observed in Fig. 18c. These patterns are similar to those observed for the prismatic nanocrystals (Sciammarella *et al.*, 2009). The gray level intensity decays from 255 to 20 within 75 nm: this quantity corresponds to the radius of the nanosphere.

An alternative way to determine diameter from experimental data is based on the WGM properties that relate the diameter or radius of the nanosphere to the standing waves which in turn are characterized by the number of zero nodes or the number of maxima. These numbers depend on the index of refraction and on the radius of the nanosphere. The numerical solution of the WG mode of polystyrene sphere developed by Pack (Pack, 2001)



(a)



(b)

Fig. 19. Relationships between nanosphere radius and a) equatorial wavelength of the WG mode; b) normalized equatorial wavelength of the WG mode.

was utilized in this study as the resonance modes occur approximately at the same wavelength, λ=386 nm. Figure 19a shows the equatorial wave length of the WG plotted vs. the particle radius. The experimentally measured wavelengths are plotted in a graph that includes also the numerically computed value (radius of 700 nm). A very good correlation (R^2=0.9976) was obtained showing that the edge detection gradient utilized yields values that are consistent with the numerically computed WG wavelengths. The graph shown in Fig. 19b – the ratio of equatorial wavelength to radius is plotted vs. the radius – seems to provide a correlation with a better sensitivity. The correlation was again good (R^2=0.9994).

9. Summary and conclusions

This paper illustrated a new approach to investigations at the nano-scale based on the use of evanescent illumination. The proposed methodology allows to measure topography of

nano-sized simple objects such as prismatic sodium-chloride crystals and spherical polystyrene particles. The basic foundation of the study is that the observed images are lens holographic interferograms of phase objects. The process of formation of those interferograms can be explained in view of the self-luminosity of the observed nano-objects caused by electromagnetic resonance. Another aspect to be considered is the possibility for the generated wave fronts to go through the whole process of image formation without the restrictions imposed by diffraction-limited optical instruments.
Remarkably, nanometer resolution was achieved using an experimental setup including just a conventional optical microscope. Besides the nano-size objects, the diffraction pattern of a 6 μm microsphere illuminated by the evanescent field was also recorded by the sensor. Although the sphere was in the range of the objects that can be visualized by classical microscopy, the diffraction pattern was described by an expansion of the electromagnetic field in a series of Bessel functions and coincided with the typical pattern of a super-resolving pupil. This is again a consequence of electromagnetic resonance phenomena taking place in regions with dimensions smaller than the wavelength of the light.
A great deal of theoretical developments is required to substantiate the experimentally observed properties of the illumination resulting from evanescent waves. However, it can be said that, from the point of view of direct application of these properties, it is possible to access the nano-range utilizing far-field observations. The results presented in the paper indicate that evanescent illumination is the key to making Experimental Mechanics methodologies well suitable for nano-engineering applications.

10. References

Ayyagari, R.S. and Nair, S. (2009). Scattering of P-polarized evanescent waves by a spherical dielectric particle. *Journal of the Optical Society of America, Part B: Optical Physics*, Vol. 26, pp. 2054-2058.

Bohren, C.F. and Huffman, D.R. (1998). *Absorption and Scattering of Light by Small Particles*. Wiley, New York (USA), ISBN: 978-0471293408.

Born, M. and Wolf, E. (1999). *Principles of Optics*. VII Edition. Cambridge University Press, Cambridge (UK), ISBN: 978-0521642224.

Bouchal, Z. (2003). Non diffracting optical beams: physical properties, experiments, and applications. *Czechoslovak Journal of Physics*, Vol. 53, pp. 537-578.

Brillouin, L. (1930). Les électrons dans les métaux et le classement des ondes de de Broglie correspondantes. *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences,* Vol. 191, pp. 292-294.

Burch, J.W., Gates, C., Hall, R.G.N. & Tanner, L.H. (1966). Holography with a scatter-plate as a beam splitter and a pulsed ruby laser as light source. *Nature*, Vol. 212, pp. 1347-1348, 1966.

Durnin, J., Miceley, J.J. & Eberli, J.H. (1987). Diffraction free beams. *Physical Review Letters*, Vol. 58, pp. 1499-1501.

General Stress Optics Inc. (2008) *Holo-Moiré Strain Analyzer Version 2.0*, *Chicago (IL), USA*. http://www.stressoptics.com

Guillemet, C. (1970) *L'interférométrie à ondes multiples appliquée à détermination de la répartition de l'indice de réfraction dans un milieu stratifié*. Ph.D. Dissertation, University of Paris, Paris (France).

Gutiérrez-Vega, J.C., Iturbe-Castillo, M.D., Ramirez, G.A., Tepichin, E., Rodriguez-Dagnino, R.M., Chávez-Cerda, S. & New, G.H.C. (2001). Experimental demonstration of optical Mathieu beams. *Optics Communications*, Vol. 195, pp. 35-40.

Hernandez-Aranda, R.I., Guizar-Sicairos, M. & Bandres, M.A. (2006). Propagation of generalized vector Helmholtz–Gauss beams through paraxial optical systems. *Optics Express*, Vol. 14, pp. 8974–8988.

Hudgins, R.R., Dugourd, P., Tenenbaum, J.N. & Jarrold, M.F. (1997). Structural transitions of sodium nanocrystals. *Physical Review Letters*, Vol. 78, pp. 4213-4216.

Jackson, J.D. (2001). *Classical Electrodynamics*, *Third Edition*. John Wiley & Sons, New York (USA), ISBN: 978-0471309321.

Johnson, B.R. (1993). Theory of morphology-dependent resonances - shape resonances and width formulas. *Journal of the Optical Society of America, Part A: Optics Image Science and Vision*, Vol. 10, pp. 343-352.

Mugnai, D., Ranfagni, A. & Ruggeri, R. (2001). Pupils with super-resolution. *Physics Letters*, Vol. A 311, pp. 77-81.

Mugnai, D., Ranfagni, A. & Ruggeri, R. (2004). Beyond the diffraction limit: Super-resolving pupils. *Journal of Applied Physics*, Vol. 95, pp. 2217-2222.

Pack, A. (2001). *Current Topics in Nano-Optics*. PhD Dissertation. Chemnitz Technical University, Chemnitz (Germany).

Sciammarella, C.A. (2008). Experimental mechanics at the nanometric level. *Strain*, Vol. 44, pp. 3-19.

Sciammarella, C.A. and Lamberti, L. (2007). Observation of fundamental variables of optical techniques in nanometric range. In: *Experimental Analysis of Nano and Engineering Materials and Structures* (E.E. Gdoutos Ed.). Springer, Dordrecht (The Netherlands), ISBN: 978-1402062384.

Sciammarella, C.A., Lamberti, L. & Sciammarella, F.M (2009). The equivalent of Fourier holography at the nanoscale. *Experimental Mechanics* , Vol. 49, pp. 747-773.

Sciammarella, C.A., Lamberti, L. & Sciammarella, F.M. (2010). Light generation at the nano scale, key to interferometry at the nano scale. *Proceedings of 2010 SEM Annual Conference & Exposition on Experimental and Applied Mechanics*. Indianapolis (USA), June 2010.

Spencer, R.C. and Anthony, S.A. (1968). Real time holographic moiré patterns for flow visualization. *Applied Optics*, Vol. 7, p. 561.

Stroke, G.W. (1969). *An Introduction to Coherent Optics and Holography*, *2nd Edition*. Academic Press, New York (USA), ISBN: 978-0126739565.

Tanner, L.H. (1974). The scope and limitations of three-dimensional holography of phase objects. *Journal of Scientific Instruments*, Vol. 7, pp. 774-776.

The MathWorks Inc. *MATLAB® Version 7.0*, Austin (TX), USA, 2006. http://www.mathworks.com

Toraldo di Francia, G. (1952). Super-gain antennas and optical resolving power. *Nuovo Cimento*, Vol. S9, pp. 426-435.

Toraldo di Francia, G. (1958). *La Diffrazione della Luce*. Edizioni Scientifiche Einaudi, Torino (Italy).

Vanderplaats, G.N. (2001). *Numerical Optimization Techniques for Engineering Design*, 3rd Edn. VR&D Inc., Colorado Springs (USA), 2001, ISBN: 978-0944956014.

Vigoureux, J.M. (2003). De l'onde évanescente de Fresnel au champ proche optique. *Annales de la Fondation Luis de Broglie*, Vol. 28, pp. 525-547.

Yu, F.T.S. (2000). *Entropy and Information Optics*. Marcel Dekker, New York (USA), ISBN: 978-0824703639.

Part 4

Holographic Applications

10

Quantitative Analysis of Biological Cells Using Digital Holographic Microscopy

Natan T. Shaked, Lisa L. Satterwhite, Matthew T. Rinehart and Adam Wax
Department of Biomedical Engineering, Fitzpatrick Institute for Photonics, Duke University, Durham, North Carolina 27708, USA

1. Introduction

Biological cells are microscopic dynamic objects, continuously adjusting their three-dimensional sizes, shapes and other biophysical features. Wide-field microscopy of cell dynamics can provide a powerful research tool for cell biology studies, as well as a potential means for medical diagnosis and monitoring of diseases. Biological cells, however, are mostly-transparent objects, and thus imaging them with conventional intensity-based light microscopy fails to provide adequate optical contrast between the cell and its environment. Although exogenous contrast agents such as fluorescent dyes can be used to solve this problem, they might be cytotoxic in the long run and there is a possibility they will influence the cellular behavior. Additionally, fluorescent dyes tend to photobleach, potentially limiting the imaging time.

The contrast problem when imaging biological cells can also be solved by using phase microscopy, which records the optical path delays of light passing through the cells and subsequently obtains information on the cellular structure and dynamics without using any exogenous labelling. Since detectors are sensitive to intensity only, the phase of the light that has interacted with the cells must first be converted to intensity variations for detection. Widely used methods to achieve this include phase contrast microscopy and differential interference contrast (DIC) microscopy. However, these techniques are not inherently quantitative and present distinct imaging artifacts that typically prevent straightforward extraction of the entire optical path delay profile of the cell.

Wide-field digital interferometry (WFDI) is a label-free holographic technique that is able to record the entire complex wavefront (amplitude and phase) of the light which has interacted with the sample (Cuche et al., 1999). From the recorded complex field, one can obtain full quantitative phase profiles of cells as well as correct for out-of-focus image features by post-processing. WFDI microscopy (also called digital holographic microscopy) has been applied to various types of biological cell systems and has recorded a diverse range of cellular phenomena (Marquet et al., 2005; Ikeda et al., 2005; Shaked et al., 2009 b; Shaked et al., 2010 b-d). Section 2 reviews the principle of WFDI for obtaining phase profiles of cells, starting from the experimental setup, and ending in digital processing for obtaining the final quantitative phase profile of the cell.

Although WFDI is a quantitative recording technique, simple quasi-three-dimensional holographic visualization of the cell phase profile need not be the end of the process.

Quantitative analysis should permit extraction of numerical parameters which are useful for cytology or medical diagnosis. Using a transmission-mode interferometric setup, the resulting phase profile represents the multiplication between the refractive index differences and the thickness of the sample. These coupled parameters, the refractive index and the thickness, are not distinct when acquiring the phase profile of a dynamic cell. To allow quantitative cell analysis by WFDI, this fact must be considered during the system development and the following quantitative data analysis. Section 3 first deals with the interpretation of the resulting phase profile of the cell. As also reviewed in this section, many morphological parameters that are useful for cell biologists (such as cell volume, cell force distribution, etc.) are based on the geometric thickness profile of the cell rather than on its phase profile. Therefore, we review methods to decouple the cell thickness from refractive index using the cell phase profile obtained by WFDI. As will be shown, for certain cells, in which a constant refractive index can be assumed for the entire cell contents, such as red blood cells, the thickness profile can be directly obtained from the phase profile. In contrast, for other types of cells containing inner organelles with different refractive indices (e.g. nuclei, mitochondria, etc.), certain parameters such as area, dry-mass, and relative volume can still be calculated directly from the phase profile. Measurements of these parameters are presented for experiments with articular cartilage chondrocytes. If, however, a complete thickness profile is required, more complex experimental measurements are typically employed. Decoupling cell refractive index and thickness can be accomplished, for example, by measuring the phase profiles of the same cell immersed in two different growth media with distinct refractive indices. Alternatively, when the thickness profile is measured by another method (such as confocal microscopy), it is possible to use WFDI to calculate the refractive index of the cell inner organelles.
Finally, we show in Section 4 that the phase profile is still useful for quantitative analysis of cells even in cases where decoupling of thickness and refractive index is not possible or desired. This operation is carried out by defining new numerical phase-profile-based parameters, which can characterize certain cell processes of value to cell biologists. Experimental demonstrations of this approach will be presented using cardiomyocytes (heart muscle cells) undergoing temperature changes. These cells contain a significant number of subcellular organelles with varying refractive indices. In addition, the cell dynamics are characterized by a rapid contraction of the cell followed by restoration to the resting point. Capturing the dynamics of these cells during tremperature change by measuring the phase profiles with WFDI is shown to be a suitable method for obtaining quantitative parameters for biological studies, even without the need for decoupling cell thickness from refractive index.

2. Acquiring the phase profile of biological cells by WFDI

Figure 1(a) presents a possible scheme of a single-exposure WFDI setup that is based on the Mach-Zehnder interferometer and an off-axis holographic geometry. In this optical setup, light from a coherent source (HeNe laser, for example) is first spatially filtered using a pair of spherical lenses and a confocally-positioned pinhole, and then split into reference and object beams by beam splitter BS_1. The object beam is transmitted through the sample and magnified by a microscope objective. The reference beam is transmitted through a compensating microscope objective (typically similar to the object-beam objective) and then

combined with the object beam at an angle. The combined beams are projected onto a digital camera by lens L_2, which is postioned in a 4f configuration with each of the microscope objectives, meaning that the distance between each of the microscope objectives and lens L_2 is equal to the summation of their focal lengths. This configuration allows projection of the amplitude and phase distribution of the sample onto the camera. The combination of the sample and reference beams creates a high spatial frequency off-axis hologram of the sample on the digital camera. The digital off-axis hologram acquired by the camera is the intensity of the summation of the object and reference waves, and can be mathematically expressed as follows:

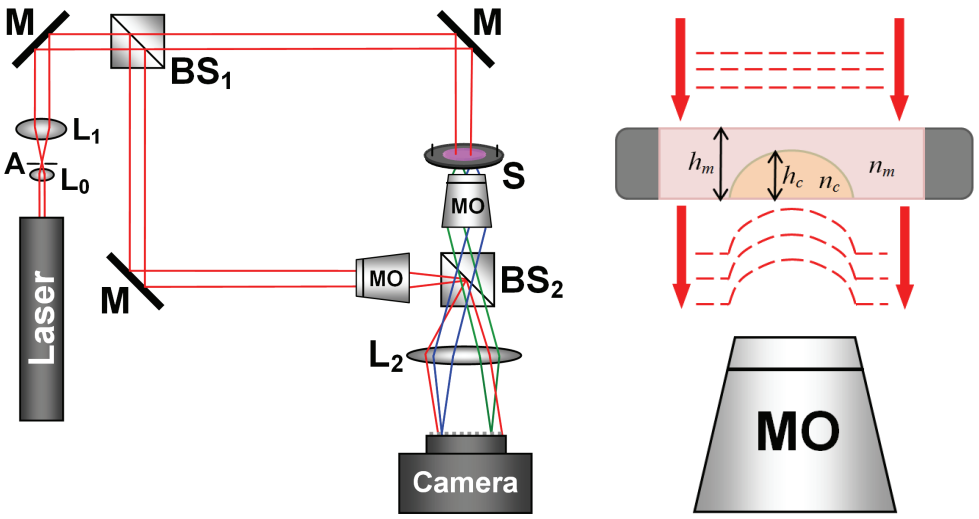
$$H(x,y) = \left|E_s + E_r\right|^2 = \left|E_s\right|^2 + \left|E_r\right|^2 + \left|E_s\right|\left|E_r^*\right|\exp[j(\phi(x,y) + qx)], \tag{1}$$

where E_s and E_r are respectively the sample and reference field distributions, $\phi(x,y)$ is the spatially-varying phase associated with the sample, q is the fringe frequency due to the angular shift between the sample and reference fields, x is the direction of the angular shift (assuming linear horizontal fringes in the off-axis hologram).

The common digital processing method applied to the off-axis hologram starts with a digital two-dimensional Fourier transform. The resulting spatial-frequency contents include reference-field and sample-field autocorrelations (as a result of transforming the first two elements of Eq. (1)) that are located around the origin of the spatial spectrum, and two mathematically conjugated cross-correlation terms (as a result of transforming the exponential term in Eq. (1)), each located at a different side of the spatial spectrum. The exact locations of the cross-correlation terms are dependent on the angle between the object and reference beams. Looking at the spectrum profile, it is easy to isolate only one of the cross-correlation terms, center it, and perform a digital two-dimensional inverse Fourier transform on the result, yielding $\left|E_s\right|\left|E_r^*\right|\exp[j(\phi(x,y)]$. Assuming a weak amplitude modulation due to the transparancy of biological cells in culture, the phase argument of the result $\phi(x,y)$ is the phase profile of the sample.

An alternative method for isolating the phase $\phi(x,y)$ is to perform digital spatial filtering in the image domain rather than in the spectral domain; this is easy to implement automatically, even for dynamic samples (Shaked et al., 2009 a; Shaked et al., 2010 d). Assuming a linear horizontal fringe pattern, the background fringe frequency q in Eq. (1) can be calculated by summing the fringe pattern columns of a small part of the background and fitting the resulting vector to a sine wave. Then, Eq. (1) should be digitally multiplied by $\exp(-jqx)$, and $(\left|E_r\right|^2 + \left|E_s\right|^2)\exp(-jqx)$ can be removed by measuring $\left|E_r\right|^2$ and $\left|E_s\right|^2$, or, if the sample is dynamic, it is possible to simply remove only high spatial frequencies until the optimal result for the phase profile is obtained.

To understand the meaning of the measured phase profile, let us look at Figure 1(b), which presents the sample chamber in detail. As can be seen in this image, the *in vitro* cell is typically adhered to the bottom coverslip and is immersed in cell growth medium. The top coverslip of the chamber ensures a constant cell medium height accross the chamber and thus a constant physical chamber thickness. Based on this chamber, the spatially-varying phase measured by WFDI is proportional to the optical path delay (OPD) profile of the sample and defined as follows:



(a) (b)

Fig. 1. (a) Off-axis WFDI phase-microscopy system. A = Pinhole; L_0, L_1, L_2 = Lenses; BS_1, BS_2 = Beam splitters; M = Mirror; S = Sample; MO = Microscope objective; (b) Detailed scheme of the sample chamber.

$$\begin{aligned}\phi(x,y) &= \frac{2\pi}{\lambda}\left[\bar{n}_c(x,y)h_c(x,y) + n_m\left(h_m - h_c(x,y)\right)\right] \\ &= \frac{2\pi}{\lambda}\left[\left(\bar{n}_c(x,y) - n_m\right)h_c(x,y) + n_m h_m\right] \\ &= \frac{2\pi}{\lambda}\left[OPD_c(x,y) + OPD_m\right],\end{aligned} \tag{2}$$

where λ is the illumination wavelength, $\bar{n}_c(x,y)$ is the spatially varying integral refrative index, n_m is the medium refractive index, $h_c(x,y)$ is the spatially varying thickness profile of the cell, and h_m is the height of the cell medium. For each spatial point (x,y), the integral refrative index $\bar{n}_c$ is defined as follows (Rappaz et al., 2005):

$$\bar{n}_c = \frac{1}{h_c}\int_0^{h_c} n_c(z)dz, \tag{3}$$

where $n_c(z)$ is a function representing the intracellular refractive index along the cell thickness. The value of $OPD_m = n_m h_m$ can be measured in advance in places where there are no cells located, and then subtracted from the total OPD measurement. However, $OPD_c = \left(\bar{n}_c(x,y) - n_m\right)h_c(x,y)$ contains two coupled parameters: the integral refractive index profile of the cell and the cell thickness profile (under the assumption that n_m is known). These parameters might not be distinct when acquiring the phase profile of a dynamic cell, and this fact must be considered during development of the WFDI optical system capturing the cell phase profile and in the quantitative data analysis that follows. Local changes in the cell refractive index along the cell thickness may occur during various dynamic processes,

such as action potential propagation, or by transverse movement of the inner organelles of the cell. Independently or not, thickness changes can occur due to any morphological change of the cell such as membrane fluctuations and cell swelling.

3. Thickness – refractive index conjugation in the phase profile

Various morphological parameters that are useful for cell biologists, including cell volume and cell force distribution, are based on the thickness profile of the cell rather than on the phase profile. Many methods have been developed to decouple thickness from refractive index difference using the cell phase profile. Popescu et al. (2005, 2008 a) and Rappaz et al. (2008 a) have shown that for certain cells, such as mature red blood cells, in which a constant refractive index can be assumed for the entire cell contents (e.g. $\overline{n}_c \cong 1.395$), the thickness profile can be directly obtained from the phase profile. Since WFDI is able to record the quantitative thickness profile of the red blood cell, it is possible to measure rapid height changes, such as membrane fluctuations, that can indicate various medical conditions and blood diseases (Park et al., 2008; Park et al., 2010). Figure 2 shows the phase profiles of rat red blood cells obtained by WFDI in our laboratory, and the associated thickness profile for an arbirary cell in the field of view (FOV). Note, however, that this method of decoupling cell thickness from refractive index in WFDI-base phase profiles is limited to homogeneous cell types that do not contain nuclei or other organelles with varying refractive indices.
Other studies (Barer, 1952; Popescu et al., 2008 b; Rappaz et al., 2009) have shown that for heterogeneous cells that contain organelles with different refractive indices, certain parameters such as cell area and dry mass can be obtained directly from the phase profile. Cell area S_c is simply defined as the number of pixels, for which the OPD is above the background OPD, multiplied by the pixel area. After S_c is known, cell dry mass can be calculated by the following formula:

$$M = \frac{1}{\alpha}\int_{S_c} OPD_c(x,y)ds = \frac{S_c}{\alpha}\langle OPD_c \rangle, \tag{4}$$

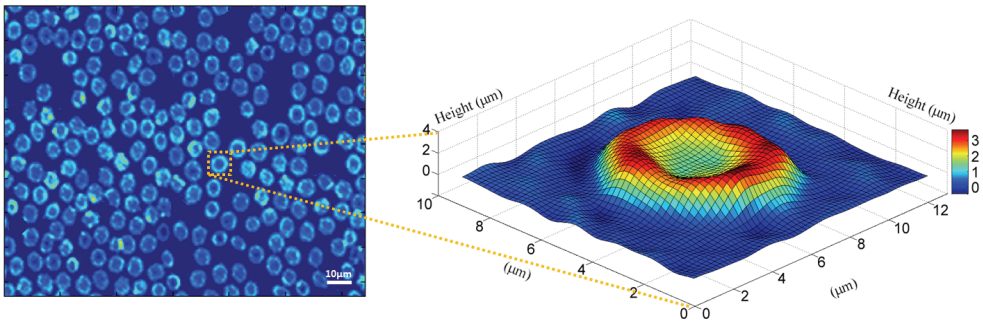


Fig. 2. WFDI-based phase profile of rat whole blood under 40x magnification, demonstrating the valuable quantitative morphological data that can be obtained by WFDI in a single exposure, and without any type of sample preparation or labeling. Cell height profile is shown on the right. This quantitative profile can be derived for each of the cells in the FOV.

where α is the refractive increment constant and can be approximated as 0.18-0.21 ml/g (Barer, 1952), and where $\langle OPD_c \rangle$ is the average OPD over the entire cell area. In a similar way, dry mass surface density can be calculated as follows:

$$\sigma_M(x,y) = \frac{1}{\alpha} OPD_c(x,y) . \quad (5)$$

In addition, if the cell volume transiently increases in an isotropic way (for example, due to swelling), relative volume can still be calculated in a good approximation. For example, we have shown that cell swelling in articular chondrocytes can be analyzed quantitatively without the need to decouple the thickness from refractive index in the WFDI-based phase measurement (Shaked et al., 2010 d). Articular chondrocytes are the cells that compose the cartilage, the connective tissue that distributes mechanical loads between bones and provides almost frictionless surfaces in the joints. The phenotypic expression and metabolic activity of these cells are strongly influenced by shape and volume changes occurring due to mechanical and osmotic stresses. Chondrocytes exhibit rapid swelling or shrinking followed by an active volume recovery in response to osmotic stress. Thus instantaneous evaluation of the chondrocyte volumetric adaptation to such stresses can provide important information on the structure–function relationships in these cells. We induced hypoosmotic stress on *in vitro* chondrocytes by changing the cell medium (Shaked et al., 2010 d). Due to the stress, the cells started swelling and ultimately burst. We recorded the dyanmic phase profiles of the chondrocytes during this phenomenon by WFDI. Figure 3(a) shows the phase profile of one chondrocyte in the monolayer at three different time points. During cell swelling, the phase profile looks wider and lower. Figure 3(b) shows a two-dimensional view of the phase profiles of several cells in the monolayer, whereas Fig. 3(c) shows a DIC microscopy image of the sample. This demonstrates that the contrast mechanism in DIC microscopy does not yield quantitative information while the contrast in WFDI allows direct quantification of the OPD and various numerical parameters at each spatial point on the cell. In addition, as we have shown (Shaked et al., 2010 d), since WFDI captures the entire wavefront, it is possible to correct for out of focus effects in the sample using only digital Fresnel propagation in post-processing and thus avoiding mechanical sample adjustment. This cannot be accomplished using a non-quantitative technqiue such as DIC microscopy.

Based on these dynamic quantitative WFDI-based phase profiles, we calculated relative volume (according to methods described by Popescu et al. (2008 b) and based on the assumption of isotropic volume change); relative dry mass according to Eq. (4); relative area; and relative average phase. All parameters were calculated as the fractional change from the initial value. Figure 3(d) presents the temporal changes of these parameters during the single-cell hypoosmotic swelling (for the cell illustrated in Fig. 3(a)). As can be seen from these graphs, the chondrocyte volume and area increased by 46% and 52%, respectively, during swelling and maintained an approximately constant dry mass. Figure 3(e) shows the parameter graphs for the hypoosmotic swelling of another single chondrocyte that gained in volume and area until bursting, at which point its dry mass decreased. This observation provides experimental support of the dry mass calculation (Eq. (4)) that is based on the chondrocyte phase profile. The small jumps that can be seen on the graphs before the chondrocyte bursts are further validation of Eq. (4). These jumps correspond exactly to time points at which intracellular debris from other previously burst chondrocytes enter the field of view (FOV). Based on the high temporal resolution of our measurements, we have calculated the chondrocyte volume just prior to bursting as V_L=1.28 times the initial cell

volume V_0. Figure 3(f) shows the time dependence of the relative area, dry mass, and average phase of the cell monolayer visualized in Fig. 3(b). The graphs illustrate the trends in these parameters that occur during the dynamic response of the monolayer. Different chondrocytes start swelling at different time points, swell to various extents, and burst at different time points. Individual cell swelling and bursting results in a decrease in the average phase value. The rupture of an individual cell is characterized by a loss of dry mass and an increase of viewable area until the chondrocyte intracellular debris leaves the FOV. New chondrocytes and intracellular debris entering the FOV result in an increase in dry mass and area. It was demonstrated that the values of all three parameters decrease over time due to the rupture of most chondrocytes in the monolayer; this results in an approximately uniform distribution of intracellular debris in the chamber (Shaked et al., 2010 d).

Note that in the chondrocyte experiment, we did not decouple thickness from refractive index since the calculated parameters did not require this operation. If, however, a complete thickness profile is required, more involved experimental measurements are typically employed. Rappaz et al. (2008 a, 2009) used two types of cell media with distinct refractive indices and measured two phase profiles of the same cell. The cell is first measured in the presence of a cell medium with refractive index n_m , yielding a measured cellular OPD of:

$$OPD_{c,1}(x,y) = \left(\bar{n}_c(x,y) - n_m\right) \cdot h_c(x,y). \tag{6}$$

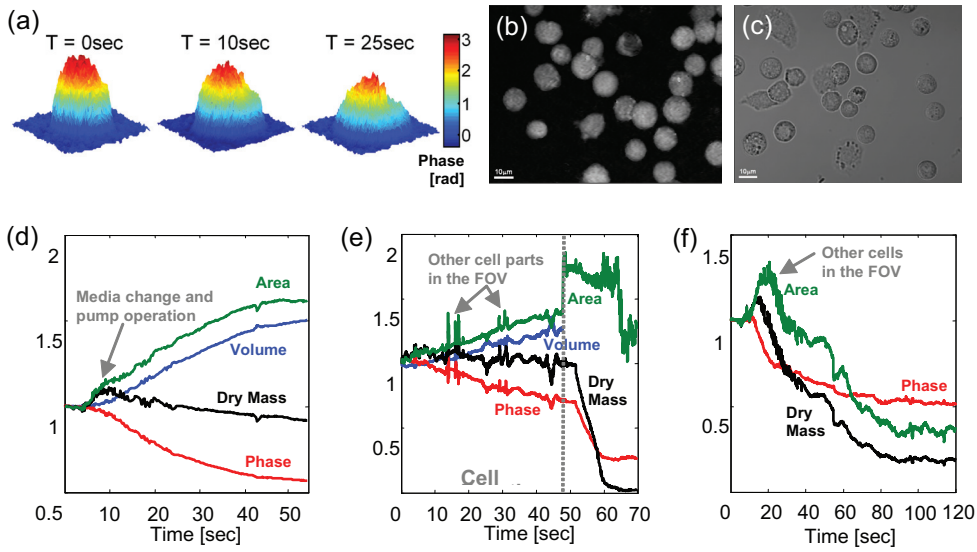


Fig. 3. Articular chondrocyte fast dynamics due to hypoosmotic pressure: (a) WFDI-based surface plots of the phase profiles at several different time points; (b) WFDI-based phase profile of the cell monolayer, acquired at 120 frames per second; (c) Phase image of the monolayer obtained by DIC microscopy. (d)-(f) WFDI-based graphs of the relative change in various cell morphological parameters during: (d) single-cell swelling as partially visualized in (a); (e) single-cell swelling and bursting; and (f) cell monolayer dynamics as partially visualized in (b). (Shaked et al., 2010 d).

Then, the current cell medium is replaced by another medium with the same osmolarity, to avoid cell volume changes, but with a different refractive index of $n_m + \Delta n$, yielding a cell OPD of:

$$OPD_{c,2}(x,y) = \left(\overline{n}_c(x,y) - \left(n_m + \Delta n_m\right)\right) \cdot h_c(x,y). \tag{7}$$

The cell thickness profile can be obtained by subtracting Eqs. (6) and (7) which yields:

$$h_c(x,y) = \frac{OPD_{c,1}(x,y) - OPD_{c,2}(x,y)}{\Delta n_m}. \tag{8}$$

Substituting Eq. (8) in Eq. (6) and (7) yields the integral refractive index of the cell as follows:

$$\overline{n}_c(x,y) = \frac{OPD_{c,1}(x,y)\Delta n_m}{OPD_{c,1}(x,y) - OPD_{c,2}(x,y)} + n_m. \tag{9}$$

Despite the simplicity of this two-exposure method, it is effective only if the cell is not highly dynamic and the changes between the consecutive phase measurements are minimal. In other cases, this method is not useful for measuring the correct thickness profile of the cell.

Alternatively, methods of scanning the cell from different points of view can be employed to obtain an intracellular refractive index map (Charrière et al. 2006, Choi et al., 2007). Briefly, phase profiles of the cell are measured by WFDI at different angles, by either rotating the sample or changing the illumination direction, and are then processed by a tomographic algorithm (e.g. the filtered backprojection algorithm) to obtain a three-dimensional refractive index map $n_c(x,y,z)$ of the cell. Since the obtained refractive index map is three-dimensional and not only the integral refractive index $\overline{n}_c(x,y)$ across a plane of view, it can be presented slice by slice using any pair of dimensions. This method is more complicated than simple WFDI, since it typically requires mechanical scanning with dedicated hardware, and it also assumes that the cell is static during the scan time; this precludes acquiring three-dimensional refractive index maps of highly dynamic cells.

Park et al. (2006) have proposed a system integrating WFDI and epi-fluorescence microscopy, which can in principle detect organelle locations in real time. If the organelle refractive indices and sizes are known in advance, then the cell thickness profile can be calculated. Rappaz et al. (2008 b) have proposed simultaneous measurement of cell thickness and refractive index by using two illumination wavelengths and a dispersive extracellular dye in the medium.

Alternatively, phase profile measurements can be used in a complementary way: rather than measuring or assuming a certain refractive index and calculating the cell thickness profile, the cell thickness can be measured by another method and then used in combination with the phase measurement obtained by WFDI to calculate the refractive indices of cellular organelles. For example, confocal microscopy has been used in combination with WFDI microscopy to measure refractive indices of cell organelles (Curl et al., 2005; Lue et al., 2009), and cell height measurments obtained by shear-force feedback topography have been combined with WFDI-based phase measurements (Edward et al., 2009). Another approach is to obtain the cell thickness by restraining the cell mechanically to a known thickness in the direction perpendicular to the image plane. This can be performed, for example, by

attaching another coverslip to the sample (Kemper et al., 2006) or using a dedicated microchannel device (Lue et al., 2006). This method, however, applies pressure to the cell and might disturb the behavior of the cell or affect the phenomena of interest. Kemper et al. (2007), Kemmler et al. (2007), and Tychinsky et al. (2008) have shown that for cells of a known uniform shape in suspension, the transverse viewable area of the cell can be used to evaluate cell thickness. For example, if the cell shape is a perfect sphere, its width is equal to its height. In all of these specific cases, the integral refractive index can be calculated using the phase profile obtained by WFDI since the cellular thickness is known.

4. Whole-cell-analysis based on WFDI phase profiles

In this section, we show that the WFDI-based phase profiles are useful for quantitative analysis of cells, even in cases where decoupling of thickness and refractive index is not possible or desired. This typically happens for highly-dynamic heterogeneous-refractive-index cells, such as cardiomyocytes (heart muscle cells). By coordinated contraction, these cells control blood flow through the blood vessels of the circulatory system. The dynamic behavior of cardiomyocytes is characterized by a rapid contraction of the cell followed by restoration to equilibrium. Contrary to cells with homogenous refractive index, cardiomyocytes contain organelles with varying refractive indices distributed across the cell interior. These mainly include myofibrils of highly organized sarcomeric arrays of myosin and actin, nuclei, and mitochondria. Using confocal dual-channel fluorescence microscopy, we have demonstrated that the cardiomyocyte organelles of different refractive indices are in motion during the entire beating cycle of the cell (Shaked et al., 2010 a). For this reason, it is not possible to accurately decouple refractive index from thickness using the phase-measurement of the entire cardiomyocyte obtained only from single-exposure WFDI. Furthermore, alternative approaches described previously that require more than one exposure (e.g. tomographic scanning or medium-exchange differential measurements; see Section 3) can result in loss of dynamic information when recording these cells due to their rapid dynamic nature. This limitation precludes calculating the cell thickness profiles from the phase measurements obtained by single-exposure WFDI during the cell beating cycle. In spite of this fact, we have shown that the dynamic WFDI-based phase profiles of the whole cell are still useful for numerical analysis of the cells (Shaked et al., 2010 a). This has been done by identifying certain numerical parameters that quantify specific processes of interest to cell biologists. We have validated the utility of the proposed parameters by showing they are sensitive enough to detect modification of cardiomyocyte contraction dynamics due to temperature change.

In order to numerically quantify the dynamic phase profile of the cells, without the need to extract the thickness profile, we first define the phase-average displacement (PAD) as follows (Shaked et al., 2010 a):

$$\Delta\varphi_t(x,y) = \varphi_t(x,y) - \varphi_0(x,y)\,, \tag{10}$$

where $\varphi_t(x,y)$ is the spatially varying phase at time point t, and $\varphi_0(x,y)$ is the spatially varying phase at the resting time point of the cell; if such a time point is not known, $\varphi_0(x,y)$ is defined as the time average of the entire phase-profile $\varphi_0(x,y) = \langle \varphi_t(x,y) \rangle_t$. Using Eq. (10), we define the positive and negative mean-square phase-average displacements (MS-PAD+ and MS-PAD–, respectively) as follows:

$$\Delta\varphi_{MS}^{+}(x,y)=\left\langle\left(\Delta\varphi_t(x,y)\right)^2:\Delta\varphi_t(x,y)\geq 0\right\rangle_t,$$
$$\Delta\varphi_{MS}^{-}(x,y)=\left\langle\left(\Delta\varphi_t(x,y)\right)^2:\Delta\varphi_t(x,y)<0\right\rangle_t, \tag{11}$$

and the spectral-domain MS-PAD as follows:

$$\Delta\varphi_{MS}(f_x,f_y)=\left\langle\left|\Delta\varphi_t(f_x,f_y)\right|^2\right\rangle_t, \tag{12}$$

where $\Delta\varphi_t(f_x,f_y)$ is obtained by a two-dimensional spatial Fourier transform of $\Delta\varphi_t(x,y)$. Using Eqs. (11) and (12), we define the following parameters to describe the MS-PAD global contributions:

$$\eta_1^{+}=\sqrt{\left\langle\Delta\varphi_{MS}^{+}(x,y)\right\rangle_{(x,y)}},$$
$$\eta_1^{-}=\sqrt{\left\langle\Delta\varphi_{MS}^{-}(x,y)\right\rangle_{(x,y)}}, \tag{13}$$
$$\eta_2=\sqrt{\left\langle\Delta\varphi_{MS}(f_x,f_y)\right\rangle_{(f_x,f_y)}},$$

where $\langle\bullet\rangle_{(x,y)}$ and $\langle\bullet\rangle_{(f_x,f_y)}$ define an area averaging.

Let us also define the phase instantaneous displacement (PID) as follows:

$$\Delta\varphi_{t,\tau}(x,y)=\varphi_{t+\tau}(x,y)-\varphi_t(x,y). \tag{14}$$

where τ defines the time duration between time point t and time point $t+\tau$. Using Eq. (14), we define the positive and negative mean-square phase instantaneous displacements (MS-PID+ and MS-PID–, respectively) as follows:

$$\Delta\varphi_{MS,\tau}^{+}(x,y)=\left\langle\left(\Delta\varphi_{t,\tau}(x,y)\right)^2:\Delta\varphi_{t,\tau}(x,y)\geq 0\right\rangle_t,$$
$$\Delta\varphi_{MS,\tau}^{-}(x,y)=\left\langle\left(\Delta\varphi_{t,\tau}(x,y)\right)^2:\Delta\varphi_{t,\tau}(x,y)<0\right\rangle_t, \tag{15}$$

as well as the spectral-domain MS-PID as follows:

$$\Delta\varphi_{MS,\tau}(f_x,f_y)=\left\langle\left|\Delta\varphi_{t,\tau}(f_x,f_y)\right|^2\right\rangle_t, \tag{16}$$

where $\Delta\varphi_{t,\tau}(f_x,f_y)$ is obtained by a spatial Fourier transform of $\Delta\varphi_{t,\tau}(x,y)$. By using Eqs. (15) and (16), we define parameters describing the MS-PID global contributions:

$$\gamma_{1,\tau}^{+}=\sqrt{\left\langle\Delta\varphi_{MS,\tau}^{+}(x,y)\right\rangle_{(x,y)}},$$
$$\gamma_{1,\tau}^{-}=\sqrt{\left\langle\Delta\varphi_{MS,\tau}^{-}(x,y)\right\rangle_{(x,y)}}, \tag{17}$$
$$\gamma_{2,\tau}=\sqrt{\left\langle\Delta\varphi_{MS,\tau}(f_x,f_y)\right\rangle_{(f_x,f_y)}}.$$

As the cell phase profile is strongly associated with its dry mass (Barer, 1952; Popescu et al., 2008 b; Rappaz et al., 2009), the parameters defined above are associated with motion of different intracellular structures including organelles. For example, the spectral-domain MS-PID has been used to describe cancer cell dynamics (Popescu et al., 2008 b). Note that all the parameters defined above are based on the entire dynamic phase profile, without the need to first extract the physical thickness profile.

The analysis tools described above are sensitive enough to detect the subtle changes in the dynamic phase profile of a beating cardiomyocyte (Shaked et al., 2010 a). Figures 4(a) and 4(e) compare the phase profiles of a cardiomyocyte during a single beating cycle at 30°C and 23°C, respectively. The cell in Fig. 4(e) show more motion along a greater portion of its length in comparison to the cell shown in Fig. 4(a), especially in the recovery stage (defined as the time interval between the point of maximum stretch of the cell and the point when the cell returned to the resting stage). The relatively slow recovery phase observed at lower temperatures is consistent with temperature-induced inhibition of calcium-regulated contraction, as discussed at the end of this section. The dynamic differences during the beating cycle of the same cell at 30°C in comparison to 23°C are more obvious from the PAD and PID profiles shown in Figs. 4(b-d,f-h) (also see Media 4 in Shaked et al., 2010 a). As seen in Figs. 4(b,f), the PAD profiles of cardiomyocytes provide a means of tracking the differential changes in the current WFDI phase profile compared to the associated phase profile in the cell resting point (defined by Eq. (10)), which reveals the dry-mass movement inside the cell. Thus the PAD profiles can be used to study cell function during the beating cycle. Furthermore, the global contribution of the dynamic PAD profiles characterizes the entire beating cell cycle and is the basis for the MS-PAD profiles, which are described by the single-value η parameters. As can be seen in Figs. 4(c,d,g,h), the PID profiles give an indication of the instantaneous movement of dry mass in the cell associated with organelles of different sizes over different time periods during the cell beating cycle, as defined by the time parameter τ. The subtraction operation defined by Eq. (14) creates a differential measurement between the current phase profile and a previous phase profile that is time-shifted by τ. Phase contributions from this previous frame are thus canceled and the uncorrelated regions between the two frames are revealed. The MS-PID profiles are calculated based on the dynamic PID profiles, yielding the single-valued γ parameters characterizing the cell over various time periods during the experiment.

Figures 5(a,b) and (c,d) show the MS-PAD profiles obtained for the dynamic PAD profiles shown in Figs. 4(b) and (f) at 30°C and 23°C, respectively. These MS-PAD profiles were used to calculate the η parameters as defined by Eq. (4), and for this particular cell yielded $\eta_1^+ = 0.10998$ and $\eta_1^- = 0.10941$ at 30°C, and $\eta_1^+ = 0.10249$ and $\eta_1^- = 0.10251$ at 23°C. In a similar way, the associated spectral-domain PAD profile yielded $\eta_2 = 16.8186$ at 30°C and $\eta_2 = 16.1487$ at 23°C.

Figures 5(e,f) and (g,h) show the MS-PID profiles for $\tau = 8.3$ msec (single-frame separation at 120 fps) at 30°C and 23°C, respectively. Media 5 in Shaked et al. (2010 a) presents the associated dynamic MS-PID profiles as a function of the time parameter τ. Based on these dynamic MS-PID profiles and on the spectral domain MS-PID profile, we calculated the γ parameters as defined by Eq.(17). Figures 5(i-k) show the dependency of the various γ parameters on τ at the two temperature levels. As can be seen from these graphs, the difference between the two temperature levels is more pronounced for higher τ values than for lower τ values. This result is expected due to the fact that the change between the two temperature levels is more pronounced in the recovery phase of the cardiomyocyte beating cycle.

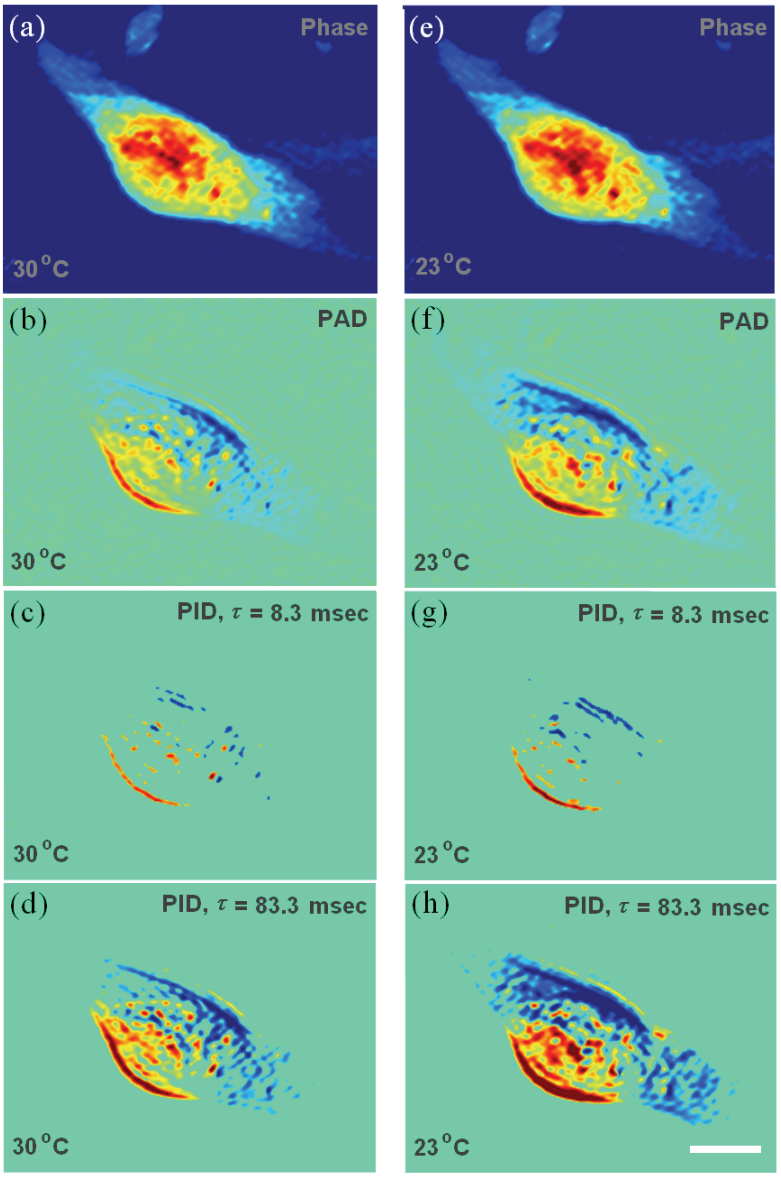


Fig. 4. Example of numerical analysis applied on a WFDI-based phase profile of a cardiomyocyte during beating at two different temperatures: (a-d) at 30°C, (e-h) at 23°C. (a,e) Phase profile; (b,f) PAD profile; (c,g) PID profile for τ = 8.3 msec (single-frame separation at 120 fps); (d,h) PID profile for τ = 83.3 msec (ten-frame separation at 120 fps). In (b-d,f-h): 'hot' colors represent positive values, 'cold' colors represent negative values, and cyan represents zeros. White horizontal scale bar represents 10 μm. Dynamics, 120 fps for 1 sec: see Media 4 in (Shaked et al., 2010 a).

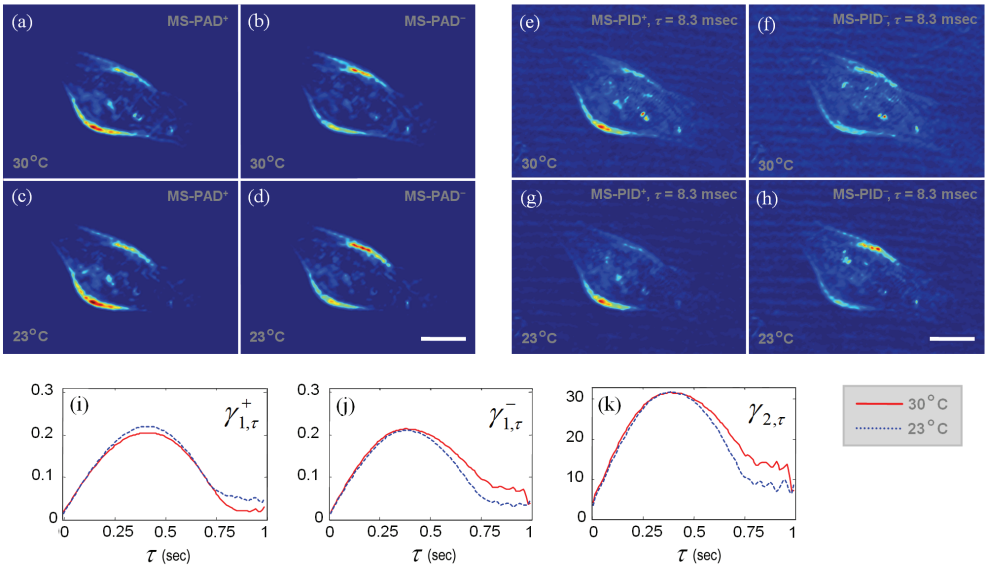


Fig. 5. Calculation of different numerical parameters based on the dynamic phase profiles shown in Fig. 4. (a-d) MS-PAD profiles (based on which the η parameters are defined): (a) MS-PAD+ at 30°C (which yields $\eta_1^+ = 0.10998$); (b) MS-PAD– at 30°C (which yields $\eta_1^- = 0.10941$); (c) MS-PAD+ at 23°C (which yields $\eta_1^+ = 0.10249$); (d) MS-PAD– at 23°C (which yields $\eta_1^- = 0.10251$). (e-h) MS-PID profiles for τ = 8.3 msec; varying τ: Media 5 in (Shaked et al., 2010 a): (e) MS-PID+ at 30°C; (f) MS-PID– at 30°C; (g) MS-PID+ at 23°C; (h) MS-PID– at 23°C. (i-k) Graphs showing the dependency of the γ parameters on τ. Solid red lines represent the measurements done at 30°C. Dashed blue lines represent the measurements done at 23°C. (i) $\gamma_{1,\tau}^+$ (which is based on MS-PID+); (j) $\gamma_{1,\tau}^-$ (which is based on MS-PID–); (k) $\gamma_{2,\tau}$ (which is based on the spectral-domain MS-PID). The white horizontal scale bars represent 10 μm. (Shaked et al., 2010 a).

The numerical analysis described above was performed on the WFDI phase profiles of 18 individual cardiomyocytes at 30°C (Shaked et al. 2010 a). The values obtained for each of the γ and η parameters were averaged over 3-4 beating cycles and normalized by the viewable area of the cell. Afterwards, the experiments were repeated at 23°C. Figure 6 summarizes the results obtained. Statistical significance between the two groups of cells (at 30°C and at 23°C) was seen for all γ and η parameters as indicated by low p-values, which were calculated by the two-sided Wilcoxon rank-sum test (Conover, 1999). These results demonstrate that these unique whole-cell-based numerical parameters can be used to discriminate between different dynamic behaviors of cardiomyocytes, and thus can be used to quantitatively study dynamic phenomena in these cells.

As can also be seen in Fig, 6, there is an apparent advantage for using the negative parameters $\gamma_{1,\tau}^-$ for discriminating between the two groups of cells. Higher values in these parameters represent increased levels of MS-PID–. In the recovery phase of the cell, it is more likely to have more cell points with negative MS-PID than positive MS-PID, since the phase profile in the cell contractile region decreases. This implies that there is a larger influence of the ambient temperature on the recovery phase of the cell beating, as compared

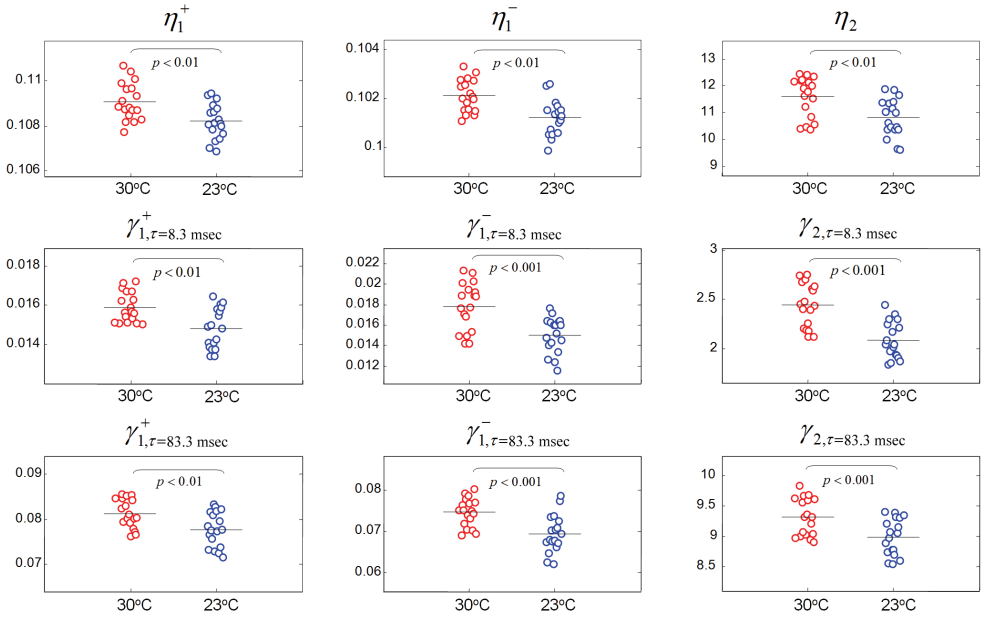


Fig. 6. Values of the γ and η parameters that are based on the whole-cell phase profiles, demonstrating that these parameters discriminate between cardiomyocytes beating at 30°C and 23°C (18 cells in each group, 3-4 beating cycles per each cell). Each circle represents a different cardiomyocyte, and the horizontal line at each group represents the average value for all cells in the group. (Shaked et al., 2010 a).

to the contraction phase. These results are supported by previous studies performed by other methods, where temperature had a profound effect on the biochemistry of contraction in the myocardium of the intact heart and in cardiomyocytes *in vitro* (Covell et al., 1966; Badeer, 1967; Engel et al., 1995; Fu et al., 2005).

The whole-cell analysis tools presented here capture intermediate events associated with dry mass movement over different time scales during the cardiomyocyte beating cycle. These intermediate events cannot be well discriminated by directly visualizing the dynamic phase profiles of the cell. In contrast, the single-valued η and γ parameters can uniquely characterize cell function, as demonstrated for temperature change. We believe that these numerical tools will also be useful for analyzing various fast dynamic behaviors in other biological cells, including intracellular and extracellular membrane fluctuations and reorganization of the cell cytoskeleton. More details on this stubject can be found in (Shaked et al., 2010 a).

5. Conclusion

We have started by presenting the principles of WFDI phase microscopy for quantitative holgoraphic imaging of biological cells. After illustrating the transmissive optical geometry employed, we reviewed methods of recovering phase information from the captured interferometric intensity pattern (which is the digital hologram of the sample). Phase

profiles can then be converted to OPD, which contain the specimen coupled refractive index and physical thickness.

We have also discussed methods of calculating meaningful biological parameters from the OPD measurement. While it is straightforward to decouple physical thickness from refractive index in homogeneous samples where one of the two parameters can be determined ahead of time or taken from literature values and assumed constant, parameters such as dry mass, relative area, and relative volume can be calculated even for heterogeneous samples without decoupling physical thickness and refractive index. We demonstrated the utility of this technique by analyzing individual articular chondrocytes under hypoosmotic loading and characterized dynamic changes in the dry mass, area, and relative volume. Several other systems made use of multiple measurements at varying angular projections, immersion media, or wavelengths in an attempt to decouple refractive index from physical thickness; however all of these techniques trade off system complexity and temporal resolution in order to extract these parameters.

Finally, we presented whole-cell analysis that makes use of differential phase profiles referenced to either an initial physiological state (phase-average displacement, PAD) or a finite time delay (phase-instantaneous displacement, PID). Examining the PAD and PID over the course of a dynamic experiment provides information about the motion of intracellular organelles. Furthermore, analysis of global contribution parameters γ and η has been shown to be effective at discriminating dynamic cellular behavior as influenced by environmental variables such as temperature.

The development of WFDI phase microscopy has introduced a variety of optical techniques for obtaining label-free quantitative information about the structure of cells. While phase contrast and DIC microscopy have previously utilized OPDs induced by cellular structures to provide relative image contrast and qualitative measurements, quantification of phase profiles by WFDI represents a major development that allows more rigorous studies of cellular behavior over time. Furthermore, WFDI phase microscopy can in many cases obviate the need for exogenous contrast agents such as fluorophores whose drawbacks may include photobleaching over time, cytotoxicity, and potential modification of intracellular behavior in response to the fluorophore or bound antibody. While WFDI is a whole-cell hologaphic approach and thus does not provide contrast arising from molecular specificity, it is a label-free quantitative technology that complements labeled-based (e.g. fluorescence) imaging for investigating and quantifying cellular dynamics.

WFDI phase microscopy has been shown to be highly effective for high speed cellular dynamics studies without the need for special sample preparation. The presented set of holographic techniques has significant potential applications for characterization of various cellular phenomena and disease conditions.

6. Acknowledgments

This work was supported by National Science Foundation (NSF) grant CBET-0651622. N.T.S. greatly acknowledges the support of the Bikura Postdoctoral Fellowship from Israel.

7. References

Badeer, H. S. (1967) Effect of hypothermia on the contractile 'capacity' of the myocardium, *Journal of Thoracic and Cardiovascular Surgery* 53, 5, (May 1967) 651-656

Barer, R. (1952) Interference microscopy and mass determination, *Nature* 169, 4296, (March 1952) 366-367

Charrière, F.; Marian, A.; Montfort, F.; Kuehn, J.; Colomb, T.; Cuche, E.; Marquet, P. & Depeursinge, C. (2006) Cell refractive index tomography by digital holographic microscopy, *Opt. Lett.* 31, 2, (January 2006) 178-180

Choi, W.; Fang-Yen, C.; Badizadegan, K.; Oh, S.; Lue, N.; Dasari, R.R. & Feld, M. S. (2007) Tomographic phase microscopy, *Nature Methods* 4, (August 2007) 717-719

Conover, W. J. (1999) *Practical Nonparametric Statistics*, 3rd Edition, John Wiley, pp. 271-276.

Covell, J.W.; Ross Jr., J.; Sonnenblick, E. H. & Braunwald, E. (1966) Comparison of the force-velocity relation and the ventricular function curve as measures of the contractile state of the intact heart, *Circulation Research* 19, (March 1966) 364-372

Cuche, E.; Marquet, P. & Depeursinge, C. (1999) Simultaneous amplitude-contrast and quantitative phase-contrast microscopy by numerical reconstruction of Fresnel off-axis holograms, *Appl. Opt.* 38, 34, (December 1999) 6994-7001

Curl, C.L.; Bellair, C.J.; Harris, T.; Allman, B.E.; Harris, P.J.; Stewart, A.G.; Roberts, A.; Nugent, K.A. & Delbridge, L.M. (2005) Refractive index measurement in viable cells using quantitative phase-amplitude microscopy and confocal microscopy, *Cytometry* 65, 1, (May 2005) 88-92

Edward, K.; Farahi, F. & Hocken, R. (2009) Hybrid shear force feedback/scanning quantitative phase microscopy applied to subsurface imaging, *Opt. Express* 17, 21, (October 2009) 18408-18418

Engel, J.; Sowerby, A.J.; Finch, S.A.; Fechner, M. & Stier, A. (1995) Temperature dependence of Ca^{2+} wave properties in cardiomyocytes: implications for the mechanism of autocatalytic Ca^{2+} release in wave propagation, *Biophysical J.* 68, 1, (January 1995) 40-45

Fu, Y.; Zhang, G.-Q.; Hao, X.-M.; Wu, C.-H.; Chai, Z. & Wang, S.-Q. (2005) Temperature dependence and thermodynamic properties of Ca^{2+} sparks in rat cardiomyocytes, *Biophysical J.* 89, 4, (October 2005) 2533-2541

Ikeda, T.; Popescu, G.; Dasari, R.R. & Feld, M.S. (2005) Hilbert phase microscopy for investigating fast dynamics in transparent systems, *Opt. Lett.* 30, 10, (May 2005) 1165-1167

Kemmler, M.; Fratz, M.; Giel, D.; Saum, N.; Brandenburg, A. & Hoffmann, C. (2007) Non-invasive time-dependent cytometry monitoring by digital holography, *J. Biomed. Opt.* 12, 6, (Nov-Dec 2006) 064002

Kemper, B.; Kosmeier, S.; Langehanenberg, P.; von Bally, G.; Bredebusch, I.; Domschke, W. & Schnekenburger, J. (2007) Integral refractive index determination of living suspension cells by multifocus digital holographic phase contrast microscopy, *J. Biomed. Opt.* 12, 5, (Sept-Oct 2007) 054009

Kemper, B.; Carl, D.; Schnekenburger, J.; Bredebusch, I.; Schafer, M.; Domschke, W. & von Bally, G. (2006) Investigation of living pancreas tumor cells by digital holographic microscopy, *J. Biomed. Opt.* 11, 3, (May-Jun 2006) 34005

Lue, N.; Choi, W.; Popescu, G.; Yaqoob, Z.; Badizadegan, K.; Dasari, R.R. & Feld, M.S. (2009) Live cell refractometry using Hilbert phase microscopy and confocal reflectance microscopy, *Journal of Physical Chemistry A*, 113, (October 2009) 13327-13330

Lue, N.; Popescu, G.; Ikeda, T.; Dasari, R.R.; Badizadegan, K. & Feld, M.S. (2006) Live cell refractometry using microfluidic devices, *Opt. Lett.* 31, 18, (September 2006) 2759-2761

Marquet, P.; Rappaz, B.; Magistretti, P.J.; Cuche, E.; Emery, Y.; Colomb, T. & Depeursinge, C. (2005) Digital holographic microscopy: a noninvasive contrast imaging technique allowing quantitative visualization of living cells with subwavelength axial accuracy, *Opt. Lett.* 30, 5, (March 2005) 468-470

Park, Y.K.; Diez-Silva, M.; Fu, D.; Popescu, G.; Choi, W.; Barman, I.; Suresh, S. & Feld, M.S. (2010) Static and dynamic light scattering of healthy and malaria-parasite invaded red blood cells, *Journal of Biomedical Optics Letters* 15, 2, (March-April 2010) 020506

Park, Y. K.; Diez-Silva, M.; Popescu, G.; Lykorafitis, G.; Choi, W.; Feld, M.S. & Suresh, S. (2008) Refractive index maps and membrane dynamics of human red blood cells parasitized by Plasmodium falciparum, *Proc. Natl. Acad. Sci. USA. (PNAS)* 105, (2008) 13730

Park, Y.K.; Popescu, G.; Badizadegan, K.; Dasari, R.R. & Feld, M.S. (2006) Diffraction phase and fluorescence microscopy, *Opt. Express* 14, 18, (September 2006) 8263-8268

Popescu, G.; Park, Y.K.; Choi, W.; Dasari, R.R.; Feld, M.S. & Badizadegan, K. (2008 a) Imaging red blood cell dynamics by quantitative phase microscopy, *Blood Cells, Molecules and Diseases* 41, 1, (July-Aug 2008) 10-16

Popescu, G.; Park, Y.K.; Lue, N.; Best-Popescu, C.A.; Deflores, L.; Dasari, R.R.; Feld, M.S. & Badizadegan, K. (2008 b) Optical imaging of cell mass and growth dynamics, *The American Journal of Physiology - Cell Physiology* 295, (June 2008) C538-C544

Popescu, G.; Ikeda, T.; Badizadegan, K.; Dasari, R.R. & Feld, M.S. (2005) Erythrocyte structure and dynamics quantified by Hilbert phase microscopy, *J. Biomed. Opt. Lett.* 10, (2005) 060503

Rappaz, B.; Cano, E.; Colomb, T.; Kühn, J.; Simanis, V.; Depeursinge, C.; Magistretti, P.J. & Marquet, P. (2009) Noninvasive characterization of the fission yeast cell cycle by monitoring dry mass with digital holographic microscopy, *J. Biomed. Opt.* 14, 3, (May-June 2009) 034049

Rappaz, B.; Barbul, A.; Emery, Y.; Korenstein, R.; Depeursinge, C.; Magistretti, P.J. & Marquet, P. (2008 a) Comparative study of human erythrocytes by digital holographic microscopy, confocal microscopy, and impedance volume analyzer, *Cytometry A.* 73, 10, (October 2008) 895-903

Rappaz, B.; Charrière, F.; Depeursinge, C.; Magistretti, P.J. & Marquet, P. (2008 b) Simultaneous cell morphometry and refractive index measurement with dual-wavelength digital holographic microscopy and dye-enhanced dispersion of perfusion medium, *Opt. Lett.* 33, 7, (April 2008) 744-746

Rappaz, B.; Marquet, P.; Cuche, E.; Emery, Y.; Depeursinge, C. & Magistretti, P. (2005) Measurement of the integral refractive index and dynamic cell morphometry of living cells with digital holographic microscopy, *Opt. Express* 13, 23, (November 2005) 9361-9373

Shaked, N. T.; Satterwhite, L.L.; Bursac, N. & Wax, A. (2010 a) Whole-cell-analysis of live cardiomyocytes using wide-field interferometric phase microscopy, *Biomedical Optics Express* 1, 2, (September 2010) 706-719

Shaked, N. T.; Zhu, Y.; Badie, N.; Bursac, N. & Wax, A. (2010 b) Reflective interferometric chamber for quantitative phase imaging of biological sample dynamics, *J. of Biomed. Opt. Lett.* 15, 3, (May-June 2010) 030503

Shaked, N. T.; Newpher, T.M.; Ehlers, M.D. & Wax, A. (2010 c) Parallel on-axis holographic phase microscopy of biological cells and unicellular microorganism dynamics, *Appl. Opt.* 49, 15, (May 2010) 2872-2878

Shaked, N. T.; Finan, J.D.; Guilak, F. & Wax, A. (2010 d) Quantitative phase microscopy of articular chondrocyte dynamics by wide-field digital interferometry, *J. Biomed. Opt. Lett.* 15, 1, (Jan-Feb 2010) 010505

Shaked, N. T.; Zhu, Y.; Rinehart, M.T. & Wax, A. (2009 a) Two-step-only phase-shifting interferometry with optimized detector bandwidth for microscopy of live cells, *Opt. Express* 17, 18, (August 2009) 15585-15591

Shaked, N. T.; Rinehart, M.T. & Wax, A. (2009 b) Dual-interference-channel quantitative-phase microscopy of live cell dynamics, *Opt. Lett.* 34, 6, (March 2009) 767-769

Tychinsky, V.P.; Kretushev, A.V.; Klemyashov, I.V.; Vyshenskaya, T.V.; Filippova, N.A.; Raikhlin, N.T. & Shtil, A.A. (2008) Quantitative real-time analysis of nucleolar stress by coherent phase microscopy, *J. Biomed. Opt.* 13, 6, (Nov-Dec 2008) 064032

11

Digital Holography and Cell Studies

Kersti Alm[1], Helena Cirenajwis[2],
Lennart Gisselsson[1], Anette Gjörloff Wingren[3], Birgit Janicke[1],
Anna Mölder[1], Stina Oredsson[2] and Johan Persson[1]
[1]*Phase Holographic Imaging AB, Lund,*
[2]*Department of Biology, Lund University, Lund,*
[3]*Department of Biomedical Laboratory Science, Health and Society,*
Malmö University, Malmö,
Sweden

1. Introduction

Digital holographic microscopy (DHM) is a novel high-resolution imaging technique that offers real-time imaging and quantitative measurements of physiological parameters. It has developed into a broad field, and one of many interesting applications is to study cells without staining or labeling them and without affecting them in any way. Digital holography makes it possible to easily measure cell properties that previously have been very difficult to study in living cells, such as cell thickness, volume, and cell refractive index (Marquet et al., 2005; Rappaz et al. 2005; Mölder et al., 2008; El-Schish et al., in press; Persson et al., in press). Living, dying or dead cells as well as fixed cells can be studied. The first DHM images showing living cells were published in 2003 and 2004 (You et al., 2003; Carl et al., 2004), making this field of research rather new. Two of the most interesting functions of DHM is 3-D imaging of objects and to make in-focus measurements over time. Digital holography has been used to study a wide range of cells, *e.g.* protozoa, bacteria and plant cells as well as several types of mammalian cells such as nerve cells and tumor cells (Emery et al., 2007; Kemper et al., 2006; Moon and Javidi 2007). It has also been applied for studies of cell proliferation, cell movement and cell morphology (Kemper et al., 2009; Yu et al., 2009). Movement in both 2-D and 3-D has been studied (Langehanenberg et al., 2009; Persson et al., in press). In addition, cell viability status can be determined using DHM (Kemper et al., 2006; Kemmler et al., 2007). Interestingly, it is possible to study both single cells and entire populations simultaneously, allowing for very detailed studies. In this chapter we will compare DHM with previously used techniques and discuss the benefits and drawbacks of digital holography cell measurements. We will also present cell studies made possible by DHM.

2. Why digital holographic microscopy?

Cell imaging plays a crucial role in the understanding of cell biology. Cells are almost invisible in standard light microscopes as they do not absorb light. Cells shift the phase of the light and different light microscopy methods, such as phase contrast (Zernike, 1942) and Nomarski's

differential interference contrast (DIC) (Nomarski, 1955), have been developed to transform phase information into amplitude or intensity information. Some advantages of DIC are better contrast and sharpness of the images. These light microscopy methods can only provide qualitative information as quantitative information cannot be calculated. In a recent publication, Petibois has made a review of imaging methods for cells (Petibois, 2010). Another method that is used for measurements of unlabeled adherent cells in real-time is electric impedance. The instrument xCELLigence was developed by Roche Diagnostics (Basel, Switzerland). It is a real-time cell analyzer that measures living cells without labeling. Electric impedance allows measurements of cellular processes such as proliferation, cytotoxicity, invasion, migration and cell viability (Atienza et al., 2006; Boyd et al., 2008; Ge et al., 2009) and give results comparable to imaging methods in some ways.

Different staining methods have been developed that enable cells to absorb light. Dyes such as methylene blue which stains *e.g.* nucleic acids, haematoxylin which stains cell nuclei, eosin which stains cytoplasm and silver stains which stain proteins and DNA have been used since the mid 19th century, while fluorescent stains, such as rhodamine which stains mitochondria, neutral red which stains lysosomes, acridine orange and DAPI, which stain nucleic acids were developed and have been widely used since the beginning of the 20th century (Kricka & Fortina, 2009). New dyes that stain specific parts of the cells and new labels that fluoresce at certain wavelengths are developed continuously. The method to label cells with green fluorescence protein (GFP) was a major breakthrough in the 1990´s. GFP is less toxic than most commonly used dyes and the DNA code of GFP can be transfected into the cell's genome (Chalfie et al., 1994). When the gene contains the GFP DNA is activated, GFP will be produced by the transfected cells. As GFP affects cells less than traditional stains it gives more accurate results. However, GFP is a rather large molecule and the size probably causes steric problems. Fluorescence microscopy as well as traditional light microscopy may cause phototoxicity, and several researchers have attempted to develop non-damaging microscopy methods (Hoebe et al., 2007; Frigault et al., 2009; Logg et al., 2009).

The search for a method to study cells accurately without labeling or staining them has resulted in several interferometric quantitative microscopy techniques utilizing the phase properties of coherent light to image a sample. One of them is digital holography. In 1948, Dennis Gabor invented a way to encode the phase of the light both as information and as a record containing all the information in a single recording, *i.e.* the hologram (Gabor, 1948). Holograms are commonly used as pieces of art and are displayed as illuminated 3-D images. Gabor's findings were the base for the development of digital holography during the 1990s (Schnars & Jueptner 1994; Cuche et al., 1999), where the information is collected on a digital sensor and then fed to a computer. Using DHM it is possible to measure cell shape, volume and dry mass without any labeling and with a very low intensity light source (Rappaz et al., 2005; Mölder et al., 2008; Rappaz et al., 2008) where the intensity is well below what is considered photo-toxic. In Fig. 1, an L929 mouse fibroblast cell culture is captured using both phase contrast microscopy and DHM. The DHM images look similar to the phase contrast images.

DHM is a full field imaging of the phase of the light incident on the sensor. It can be used to image solid samples in reflection, or, which is most common in the case of biological samples, to image transparent samples in transmittance. The most commonly used system for DHM today uses an optical setup common to that of a Mach-Zender interferometer, but with the reference light at a slight angle to the light passing through the sample, and the

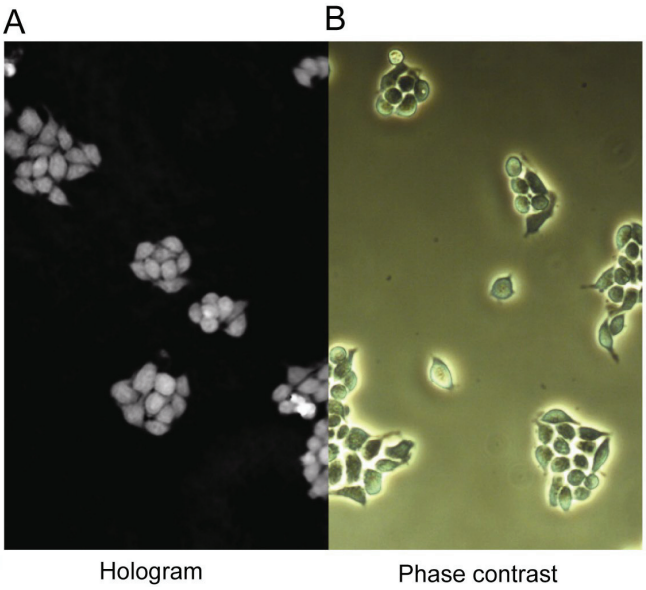


Fig. 1. A cell culture captured using digital holographic microscopy (A) and phase contrast microscopy (B). The cells are BN7005-H1D2 mouse fibroblasts which are approximately 20 µm in diameter.

image calculation is then performed using a Fresnel approximation (Cuche et al., 1999) Varieties of the setup exist (Popescu et al., 2006; Gustafsson et al., 2004) and it is also sometimes given different names (Ikeda et al., 2005). There are also variations in the calculation of the actual recognizable image. The outcome of the reconstruction of a digital hologram is two images, one representing the amplitude of the light and one representing the phase changes of the light. The amplitude image is similar to an image of the sample captured in ordinary white light. The resolution in the direction of the incident light is very high, down to the order of nanometers (Cuche et al., 1999). By combining the phase imaging technique with physical rotation of the sample, a high resolution, label-free tomographic image can be obtained (Popescu et al., 2008). One of the advantages of DHM is that the object that is studied will always be in focus, and the image can be recalculated many times if necessary to find the best focus. Images can be captured of both single cells and populations, and the images can be presented as traditional cell images as well as 3-D representations. DHM can most likely compete with many of the other methods used today within the area of cell biology and microscopy. The technique is easy to learn and simple to use, it is cheap and gives both qualitative and quantitative results. Several research groups have used DHM for cell biology studies.

3. Drawbacks of DHM

Cells shift light, and can therefore be detect with DHM. The magnitude of the phase shift depends on the refractive index of the cell and the cell thickness as well as the difference in refractive index between the cells and their surroundings. For the DHM system to be able to

measure the phase shift of the cells, their refractive index must differ from that of the background. Some cell types have a refractive index that is very close to the index of the cell culturing medium and the signal from thin parts of those cells will disappear into the background noise. This can be improved *e.g.* by using a double laser system, but there will always be thin cells that disappear into the background. As the DHM imaging is done using coherent light, the setup is very sensitive to refraction, reflection or changes in the polarization of the light, and large amounts of the effort of constructing a digital holographic setup is aimed at reducing background noise, introduced by optical components in the light path. One drawback is also that as of today, no known contrast agents exist that selectively increase the refractive index of different cells or parts of a cell. As with all label-free techniques, the sample is "as it is". Also, the phase shift is modular, and to measure absolute values, a base line must be set to identify the background. In some samples background identification is difficult, and this affects the image quality. Another drawback is the time required for reconstructing an actual image from the captured hologram, a time that is however reduced as computers successively grow more powerful. The reconstruction can be separated from the actual image capture, thus allowing image capture that is faster than the reconstruction time.

4. Cell morphology studies

Studies of cell morphology can show how the cells have been affected by different treatments or by environmental factors such as temperature or pH. These studies are usually performed using different microscopy techniques, and staining or labeling of the cells is often needed. Several researchers have studied cell morphology using DHM in different contexts using neither staining nor labeling (Rappaz et al., 2005; Mölder et al., 2008; Rappaz et al., 2008). Among others, Kemper and coworkers have shown that erythrocyte shape can be clearly visualized by DHM (Kemper et al., 2007). Kemmler and colleagues studied the morphological changes during trypsinization using oligodendrocytes from rat (Kemmler et al., 2007) and Emery and colleagues detected cell swelling and shrinking in primary mouse cortical neurons (Emery et al., 2007). Schnekenburger et al (2007) have applied DHM to study the dynamics of cytoskeleton changes in the human pancreatic tumor cell lines PaTu 8988S and PaTu 8988T. They showed that the cell shape changed visibly after Latrunculin B treatment. DHM has been used to sense, monitor and recognize microorganisms (Moon & Javidi, 2008) and to study drug-induced morphological changes in pancreatic cells (Kemper & von Bally, 2008). We have used DHM to compare cell morphology during proliferation of four different adherent cell lines (Mölder et al., 2008).

5. Nerve cell studies

Nerve cells are intensely studied in order to elucidate their growth and signaling mechanisms. Nerve cells have long, almost invisible protrusions which change shape and grow. An early study showed that it is possible to measure the thickness of the neuronal processes as well as the nerve cell body with DHM (Marquet et al., 2005). In Fig. 2, the protrusions of the nerve cell can clearly be seen.

Rappaz and collegues showed that swelling of nerve cell bodies caused by hypotonic shock could be studied non-invasively using DHM (Rappaz et al., 2005). The mechanisms of action

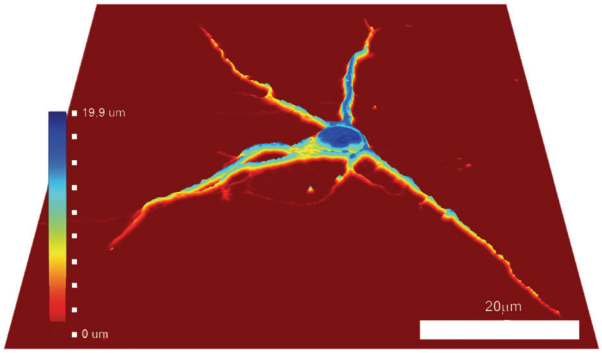


Fig. 2. A nerve cell captured using digital holographic microscopy. The protrusions are clearly seen. Both the vertical and the horizontal scale bars correspond to approximately 20 µm.

at the early stages of cell death induced by glutamate-induced excitotoxicity have been studied in primary cortical neurons from mouse (Pavillon et al., 2010). As the calcium homeostasis in a nerve cell changes, the cell volume changes as well. Pavillon and colleagues used a combination of digital holography and fluo-4 dye fluorescence signal microscopy to study the absolute volume, shape, and intracellular refractive index related to cell content, as well as intracellular calcium homeostasis simultaneously. They found a strong association between increased calcium concentration, as determined by increased fluo-4 dye fluorescence binding, and decreased quantitative phase signal during pulses of glutamate addition. The decreased phase signal was accompanied by swelling of the neurons and a surface enlargement caused by the intake of water into the cell. When the calcium homeostasis changed irreversibly, the quantitative phase changed irreversibly as well. In addition, the refractive index decreased, depending on the influx of water into the cell. Thus, using DHM in combination with fluorescence microscopy, the researchers could study cell morphology changes at the early stages of cell death caused by glutamate-induced excitotoxicity. These studies would not have been possible to perform using traditional fluorescence microscopy.

6. Differentiation studies

The differentiation process makes cells more specialized, both in shape and performance. The process is usually studied using microscopy and western blot, and different cell labels are often used. We have studied adherent 3T3L1 fibroblasts which differentiate into adipocytes after 3 days of treatment with 0.5 mM IBMX, 10 µg/ml insulin and 1 µM dexamethasone. The differentiation process is easy to monitor using DHM as the cells remain undisturbed. As is clearly seen in Fig. 3, the differentiated adipocyte cells display lipid droplets which are clearly seen in DHM, but which are not so obvious when using phase contrast microscopy. In the 3-D renderings, the lipid droplets are seen as white blebs, indicating that they are very optically dense.

We have also studied the Lund human mesencephalic neurons LUHMES, which have been induced to differentiate as described earlier (Schildknecht et al., 2009). The differentiation procedure was monitored non-invasively by capturing DHM images. Our study clearly shows that the differentiation process changed the cell shapes (Fig. 4). The differentiation process resulted in, on average, flatter cells (Table 1).

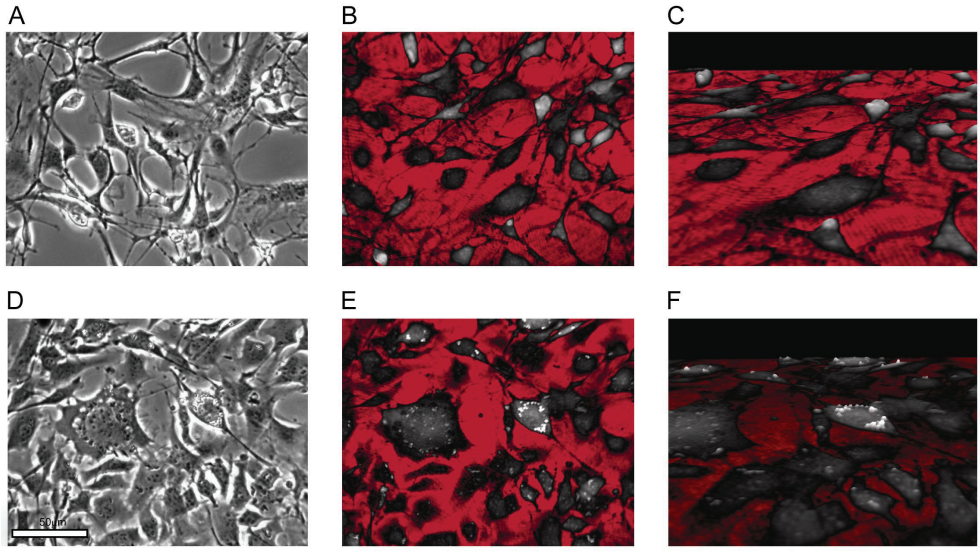


Fig. 3. 3T3L1 cells were treated with 0.5 mM IBMX, 10 µg/ml insulin and 1 µM dexamethasone for 3 days to start a differentiation process. Frames A-C show the cells in the very beginning of the differentiation process, while frames D-E show the cells after three days of treatment. Frames A and C are captured using phase contrast microscopy, while frames C-D and E-F are captured using digital holographic microscopy. In frames C and F the cells are displayed as 3-D renderings of the optical thickness measurements. The scale bar in frame D corresponds to 50 µm.

7. Stem cell studies

Tissue stem cells (TSCs) have long been known and studied for their regenerative potential, which is seen after injury and during tissue maintenance (Potten et al., 1973). Because of their ability to both self-renew and give rise to differentiated progeny, TSCs are highly exploited in the field of regenerative medicine, and were early in focus in the context of bone marrow transplantations and skin grafting (Thomas et al., 1957; Scothorne & Tough, 1952). Stem cells have also been implicated in the cell proliferation disease cancer (Reya et al., 2001). Stem cells are usually studied with flow cytometry and fluorescence microscopy. In 2007, a DHM study of sunflower and corn stem cells was performed by Moon and Javidi (Moon & Javidi, 2007). DHM was used for automated plant stem cell monitoring, sensing and identification. The authors showed that they could distinguish between the two types of stem cells by measuring morphological parameters.

8. Apoptosis studies

Apoptosis is a process of programmed cell death in vertebrates that plays a central role in development and homeostasis. Apoptosis begins with a variety of morphological changes that differ from viable cells and which are suitable for label-free quantitative and qualitative analyses by DHM. Cell membrane changes such as loss of membrane asymmetry and

	Area	Peak thickness	Volume	Avg thickness
Undifferentiated cells	116 ± 23	9.52 ± 1.48	612 ± 163	5.21 ± 0.55
Differentiated cells	118 ± 38	7.94 ± 2.15	465 ± 217	3.85 ± 0.77

Table 1. Area, peak thickness, volume and average (avg) thickness for undifferentiated and differentierad LUHMES cells, including the standard deviations (SD). The SD was based on data from 172 undifferentiated cells and 207 differentiated cells. The units used are µm² (area), µm (peak thickness, avg thickness) and µm³ (volume).

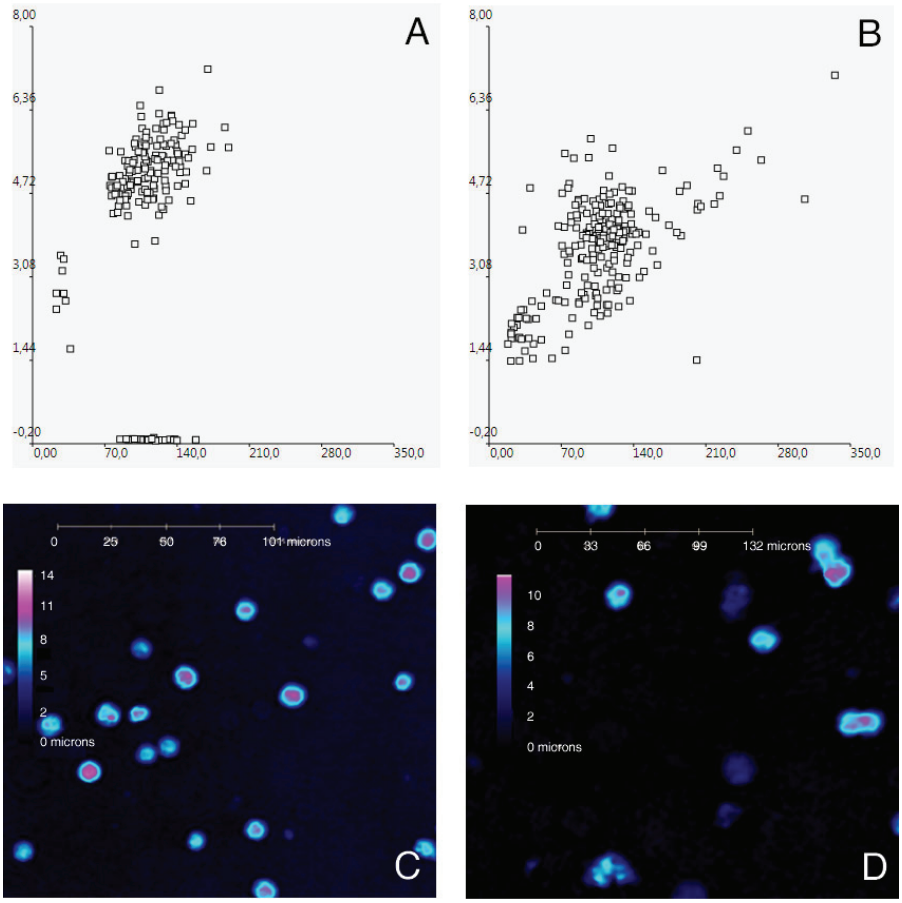


Fig. 4. Lund human mesencephalic neurons LUHMES, which have been induced to differentiate, can be analyzed for area and optical thickness. A represents cells before the differentiation process has started, while B represents cells at the end of the differentiation process. The y-axis represents the peak thickness of the cells while the x-axis represents the area in µm² of each individual object segmented in the image. Each square represents one cell. C shows the cells before the differentiation process started while D shows the cells at the end of the differentiation process.

attachment, cell shrinkage and formation of small blebs are followed by nuclear fragmentation, chromatin condensation, and chromosomal DNA fragmentation and finally the cell breaks into several apoptotic bodies (Kroemer et al., 2009). Apoptosis is usually studied with flow cytometry, fluorescence microscopy, Western blot and enzyme activity assays. DHM has been used to measure the differences in refractive index of toxin-treated and untreated adherent pancreatic cancer cells (Kemper et al., 2006). Using the refractive index to measure thickness, the study showed that cells treated with toxins or cell-death inducing drugs were thinner than control cells. DHM has also been used to follow the process of staurosporine-induced apoptosis in oligodendrocytes (Kemmler et al., 2007). More details of the apoptotic process have been shown by Colomb et al. (2008), where the apoptotic blebbing in prostate cells was clearly visualized.
We have used L929 cells treated with 200 µM etoposide to follow the apoptotic process (Fig. 5). After 12 hours of treatment (Fig. 5B), the cells were rounded up, and clearly thicker than at the beginning of the treatment. After 24 hours of treatment (Fig. 5C), the cells were very thin, and in some cases they had almost disappeared.

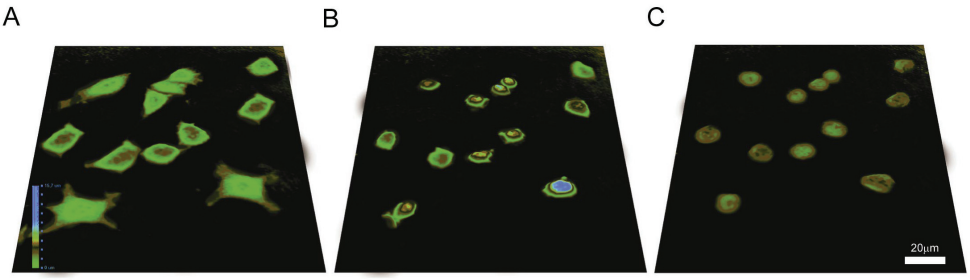


Fig. 5. L929 mouse fibroblast cells treated with 200 µM etoposide. Frame A shows untreated cells, frame B shows the cells treated for 12 hours and frame C the cells after 24 hours of treatment, at the end of the apoptotic process. The vertical color scale bar in frame A corresponds to 16 µm, showing the optical thickness of the cells. The white scale bar in frame C corresponds to 20 µm.

9. Cell division studies

The growth and division of cells follow a tightly regulated set pattern called the cell cycle. Cell cycle studies are usually performed using flow cytometry or fluorescence microscopy. The yeast *Schizosaccharomyces pombe* has long been used to study the eucaryotic cell cycle as its genome is easily accessible and easy to manipulate. Recently, changes in the cell dry mass and differences in cell density through the cell cycle have been monitored by DHM in *S. pombe* (Rappaz et al., 2009b). DHM seems to be well-suited to follow yeast through cell growth and division. The beginning and end of the cell cycle were easily detected and some steps between could also be determined. Kemper and colleagues have recently shown that they could follow mammalian cell division using DHM in, and thus measure the length of the cell cycle (Kemper et al., 2010). As yet, DHM has not been developed to perform actual cell cycle studies as the different stages of the cell cycle can not be properly identified.

10. Cell migration and motility studies

Cells move continuously both *in vivo* and *in vitro*. When cells are in culture, the movement is often random while normal cell movement in an organism is more organized. Cancer metastasis studies often involve migration or motility studies. However, cell migration studies are often tedious and difficult when using the standard filter assay methods, *e.g.* the Boyden (Boyden, 1962), Zigmond (Zigmond & Hirsch, 1973) or Dunn (Zicha et al., 1991) chambers. Time lapse studies are very useful but often require expensive set-ups with either white light or fluorescence microscopy. Two other commonly used motility assays, the wound scratch assay and the trans-endothelial migration assay (Gabbiani et al., 1984; De Becker et al., 2007) are also labor intense and cell dependent. Only a limited number of cell lines can be studied with these methods.

DHM can provide real-time information of cell movement using a much wider spectra of cells and cell lines. The earliest studies show fibroblast migration (Mann et al., 2006). Another early study utilized the 3-D ability of DHM to develop a method to follow cancer cell migration in *in vivo*-like circumstances (Dubois et al., 2006). The *in vivo*-like 3-D environment was created by using matrix gels. The 3-D ability was further utilized by Garcia-Sucerquia and coworkers in their studies of movement through liquids (Garcia-Sucerquia et al., 2006). They followed algae and protozoa and managed to accurately determine their movement through the liquid. Suspension cells are difficult to track using traditional microscopy as they quickly move out of focus. DHM has been used to follow a population of human endothelial cells in their 3-D trajectory through a solution (Sun et al., 2008). The cells were followed for 189 µm along the Z axis. Sun and colleagues also followed the human leukemic cell line HL-60 through a suspension. Langehanenberg and colleagues presented a study where fibrosarcoma cells grew in a collagen-based tissue model (Langehanenberg et al., 2009). Using DHM, they successfully followed the movement of the cells through the tissue model. Recently, we have used DHM to follow cell movement, and to show the correlation of cell movement to cell morphology (Persson et al., in press). We showed that small and large cells in a population often move in a different way compared to the medium-sized cells, and that the pattern of cell movement is cell line dependent. DHM is well suited for non-invasive time-lapse studies of movement as can be seen with these MCF10A breast cells (Fig. 6).

11. Erythrocyte studies

Erythrocytes are among the most common cell types in the body. They travel throughout the blood system to deliver oxygen to even the most remote parts of the body. In order to carry out this function, erythrocytes are robust, dense, elastic and concavely disc-shaped. Erythrocyte shape and volume can be used for clinical diagnosis purposes (Beving et al., 1991), and tests for the erythrocyte sedimentation rate are common. Modern medical cell analysis equipment uses flow cytometry technology to determine cell volume and shape (Buttarello and Plebani, 2008). The results are mostly good, although the equipment is expensive and requires expert handling. The very distinct and clear shape of erythrocytes make them well suited for DHM studies (Fig. 7). The low optical density of the cell center is clearly seen. Rappaz and colleagues monitored erythrocytes using DHM and compared the results with confocal laser scanning microscopy and an impedance volume analyzer with good results (Rappaz et al., 2008). They managed to accurately measure cell volume, surface area, diameter, refractive index and hemoglobin content, all by capturing single DHM images of the cells. In order to pass through narrow capillaries, erythrocytes must be able to

change shape quickly. Using DHM, Bernhardt and colleagues followed the settling of erythrocytes on an artificial surface and showed that the shape of the erythrocyte changed very fast as the cells settled (Bernhardt et al., 2008). Rappaz and colleagues quantified erythrocyte cell membrane fluctuations using DHM (Rappaz et al., 2009a). They measured single cells and captured images at approximately 25 images per second. The fluctuations were measured to 35.9±8.9 nm. It has earlier been very difficult to make measurements like these without affecting the cells or their environment.

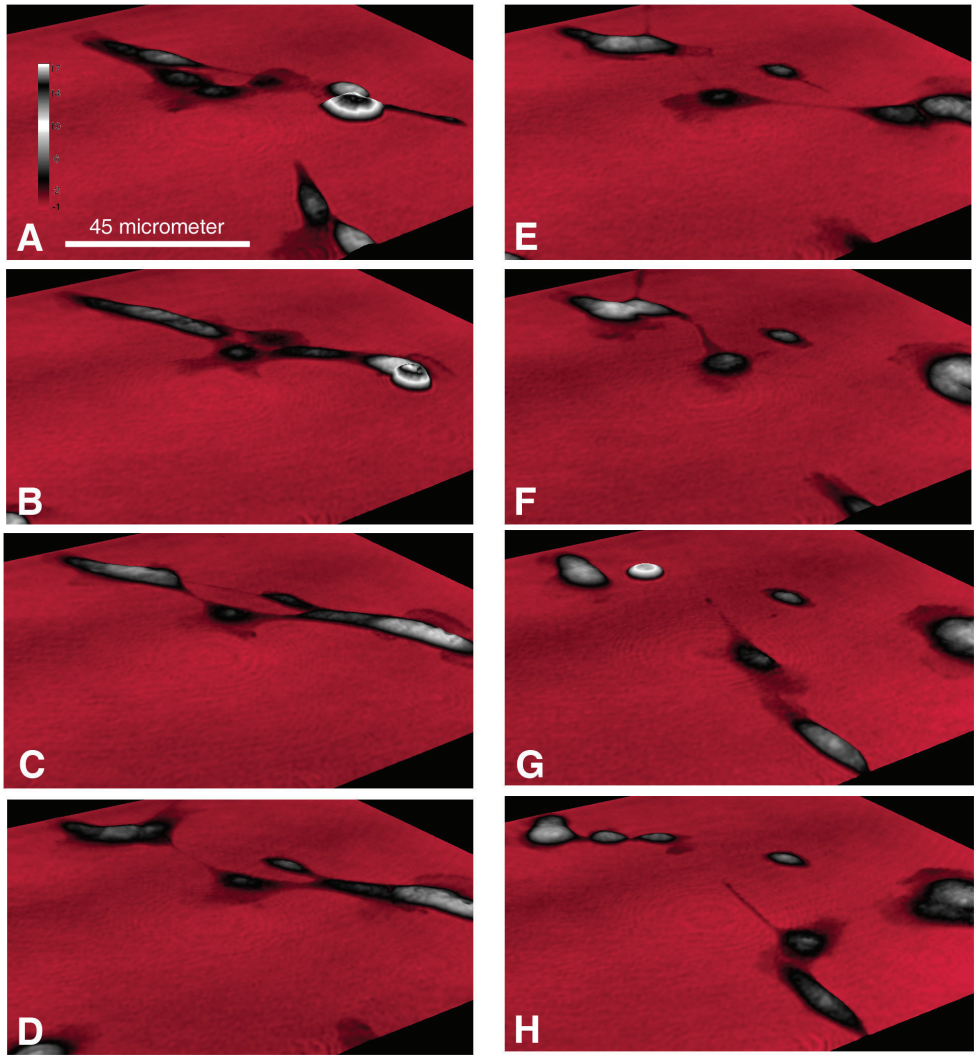


Fig. 6. MCF10A breast cells were captured every five minutes using digital holography. It was possible to see the movement of very thin cell details. Frames A-H show the movement of the cells every 100 minutes. The vertical color scale-bar in frame A corresponds to 17 µm.

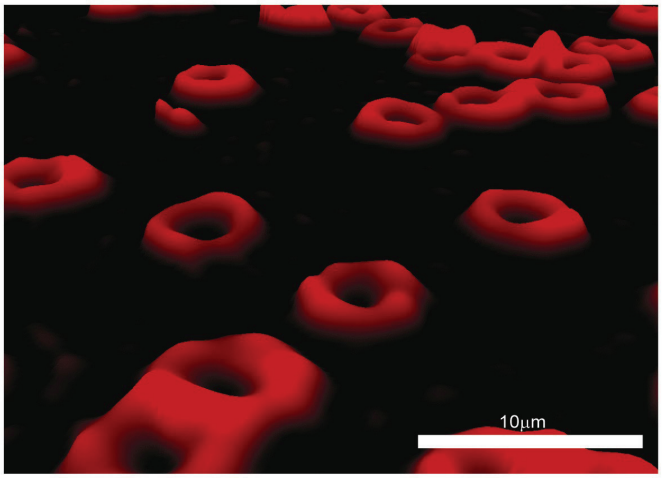


Fig. 7. Healthy, fresh human erythrocytes as captured using digital holographic microscopy. The cells are 2-3 μm thick.

12. Tissue studies

Taking it one step further, DHM can be used for tomographic measurements (Massatsch et al., 2005) that enable analysis of sub-cellular motion from deep within living tissues (Jeong et al., 2007a), along with valuable 3-D-reconstruction of the object (volumetric visualization) (Jeong et al., 2007b). Sun and colleagues showed that they could follow the movement of blood cells *in vivo* in a tadpole blood vessel (Sun et al., 2008). This shows that DHM can be a useful tool for *in vivo* studies, enabling researchers to directly follow the effect of a treatment on *e.g.* cell morphology or motility and migration. DHM has also been used for studies on spheroids (Yu et al., 2003; Jeong et al., 2007b). Spheroids are clusters of cells that serve as *in vitro* tumor models which can be used for *e.g.* studies of novel anti cancer drugs.

13. Conclusions

DHM is a very versatile technique that can be used for studies of cell types ranging from pollen and protozoa to nerve cells and even tissue. DHM can aid researchers in detecting cell changes in unlabeled cells growing as undisturbed as possible, whether in a cell culture flask or in their usual tissue environment. When DHM is combined with fluorescence microscopy, results concerning cell morphology and/or motility can be combined with a broad variety of fluorescence labeling tools, thus adding extra information to the studies of cell function. Until now, most experiments using DHM have been performed to prove that the technique is useful. Now the time has come to apply the technique on medical and biological research.

14. Acknowledgments

We wish to thank Thomas Deierborg and Amelie Gormand for supplying cells.

15. References

Atienza, J.M.; Yu, N.; Kirstein, S.L.; Xi, B.; Wang, X.; Xu, X. & Abassi,Y.A. (2006). Dynamic and label-free cell-based assays using real-time cell electronic sensing system. *ASSAY and Drug Development Technologies,* 4(5):597-607

Beving, H.; Tornling, G. & Olsson, P. (1991). Increased erythrocyte volume in car repair painters and car mechanics. *British Journal of Industrial Medicine,* 48:499-501

Bernhardt, I.; Ivanova, L.; Langehanenberg, P.; Kemper, B. & von Bally G. (2008). Application of digital holographic microscopy to investigate the sedimentation of intact red blood cells and their interaction with artificial surfaces. *Bioelectrochemistry,* 73:92–96

Boyd, J.M.; Huang, L.; Xie, L.; Moe, B.; Gabos, S. & Li, XF. (2008). A cell-microelectronic sensing technique profiling cytotoxicity of chemicals. *Analytica Chimica Acta,* 615(1):80-87.

Boyden, S. (1962) . The chemotactic effect of mixtures of antibody and antigen on polymorphonuclear leucocytes. *The Journal of Experimental Medicine,* 115:453-466

Buttarello, M. & Plebani, M.D. (2008). Automated Blood Cell Counts. *American Journal of Clinical Pathology,* 130:104-116

Carl, D.; Kemper, B.; Wernicke, G. & von Bally, G. (2004). Parameter-optimized digital holographic microscope for high-resolution living-cell analysis. *Applied Optics,* 43(36):6536-6544

Chalfie, M.; Tu, Y.; Euskirchen, G.; Ward, W. & Prasher, D. (1994). Green fluorescent protein as a marker for gene expression. *Science,* 263 (5148): 802–805

Colomb, T.; Charrirere, F. Kühn, J.; Marquet, P. & Depeursinge, C. (2008). Advantages of digital holographic microscopy for real-time full field absolute phase imaging. In: *Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing XV*. Proceedings of SPIE Volume: 6861, Conchello, J.-A.; Cogswell, C.J.; Wilson,T. & Brown, T.G., (Eds.)

Cuche, E.; Marquet, P. & Depeursinge, C. (1999). Simultaneous amplitude-contrast and quantitative phase-contrast microscopy by numerical reconstruction of Fresnel off-axis holograms. *Applied Optics*, 38, 6994–7001.

De Becker, A.;Van Hummelen, P.; Bakkus, M.; Vande Broek, I.; De Wever, J.; De Waele, M. & Van Riet, I. (2007). Migration of culture-expanded human mesenchymal stem cells through bone marrow endothelium is regulated by matrix metalloproteinase-2 and tissue inhibitor of metalloproteinase-3. *Heamatologica,* 92:440-449

Dubois, F.; Yourassowsky, C.; Monnom, O.; Legros, J.-C.; Debeir, O.; Van Ham, P.; Kiss, R. & Decaestecker, C. (2006). Digital holographic microscopy for the three-dimensional dynamic analysis of in vitro cancer cell migration. *Journal of Biomedical Optics,* 11:054032

El-Shish, Z.; Mölder, A.; Sebesta, M.; Gisselsson, L.; Alm, K. & Gjörloff Wingren, A. (in press). Digital holographic microscopy – innovative and non-destructive analysis of living cells. In: *Microscopy, Science, Technology, Applications and Education,* Méndez-Vilas,A & Díaz, J., (Eds.)

Emery, Y.; Cuche, E.; Colomb, T.; Depeursinge, C.; Rappaz, B.; Marquet, P. & Magistretti, P. (2007). DHM (Digital Holography Microscope) for imaging cells. *Journal of Physics: Conference Series*, 61:1317–1321.

Frigault, M.M.; Lacoste, J.; Swift, J.L. & Brown, C.M. (2009). Live-cell microscopy – tips and tools. *Journal of Cell Science,* 122, 753-767

Gabor, D. (1948). A new microscopic principle. *Nature,* 161:777-778

Gabbiani, G.; Gabbiani, F.; Heimark, R.F. & Schwartz, S.M. (1984). Organization of actin cytoskeleton during early endothelial regeneration *in vitro*. *Journal of Cell Science,* 66:39-50

Garcia-Sucerquia, J.; Xu, W.; Jericho, S.K.; Klages, P.; Jericho, M.H. & Kreuzer, H.J. (2006). Digital in-line holographic microscopy. *Applied Optics,* 45 (5):836-850

Ge, Y.; Deng, T. & Zheng, X. (2009). Dynamic monitoring of changes in endothelial cell-substrate adhesiveness during leukocyte adhesion by microelectrical impedance assay. *Acta Biochimica et Biophysica Sinica,* 41(3):256-262.

Gustafsson, M.; Sebesta, M.; Bengtsson, B.; Pettersson, S.G.; Egelberg, P. & Lenart, T. (2004). High-resolution digital transmission microscopy - a Fourier holography approach. *Optics and Lasers in Engineering*, 41:553-563

Hoebe, R.A.;Van Oven, C.H.; Gadella Jr, T.W.J.; Dhonukshe, P.B.; Van Noorden, C.J.F. & Manders, E.M.M. (2007). Controlled light-exposure microscopy reduces photobleaching and phototoxicity in fluorescence live-cell imaging. *Nature Biotechnology,* 25(2):249-253

Ikeda, T.; Popescu, G.; Dasari, R.R. & Feld, M.S. (2005). Hilbert phase microscopy for investigating fast dynamics in transparent systems. Optics Letters, 30(10):1165-1167

Jeong, K. (2005). Fourier-domain holographic optical coherence imaging of tumor spheroids and mouse eye, *Applied Optics,* 44:1798-1806

Jeong, K.; Turek, J.J. & Nolte, D.D. (2007a). Fourier-domain digital holographic optical coherence imaging of living tissue. *Applied Optics,* 46(22):4999-5008

Jeong, K.; Turek, J.J. & Nolte, D.D. (2007b). Volumetric motility-contrast imaging of tissue response to cytoskeletal anti-cancer drugs. *Optics Express,* 15(21): 14057-14065

Kemmler, M.; Fratz, M.; Giel, D.; Saum, N.; Brandenburg A. & Hoffman C. (2007). Noninvasive time-dependent cytometry monitoring by digital holography. *Journal of Biomedical Optics,* 12(6): 064002

Kemper, B.; Carl, D.; Schnekenburger, J.; Bredebusch, I.; Schäfer, M.; Domschke, W. & von Bally, G. (2006). Investigations on living pancreas tumor cells by digital holographic microscopy. *Journal of Biomedical Optics,* 11:034005.

Kemper, B.; Langehanenberg, P. & von Bally, G. (2007). Digital holographic microscopy. A new method for surface analysis and marker-free dynamic cell imaging. *Biophotonics,* 2:41-44

Kemper, B. & von Bally, G. (2008). Digital holographic microscopy for live cell applications and technical inspection. *Applied Optics,* 47(4): A52-A61

Kemper, B.; Langehanenberg, P.; Vollmer, A.; Ketelhut, S. & von Bally, G. (2009). Label-free 3D migration monitoring of living cells. *Imaging and Microscopy,* 11(4):26-28.

Kemper, B.; Bauwens, A.; Vollmer, A.; Ketelhut, S.; Langehanenberg, P.; Muthing, J.; Karch, H. & von Bally, G. (2010). Label-free quantitative cell division monitoring of

endothelial cells by digital holographic microscopy. *Journal of Biomedical Optics,* 15:036009

Kricka, L.J. & Fortina, P. (2009). Analytical ancestry:"Firsts" in fluorescent labeling of nucleosides, nucleotides and nucleic acids. *Clinical Chemistry*, 55(4):670-683

Kroemer, G.; Galluzzi, L.; Vandenabeele, P.; Abrams, J.; Alnemri, E.S.; Baehrecke, E.H.; Blagosklonny ,M.V.; El-Deiry, W.S.; Golstein, P.; Green, D.R.; Hengartner, M.; Knight, R.A.; Kumar, S.; Lipton, S.A.; Malorni ,W.; Nunez, G.; Peter, M.E.; Tschopp, J.; Yuan, J.; Piacentini, M.; Zhivotovsky, B. & Melino, G.(2009). Classification of cell death: recommendations of the Nomenclature Committee on Cell Death 2009. *Cell Death and Differentiation*, 16:3-11.

Langehanenberg, P.; Lyubomira, I.; Bernhardt, I.; Ketelhut, S.; Vollmer, A.; Georgiev, G.; von Bally, G. & Kemper, B. (2009). Automated three-dimensional tracking of living cells by digital holographic microscopy. *Journal of Biomedical Optics,*14: 014018.

Logg, K.; Bodvard, K.; Blomberg, A. & Mikael, K. (2009). Investigations on light-induced stress in fluorescence microscopy using nuclear localization of the transcriptionfactor Msn2p as a reporter. *FEMS Yeast Research,* 9:875–884

Mann, C.J.; Yu, L. & Kim, M.K. (2006). Movies of cellular and sub-cellular motion by digital holographic microscopy. *Biomedical Engineering Online,* 5:21

Marquet, P.; Rappaz, B. & Magistretti, P.J. (2005). Digital holographic microscopy: a noninvasive contrast imaging technique allowing quantitative visualization of living cells with subwavelength axial accuracy. *Optics Letters,* 30:468-470.

Massatsch, P.; Charriére, F.; Cuche, E.; Marquet, P. & Depeursinge, C.D. (2005). Time-domain optical coherence tomography with digital holographic microscopy. *Applied Optics*, 44(10): 1806-1812

Moon, I. & Javidi, B. (2007). Three-dimensional identification of stem cells by computational holographic imaging. *Journal of the Royal Society Interface,* 4(13):305-313

Moon, I. & Javidi, B. (2008). 3-D visualization and identification of biological microorganisms using partially temporal incoherent light in-line computational holographic imaging. *IEEE Transactions on Medical Imaging,* 27(12): 1782-1790

Mölder, A.; Sebesta, M.; Gustafsson, M.; Gisselsson, L.; Gjörloff-Wingren, A. & Alm, K. (2008). Non-invasive, label-free cell counting and quantitative analysis of adherent cells using digital holography. *Journal of Microscopy,* 232(2):240-247.

Nomarski, G. (1955). Differential microinterferometer with polarized light. *J. Phys Radium,* 16:9-13

Pavillon, N.; Benke, A.; Boss, D.; Moratal, C.; Kühn, J.; Jourdain, P.; Depeursinge, C.; Magistretti, P.J. & Marquet, P. (2010). Cell morphology and intracellular ionic homeostasis explored with a multimodal approach combining epifluorescence and digital holographic microscopy. *Journal of Biophotonics*, 3(7), 432-436

Persson, J.; Mölder, A.; Pettersson, S.-G. & Alm, K. (In press). Cell motility studies using digital holographic microscopy. In: *Microscopy, Science, Technology, Applications and Education*, Méndez-Vilas, A. & Díaz, J., (Eds.)

Petibois, C. (2010). Imaging methods for elemental, chemical, molecular, and morphological analysis of single cells. *Analytical and Bioanalytical Chemistry*, 397:2051-2065

Popescu, G.; Ikeda, T.; Dasari, R.R. & Feld, M.S. (2006). Diffraction phase microscopy for quantifying cell structure and dynamics. *Optics Letters*, 31(6):775-777

Popescu, G.; Park, Y.; Choi, W.; Dasari, R.R.; Feld, M.S. & Badizadegan, K. (2008). Imaging red blood cell dynamics by quantitative phase microscopy. *Blood Cells, Molecules, and Diseases*, 41(1): 10-16

Potten, C.S. & Hendry, J.H. (1973). Clongenic cells and stem cells in epidermis. *International Journal of Radiation Biology,* 24(5):537-540

Rappaz, B.; Marquet, P.; Cuche, E.; Emery, Y.; Depeursinge, C. & Magistretti, PJ. (2005) Measurement of the integral refractive index and dynamic cell morphometry of living cells with digital holographic microscopy. *Optics Express,* 13: 9361-9373

Rappaz, B.; Barbul, A.; Emery, Y.; Korenstein, R.; Depeursinge, C.; Magistretti, P.J. & Marquet, P. (2008). Comparative study of human erythrocytes by digital holographic microscopy, confocal microscopy and impedance volume analyzer. *Cytometry Part A,* 73A:895-903

Rappaz, B.; Cano, E.; Colomb, T.; Kuhn, J.; Simanis, V.; Magistretti, P.J & Marquet, P. (2009a). Noninvasive characterization of the fission yeast cell cycle by monitoring dry mass with digital holographic microscopy. *Journal of Biomedical Optics*, 14(3):034049.

Rappaz, B.; Barbul, A.; Hoffmann, A.; Boss, D.; Korenstein, R.; Depeursinge, C.; Magistretti, P. & Marquet, P. (2009b). Spatial analysis of erythrocyte membrane fluctuations by digital holographic microscopy. *Blood Cells, Molecules and Diseases*, 42(3):228-232.

Reya, T.; Morrison, S.J.; Clarke, M.F. & Weissman, I.L. (2001). Stem cells, cancer, and cancer stem cells. *Nature*, 414:105-111

Schildknecht, S.; Pöltl, D.; Nagel, D.M.; Matt, F.; Scholz, D.; Lotharius, J.; Schmieg, N.; Salvo-Vargas, A. & Leist, M.(2009). Requirement of a dopaminergic neuronal phenotype for toxicity of low concentrations of 1-methyl-4-phenylpyridinium to human cells. *Toxicology and Applied Pharmacology*, 241:23-35

Schnars, U. & Juptner, W. (1994). Direct recording of holograms by a CCD target and numerical reconstruction. *Applied Optics,* 33:179-181

Schnekenburger, J.; Bredebusch, I.; Domschke, W.; Kemper, B.; Langehanenberg, P. & von Bally G. (2007). Digital holographic imaging of dynamic cytoskeleton changes. *Medical Laser Application,* 22:165–172.

Scothorne, R.J. & Tough, J.S. (1952). Histochemical studies of human skin autografts and homografts. *The British Journal of Plastic Surgery,* 5(3):161-170

Sun, H.; Song, B.; Dong, H.; Reid, B.; Player, M.A.; Watson, J. & Zhao, M. (2008). Visualization of fast-moving cells in vivo using digital holographic video microscopy. *Journal of Biomedical Optics,* 13(1):014007

Thomas, E.D.; Lochte, H.L. Jr; Lu, W.C. & Ferrebee, J.W. (1957). Intravenous infusion of bone marrow in patients receiving radiation and chemotherapy. *The New England Journal of Medicine,* 257(11):491-496

Yu, P.; Mustafa, M.; Turek, J.J.; French, P.M.W.; Melloch, M.R. & Nolte, D.D. (2003). Holographic optical coherence imaging of tumor speroids. *Applied Physics Letters*, 83(3):575-577

Yu, L.; Mohanty, S.; Zhang, J.; Genc, S.; Kim, M.K.; Berns, M.W. & Chen, Z. (2009). Digital holographic microscopy for quantitative cell dynamic evaluation during laser microsurgery. *Optics Express,* 17(14):12031-12038.

Zernike, F. (1942). Phase contrast, a new method for the microscopic observation of transparent objects. *Physica,* 9:686-698

Zicha, D.;Dunn, G. A. & Brown, A.F.; (1991) . A new direct-viewing chemotaxis chamber. *Journal of Cell Science,* 99(4):769-775.

Zigmond, S.H. & Hirsch, J.G. (1972). Leukocyte locomotion and chemotaxis: New methods for evaluation, and demonstration, of a cell-derived chemotactic factor. *The Journal of Experimental Medicine,* 137:387-410

12

Fabrication of Two- and Three-Dimensional Photonic Crystals and Photonic Quasi-Crystals by Interference Technique

Ngoc Diep Lai[1], Jian Hung Lin[2], Danh Bich Do[3], Wen Ping Liang[4], Yu Di Huang[5], Tsao Shih Zheng[6], Yi Ya Huang[7], Chia Chen Hsu[8]
[1]*Laboratoire de Photonique Quantique et Moléculaire, UMR CNRS 8537, Ecole Normale Supérieure de Cachan,*
[2,3,5,6,8]*Department of Physics, National Chung Cheng University, Chiayi 621,*
[3]*Department of Physics, Hanoi National University of Education, Hanoi,*
[4,7,8]*Graduate Institute of Opto-Mechatronics, National Chung Cheng University, Chiayi 621,*
[8]*Department of Photonics, National Sun Yat-Sen University, Kaohsiung 804*
[1]*France*
[2,3,4,5,6,7,8]*Taiwan*
[3]*Vietnam*

1. Introduction

Photonic crystals (PCs) are composed of periodic dielectric or metallo-dielectric micro- or nano-structures that affect the propagation of electromagnetic waves in the same way as the periodic potential in a semiconductor crystal affects the electron motion (Joannopoulos et al., 1995; Sakoda, 2001). The wavelength ranges of disallowed propagation of electromagnetic waves of PCs are called photonic band gaps (PBG), which gives rise to distinct optical phenomena such as inhibition of spontaneous emission, high-reflecting omni-directional mirrors and low-threshold PC laser (Yablonovitch, 1987). Besides, the introduction of point or line defects into PCs offers many other potential applications, such as low loss waveguides, cavity resonators, and nanolasers, etc. (Noda et al., 2007; Rinne et al., 2008). The PCs with unique PBG properties therefore can be applied in a wide range of photonic and electronic devices (Inoue & Ohtaka, 2004). The major challenge for PCs study is the fabrication of these structures (Sibilia et al., 2008), with sufficient precision to prevent scattering losses blurring the crystal properties and with processes that can be robustly mass produced. Various techniques have been proposed to fabricate templates for PCs such as self-assembly of colloidal particles (Wong et al., 2003; Wu et al., 2008), holographic lithography (HL) (Berger et al., 1997; Campbell et al., 2000), and direct laser writing (Deubel et al., 2004; Sun et al., 1999), etc. Holographic lithography, in particular, is a very promising and inexpensive technique to fabricate large-area and defect-free PC templates. HL also allows to fabricate structures with unusual high levels of symmetry, called photonic quasi-crystals (PQCs) (Wang et al., 2003),

which recently attract a lot of attention since they have new optical properties that cannot be obtained with conventional PCs.

This chapter presents the fabrication and the characterization of two- and three-dimensional (2D and 3D) PCs and PQCs templates. The fabrication is based on the multiple-exposure of two- and/or three-beam interference patterns. The optical properties of fabricated structures, such as pseudo PBG and diffraction patterns, are characterized and compared with theoretical calculations.

In Section 2, we demonstrate the fabrication of multi-dimensional periodic structures by the use of a new proposed multiple-exposure two-beam interference technique (Lai et al., 2005a). We show that different 1D, 2D and 3D structures could be created by applying multiple exposure of the two-beam interference pattern at appropriate orientations. Various 3D structures with a periodicity as small as 400 nm are experimentally obtained in agreement with theoretical calculations. This fabrication method presents a number of advantages over other holographic techniques based on three- or four-beam interference technique, such as, simple, low cost, and flexible, etc.

In Section 3, we propose various methods to embed desired defects into 2D and 3D periodic structures. Firstly, we combine the interference, presented in Section 2, with mask lithography technique in order to add long defects into 2D structures (Lai et al., 2005b). This combination has advantages such as simple and short fabrication time. Secondly, we demonstrate that a combination of the interference with the multi-photon polymerization direct laser writing permits to introduce a tiny and arbitrary defect into not only 2D but also 3D structures. Moreover, the defects could be patterned precisely into periodic structures with desired position and orientation by employing a double scanning microscopy technique (Lai et al., 2006a).

In Section 4, we investigate theoretically and experimentally the fabrication and the optical property of the QPCs. The fabrication also employs the multiple-exposure idea but it is applied to three-beam interference technique (Lai et al., 2006b). The fabrication of quasi-periodic structures with a rotation symmetry as high as 60-fold is demonstrated and confirms the theoretical calculation (Lai et al., 2007). Moreover, we calculate the optical property of fabricated quasi-periodic structures and demonstrate that PQC possesses an isotropic PBG that could not be achieved with traditional PCs.

In Section 5, we explore diverse non-desired effects existing in above mentioned interference technique, e.g., high absorption and photoacid diffusion or low concentration of the used photoresists, to fabricate 3D controllable thickness and 2D nano-vein structures. Here again, multiple-exposure two-beam interference technique is used for fabrication. The photoresist with high absorption at irradiated wavelength allows us to obtain 3D arbitrary structures by assembling multiple 1D or 2D structures. This technique is rapid and can produce very large structure with no limitation of number of layers, and with variable period and flexible design (Lai et al., 2010). Besides, by combining different kinds of photoresists, such as negative and positive ones, we demonstrate that 2D structures with nano-connection could be fabricated (Lai et al., 2009), which allows one to obtain better PBG, compared to traditional PCs.

Finally, we will make some conclusions of this work and show what this study brings to the fabrication of PCs and PQCs. We will also present some prospects.

2. Fabrication of PCs templates by multiple-exposure two-beam interference

2.1 Theoretical calculations

In this Section, we will demonstrate theoretically how a multi-dimensional interference pattern can be created by using multiple-exposure of two-beam interference pattern. Figure 1(a) shows a typical two-beam interference configuration. We assume that two laser beams of the same monochromatic plane wave source propagate in (x, z)-plane. In the overlapping area of two beams, they interfere and their total intensity is modulated in one-dimensional, as shown in Fig. 1b. If the plane containing two beams is rotated by an angle α around the z-axis (Fig. 1c), the 1D structure remains the same but its orientation changes with respect to α. When combining different 1D structures , which are oriented in different α angles, a 2D structure is obtained. Figure 1d shows an example of a 2D square structure created by two exposures at $\alpha = 0^\circ$ and $\alpha = 90^\circ$, respectively. In order to generate 3D periodic structures, we introduce a second rotation angle β as illustrated in Fig. 1e. The intensity distribution of a two-beam interference pattern in a sample oriented at angles α and β is expressed as

$$I_\alpha = 2E_0^2 \cos^2[kz \sin\theta \sin\beta + k \cos\theta \cos\beta(x \cos\alpha + y \sin\alpha)], \quad (1)$$

where $E_{10} = E_{20} = E_0$ are amplitudes of the electric fields of beams 1 and 2, respectively, k is the wave number, α and β are the rotation angles, and θ is the semi-angle between two beams. With a chosen couple of angles (α, β), the 1D structure could be oriented in any direction in space. By combining multiple 1D structures with appropriate (α, β)-angles, we can fabricate any desired 3D periodic structure. Figure 1f shows as an example for a hexagonal close-packed-like structure structure created by three exposures of a two-beam interference pattern at $(\alpha,\ \beta) = (0^\circ, 30^\circ), (120^\circ, 30^\circ)$ and $(240^\circ, 30^\circ)$, respectively.

Note that this interference technique allows us to create any 2D and 3D structures, which cannot be created by three- or four-beam interference technique. Moreover, the lattice constants of the new created 3D structures are close in three dimensions for any value of θ-angle between two laser beams, that is difficult to be achieved by the commonly used one-exposure multi-beam interference technique (Campbell et al., 2000). The lattice constant of a two-beam interference pattern is determined by $\Lambda = \lambda/2\sin\theta$, where λ is the excitation wavelength.

2.2 Optical arrangement and fabrication process

Figure 2 shows the experimental setup of two-beam interference used to fabricate 1D, 2D and 3D periodic structures. A laser beam was first spatially cleaned and then extended to have a uniform and large profile. A 50/50 beam splitter was used to obtain two laser beams of the same profile, same polarization, and same intensity. Two mirrors were used to reorient the two beams in the position of the sample. These two beams interfered and their total intensity was periodically modulated in 1D in x-direction. The angle between two laser beams is denoted as 2θ and could be easily controlled by rotation of two mirrors. A sample was fixed in a double rotation stage, which could be rotated around the z-axis by an angle α and around the y-axis by an angle β.

Depending on the photoresist used and structure to be fabricated, we have used different kinds of excitation lasers, such He-Cd emitting at 325 nm and at 442 nm or Argon laser emitting at 514 nm. Those lasers are very stable and possess a long coherence length. According to excitation laser wavelength, different types of photoresists, negative and positive, have been used, such as SU8 and JSR (negative photoresists) and AZ4620 and S1818 (positive photoresists).

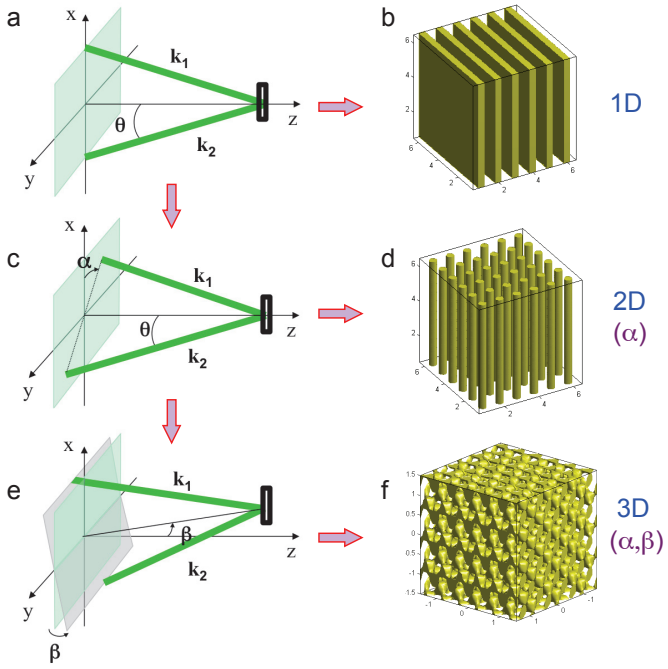


Fig. 1. Reorientation of two-beam interference pattern with multiple exposure allows to create one-, two- and three-dimensional micro and nanostructrures. (c, d) 2D structures are obtained by two or three exposures at different orientation angles α. (e, f) Desired 3D structures are obtained by three or multiple exposures at different orientation angles α and β. Simulation examples of 1D (b), 2D square (d), and 3D hexagonal close-packed-like (f) structures.

The polymeric structure was fabricated following the procedures: i) preparation of thin film sample by spin-coating photoresist on glass substrate and pre-baking to remove the solvent; ii) exposure of interference pattern; and iii) post-baking and developping sample.

2.3 Fabrication results

Figure 3 shows the experimental result of various periodic structures obtained by one exposure, two exposures and three exposures of the two-beam interference pattern. The structures are very uniform for a large area corresponding to the laser beam size of about 1 cm^2. A 1D structure was obtained by one exposure at $\alpha = 0^\circ$ as shown in Fig. 3a. 2D square (Fig. 3b) and hexagonal (Fig. 3c) structures were obtained with two exposures at $\alpha = 0^\circ$ and 90° and $\alpha = 0^\circ$ and 60°, respectively. Note that the hexagonal structure obtained by a double-exposure (Fig. 3c) consists of ellipsoidal dots (or holes). When applying three exposures with symmetrical orientations at $\alpha = 0^\circ$, $\alpha = -60^\circ$ and $\alpha = 60^\circ$, we obtained a 2D hexagonal structure with circular dots (or holes) as shown in Fig. 3d.

These structures were fabricated by using a negative SU8 photoresist (MicroChem. Corp.), with moderate film thickness (1 μm) and a laser beam at $\lambda = 325$ nm. Similar results were also obtained by using positive AZ4620 or S1818 photoresist and a laser beam at $\lambda = 325$ nm

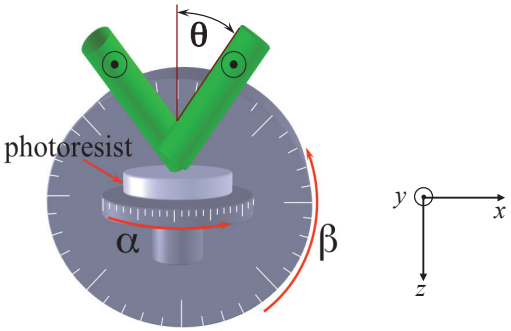


Fig. 2. Optical arrangement of multiple-exposure two-beam interference technique. Two laser beams of the same profile, polarization, and intensity were obtained by a 50/50 beam splitter. Two beams are then combined by using two mirrors, forming an angle 2θ. These beams interfere in the position of sample that is fixed in a double rotation stage: one can rotate around the z-axis, with an angle α, and the other can rotate around the y-axis, with an angle β.

or $\lambda = 442$ nm. Note that the structure obtained by using a negative photoresist is an exact duplication of the interference pattern, while the one obtained by a positive photoresist is an inverted duplication. These results are due to the chemical reaction happening when the resist absorbs the light, for which the negative resist formes a long polymer chain and the positive resist tranforms from polymer chain to momoners. An intermediate situation in which the structure is not a duplication of the interference pattern will be presented in Section 5.3.

Note that using the same photoresist, either positive or negative, we were also able to fabricate 2D periodic structures with either material-dots or air-holes (Lai et al., 2005a), by controlling the exposure-dose, i.e., exposure-time or laser power. It is also important to know that by using this interference technique, one can easily adjust the lattice constant of the fabricated structure. The lattice constant can be as small as the half-wavelength of the excitation laser ($\Lambda_{\min} = \lambda/2$). In this work, we have fabricated periodic structures with lattice constant as small as 400 nm, obtained by using $\theta = 24^\circ$ and $\lambda = 325$ nm.

Figure 4 shows a 3D structure with a period of 400 nm. In oder to fabricate a thick 3D structure, we need to choose a photoresist with low absorption at irradiation wavelength. We have therefore used another negative photoresist, namely JSR (thickness = 6 μm), and an irradiation laser at $\lambda = 325$ nm. To obtain 3D periodic structures, we varied both α and β for different exposures. Many kinds of 3D structures were successfully fabricated by choosing appropriate α and β angles. An example of 3D structure shown in Fig. 4 was obtained by three exposures at $(\alpha, \beta) = (90^\circ, 0^\circ), (0^\circ, 45^\circ)$ and $(180^\circ, 45^\circ)$, respectively. The experimental results show that 3D structures are uniform in large area and agree well with the simulation results. Moreover, comparing the lattice constants in top view and side view of 3D structures, we confirm that they are close in three dimensions as predicted by theory, which is impossible to achieve with four- or five-beam interference technique.

Figure 5a shows another 3D periodic structure of the same period $\Lambda = 400$ nm. This structure, called hexagonal close-packed-like, was obtained by three exposures at $(\alpha, \beta) = (0^\circ, 30^\circ), (120^\circ, 30^\circ)$ and $(240^\circ, 30^\circ)$, respectively. With these 3D structures of small lattice constant, we expect to obtain a PBG in visible range. In order to characterize the optical properties of fabricated 3D structures, we simply sent a white light parallel beam into the

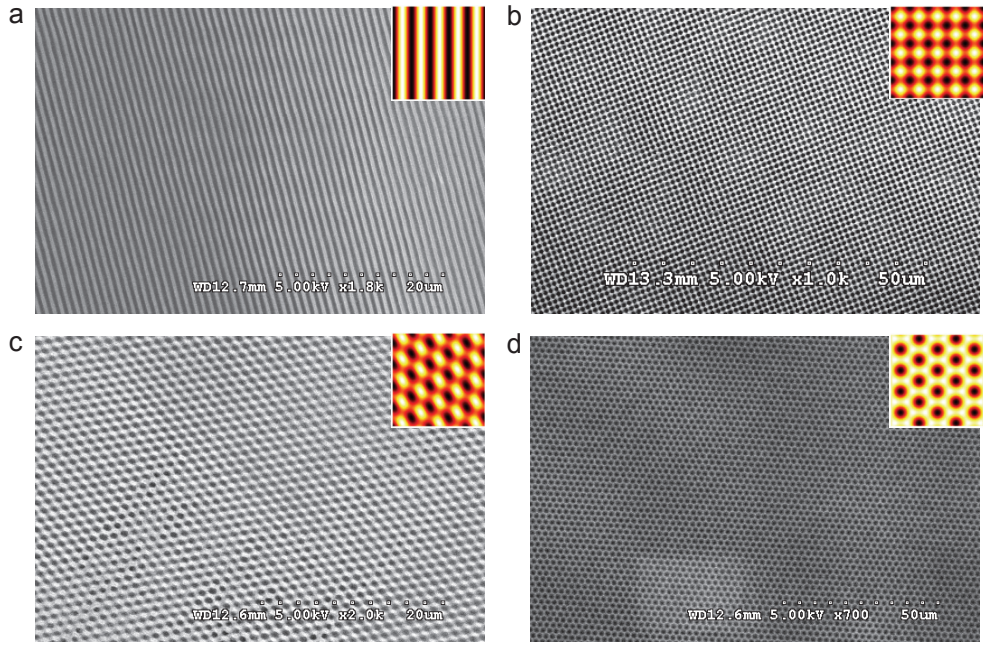


Fig. 3. SEM images of 1D and 2D periodic structures with a lattice $\Lambda = 1.1\ \mu$m. (a) 1D structure obtained by one exposure. (b) 2D square structure obtained by two exposures at $\alpha = 0^\circ$ and $\alpha = 90^\circ$. (c) 2D hexagonal structure obtained by two exposures at $\alpha = 0^\circ$ and $\alpha = 60^\circ$. (d) 2D symmetric hexagonal structure obtained by three exposures at $\alpha = 0^\circ$, $\alpha = -60^\circ$ and $\alpha = 60^\circ$. Insets show simulation results of corresponding structures.

periodic polymeric structure and detected the reflection and transmission spectra. A typical result of spectra measured at normal incidence angle is shown in Fig. 5b. It is clear that there exists a reflection peak and a transmission dip at the same wavelength region, centered at 628 nm. When changing the incidence angle of the white light beam, this reflection peak (or transmission dip) was shifted to shorter wavelength and new reflection peaks (or transmission dips) appeared again in the visible domain. These peaks obviously correspond to a pseudo PBG of our fabricated periodic structures. However, theoretical calculation of the PBG of such new 3D holographic structure is still a challenge in order to confirm this experimental observation.

2.4 Discussion and conclusion

A simple optical interference method for fabricating 2D and 3D periodic structures is theoretically and experimentally demonstrated. This technique presents many advantages over others using multi-beam (three-, four-, or five-beam) interference: i) easy to fabricate desired 2D and 3D structures by simply rotating the sample for multiple exposures, ii) best contrast between the minimal and maximal intensities of interference pattern due to the identical polarization of two laser beams in the interference area, allowing one to fabricate 3D structrure with small lattice constant, iii) new 3D periodic structures having the same period

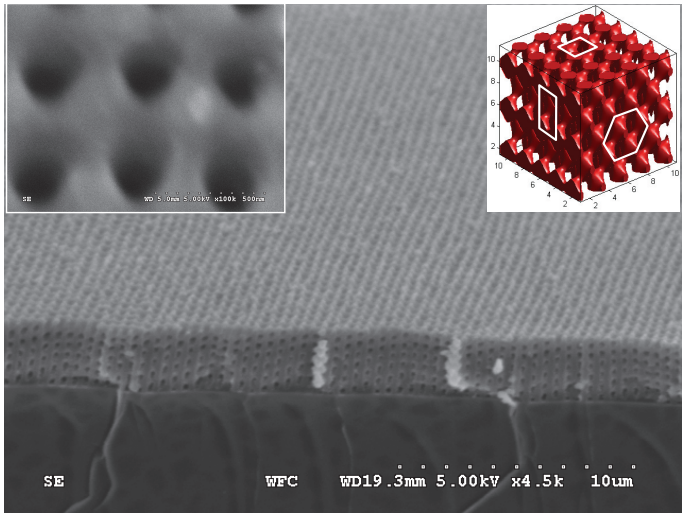


Fig. 4. SEM image of a 3D square/hexagonal periodic structure with a small lattice $\Lambda = 400$ nm, fabricated by three exposures with angles $(\alpha, \beta) = (90°, 0°), (0°, 45°)$ and $(180°, 45°)$, respectively. Left inset shows a zoom in of the top surface of the structure. Right inset shows a theoretical calculation for comparison.

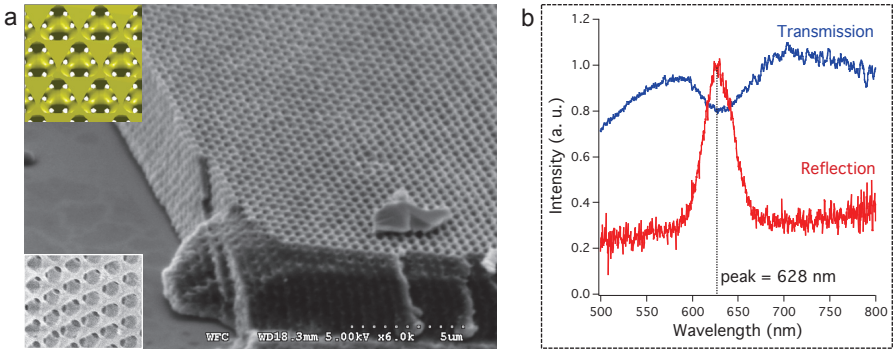


Fig. 5. (a) SEM image of a 3D hexagonal close-packed-like periodic structure with a small lattice $\Lambda = 400$ nm, fabricated by three exposures with angles $(\alpha, \beta) = (0°, 30°), (120°, 30°)$ and $(240°, 30°)$, respectively. Insets show a zoom in of the top surface of the structure and a corresponding simulation result. (b) Transmission and reflection spectra showing a clear photonic bandgap in visible range of the fabricated 3D structure.

in three dimensions, which can't be obtained by one-exposure of multi-beam interference technique.

Furthermore, this two-beam interference technique allows to fabricate not only periodic but also quasi-periodic structures (Gauthier & Ivanov, 2004). These structures recently attract a lot of attention and will be discussed in more detail in Section 4. Besides, we also can combine this interference technique with other fabrication techniques to fabricate large periodic structure containing desired defect, as it will be demonstrated in Section 3.

Finally, it is worth to note that although one theoretically can fabricate a periodic structure with a $\Lambda_{min} = 162.5$ nm (the excitation wavelength is 325 nm), it's still difficult to obtain a good 3D periodic structure with periodicity smaller than 300 nm. This difficulty can be explained by the fact that when the structure lattice is small, the photoacid diffusion of used photoresist (Juodkazis et al., 2005) makes the structure features connecting all together, and they cannot be separated by the developping process. This is true for all optical lithography techniques used to fabricate photoresist-based nanostructures. Indeed, multiphoton polymerization direct laser writing, for example, can create individual spots with a feature size down to 120 nm (Kawata et al., 2001), but structures were connected together when spots were closely created (Park et al., 2005). Improvement of photoresist quality is a challenge in order to fabricate 3D polymeric structures with nano lattice constant.

3. Incorporation of well-defined defects into periodic microstructures

The use of interference technique allows to fabricate large and uniform periodic structures containing no defect. However, there are considerable interests to embed desired defects into 2D and 3D periodic structures. The introduction of point or line defects into PCs offers various potential applications, e.g., low-loss waveguides, cavity resonators and microlasers, etc. (Noda et al., 2007; Rinne et al., 2008).
In this Section, we present different techniques,that can be combined with two-beam interference technique to embed desired defects into 2D and 3D periodic structures.

3.1 Creation of large defect by mask optical lithography

In order to rapidly and simply introduce arbitrary defects into 2D PCs, we have used a combination of HL technique and mask-photolithography technique as an efficient method to produce large-area PC templates with desired long defects. To fabricate periodic structures, we used multiple-exposure of two-beam interference pattern to produce various 2D periodic square or hexagonal structures. The sample was exposed but not developped at this stage. Desired defects were then introduced in these pre-fabricated structures by irradiating the sample with a large and uniform beam of the same laser through a mask in which the designs of desired defects were patterned [Fig. 6a]. The sample was finally developped to obtain a 2D periodic structure containing desired defect.
Note that solid or air-gap type of defect can be obtained depending on the choice of photoresist. For a negative photoresist, a solid type of defect is obtained. On the other hand, for a positive photoresist, an air-gap defect is obtained. Figure 6b shows an example of a 2D periodic square structure containing a line defect. This structure was obtained by using a positive photoresist AZ-4620 (Clariant Corp.). This photoresist absorbed a UV light pattern (interference pattern and mask) and became soluble. After developing process, we obtained a periodic structure made of AZ photoresist embedded with desired air-defects.
Using this combination technique, we can generate whatever types of defect (point defect, waveguide, Mach-Zehnder, etc.) with the size varying from several micrometers to several centimeters into periodic structures. Furthermore, we can control their positions and orientations with respect to periodic structure. Indeed, by using the multiple-exposure of two-beam interference, we were able to control perfectly the orientation of the periodic 2D structures. With well-known orientation of the defect design in the mask, we therefore could embed the defects into the periodic structures in the desired directions. The defect shown in Fig. 6b was created, for example, at 45° with respect to the orientation of the square structure.

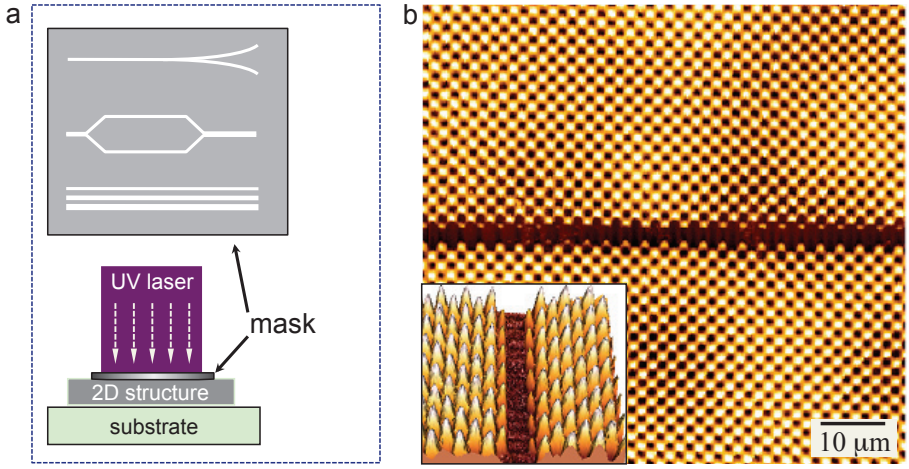


Fig. 6. (a) Creation of large defects into 2D periodic structures by optical mask-lithography. 2D structures are fabricated by two-beam interference technique. Before developping 2D structures, desired defects, which are designed in a mask, are introduced into 2D structures by an uniform beam. The sample is finally developed to obtain the desired structure. (b) AFM images of a 2D square periodic structure embedded with a line defect. The line defect is oriented at 45° with respect to the orientation of the square structure.

3.2 Introduction of small arbitrary defects by multiphoton polymerization technique

In order to create small defects with flexible design, we have combined the two-beam interference technique with the multiphoton polymerization direct laser writing technique (Deubel et al., 2004; Sun et al., 1999). The periodic structure was first created by the multiple-exposure two-beam interference. The tiny desired defect was then introduced into the periodic structure by the direct laser writing process. For the later step, a Ti:Sapphire laser with 830 nm - central wavelength, 100 fs - pulse duration, and 80 MHz - repetition rate was used as the light source to induce multiphoton polymerization effect. The laser beam was tightly focused into the sample by using a high numerical aperture (NA) objective lens (OL, NA = 0.95). The multiphoton absorption could be realized only at focusing region, which then induces the polymerization or depolymerization effect. The diameter of polymerized/depolymerized spot at focal region of OL can be continuously controlled by adjusting the laser power or exposure time. In our case, the smallest diameter of polymerized/depolymerized spots was about 200 nm. The sample could be translated in three dimensions (x, y, z) by a piezoelectric translation (PZT) stage with 10 nm-resolution [Fig. 7a]. The desired defects were designed by a computer software that commanded the voltage applied to the PZT stage.

Here again, either negative or positive photoresist can be used to fabricate various types of defects. Figure 7b shows a 2D periodic hexagonal structure containing some arbitrary "letters" defects, obtained by using a negative SU8 photoresist. The result shows that we can fabricate any periodic structure containing defect as desired. Figure 8 shows the result of a 3D structure containing some “line” defects. The 3D structure was obtained by three exposures at $(\alpha, \beta) = (0°, 30°), (60°, 0°)$ and $(180°, 0°)$. The defects were introduced into the 3D structure at different z-positions as indicated in Fig. 8a. The electronic microscope doesn’t allow to

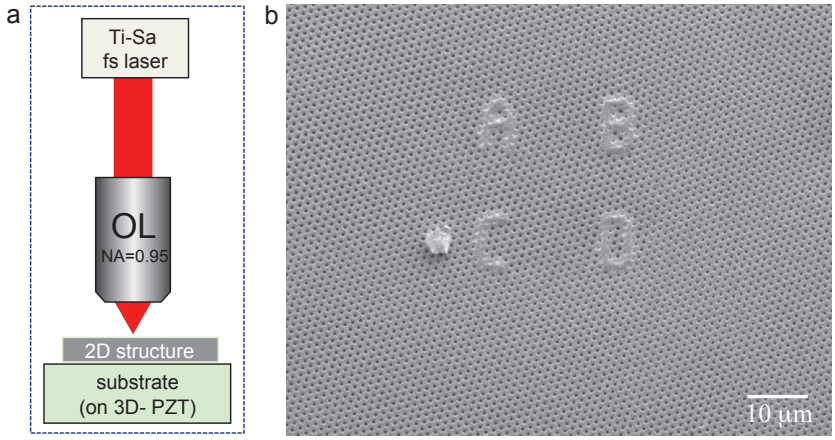


Fig. 7. (a) Simple optical arrangement of a multiphoton laser microfabrication system used to pattern arbitrary defects into periodic structures. 2D structures are fabricated by two-beam interference technique and desired defects are patterned by multiphoton polymerization direct laser writing technique. (b) SEM images of a 2D hexagonal periodic structures ($\Lambda = 1.15\ \mu$m) embedding different letters defects. Note that the small particle near the letter C is a small piece of glass accidentally dropped in the structure during the cutting process of the sample.

observe all defects, in particular those created inside the 3D structure. However, thanks to the fabrication technique and the 3D structure design, we confirm the creation of defects (Z4 for example) at the predicted position and orientation inside 3D periodic structure.
Note that, since the defect was introduced into the pre-fabricated periodic structure before the developping step (the periodic structure information is unknown yet), the position and orientation of fabricated defect with respect to the periodic structures were not totally controlled. It requires a technique to precisely incorporate the desired defects into the periodic structures.

3.3 Precise introduction of desired defects by a double-scanning technique

We have therefore developped a double-step laser scanning technique to embed arbitrary defect in periodic structures with desired position and orientation. We first determined the position and orientation (the mapping step) of periodic structure, and then embedded the desired defects (the fabrication step) into the periodic structure with well-known position and orientation. For both steps, we have used the same direct laser writting system but with different excitation laser powers.
The main concept of the mapping step is based on the shrinkage effect of the negative photoresist. In fact, when excited by two-beam interference pattern, the SU8-momomers moved from the low intensity area to the high intensity area to form linked polymer chains. This induced a 2D surface relief grating (SRG) of the photoresist sample. The SRG image obtained by atomic force microscope, without developping process, is shown in Fig. 9a indicating a modulation of sample surface of about 30 nm. When using a low excitation power of femtosecond laser and detecting the transmission light, we obtained a similar image of the 2D non-developped structure [Fig. 9b]. In this step, the excitation power was chosen to be

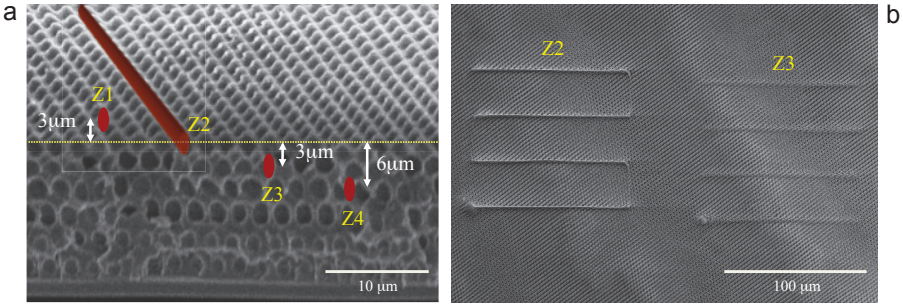


Fig. 8. Introduction of arbitrary defect into a 3D periodic structure by multiphoton polymerization direct laser writing technique. (a) SEM image of the 3D structure and illustration of recording defects at different vertical positions. 3D structures were created by three exposures of two-beam interference pattern. Desired defects were introduced into 3D structures by direct laser writing technique. (b) SEM image of a 3D hexagonal/hexagonal periodic structures ($\Lambda = 2\ \mu$m) embedding several line defects. Only defects on the surface (Z2) and 3 μm below of the structure surface (Z3) can be observed by SEM. The Z1 is not fabricated since the focusing point of the direct laser writing technique was out of the structure surface. The Z4 is not observed because it is too deep inside the sample.

lower than the threshold of multiphoton absorption effect so that the mapping process didn't induce any polymerization effect.
Using the scanned image, the position and orientation of the periodic structures were then determined. By keeping the sample unmoved, the desired defects were patterned in the designed positions and orientations with respect to the periodic structures [Fig. 9b]. For this fabrication step, the excitation laser power was increased to induce the multiphoton absorption polymerization at the focused beam. The defect was then created by scanning the focusing point following the defect design with accurate position and orientation. Figure 9c shows a SEM image of a 2D square periodic structure containing some 90°-bending defects, which agree well with the defects design shown in Fig. 9b.
In a similar way, we expect that this double-scanning technique could be applied to introduce precisely desired defects into 3D periodic structures. In this case, the surface of 3D periodic structure is also modulated due to the shrinkage effect as in the case of 2D periodic structures. By scanning only the structure surface, we can get informations (position and orientation) of the surface of the studied 3D structures. Because no intermediate development is needed before mapping, the fabricated defect-free periodic structures should be exactly identical to its design. The position and orientation of structures below the surface at any z-position could be therefore determined by using the design of fabricated periodic structures. The desired defects can then be introduced at accurate positions and orientations inside 3D periodic structures.

4. Study of 2D and 3D quasi-periodic structures

Multi-beam interference technique (Wang et al., 2003) is a commonly adopted HL method to fabricate a 2D quasi-periodic structure (QPS). This technique however demands a careful alignment of multiple laser beams, which is a challenge for experimental fabrication. In this Section, we demonstrate theoretically and experimentally that combination of three-beam

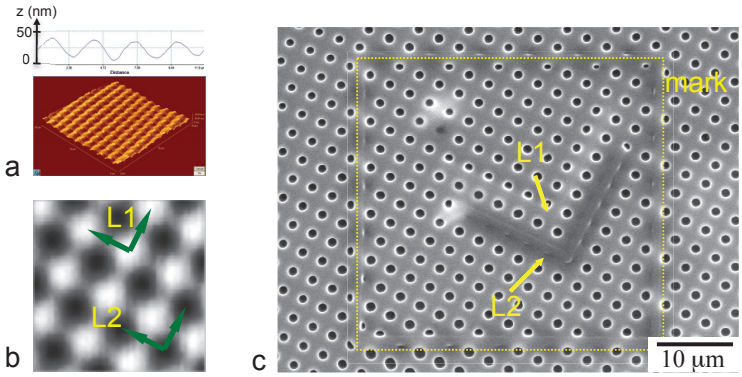


Fig. 9. Precise introduction of desired defect into periodic structure. 2D square structure is obtained by exposing the sample by a double exposure of a two-beam interference pattern at $\alpha = 0^\circ$ and $\alpha = 90^\circ$. (a) AFM image of a structure obtained without developing step. The structure can be seen thanks to the shrinkage effect of negative photoresist. (b) The structure is mapped by using a direct laser writing at low excitation power, and its position and orientation are identified. (c) SEM image of a 2D structure containing well-defined defect. Note that the defect L1 is invisible since it is overlapped with the solid part of the periodic structure. The large square frame around defect is added intentionnally for identification of defect.

interference technique and multiple-exposure idea is a simple and useful method to fabricate various kinds of 2D and 3D QPS with very high rotation symmetry level.

4.1 Optical arragement of multiple-exposure three-beam interference

Figure 10 shows the experimental arrangement used to fabricate 2D and 3D QPSs. To avoid alignment complexity and inaccuracy as well as mechanical vibration instability, a single multi-surface prism was employed. A large and uniform beam is sent in normal incidence upon three surfaces of the prism, noted as A_1, A_2, and A_3 [Fig. 10a]. After passing through the prism, three sub-beams corresponding to three surfaces are obtained and deviated symetrically around the O-axis. Each beam makes an angle θ with respect to the symmetrical O-axis. They are then overlapped at one point in which the total intensity is periodically modulated due to their interference. The interference pattern of three laser beams is a 2D hexagonal structure (6-fold symmetry). A photoresist sample could be placed at the cross-point of three beams and the interference pattern could be recorded to the sample.
In order to fabricate 2D QPSs, the sample holder is aligned perpendicular to O-axis [Fig. 10b] and the sample could be rotated around this axis by an angle α for different exposures. To fabricate 3D QPSs, the sample surface (N-axis) is oriented by an angle β with respect to O-axis [Fig. 10c] and the sample could be rotated around the z-axis by an angle α for different exposures. By this way, we can easily make multi-exposure of the interference pattern to fabricate desired multi-fold symmetry 2D QPSs. Indeed, by making for example a double-exposure and a triple-exposure of the three-beam interference pattern with appropriate rotation angles, we can fabricate 2D and 3D twelve-fold QPSs, respectively. Insets of Fig. 10b and Fig. 10c show simulation results of those QPSs.

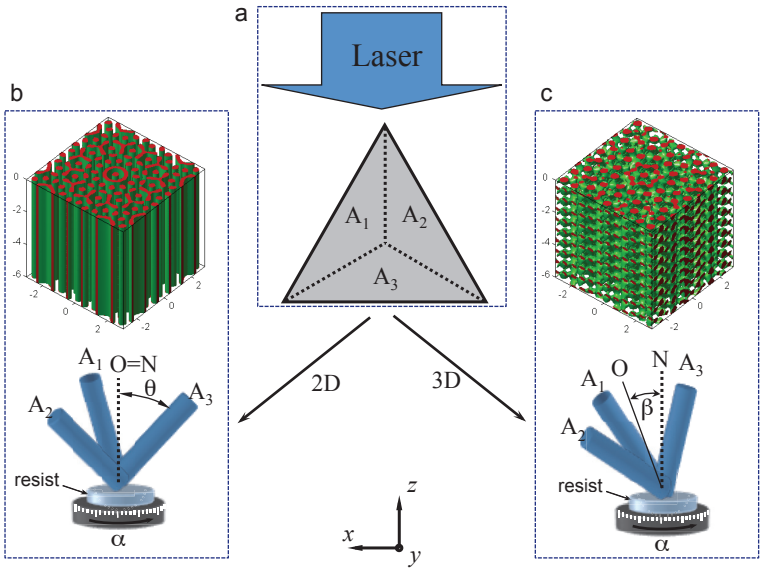


Fig. 10. Fabrication of quasi-periodic structure by multiple-exposure of three-beam interference pattern. (a) A large and uniform laser beam comes into a multi-surface prism along the z-axis and it is divided into three sub-beams corresponding to three surfaces denoted as A_1, A_2, and A_3, respectively. Three beams after passing through the prism changed their directions and overlapped at one point in the z-axis in which a photoresist sample was placed for the fabrication. (b) If the symmetrical axis (O) of the three beams is parrallel to the normal axis (N) of the sample, we can fabricate different 2D quasi-periodic structures by exposing the sample at different rotation angles, α. (c) If O-axis makes an angle, β, with N-axis, we can fabricate different 3D quasi-periodic structures by applying multiple exposures.

4.2 Fabrication of highly rotational symmetric 2D quasi-periodic structures

To fabricate large QPSs, we have used a large and uniform laser beam emitted from a He-Cd laser (λ = 442 nm), which was spatially cleaned and extended (diameter = 2 cm) by a lenses and pinhole system. A mask with three irises was introduced in the laser beam to select three laser sub-beams of the same profile (diameter = 6 mm) and same intensity. These beams are sent to three surfaces, A_1, A_2, and A_3 of the prism to realize the three-beam interference. The sample (AZ4620 photoresist, thickness = 1 μm) is fixed in a holder and can be oriented by angles α and β.

To fabricate 2D QPSs, the β is chosen to be 0°, i.e., the O-axis and the N-axis are parallel. The sample is rotated at different α angles for multiple exposures. Figure 11a shows the experimental results of a 2D QPS obtained with three exposures of the three-beam interference pattern at α = 0°, 20° and 40°, respectively. The structure is quite uniform in a very large area (6 mm × 6 mm, corresponding to the size of the iris). Inset of Fig. 11a shows a zoom in of a particular area of the QPS in which one can clearly see a 18-fold symmetry structure, in agreement with the theoretical calculation. This symmetry is confirmed by measuring the diffraction pattern of fabricated structure. Figure 11b shows the corresponding diffraction pattern of the 2D QPS, which contains a series of circles constituting by eighteen bright spots

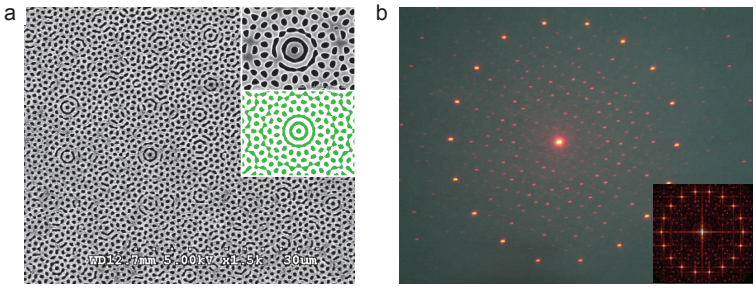


Fig. 11. (a) SEM image of a 18-fold symmetry structure that is fabricated by three exposures of three-beam interference pattern with $(\alpha, \beta) = (0°, 0°), (20°, 0°)$ and $(40°, 0°)$. Insets show theoretically and experimentally a zoom in of a particular area that shows the 18-fold symmetry level of the structure. (b) Experimental diffraction pattern of the corresponding fabricated structure. Inset shows the calculated Fourier transform of the 18-fold symmetry quasi-periodic structure.

around the zero-order diffraction spot. We also calculated the Fourier transform of this 18-fold QPS and found that the spectrum, as shown in inset of Fig. 11b, and its 18-fold rotation symmetry, inherited by the statistical symmetry of the structure, is clearly visible.
We have fabricated by this technique many other QPSs structures with the rotation symmetry level as high as 60-fold. The rule applied for the number of exposure and the symmetry level of QPSs is defined by :

$$N = 6n \tag{2}$$

and

$$\alpha_i = 60(i-1)/n, \tag{3}$$

where N is the rotation symmetry level, α_i is the angle corresponding to the ith-exposure, and n is the number of exposures.

4.3 Fabrication of 12-fold symmetry 3D quasi-periodic structures

For fabrication of 3D QPSs, we reoriented the symmetrical O-axis of three laser beams with respect to the z-axis by an angle β [Fig. 10c] and made three exposures of the three-beam interference pattern with $\alpha_i = 0°$, $20°$ and $40°$, respectively. The β-angle was chosen to be $54.7°$ for all exposures. With this β-angle, the rods of the hexagonal structure obtained by one-exposure are perpendicular to the rods of the hexagonal structure obtained by any other two exposures. The combination of three hexagonal structures oriented in three perpendicular directions resulting in a 12-fold symmetry 3D QPS, as previously seen in Fig. 10c. AZ4620 positive photoresist with a thickness = 12 μm was chosen to fabricate these structures. Figure 12a shows a SEM image of a 3D QPS fabricated by the above mentioned parameters. Quasi-periodic structure can be clearly seen in the top surface. Inset of Fig. 12a shows a side view (tilt angle $60°$) of this 3D QPS in which we can see that the structure is also quasi periodic in the z-direction. Remark that the 12-fold symmetry can be seen in the xy-plane (not xz- neither yz-plane) and other seven planes as illustrated in middle inset of Fig. 12a. These eight planes are oriented in four different directions making the structure quasi-periodic in 3D.
A beautiful diffraction pattern with very bright spots was also obtained showing a 12-fold symmetry level in the xy-plane of the fabricated structure [Fig. 12b]. To experimetally prove

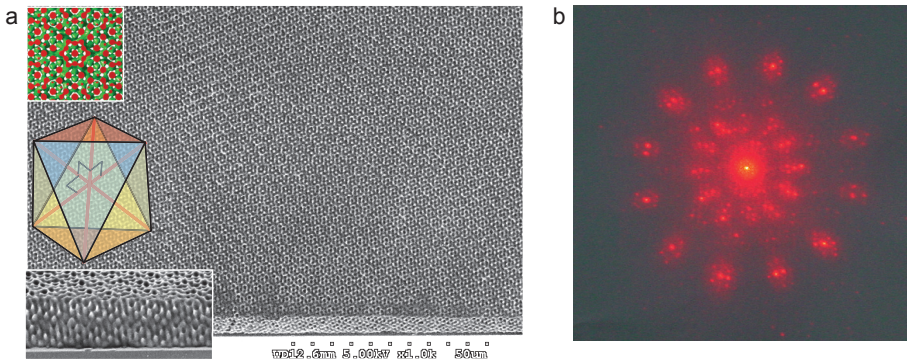


Fig. 12. (a) SEM image of a 3D quasi-periodic structure that is fabricated by three exposures of three-beam interference pattern with $(\alpha, \beta) = (0°, 54.7°), (120°, 54.7°)$ and $(240°, 54.7°)$. The upper inset shows a theoretical calculation of the top surface of the structure and the lower inset shows a zoom in of the side view of the structure. The middle inset illustrates 8 equivalent planes in which we can identify the 12-fold symmetry of structure. (b) Experimental diffraction pattern of the corresponding fabricated structure, showing a complicated distribution of a 3D quasi-periodic structure.

the symmetry level of other seven planes, it is worth to mention that the sample is not thick enough to check the quasi-structure from side view by SEM or to send a laser beam to one side of the sample to observe the diffraction pattern. However, the diffraction pattern shown in Fig. 12b is more complicated than those of 2D QPSs and it should reflect an interesting property of these 3D QPSs.
Furthermore, our calculations predicted that 12-fold symmetry 3D QPS can also be obtained by six exposures with $\beta = 54.7°$ and $\alpha_i = 0°$, $30°$, $120°$, $150°$, $240°$ and $270°$, respectively. In that case, the 12-fold symmetry structure could be seen in sixteen equivalent planes and the PBG of such structure must be more isotropic in 3D.

4.4 Calculation of transmission spectra of 2D quasi-periodic structures

There exists many calculation methods for analyzing PBG properties of PCs. To use those calculation methods, the PC structure needs to be periodic and can be represented by an unit cell structure. However, in the case of QPSs it doesn't exist such unit cell structure, therefore PBG calculation is almost impossible. In some particular cases of QPSs, such as Penrose structure (Zoorob et al., 2000), one needs to use so-called supercell structure to calculate and it takes a long simulation time and large computer memory.
In general, the optical properties of QPSs could be characterized by calculating the transmission or reflection spectra of a light beam applied to the QPSs. For the QPSs fabricated by the multiple-exposure three-beam interference technique, the 2D approximation finite element method (FEM) (Lai et al., 2006b) was used to simulate the transmission spectra for different incident angles. For that we sent a broad-band light source into a finite size quasi-structure at different incidence angles and measured the transmission spectra. The QPSs fabricated by interference technique however contains some particular areas with a rotation symmetry as shown in Fig. 11a. We therefore calculated the transmission spectra using different finite-size structures with and without this symmetrical center. Figures 13 and 14 show the transmission spectra of a 2D 24-fold QPS, for two different cases, with and

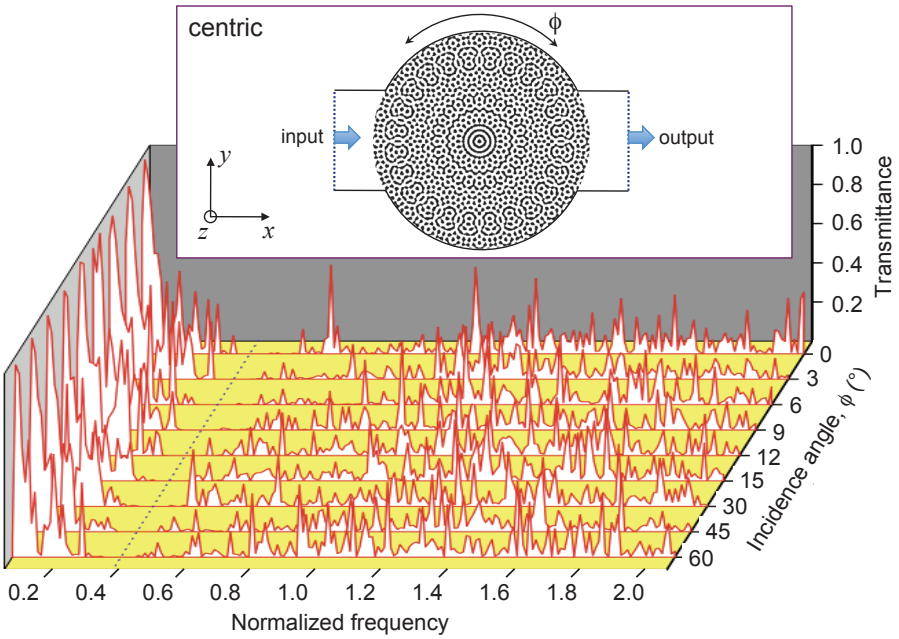


Fig. 13. Simulation result of transmission spectra (TM-mode) of a 2D 24-fold symmetry quasi-periodic structure. The structure is obtained by four exposures of three-beam interference pattern at $(\alpha, \beta) = (0°, 0°), (15°, 0°), (30°, 0°)$ and $(45°, 0°)$. The calculation is realized using a particular area containing a symmetric point, and the transmission spectra are calculated with different incident angles, ϕ, of the light beam. An isotropic photonic bandgap centered at 0.4 is clear observed.

without symmetry center. The 24-fold symmetry structure was obtained by four-exposure of three-beam interference pattern at $\alpha_i = 0°$, 15°, 30° and 45°, respectively. The used finite-size structure is shown in each corresponding calculations result. For all calculations, we assumed that the refractive index of the structure material and the air are 1.6 and 1, respectively. A transmission dip is appeared between 0.35 and 0.45 normalized frequency, which corresponds to the PBG of the structure. Because the lowest symmetry level of these QPSs is six-fold, according to a hexagonal structure obtained by one exposure, it is enough to calculate the transmission spectra only for the incidence angles ranging from 0° to 60°. The calculation result shows that the transmission spectra obtained with different light incidence angles are almost the same, corresponding to an isotropic PBG.

5. Fabrication of 2D nanovein structrures and 3D controllable-thickness structures

Through above presented work, the working principle of HL technique can be sumerized as following: a photoresist is placed in the interference region and the interference pattern is transferred into the photoresist to form a corresponding matter structure. The photoresist is subsequently developed and the periodic or quasi-periodic structures are then obtained. However, in this simple argument, some factors may affect fabricated structures are ignored. Indeed, many researches have shown that final fabricated structures obtained by HL

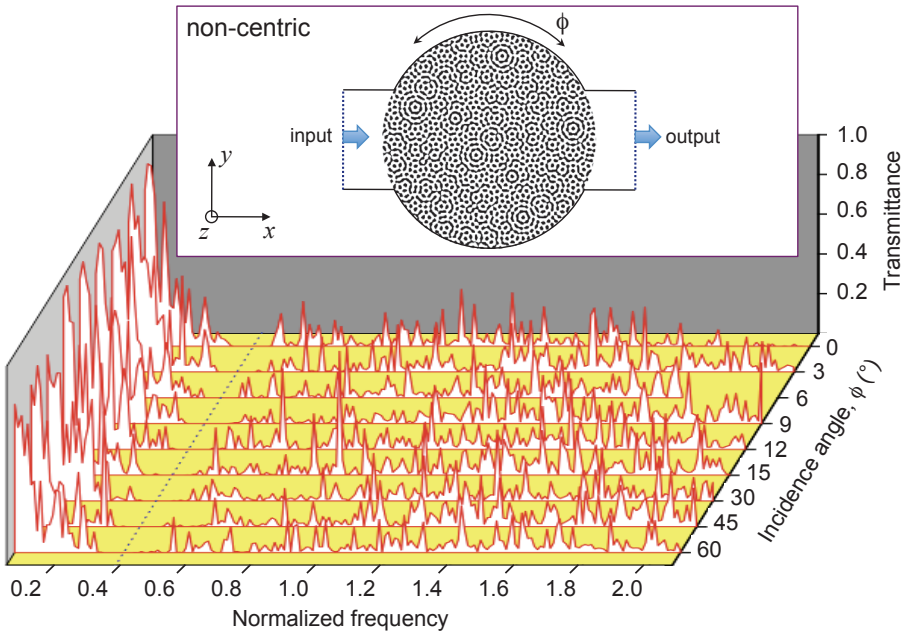


Fig. 14. Simulation result of transmission spectra (TM-mode) of a 2D 24-fold symmetry quasi-periodic structure. The structure is obtained by four exposures of three-beam interference pattern at $(\alpha,\beta) = (0°,0°),(15°,0°),(30°,0°)$ and $(45°,0°)$. The calculation is realized using a random chosen area of the structure and the transmission spectra are calculated with different incident angles, ϕ, of the light beam. An isotropic photonic bandgap centered at 0.4 is clearly observed, similar to the result shown in Fig. 13.

technique strongly depend on the developing process (Mello et al., 1995), the absorption of photoresists (Rumpf et al., 2004), the mass transport effect (Lai et al., 2006a), and the diffusion of photoacid of negative photoresist (Park et al., 2005). In this Section, we present the use of some existing effect in HL technique to obtain new interesting 2D and 3D structures.

5.1 Recording an interference pattern on a pure photoresist: The absorption effect

The first inherent effect regarding the matter used to record the interference pattern is the absorption. Due to this effect, the exposure dosage decreases along the light propagation, resulting in a non-uniform recorded structure. For photoresist whose absorption coefficient at the excitation wavelength is high, the 2D structures obtained by interference technique will consist of regular material-rods with very limited height and the fabrication of 3D structures by this method is almost impossible.

We consider for example the use of two-beam interference technique for fabrication of material structures. Including the absorption effect of the recorded material, the intensity distribution of a two-beam interference pattern in a sample oriented at angles α and β is expressed as

$$I_\alpha = 2E_0^2 \exp[-2\sigma z/\cos(\theta)]\cos^2[kz\sin\theta\sin\beta + k\cos\theta\cos\beta(x\cos\alpha + y\sin\alpha)], \qquad (4)$$

where σ is the absorption coefficient of the photoresist at the excitation wavelength.

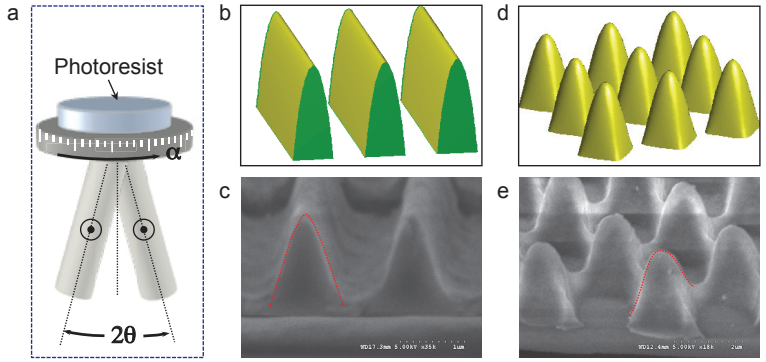


Fig. 15. (a) Experimental arrangement of two-beam interference used for testing the high absorption effect of photoresist. Sample could be rotated by an angle α between different exposures. (b-e) Theoretical calculations and experimental results of 1D (b,c) and 2D (d,e) structures, obtained by using a negative SU8 photoresist and an irradiation light at 325 nm. The structure height (or sample thickness) is limited to about 1.5 μm due to the high absorption coefficient of SU8 at 325 nm-wavelength.

Figures 15b and d show the theoretical calculation of the iso-intensity distribution of a two-beam interference pattern obtained with one-exposure (1D) and double-exposure (2D square), respectively. In this case, we used $\sigma = 1656\ \mathrm{cm}^{-1}$ at λ = 325 nm, according to the absorption coefficient, measured experimentally, of the SU8 commercial photoresist (MicroChem Corp.). Due to the high absorption effect of SU8 at the chosen wavelength, the shape of the 1D and 2D structures is different from those obtained without the absorption effect (Lai et al., 2005a;b). The 2D structure for example no longer contain the long cylinders but cylinders-like or cones instead.

The 2D simulated structure resembles the well-known Moth's eye structures (Baker, 1999; Clapham & Hutley, 1973), in which the effective index is increased continuously from 1 to the index of substrate. This kind of structures can be used as an efficient antireflective surface. The two-beam interference therefore can be an useful method to fabricate such kind of photoresist-based antireflective structures with very large area and with adjustable lattice constant and configuration (hexagonal, square, or other).

In order to observe the shape of the fabricated structure, we employed the fabrication setup shown in Fig. 15a, for that the light beam is applied in the back side of the sample, which allows to keep the structure on the substrate after fabrication. The sample is chosen to be very thick (thickness = 10 μm) to avoid the effect of film thickness on the shape of structure. Figures 15c,e show the structures obtained with one-exposure and double-exposure, respectively, which are in agreement with the theoretical predictions. As consequence of the high absorption of the used photoresist, the structure was formed for a limited height of about 1.5 μm and the surface of the cylinder-like was no longer flat instead of a curved shape surface was formed. A Moth's eye structure can be easily obtained by using a simple interference technique.

Note that the combination of high absorption effect and thin film sample (thickness of about 1.5 μm) allows to obtain neither a 2D perfect cylindrical structure nor Moth's eye structure but a microlens array structure. In addition, the 1D (or 2D) structure could be tilted in any direction by rotating the sample by an angle β, as shown in Section 2, resulting in different

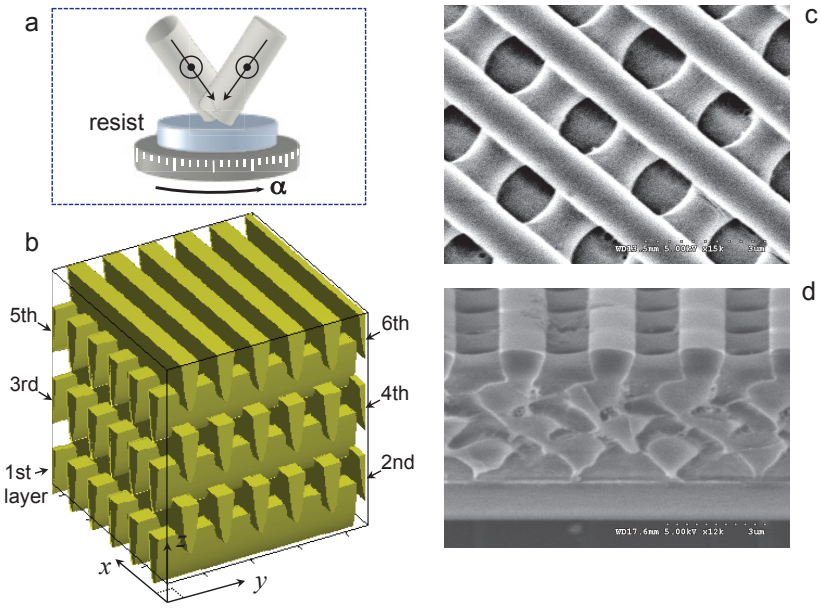


Fig. 16. (a) Experimental arrangement of two-beam interference used for fabrication of desired 3D structure with controllable thickness. (b) Simulation result of a woodpile-like 3D structure obtained by assembling multiple 1D structures. (c) and (d) Top view and side view of an example of a 3D structure consisting of 8 layers, alternatively fabricated in two orthogonal directions. Fabrication was realized by using a negative SU8 photoresist and two-beam interference at 325 nm-wavelength.

kinds of 1D grating such as Blaze-like gratings. Furthermore, 1D or 2D structures obtained by using high absorption effect could be assembled layer by layer to create various 3D structures with no limitation of number of layers.

5.2 Fabrication of desired 3D structures by holographic assembly technique

We have developped a new technique to fabricate 3D desired structure. This technique is based on the high absorption effect and the multi-exposure two-beam interference technique. In fact, under one-exposure, the combination of these two effects allows to obtain a 1D structure with a limited film thickness. By assembling multiple structured layers, layer-by-layer, we then obtained a 3D structures as desired. Each structural layer was obtained by: i) spin-coating the thin film, ii) recording the structure, and iii) post-exposure-baking the sample. Note that the fabrication of top layer doesn't affect the bottom previously recorded layer thanks to the high absorption effect. The desired 3D structure was finally obtained with only one developing process after finishing all fabrications. An example of a simulated woodpile-like 3D structure using a commercial SU8 photoresist exposed at 325 nm-wavelength is shown in Fig. 16b.

Woodpile structure is a simple but very useful lattice with a full PBG (Ozbay et al., 1994). To fabricate such structure, we assembled multiple layers of 1D periodic structure fabricated by one-exposure of two-beam interference pattern. The orientation of 1D structures in two successive layers is changed by 90°, by rotating the sample for two different exposures

(α-angle = 0° and 90°). Figure 16c and d show the top view and side view of a 3D structure obtained by the assembly of 8 layers. The fabricated structures are very uniform for large area, and the structure of each layer is well formed and has no effect on other layers.
We also successfully fabricated other types of 3D structures by rotating the sample by 60° for fabrication of each layer. Similarly, various 3D structures, such as simple cubic or body-centered cubic structures, have been fabricated by assembling multiple 2D structural layers, which are obtained by applying a double-exposure of two-beam interference pattern for each layer. By using this technique, we can therefore assembly multiple 1D and 2D structures, with hexagonal forms or hexagonal and square forms, to obtain many other kinds of 3D structures. It is also possible to assembly multiple 1D structures with different lattice constants in order to obtain bi-photonic crystals with multiple PBG. Furthermore, since the fabrication of the structure in each layer is independent and controlable, we can therefore introduce any kind of defects into desired layer, by using for example the multiphoton polymerization direct laser writing technique or mask-lithography, as presented in Section 3.

5.3 Fabrication of nanovein structures and theoretical analysis of its photonic bandgap

Recently, It has been demonstrated (Park et al., 2005) that the generation and diffusion of photoacid, which acts as a cross-linking agent, depend on the fabrication conditions such as exposure intensity, exposure time, post-exposure bake (PEB), etc. During exposure process, photoinitiator molecules release photoacids in light-exposed regions, and the PEB process accelerates the diffusion of photoacids and induces cationic polymerization of epoxy groups in a negative photoresist. Obviously, with high concentration of photoinitiator molecules and high exposure dosage, it is easy to create a dense polymerized pattern, corresponding to a large size solid structure. Decreasing the concentration of photoinitiator in the photoresist is a simple way to reduce the size of the solid structure. We have developped an original way to fabricate nanostructures by using a mixing of a negative and a positive photoresists. The addition of positive photoresist can reduce the concentration of negative photoresist and thus slow down its polymerization speed.
We mixed S1818, a kind of positive photoresist (Shipley), with SU8 negative photoresist. The fabrication process of using this kind of mixed photoresist is similar to that of pure SU8. The mixing was exposed by a two-beam interference pattern, using a He-Cd laser at α = 325 nm. As shown in Fig. 17a, when exposed by the interference pattern, S1818 polymers become monomers and move to the low intensity area while SU8 monomers move to the high intensity area and become polymers. After exposure, the sample was PEB and then developed by Acetone. The developer washed out all S1818 photoresist and SU8 unpolymerized photoresist. The final result was a solid polymerized SU8 structure. We found that the fabricated structures were solid and smooth, just as those obtained by pure SU8 resist, but the exposure dosage required for obtaining solid structure was much higher than that of pure SU8. Indeed, depending on the volume ratio (VR = ..., 2, 1, 0.5, ...) between SU8 and S1818 photoresists, the exposure time could be varied from ten seconds to few minutes. VR = 2 was an optimum ratio to achieve good structures. The increase of the exposure time required to obtain solid structure is due to the decrease of concentration of SU8 monomers and photoinitiator.
Figures 17b and c show some new structures fabricated by this photoresist mixing and by multiple-exposure two-beam interference technique. The structures are constituted by square [Fig. 17b] or rectangular [Fig. 17c] rods connecting by very narrow veins ($<$ 100 nm). They are totally different to those obtained by pure SU8 photoresist, and do not agree with the

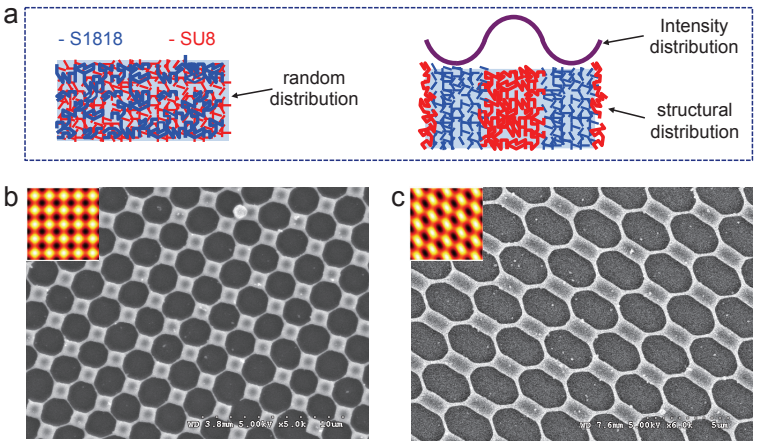


Fig. 17. (a) Illustration of the molecular distribution of the SU8/S1818 mixing. Before exposure, S1818 consists of oligomers and polymers (positive photoresist) and SU8 consists of monomers (negative photoresist). After exposed by the interference pattern, S1818 polymers become monomers and move to the low intensity area while SU8 monomers move to the high intensity area and become polymers. (b) and (c) SEM images of 2D periodic structures obtained with a double-exposure of two-beam interference pattern at different directions. The square or rectangular blocks are connected with narrow veins, thanks to the low concentration of SU8 monomers and the movement of both S1818 and SU8 photoresists. Insets illustrate the simulated structures, same as those shown in Fig. 3, for each experimental result.

theoretical calculations shown in insets of Figs. 17b,c. The result should be a consequence of low concentration of photoinitiator and SU8 monomers. The narrow veins were resulted from high irradiation dosage of each individual exposure (1D) and the square/rectangular rods were due to high irradiation dosage of the combination of two exposures (cross-points of two 1D structures).

This experimental result breaks the well-known HL rule, which usually states that the recorded photoresist structure is a duplication or inverted-duplication of the interference pattern. The result indicates that we need to use the final fabricated structure, not the interference pattern, to calculate optical properties of a PC. Moreover, we show in Fig. 18 that the novel nanovein structures are very useful for PCs applications.

As just mentioned, it is important to use the final fabricated structure as model to calculate the PBG of corresponding PC. Insets of Fig. 18d show a zoom in SEM image of a square unit of the fabricated structure and a corresponding model structure used for the simulation of PBG properties, respectively. The band structures of PCs were calculated by the standard plane-wave expansion method. We first assumed that the PC is constituted of GaAs (dielectric constant ϵ = 11.8) square rods (rod width R = 0.8a, a is the lattice constant) connected by nanoveins (vein diameter D = 0.04a). It is well known that there is no complete PBG for a standard square 2D PC made by GaAs. However, a complete PBG with a high gap ratio ($\Delta\omega/\omega$) was obtained using this novel square structure. Naturally, this result can be explained by the fact that when the dielectric rods are connected by narrow veins, we achieve a compromise between TE and TM modes, and thus a complete PBG. This argument is similar

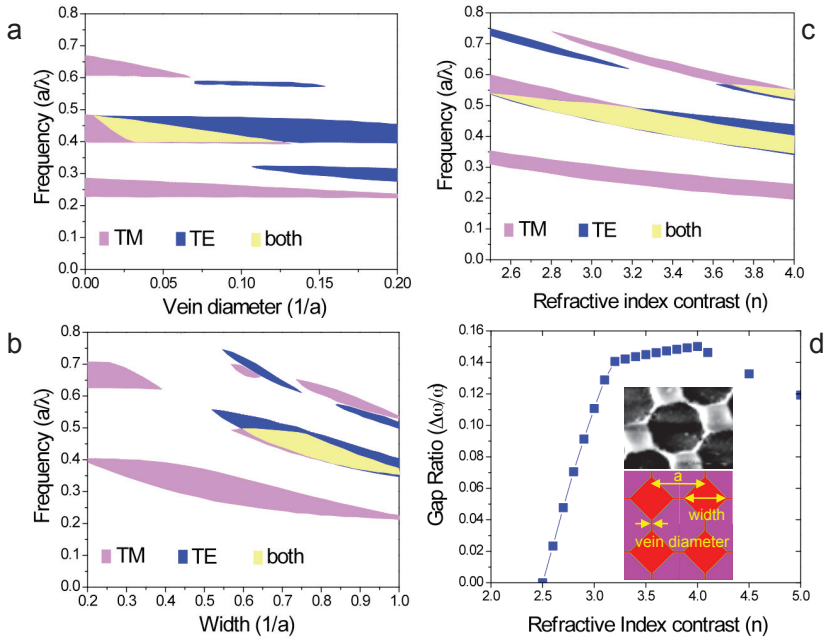


Fig. 18. Simulation results of photonic bandgap (PBG) of a square structure consisting of square rods connecting by nanoveins. (a) Gap map as a function of vein diameter (D), with square rod width R = 0.8a. (b) Gap map as a function of square rod width (R), with vein diameter D = 0.04a. In (a) and (b), the dielectric constant is assumed to be 11.8 and a is the lattice constance of structure. (c) Gap map as a function of the refractive index contrast, n. Vein diameter and square rod width are chosen to be D = 0.04a and R = 0.8a, respectively. (d) Dependence of the normalized absolute PBG ($\Delta\omega/\omega$) on the refractive index contrast. Insets show a zoom in of SEM image and a square unit used for simulations.

to that of the well-known honeycomb structure (Fu et al., 2005). Figures 18a and b show the photonic gap map as a function of R and D (ϵ = 11.8). We can see that a complete PBG can be obtained with a wide range of R (0.6a to 1a) and D (0.005a to 0.125a). The maximum gap ratio was achieved with R = 0.75a and D = 0.03a. Note that the optimum parameters are slightly different depending on the dielectric constant of the material. By scanning all possible parameters, we found that the gap ratio can reach 16% with following optimum parameters: ϵ = 16 (Germanium), R = 0.74a, and D = 0.035a. Figure 18c shows the gap map as a function of the refractive index contrast (n) when R = 0.8a and D = 0.04a. The minimum index contrast required for the TE and TM modes to overlap is 2.529 (ϵ = 6.4). Figure 18d shows the dependence of gap ratio of the absolute PBG on the refractive index contrast. The gap ratio increases rapidly when the index contrast varies from 2.53 to 3.2 and it then slightly increases until when n reaches 4. These results are interesting because the new fabricated structure allows to achieve a complete PBG, which is impossible with the traditional square structure.

5.4 Discussions and Conclusions

Interference is an ideal method to fabricate periodic multi-dimensional structures. In this Section, we recognized that there is many influences of fabrication parameters onto the recorded photoresist structures. Those effects traditionally limit the fabrication of desired structures. However, we have studied and used them as new methods to overcome the limitation and create other new and interesting structures, which cannot be obtained by traditional method.

Besides, we have also studied other influences such as mass transport or developping effects. Those effects allow to fabricate for example polymeric microlens arrays (MLAs). Indeed, we have demonstrated that 2D MLAs can be easily fabricated by multi-exposure two-beam interference technique based on the mass transport effect, i.e., diffusion of monomers from unexposed to exposed areas, of SU8 negative photoresist (Wu et al., 2009). The SU8-based MLAs were obtained without transfering to other materials. When using AZ positive photoresist, we have demonstrated a promising method to obtain plastic MLAs with shape- and structure-controlled, by using the combination of double-exposure two-beam interference and plastic replication techniques. Thanks to the developing effect of the positive photoresist, fabricated structures consisting of parabolical, elliptical or hemispherical-shaped concave holes were obtained. These structures were then transfered to plastic MLAs by employing replication and embossing techniques. The fabricated MLAs, with different structures and shapes, can be useful for different applications in optical and display systems such as correction of laser beam profile, aberration, etc.

6. Conclusion

Realization of photonic bandgap materials (photonic crystals) is in high demand for the use in many applications in opto-electronics and photonic domains. In this chapter, simple fabrication techniques based on multiple-exposure of two- or three-beam interference pattern was theoretically and experimentally demonstrated. These techniques have been applied to different kinds of negative and positive photoresists. Various 1D, 2D and 3D periodic and quasi-periodic structures with different configurations were obtained. The structures are perfectly uniform for area as large as 1 cm^2 and with a lattice constant as small as 400 nm. The photonic bandgap of 3D periodic structures is experimentally demonstrated in visible range, which is very promissing for many applications. Quasi-periodic structures with very high rotation symmetry level is demonstrated. Its photonic bandgap is theoretically demonstrated to be very isotropic, which cannot be obtained by standard photonic crystal. The experimental results are in well agreement with the theoretical calculations.

Besides, different combination techniques have been proposed in order to embed desired defects, such as waveguide, cavity, etc., into periodic structures. Each method presents its advantage, and can be chosen for the desired applications. Mask-lithography allows to rapidly introduce large defects while multiphoton polymerization direct laser writing creats tiny and arbitrary defects. Moreover, the defect were demonstrated to be embedded in correct position and orientation inside the periodic 2D and 3D structures.

Furthermore, different effects that influence the photoresist-based structures fabricated by interference technique were studied in detail. The high absorption effect limits the height of 2D structures and the thickness of 3D structure. But this effect is demonstrated to be useful for fabrication of other special structures such as moth-eye anti-reflective structure or 3D assembled structure. The control of the monomers and photoinitiators concentration of the used photoresist by mixing two kinds of photoresists allows to obtain new periodic

structures consiting of square or rectangular rods connecting by nanoveins. These structures present complete absolute photonic bandgap at low refractive index contrast, which cannot be obtained by standard photonic crystals.
These studies are very important and useful for the photonic crystal community. It provides a rapid and inexpensive fabrication of large and controllable structures. New structures fabricated by these methods show new optical properties and open the ways for interesting applications.

7. References

Baker K. M. (1999). Highly corrected close-packed microlens arrays and moth-eye structuring on curved surfaces. *Appl. Opt.*, Vol. 38 (January 2009) 352-356, ISSN 1559-128X.

Berger, V.; Gauthier-Lafaye, O. & Costard E. (1997). Photonic band gaps and holography. *J. Appl. Phys.*, Vol. 82 (July 1997) 60-64, ISSN 0021-8979.

Campbell, M.; Sharp, D. N.; Harrison, M. T.; Denning, R. G. & Turberfield A. J. (2000). Fabrication of photonic crystals for the visible spectrum by holographic lithography. *Nature*, Vol. 404 (March 2000) 53-56, ISSN 0028-0836.

Clapham, P. B. & Hutley, M. C. (1973). Reduction of lens reflection by the Moth eye principle. *Nature*, Vol. 244 (August 1973) 281-282, ISSN 0028-0836.

Deubel, M.; Freymann, G. V.; M. Wegener, M.; Pereira, S.; Busch, K. & Soukoulis, C. M. (2004). Direct laser writing of three-dimensionalphotonic-crystal templates for telecommunications. *Nature Mater.*, Vol. 3 (July 2004) 444-447, ISSN 1476-1122.

Fu, H. K.; Chen, Y. F.; Chern, R. L. & Chang, C. C. (2005). onnected hexagonal photonic crystals with largest full band gap. *Opt. Express*, Vol. 13 (October 2005) 7854-7860, ISSN 1094-4087.

Gauthier, R. & Ivanov, A. (2004). Production of quasi-crystal template patterns using a dual beam multiple exposure technique. *Opt. Express*, Vol. 12 (March 2004) 7832-7841, ISSN 1094-4087.

Inoue, I. & Ohtaka, K. (2004). *Photonic crystals: physics, fabrication and applications*, Springer, Berlin.

Joannopoulos, J. D.; Meade R. D. & Winn, J. N. (1995). *Photonic crystals: molding the flow of light*, Princeton University Press, Princeton.

Juodkazis, S.; Mizeikis, V.; Seet, K. K..; Miwa, M. & Misawa, H. (2005). Two-photon lithography of nanorods in SU-8 photoresist. *Nanotechnology*, Vol. 16 (April 2005) 846-849, ISSN 0957-4484.

Kawata, S.; Sun, H.-B.; Tanaka, T. & Takada, K. (2001). Finer features for functional microdevices. *Nature*, Vol. 412 (August 2001) 697-698, ISSN 0028-0836.

Lai, N. D.; Liang, W. P.; Lin, J. H.; Hsu, C. C. & Lin, C. H. (2005). Fabrication of two- and three-dimensional periodic structures by multi-exposure of two-beam interference technique. *Opt. Express*, Vol. 13 (November 2005) 9605-9611, ISSN 1094-4087.

Lai, N. D.; Liang, W. P.; Lin, J. H. & Hsu, C. C. (2005). Rapid fabrication of large-area periodic structures containing well-defined defects by combining holography and mask techniques. *Opt. Express*, Vol. 13 (July 2005) 5331-5337, ISSN 1094-4087.

Lai, N. D.; Lin, J. H.; Liang, W. P.; Hsu, C. C. & Lin, C. H. (2006). Precisely introducing defects into periodic structures by using a double-step laser scanning technique. *Appl. Opt.*, Vol. 45 (August 2006) 5777-5782, ISSN 1559-128X.

Lai, N. D.; Lin, J. H.; Huang, Y. Y. & Hsu, C. C. (2006). Fabrication of two- and three-dimensional quasi-periodic structures with 12-fold symmetry by interference technique. *Opt. Express*, Vol. 14 (October 2006) 10746-10752, ISSN 1094-4087.

Lai, N. D.; Lin, J. H. & Hsu, C. C. (2007). Realization of highly rotational symmetric quasi-periodic structures by multi-exposure of three-beam interference technique. *Appl. Opt.*, Vol. 46 (August 2007) 5645-5648, ISSN 1559-128X.

Lai, N. D.; Huang, Y. D.; Do, D. B.; Lin, J. H. & Hsu, C. C. (2009). Fabrication of periodic nanovein structures by holography lithography technique. *Opt. Express*, Vol. 17 (March 2009) 3362-3369, ISSN 1094-4087.

Lai, N. D.; Zheng, T. S.; Do, D. B.; Lin, J. H. & Hsu, C. C. (2010). Fabrication of desired three-dimensional structures by holographic assembly technique. *Appl. Phys. A*, Vol. 100 (July 2010) 171-175, ISSN 0947-8396.

Mello, B. A.; Costa, I. F.; Lima, C. R. A. & Cescato, L. (1995). Developed profile of holographically exposed photoresist gratings. *Appl. Opt.*, Vol. 34 (Fabruary 1995) 597-603, ISSN 1559-128X.

Noda, S; Fujita M. & Asano T. (2007). Spontaneous-emission control by photonic crystals and nanocavities. *Nature Photon.*, Vol. 1 (August 2007) 449-458, ISSN 1749-4885.

Park, S. H.; Lim, T. W.; Yang, D. Y.; Cho, N. C. & Lee, K. S. (2006). Fabrication of a bunch of sub-30-nm nanofibers inside microchannels using photopolymerization via a long exposure technique. *Appl. Phys. Lett.*, Vol. 89 (October 2006) 173133, ISSN 0003-6951.

Ozbay, E.; Abeyta, A.; Tuttle, G.; Tringides, M.; Biswas, R.; Soukoulis, C.; Chan, C. & Ho, K. (1994). Measurement of a three-dimensional photonic band gap in a crystal structure made of dielectric rods. *Phys. Rev. B*, Vol. 50 (July 1994) 1945-1948, ISSN 1098-0121.

Rinne, S. A.; Garcia-Santamaria, F. & Braun, P. V. (2008). Embedded cavities and waveguides in three-dimensional silicon photonic crystals. *Nature Photon.*, Vol. 2 (December 2007) 52-56, ISSN 1749-4885.

Rumpf, R. C. & Johnson, E. G. (2004). Fully three-dimensional modeling of the fabrication and behavior of photonic crystals formed by holographic lithography. *J. Opt. Am. Soc. A*, Vol. 21 (September 2004) 1703-1713, ISSN 1084-7529.

Sakoda, K. (2001). *Optical properties of photonic crystals*, Springer, Berlin.

Sibilia, C.; Benson, T. M.; Marciniak, M. & Szoplik, T. (Eds.) (2008). *Photonic crystals: physics and technology*, Springer, ISBN 978-88-470-0843-4, Berlin.

Sun, H. B.; Matsuo, S. & Misawa, H. (1999). Three-dimensional photonic crystal structures achieved with two-photon-absorption photopolymerization of resin. *Appl. Phys. Lett.*, Vol. 74 (February 1999) 786-788, ISSN 0003-6951.

Yablonovitch, E. (1987). Inhibited spontaneous emission in solid-state physics and electronics. *Phys. Rev. Lett.*, Vol. 58 (May 1987) 2059-2062, ISSN 0031-9007.

Wang, X.; Ng, C. Y.; Tam, W. Y.; Chan, C. T. & Sheng, P. (2003). Large-area two-dimensional mesoscale quasi-crystals. *Adv. Mater.*, Vol. 15 (September 2003) 1526-1528, ISSN 0935-9648.

Wong, S.; Kitaev, V. & Ozin, G. A. (2003). Colloidal crystal films: advances in universality and perfection. *J. Am. Chem. Soc.*, Vol. 125 (Novembre 2003) 15589-15598, ISSN 0002-7863.

Wu, C. Y.; Lai, N. D. & Hsu, C. C. (2008). Rapidly self-assembling three-dimensional opal photonic crystals. *J. Korean Phys. Soc.*, Vol. 52 (May 2008) 1585-1588, ISSN 0374-4884.

Wu, C. Y.; Chiang, T. H.; Lai, N. D.; Do, D. B. & Hsu, C. C. (2009). Fabrication of SU-8 microlens arrays by multi-exposure two-beam interference technique. *Appl. Opt.*, Vol. 48 (May 2009) 2473-2479, ISSN 1559-128X.

Zoorob, M. E.; Charlton, M. D. B.; Parker, G. J.; Baumberg, J. J. & Netti, M. C. (2000). Complete photonic bandgaps in 12-fold symmetric quasicrystals. *Nature*, Vol. 404 (April 2000) 740-743, ISSN 0028-0836.

13

Achieving Wide Band Gaps and a Band Edge Laser Using Face-Centered Cubic Lattice by Holography

Tianrui Zhai and Dahe Liu
Applied Optics Beijing Area Major Laboratory, Department of Physics, Beijing Normal University, Beijing 100875, China

1. Introduction

Complete band gaps (CBGs) is an important research topic in the study of photonic crustals (PCs). In 1990, K.M.Ho *et al.* demonstrated theoretically that a diamond structure possesses CBGs [1], since then, great interests were attracted in fabricating three-dimensional (3D) PCs in order to obtain CBGs. Several methods were reported, and CBGs were achieved in the range of microwave or sub-microwave. Theoretical analysis shows that although CBGs can be obtained by diamond structure, a strict condition has to be satisfied, i.e., the modulation of refractive index of the material used should be larger than 2.0 [2, 3]. So, many scientists paid their attention to find the materials with high the refractive index. Some of them tried to fill the templates with high refractive index materials to increase the modulation of the refractive index [4-6], and CBGs were obtained. However, the achieved CBGs in 3D PCs were mostly in microwave and infrared regions [5, 7-9].

Holography is a cheap, rapid, convenient, and effective technique for fabricating 3D structures. In 1997, holographic technique was introduced to fabricate the face centered cubic (fcc) structure [10]. Campbell *et al.* fabricated fcc structure in terms of holographic lithography [11]. Several authors reported their works in this topic [2, 3, 7, 12]. Toader and John also showed theoretically a five-beam "umbrella" configuration for synthesis of a diamond photonic crystal [13]. Because both the value and modulation of the refractive index of the holographic recording materials are commonly low so that the PCs made directly by holography would not have CBGs. For example, the epoxy photoresist is generally used in holographic lithography [11, 14-16] , but, its refractive index is n=1.6 which is a little bit too low. They may, however, be used as templates for the production of inverse replica structures by, for example, filling the void with high refractive index and burning out or dissolving the photoresist [11], and a good work was done by D.C.Meisel [17] etc. But, to find the materials with large refractive index is not easy, and the special techniques needed are very complicated and expensive. It limits the applications of PCs, especially for industrial productions.

Although band gaps can be broadened by means of multi-structures [18], CBGs in visible range had not yet been achieved, especially by using the materials with low refractive index. Therefore, it is a big challenge to fabricate 3D PCs possessing CBGs in visible range by using

the materials with low refractive index. Although some efforts had been made [18], CBGs in visible range had not yet been achieved by using the materials with low refractive index. Therefore, it is a big challenge to fabricate 3D PCs possessing CBGs in visible range by using materials with low refractive index, though it is greatly beneficial for future PC industry. As the first step for achieving this, it is important to obtain very wide band gaps. In the previous investigations [19, 20], it was evidently shown that the anisotropy of a photonic band gap in a two-dimensional photonic crystal is dependent on the symmetry of the structure, and as the order of the symmetry increases, it becomes easier to obtain a complete band gap. One would naturally ask whether or not such a method is applicable to 3D PCs. In view of this question, in this paper we proposed a holographic method for fabricating triple-diamond structure by using materials with low refractive index, and the features of the band gap in such a complex diamond structure were then studied experimentally.
On the other hand, as a new kind of laser source band edge laser has the significance in the applications of PCs, and is helpful for further understanding of the characteristics of PCs. There fore, it attracts great attentions. In this paper, the emission properties of the band edge laser in a fcc PC fabricated by holography was studied. The different emission properties of the band edge laser pumped by picosecond and nanosecond pulses were investigated experimentally. Some intriguing phenomena were revealed.

2. Complex diamond structure

2.1 Experimental method and the samples

It is known that a cell of diamond structure consists of two cells of fcc structures, and the two cells have a distance of one quarter of the diagonal length of the cell along the diagonal line. Therefore, the diamond structure can be implemented by holography through two exposures: an fcc structure is recorded in the first exposure, then, after the recording material is translated one quarter of the diagonal length along the diagonal line of the fcc structure a second exposure is made to record another fcc structure. In this way, a PC with diamond structure can be obtained.

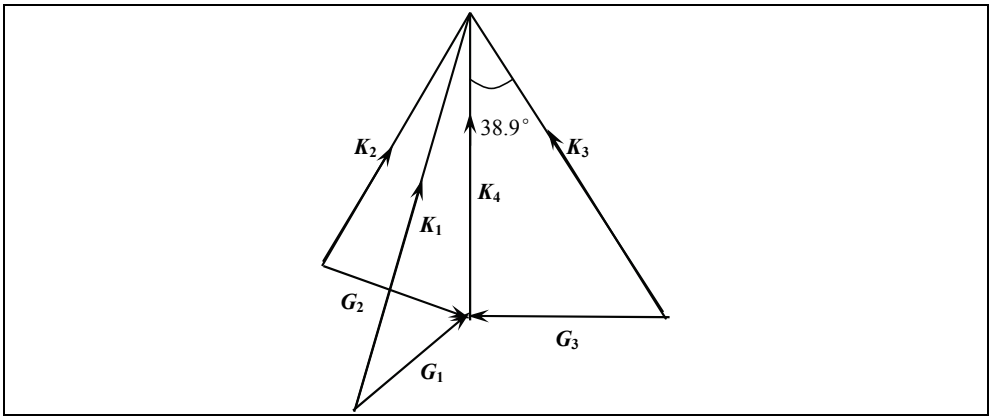


Fig. 1. Schematic optical layout for recording fcc structure.

Fig. 1 shows schematically the optical layout in our experiments. Four beams splitted from a laser beam were converged to a small area. The central beam was set along the normal

direction of the surface of the plate, while the other three outer beams were set around the central beam symmetrically with a angle of 38.9° with respect to the central one. The laser used was a diode pumped laser working at 457.9 nm with line width of 200 kHz (Melles Griot model 85-BLT -605). The polarization state of each beam was controlled to achieve the best interference result [21, 22]. The recording material was mounted on an one-dimensional (1D) translation stage driven by a stepping motor with the precision of $0.05\mu m$ per step. Also, the translation stage was mounted on a rotary stage. The holographic recording material used was dichromated gelatin (DCG) with refractive index $n = 1.52$. The maximum value of its refractive index modulation Δn can reach more than 0.2, but generally is around 0.1. It is a very small value for obtaining wide band gaps. The thickness T of the material was $36\mu m$. The DCG emulsion was coated on an optical plate glass with flatness of $\lambda/10$ and without any doping.

It is well known that there are several directions with high symmetry in the first Brillouin zone of an fcc structure [22]. During our experiments, two exposures as mentioned above (in paragraph 4) were made firstly in $\Gamma - L$ ([111]) direction to get a diamond structure. And then, a second, even a third diamond structure was implemented by changing the orientation of the recording material through rotating the rotary stage to other symmetric directions i.e. the direction of $\Gamma - X$ ([100]) and/or the direction of $\Gamma - K$ ([110]) in the first Brillouin zone of the first fcc structure. In this way, photonic crystals with one, two even three diamond structures were fabricated. It should be pointed out that when the second or the third diamond structure was recorded, the angle between any two beams should be changed to guarantee all the beams inside the medium satisfy the relation shown in Fig. 1 so that the standard diamond structures can be obtained.

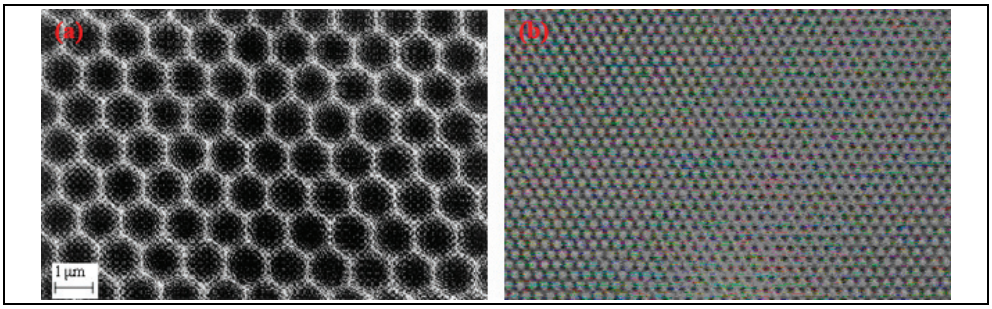


Fig. 2. (a) SEM image of an fcc structure made with photoresist of $1\mu m$ thickness. (b) Microscopic image of an fcc structure made with photoresist of $1\mu m$ thickness.

Since the hologram made with DCG is a phase hologram, there is only the distribution of the refractive index and no plastic effect inside the hologram, so, SEM image can not be obtained. To verify the structure in the hologram, a same fcc structure was implemented using same optical layout but with the photoresist of $1\mu m$ thickness. Fig. 2 shows the structure at [111] plane, (a) shows the SEM image and (b) shows the optical microscopic image taken by a CCD camera mounted on an $1280\times$ microscope, the pixel size of the CCD is $10\mu m$. The theoretical calculated diamond lattice formed by 8 beams as mentioned above is shown in Fig.3. It should be pointed out that the shape of the maximum of interference appears as an American football. In this figure the vertical bar shows a gradient of bottom to

top corresponding to the outer to inner region of a cross-sectional cut of a football. The gradient of the outer surface is related to the value of the vertical bar. Whether a diamond lattice can be obtained depend on how extend the two footballs. If the major parts of the two footballs are overlap the structure can not considered as a diamond lattice. If the two footballs can be recognized independently, it means that a diamond lattice is formed. In our experiments, to get optimal interference result the absorption coefficient α of the material was chosen as $\alpha = 1/T$ [23]. According to previous works [23], it is well known that for holographic recording materials the relation between the optical density and the exposure is nonlinear. The holographic exposure can be controlled in the region of stronger nonlinear dependence between optical density versus exposure as discussed previously [24]. In this way, two fcc lattice can be recognized so that the structure can be considered as a diamond lattice.

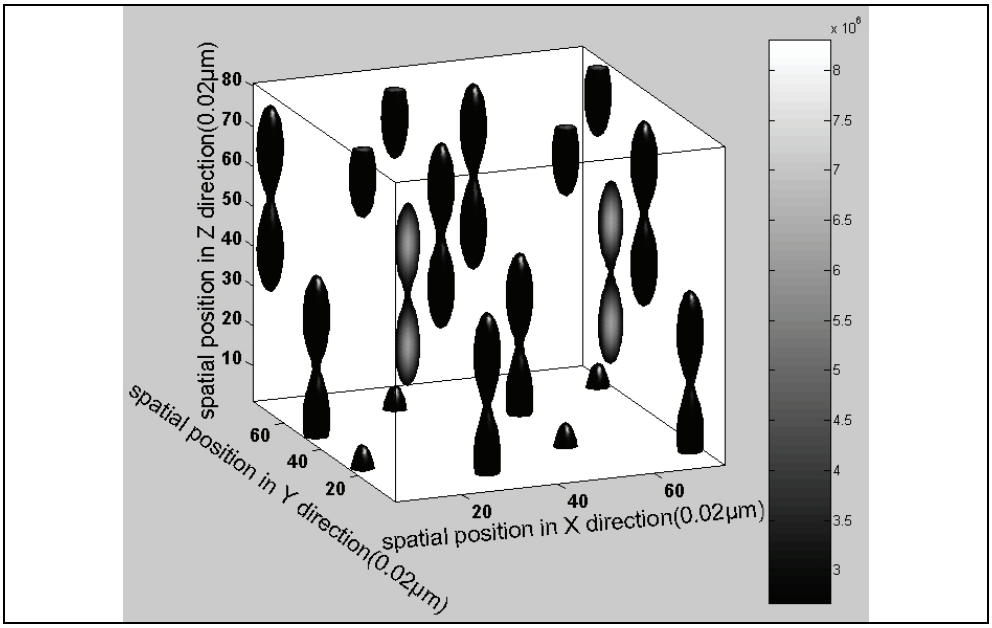


Fig. 3. Diamond lattice formed by two fcc lattice in nonlinear exposures. The vertical bar shows a gradient of bottom to top corresponding to the outer to inner region of a cross-sectional cut of a football. The gradient of the outer surface is related to the value of the vertical bar.

The transmission spectra of PCs with diamond structure made with DCG were measured. In the measurements, a J-Y 1500 monochrometer was employed. To minimize the energy loss from reflection at large incident angle, a right triangular prism was used. The measuring set-up geometry is shown in Fig. 4. The measured range of the wavelength was 390 nm~800 nm. The [1 1 1] plane of the PC measured was set firstly at an arbitrary orientation. The incident angle of the collimated white light beam was changed from $0° \sim \pm 90°$ (the incident angle is in the plane parallel to the paper surface), Then, the PC was rotated to other orientations (the orientation angel is in the plane perpendicular to the paper surface), and the incident

angle of the collimated white light beam was still changed from $0° \sim \pm 90°$. The orientation angel and the incident angle are the angles φ and θ in a spherical system indeed.

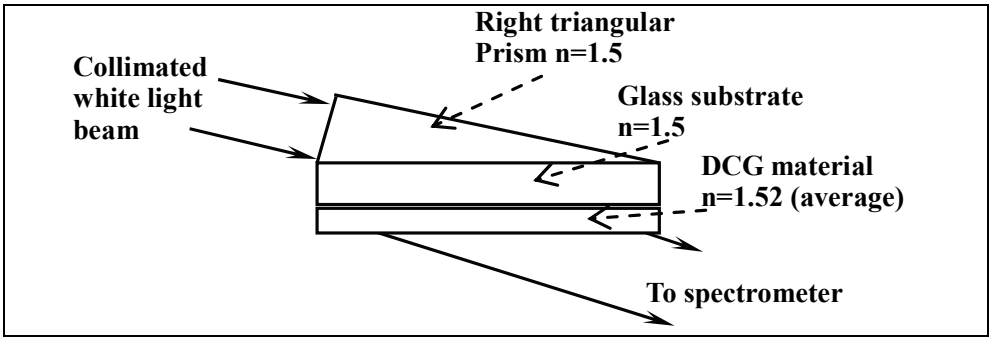


Fig. 4. Set-up geometry for measuring transmission spectra of photonic crystals.

2.2 Measured spectra and discussions

The measured transmission spectra of a fabricated triple diamond PC are shown in Fig. 5(a), (b) and (c). (a), (b) and (c) represent measured results at different orientations of the sample. And each curve in the figures gives the spectrum measured at a certain incident angel and a certain orientation. Different curves give results at different incident angels and orientations. Thus, whether or not the common gap exists can be determined by the overlap of all the curves. It can be seen that, for the PC with three diamond structures (6 fcc structures) the width of the band gaps reached 260 nm, the ratio between the width and the central wavelength of the gaps reached 50 %, and there is a common band gap with a width about 20 nm at 450~470 nm in a range of 150° of the incident angle. The common gap achieved using DCG with very low refractive index exists in a wide range which reaches 83 % of the 4π solid angle. Although a complete band gap for all directions is not obtained in our experiments, it is significant to achieve such a wide angle band gap by using a material (DCG) with very low refractive index, because this angular tuning range can satisfy most applications in practice.

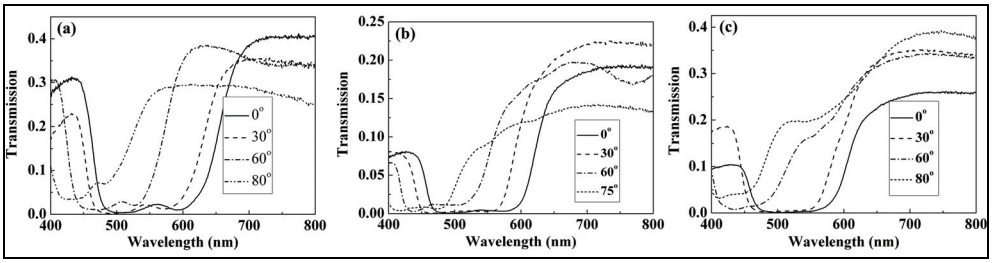


Fig. 5. Measured transmission spectra of PC including 3 diamond lattices recorded by holography in $\Gamma - L$, $\Gamma - X$ and $\Gamma - K$ directions of an fcc lattice respectively. (a), (b) and (c) correspond to the measured results. The angles appeared in (a), (b) and (c) are the incident angle which is in a plane parallel to the paper surface. (a) Orientation angle is 0°. (b) Orientation angle is 30°. (c) Orientation angle is 90°.

This interesting result comes from complex diamond structure. The photonic crystal with triple diamond structures is an actual multi-structure not a stack of several same structures. The [111] direction of the first diamond structure is actually the $\Gamma - X$ direction of the second diamond structure and the $\Gamma - K$ direction of the third diamond structure respectively. When a beam incidents normally on the PC, the beam is in the [111] direction of the first diamond structure. When the incident angle changes, the beam deviates from the [111] direction of the first diamond structure but tends to approach the [111] direction of the second or the third diamond structure. Therefore, though the incident angle changes, the light beam keeps always around the [111] direction of the other diamond structures. The narrow region in the $K - \omega$ dispersion relation of a diamond lattice may be expanded by other diamond lattices.

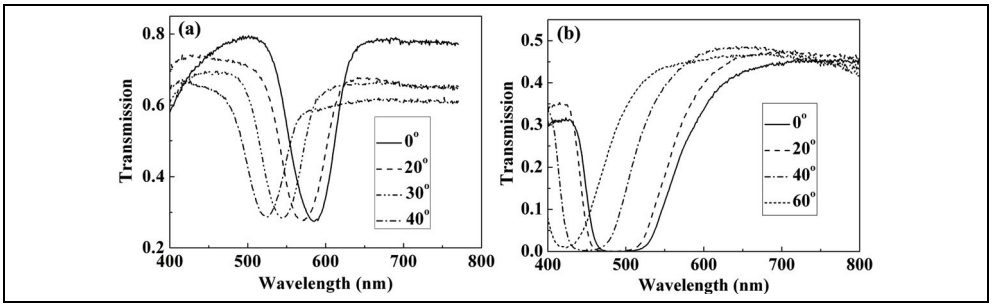


Fig. 6. Measured transmission spectra of a PC at different orientations and incident angles recorded with one and two diamond lattices. (a) Measured transmission spectra of a PC at different orientations and incident angles recorded with single diamond lattice recorded by holography in $\Gamma - L$ direction of an fcc lattice. (b) Measured transmission spectra of a PC at different orientations and incident angles recorded with two diamond lattices recorded by holography in $\Gamma - L$ and $\Gamma - X$ directions of an fcc lattice respectively.

The advantage of a multi diamond structure can be seen more clearly from the comparison to those with one and two diamond structures. Fig. 6 gives the measured transmission spectra of PCs with one and two diamond lattices at a certain orientation respectively. It can be found that with the increase of the number of diamond lattice the width of band gaps of a PC can be broadened effectively, For two diamond structures, the space angle range of the common gap is $80°$, and for the single diamond structure the common gap becomes $40°$. This means that the common gap can be enlarged or created by means of multi-structures. The physical origin of such a phenomenon can be understood from the change of structure symmetry. With the increase of the number of diamond lattice, the symmetries around the center of the structure become higher. This made it easier to obtain a broader response in many directions (common band gap in a wide range of angles), which is similar to the case of two-dimensional PCs (See Refs. [19, 20]).

2.3 Conclusion

In summary, the width of the band gaps in a PC made by the material with low refractive index is broadened greatly by multi diamond structures. A common band gap in a range of $150°$ of the incident angle is obtained. This technique will be greatly beneficial to achieve complete band gaps by the materials with low refractive index.

3. Band edge laser

Theoretical analysis has shown that a PC doped with gain materials can achieve lasing based on the characteristics of the high state density at the band edge of a PC [25-28]. Experimentally, the band edge laser in 1D [29], 2D (to dimension change to 2D) [30, 31] and 3D [32, 33] PCs have been reported. Holography has shown good potential in fabricating PCs [10, 11, 13, 34, 35]. Recent research shows that the 1D PC and quasicrystals PC, with gradual distribution of the refractive index materials made by holography using DCG, can also achieve band edge lasing [36, 37]. Using DCG, the present authors fabricated a 3D lattice structure which achieved wide band gaps [38], and sometimes achieved complete band gaps [39, 40]. Based on these studies, a band edge laser in an fcc PC was achieved in our laboratory. Although many studies on band edge lasers have been reported [25, 41-43], the different output properties of band edge lasers pumped by picosecond (ps) and nanosecond (ns) pulses have not yet been investigated. However, because the pump duration is related directly to the behavior of band edge laser emission, this kind of research is significant for understanding further the mechanism of the band edge laser. Recently, we found that the band edge laser showed obvious different emission properties depending on whether it was pumped by a ps laser or by a ns laser. Therefore, the investigation of such a problem becomes important. The following outlines our work.

3.1 Experimental method and the samples

The set-up geometry for fabricating the fcc structure is the same as that in Ref. [38]. In our experiments, the laser was a diode pumped laser working at 457.9 nm with line width of 200 kHz (Melles Griot model 85-BLT-605). The holographic recording material was DCG. It was coated on an optical glass plate with a flatness of $\lambda/10$ without any doping. DCG is a phase type holographic recording material with a refractive index of 1.52 and a modulation of refractive index less than 0.1. The thickness of the DCG is 36 μm. After exposure, the DCG plate was developed in running water at 20°C for two hours, and then soaked in a Rhodamine 6G solution bath with a concentration of 0.125 mg per milliliter water at the same temperature for 60 min, enabling the dye molecules to diffuse deeply into the emulsion of the gelatin. The DCG plate was dehydrated in turn by soaking it in 50 %, 75 % or 100 % isopropyl alcohol with the same concentration of Rhodamine 6G at 40°C for 15 min each. After dehydration, the DCG plate was baked at 100°C for 60 min in an oven.

3.2 Measured spectra and discussions

Figure 7(a) shows the measured transmission spectrum (the black line) of the fcc structure without dye doping and the absorption spectrum of Rhodamine 6G (the red line) respectively using a J-Y 1500 monochromator. It can be seen that the fcc structure has band gaps with a width (FWHM-full width at half maximum) of 80 nm. The center of the absorption spectrum of Rhodamine 6G is located just at the shorter wavelength edge of the band of the fcc structure. Fig. 7(b) shows the photoluminescence of Rhodamine 6G excited by a 532 nm laser beam. It can be seen that the whole photoluminescence also locates at the shorter wavelength edge of the band of the fcc structure.

Figure 8 shows the set-up geometry for measuring the emission of the band edge laser. In our experiments, the measured structure was pumped by a ns pulsed Nd:YAG laser with a pulse duration of 8 ns and a ps pulsed Nd:YAG laser with a pulse duration of 30 ps. The pulse laser with a repetition rate of 10 Hz is the most common commercial equipment in

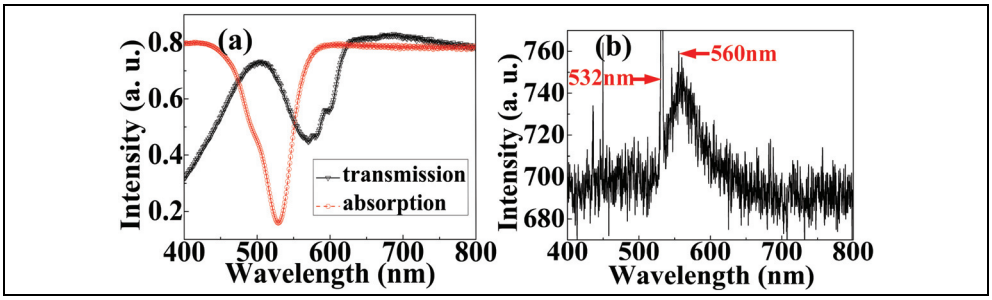


Fig. 7. (a) Transmission spectrum of the fcc structure and absorption spectrum of Rhodamine 6G. (b) Photoluminescence of Rhodamine 6G excited by a 532 nm laser beam.

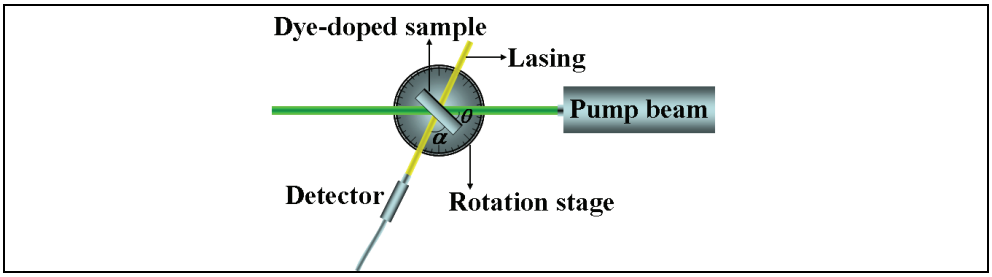


Fig. 8. Set-up geometry for measuring the emission of the band edge laser. θ is the angle of the pump beam with respect to the surface of the sample. α is the angle of the emission beam with respect to the surface of the sample.

wide use, for example, in the work of [36] and [37]. Therefore, such lasers running at 532 nm and repetition rates of 10 Hz were used. During measurements, the laser emission was collected by a lens, which was focused into a fiber and fed to a spectrometer. It should be clarified that the aim of the present work is to find the different emission characteristics of a band edge laser pumped by a ns laser and a ps laser. It is not to investigate the relationship between the band gap and the incident angle. In our experiments, lasing can be found in certain ranges of the incident and the output angles (corresponding to different directional band gaps). The following experimental results are the best for the corresponding conditions.

Figure 9 shows the emission spectra of the band edge laser pumped by a ps laser. For the case pumped by a ps laser, the size of the pump beam is 3 mm and the maximum pulse energy is 40 mJ. It can be seen that there are several modes; each of them with a width less than 0.4 nm. The threshold value is about 1.65 GW/mm^2. The lasing modes are not stable and no one is dominant among the several modes. The results shown in Fig. 9 are only two of the measured spectra at different pump power intensities. Furthermore, the measured spectra recorded at different times under the same pump power intensity are quite different, and the heights and the positions of the peaks in each spectrum are different (See Fig. 10).

Figure 11(a) shows the emission spectra of the band edge laser pumped by a ns laser. For the case pumped by the ns laser, the size of the pump beam is 13 mm and the maximum pulse energy is 1.5 J. It can be seen that the lasing threshold is around 6 MW/cm^2. This is only an estimated value since the lasing emission around the threshold is unstable. The threshold is

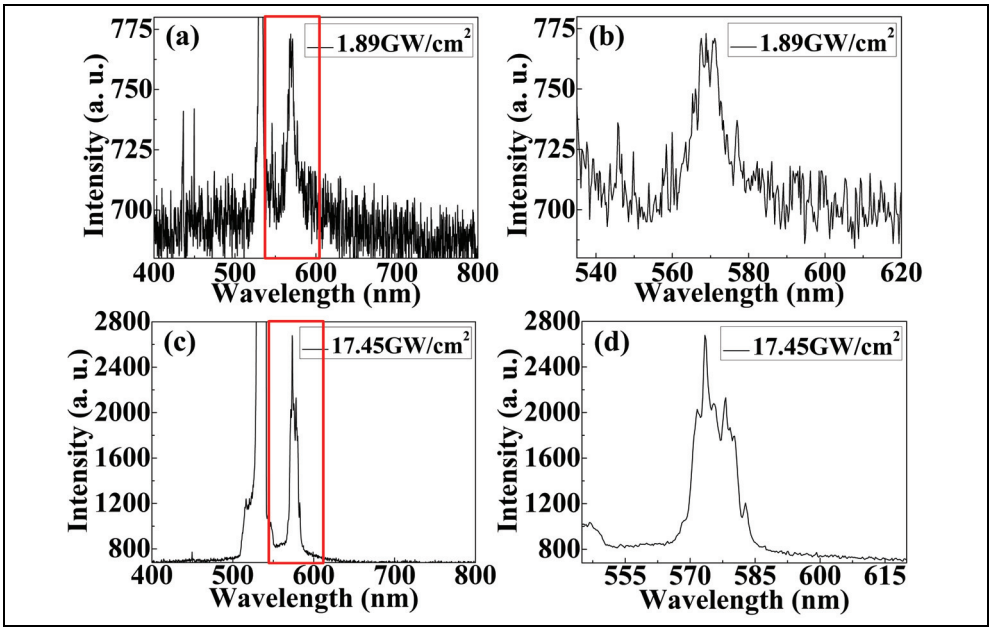


Fig. 9. Measured emission spectra of the band edge laser at different pump power intensity pumped by a ps laser. θ≈65° and α≈50°. (a) and (c) are the spectra measured at different pump power intensities, (b) and (d) are the enlargement of the inset (red box) in (a) and (c).

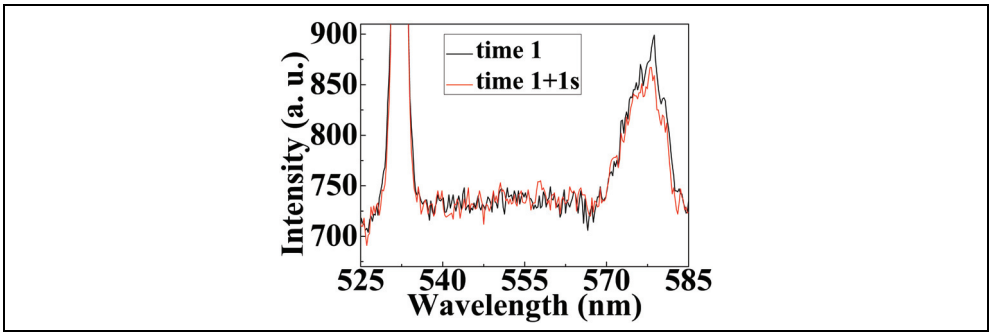


Fig. 10. The measured spectra recorded at different times under same pump power intensity.

much lower than that pumped by a ps laser. The right inset in Fig. 11(a) shows the intensity (green line) and the line width (blue line) vs. pump power intensity. From Fig. 11(b) it can be seen that when the pump power intensity is in a certain range (higher than the lasing threshold), there are two emission modes in competition. The two peaks in Fig. 11(b) are both located in the range of the gain profile. If there is no competition, the two peaks will grow with the increase of the pump power intensity. However, in our observations, only one peak grows when the pump power intensity increases, and the height of the other is always kept very low. Obviously, it is the phenomenon of competition. In addition, it is

found that one emission mode is higher when the pump power intensity is higher. When the pump power intensity is high enough there is only one mode in the emission that becomes dominant. This mode has a width of several nanometers, and the intensity of the emission is very high corresponding to the photoluminescence. Our experimental observations show that the dominant emission mode is quite stable, i.e., for every pump pulse, the frequency position and the output intensity always remain unchanged.

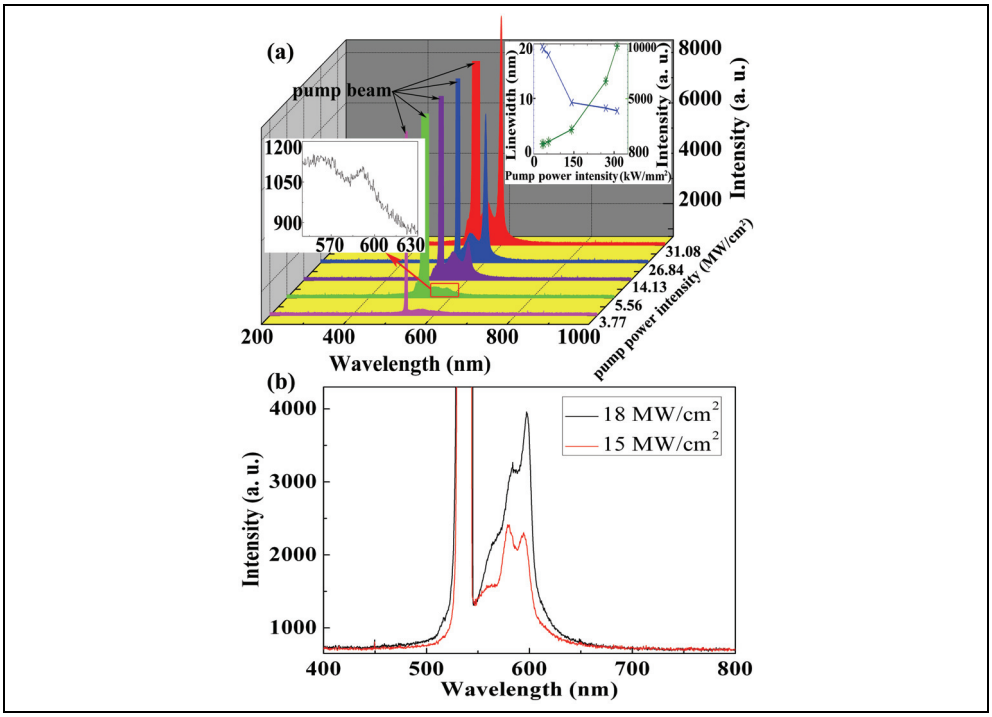


Fig. 11.(a) The emission spectra of dye-doped fcc PC as a function of pump power intensity pumped by a ns laser. In the right inset, the green line corresponds to the emission intensity of lasing, and the blue line corresponds to the line width of the lasing emission. It should be noted that the pump beam is saturated on the CCD used. (b) Measured spectra when the pump power intensity is in a certain range (higher than the lasing threshold).

There are many studies on the theoretical investigation of band edge lasers [25, 41-43]. These theoretical works are strictly quantitative analyses, and are suitable for most band edge lasers. Therefore, we will not repeat them here, but will provide a brief analysis on the physical mechanism to discuss the difference between the cases pumped by a ps laser and by a ns laser.

It is known that one of the important features of a PC is the high density of the mode at the band edge. If gain material exists in a PC, and the range of the gain is at the band edge, it will provide a lasing condition. The modes just inside the gain range may be lasing. As observed in Figs. 9 and 11, there are multi modes for the case pumped by a ps laser, but there is only one dominant mode for the case pumped by a ns laser. The reason is that in the

case pumped by a ns laser, the pulse width is about 10 ns, i.e., the interaction between the pump beam and the material will last for about 10 ns, which is a relatively long period, and the "mode competition" will have finished. As a result, one mode becomes dominant in the emission and is very stable. However, in the case pumped by a ps laser, the interaction between the pump beam and the material will only last for about 30 ps, and the "mode competition" cannot finish in the period of one pulse. Therefore, multi modes will appear in the emission. It is also found that the lasing threshold pumped by a ns laser is about three orders of magnitude less than that pumped by a ps laser. The reason is that lasing needs a build up time. For the case pumped by a ps laser if the pump power intensity is lower, and the pump pulse is over, the lasing has not built up. To build up the lasing needs higher pump power intensity. However, for the case pumped by a ns laser, the process of building up has a long enough duration. Thus, much lower pumped power intensity can build up the lasing. It should be pointed out that, the life time of the used energy level of Rhodamine 6G is around the magnitude of 10^{-7} s [44, 45], which is only one order of magnitude of the pulse width when a ns laser is used (10 ns), but is four orders of magnitude longer than the pulse width when a ps laser is used (30 ps).

3.3 Conclusion

In summary, when the band edge laser is pumped by a ps laser or by a ns laser it shows different emission characteristics. The number of emission mode, the line width, the lasing threshold and the stability of the lasing are obviously different.

4. References

[1] K. M. Ho, C. T. Chan, and C. M. Soukoulis, "Existence of a photonic band gap in periodic structures," Phys. Rev. Lett. 65, 3152 (1990).

[2] T. Y. M. Chan, O. Toader, and S. John, "Photonic band gap templating using optical interference lithography," Phys. Rev. E 71, 46605 (2005).

[3] D. N. Sharp, A. J. Turberfield, and R. G. Denning, "Holographic photonic crystals with diamond symmetry," Phys. Rev. B 68, 205102 (2003).

[4] D. C. Meisel, M. Wegener, and K. Busch, "Three-dimensional photonic crystals by holographic lithography using the umbrella configuration: Symmetries and complete photonic band gaps," Phys. Rev. B 70, 165104 (2004).

[5] A. Blanco, E. Chomski, S. Grabtchak, S. John, S. W. Leonard, J. P. Mondia, M. Ibisate, C. Lopez, F. Meseguer, and H. Miguez, "Large-scale synthesis of a silicon photonic crystal with a complete three-dimensional bandgap near 1. 5 micrometres," Nature 405, 437 (2000).

[6] Y. Vlasov, "On-chip natural assembly of silicon photonic bandgap crystals," Nature 414, 289 (2001).

[7] L. Wu, Y. Zhong, C. T. Chan, K. S. Wong, and G. P. Wang, "Fabrication of large area two- and three-dimensional polymer photonic crystals using single refracting prism holographic lithography," Appl. Phys. Lett. 86, 241102 (2005).

[8] S. Y. Lin, J. G. Fleming, D. L. Hetherington, B. K. Smith, W. Zubrzycki, S. R. Kurtz, J. Bur, R. Biswas, K. M. Ho, and M. M. Sigalas, "A three-dimensional photonic crystal operating at infrared wavelengths," Nature 394, 251 (1998).

[9] S. Noda, K. Tomoda, N. Yamamoto, and A. Chutinan, "Full Three-Dimensional Photonic Bandgap Crystals at Near-Infrared Wavelengths," Science 289, 604 (2000).
[10] V. Berger, O. Gauthier-Lafaye, and E. Costard, "Photonic band gaps and holography," J. Appl. Phys. 82, 60 (1997).
[11] M. Campbell, D. N. Sharp, M. T. Harrison, R. G. Denning, and A. J. Turberfield, "Fabrication of photonic crystals for the visible spectrum by holographic lithography," Nature 404, 53 (2000).
[12] Y. C. Zhong, S. A. Zhu, H. M. Su, H. Z. Wang, J. M. Chen, Z. H. Zeng, and Y. L. Chen, "Photonic crystal with diamondlike structure fabricated by holographic lithography," Appl. Phys. Lett. 87, 061103 (2005).
[13] O. Toader, T. Y. M. Chan, and S. John, "Diamond photonic band gap synthesis by umbrella holographic lithography," Appl. Phys. Lett. 89, 101117 (2006).
[14] T. Kondo, S. Matsuo, S. Juodkazis, and H. Misawa, "Femtosecond laser interference technique with diffractive beam splitter for fabrication of three-dimensional photonic crystals," Appl. Phys. Lett. 79, 725 (2001).
[15] Y. V. Miklyaev, "Three-dimensional face-centered-cubic photonic crystal templates by laser holography: fabrication, optical characterization, and band-structure calculations," Appl. Phys. Lett. 82, 1284 (2003).
[16] I. Divliansky, T. Mayer, K. Holliday, and V. Crespi, "Fabrication of three-dimensional polymer photonic crystal structures using single diffraction element interference lithography," Appl. Phys. Lett. 82, 1667 (2003).
[17] D. C. Meisel, M. Diem, M. Deubel, F. Pérez-Willard, S. Linden, D. Gerthsen, K. Busch, and M. Wegener, "Shrinkage precompensation of holographic three-dimensional photonic-crystal templates," Adv. Mater. 18, 2964 (2006).
[18] L. Cui, F. Wang, J. Wang, Z. Wang, and D. Liu, "The rule for broadening of band-gaps in biperiodic photonic crystals," Phys. Lett. A 324, 489 (2004).
[19] M. E. Zoorob, M. B. D. Charlton, G. J. Parker, J. J. Baumberg, and M. C. Netti, "Complete photonic bandgaps in 12-fold symmetric quasicrystals," Nature 404, 740 (2000).
[20] X. Zhang, Z. Q. Zhang, and C. T. Chan, "Absolute photonic band gaps in 12-fold symmetric photonic quasicrystals," Phys. Rev. B 63, 81105 (2001).
[21] H. M. Su, Y. C. Zhong, X. Wang, X. G. Zheng, J. F. Xu, and H. Z. Wang, "Effects of polarization on laser holography for microstructure fabrication," Phys. Rev. E 67, 56619 (2003).
[22] X. Wang, J. F. Xu, H. M. Su, Z. H. Zeng, Y. L. Chen, H. Z. Wang, Y. K. Pang, and W. Y. Tam, "Three-dimensional photonic crystals fabricated by visible light holographic lithography," Appl. Phys. Lett. 82, 2212 (2003).
[23] D. Liu, W. Tang, W. Huang, and Z. Liang, "Relationship between the diffraction efficiency of a reflection hologram and the thickness and absorption of the recording medium," Opt. Eng. 31, 809 (1992).
[24] H. Smith, and R. Bartolini, *Holographic recording materials* (Springer-Verlag Berlin, Berlin, 1977), Vol. Ch.3.
[25] J. Dowling, M. Scalora, M. Bloemer, and C. Bowden, "The photonic band edge laser: A new approach to gain enhancement," J. Appl. Phys. 75, 1896 (1994).

[26] S. Nojima, "Enhancement of optical gain in two-dimensional photonic crystals with active lattice points," Jpn. J. Appl. Phys. 37, L565 (1998).
[27] K. Sakoda, K. Ohtaka, and T. Ueta, "Low-threshold laser oscillation due to group-velocity anomaly peculiar to two-and three-dimensional photonic crystals," Opt. Express 4, 481 (1999).
[28] L. Florescu, K. Busch, and S. John, "Semiclassical theory of lasing in photonic crystals," J. Opt. Soc. Am. B 19, 2215 (2002).
[29] M. Tocci, M. Scalora, M. Bloemer, J. Dowling, and C. Bowden, "Measurement of spontaneous-emission enhancement near the one-dimensional photonic band edge of semiconductor heterostructures," Phys. Rev. A 53, 2799 (1996).
[30] M. Meier, A. Mekis, A. Dodabalapur, A. Timko, R. Slusher, J. Joannopoulos, and O. Nalamasu, "Laser action from two-dimensional distributed feedback in photonic crystals," Appl. Phys. Lett. 74, 7 (1999).
[31] M. Imada, S. Noda, A. Chutinan, T. Tokuda, M. Murata, and G. Sasaki, "Coherent two-dimensional lasing action in surface-emitting laser with triangular-lattice photonic crystal structure," Appl. Phys. Lett. 75, 316 (1999).
[32] M. Shkunov, Z. Vardeny, M. DeLong, R. Polson, A. Zakhidov, and R. Baughman, "Tunable, gap-state lasing in switchable directions for opal photonic crystals," Adv. Funct. Mater. 12, 21 (2002).
[33] W. Cao, A. Munoz, P. Palffy-Muhoray, and B. Taheri, "Lasing in a three-dimensional photonic crystal of the liquid crystal blue phase II," Nature Mater. 1, 111 (2002).
[34] L. Z. Cai, X. L. Yang, and Y. R. Wang, "All fourteen Bravais lattices can be formed by interference of four noncoplanar beams," Opt. Lett. 27, 900 (2002).
[35] L. Yuan, G. P. Wang, and X. Huang, "Arrangements of four beams for any Bravais lattice," Opt. Lett. 28, 1769 (2003).
[36] M. Kok, W. Lu, J. Lee, W. Tam, G. Wong, and C. Chan, "Lasing from dye-doped photonic crystals with graded layers in dichromate gelatin emulsions," Appl. Phys. Lett. 92, 151108 (2008).
[37] M. Kok, W. Lu, W. Tam, and G. Wong, "Lasing from dye-doped icosahedral quasicrystals in dichromate gelatin emulsions," Opt. Express 17, 7275 (2009).
[38] Z. Ren, Z. Wang, T. Zhai, H. Gao, D. Liu, and X. Zhang, "Complex diamond lattice with wide band gaps in the visible range prepared by holography using a material with a low index of refraction," Phys. Rev. B 76, 035120 (2007).
[39] Z. Ren, T. Zhai, Z. Wang, J. Zhou, and D. Liu, "Complete Band Gaps in the Visible Range Achieved by a Low-Refractive-Index Material," Adv. Mater. 20, 2337 (2008).
[40] T. Zhai, Z. Wang, R. Zhao, J. Zhou, D. Liu, and X. Zhang, "A D_{nv} point group structure possessing complete band gap based on gradual heterostructure and self-simulating sphere," Appl. Phys. Lett. 93, 201902 (2008).
[41] V. Kopp, B. Fan, H. Vithana, and A. Genack, "Low-threshold lasing at the edge of a photonic stop band in cholesteric liquid crystals," Opt. Lett. 23, 1707 (1998).
[42] C. O. Cho, J. Jeong, J. Lee, H. Jeon, I. Kim, D. H. Jang, Y. S. Park, and J. C. Woo, "Photonic crystal band edge laser array with a holographically generated square-lattice pattern," Appl. Phys. Lett. 87, 161102 (2005).

[43] G. Vecchi, F. Raineri, I. Sagnes, A. Yacomotti, P. Monnier, T. Karle, K. Lee, R. Braive, L. Le Gratiet, and S. Guilet, "Continuous-wave operation of photonic band-edge laser near 1.55 um on silicon wafer," Opt. Express 15, 7551 (2007).
[44] S. John, and G. Pang, "Theory of lasing in a multiple-scattering medium," Phys. Rev. A 54, 3642 (1996).
[45] O. Svelto, *Principles of Lasers* (Heyden, New York, 1976), pp. 227.

14

Accurate Axial Location for Particles in Digital In-Line Holography

Zhi-Bin Li[1], Gang Zheng[2], Li-Xin Zhang[2], Gang Liu[1] and Fei Xia[1]
[1]*Shanghai University of Electric Power*
[2]*University of Shanghai for Science and Technology*
China

1. Introduction

The potential of in-line holography to analyze flow by means of particles has been noted since its very beginnings. The development of optical holography applications in fluid mechanics established the capability of holography to provide both particle size and tridimensional position (Adams et al., 2000). In classical holography, the diffraction pattern produced by a particle field is recorded on a holographic plate or film with a collimated laser beam. The reconstruction of the 3D particle images is performed by illuminating developed hologram with the reference beam. In digital holography, the diffraction pattern is directly recorded on a charge-coupled device (CCD) camera and the reconstruction process is performed numerically, plane by plane to produce a region of the reconstructed volume. So the sizes and locations of particles are computed from the reconstructed volume by an automated process.

Recently, this technique has been used to develop a new holographic particle image velocimetry (HPIV) system (Meng et. al., 2004). It is particularly interesting for high-speed phenomena analysis, as in fluid mechanics. But it has become the bottleneck in flow field measurement due to the lower axial positioning accuracy of particle. Many contributions have been made in order to improve the measurement accuracy of the localization of particles. Generally, the two main steps of the numerical processing are a numerical reconstruction step to obtain a synthesized 3D image with focused particles, and a segmentation step to extract locations and sizes of the particles from this 3D distribution.

However, the direct light and twin-images of particles can not be separated from the focused-images of particles in the numerical reconstruction process due to inline optical system, which resulting in the decline in quality for focused-images of particles. The impact of direct light and twin-images of particles can be removed and improved through all kinds of different methods (Kreis et al., 1997 and Takaki et al., 1999)). As an important parameters of imaging system, the depth of focus (DOF) is tens of times or even hundreds of times more than particle diameter in digital in-line Fresnel holographic system (Dubois et al., 2006). However, the DOF of inline digital holographic system is characterized differently in different literatures (Liebling et al., 2004). It is very difficult to give precise expression to DOF of inline digital holographic system. So approximations have to be applied in the analysis of the DOF. DOF is the main reason for the lower axial positioning accuracy of

particle. Because there exists no definitive criterion for finding the focal plane of a scene or finding the focal distance for a region within a scene, many auto-focusing evaluation methods have been proposed to reduce the impact of DOF. The best focus plane can be determined by a maximum value of distribution curve along the optical axis through the light intensity of the particle center, the complex amplitude or the variance of imaginary part (Shen et al., 2005 and Ferraro et al., 2003). And these algorithms employ focus measures such as self-entropy (Gillespie et al., 1989), wavelet analysis (Liebling et al., 2004), the integrated amplitude modulus (Dubois et al., 2006), phase changes (Ma et al., 2004) and gray-level variance (Ferraro et al, 2003). But due to the impact of the depth of focus, these algorithms have property of less single-peak and poor stability.

In this paper, the mathematical formula of the DOF in the inline digital holographic system is derived and analyzed in detail with a new Fresnel number defined by the defocused parameter. Then in order to reduce the impact of DOF, an auto-focusing algorithm based on the most gradient for accurate location of particle was proposed. At the same time, another threshold parameter is introduced to improve the precision and stability of auto-focusing algorithm.

2. The influence of depth of focus for accurate axial location of particle

2.1 DOF of inline digital holographic system

According to the principle of holography, the hologram is recorded and reconstructed accurately only for a planar object. However, at certain conditions, a minute 3D object can be reconstructed with high accuracy. So the depth of focus of the digital holographic system can be defined the axial depth of the clearly reconstructed object. Although there is only the best clearly reconstructed object in theory, the defocused reconstructed object can be considered the same as the focused object within area of the DOF. Therefore, the DOF is an error of axial depth for clearly reconstructed object.

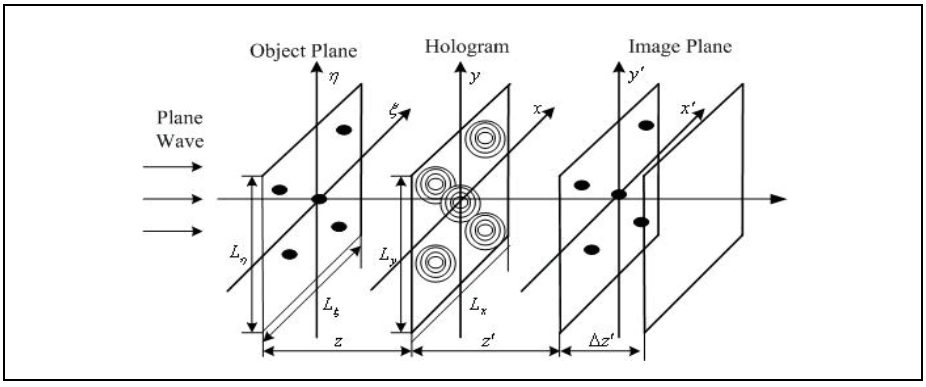


Fig. 1. Recording and reconstruction of in-line holography.

The particle hologram is recorded with the classical in-line holography arrangement. The principle of recording and reconstruction of hologram is shown in figure 1. Suppose that the monochromatic plane parallel wave of laser beam from left to right illuminate the particles on the object plane $\eta - \xi$, part of the wave is diffracted by particles and the remaining part

passes the setup without being diffracted saving as a reference wave, So we can have a hologram from CCD recording plane $x-y$ located at a distance z from the object plane on which the axisymmetric interference fringes for all the particles are recorded overlapping each other. The reconstructed particle image on image plane $x'-y'$ located at a distance z' from the hologram plane can be computed at the true depth where $z'=z$ to obtain the clearest image that is the shade of particle. In figure 1, L_x and L_y denote the length and width of CCD, L_ξ and L_η denote the length and width of particle field recorded. The distance $\Delta z'$ from the best clearly focus plane to actual image plane denotes the maximum axial depth which we can tolerant.

Now, taking the best focus plane as reference, i.e. $\frac{1}{z'}-\frac{1}{z}-\frac{1}{f}=0$, the defocus parameter $1/F$ can be expressed as following

$$\frac{1}{F}=\frac{1}{z'}-\frac{1}{z}-\frac{1}{f} \tag{1}$$

Because the particles are illuminated by plane wave in in-line holographic system without lens, as shown in figure 1, Eq. (1) can be written as

$$\frac{1}{F}=\frac{1}{z'}-\frac{1}{z} \tag{2}$$

where $1/F=0$ for perfectly focused image and $1/F\neq 0$ for defocused image. When $z'=z$, the perfectly focused image of particle can be obtained. Hence, the clearly reconstructed image of particle can be obtained from the location $z'=-z\pm\Delta z'$. So the expression of defocus parameter can be written as

$$\frac{1}{F}=\frac{1}{-z\pm\Delta z'}+\frac{1}{z}=\frac{\pm\Delta z'}{z(-z\pm\Delta z')} \tag{3}$$

A square aperture with a side length of L is assumed for the CCD. According to the definition of Fresnel number (Goodman, 2006), a new parameter N_F is defined as

$$N_F=\frac{W^2}{\lambda F} \tag{4}$$

where $W=\frac{L}{2}$.

In accordance with the definition of numerical aperture in traditional optical imaging system, if lens is not used in the digital holographic recording system and there exists a certain condition which meets $L\ll z$, the numerical aperture NA of image system can be expressed as

$$NA=\frac{L}{2z}=\frac{W}{z} \tag{5}$$

From Eq. (4) and Eq. (5), Eq. (6) can be derived and written as

$$\frac{1}{F}=N_F\frac{\lambda}{z^2(NA)^2} \tag{6}$$

From Eq. (3), it is apparent that $\frac{z}{F} << 1$, in such a way that the Eq.(3) becomes

$$\mp\Delta z' \approx \frac{z^2}{F} = N_F \frac{\lambda}{(NA)^2} \tag{7}$$

The maximum axial depth which we can tolerant, that is depth of focus, can be expressed as

$$2\Delta z' = 2N_F \frac{\lambda}{(NA)^2} \tag{8}$$

In other words, Eq. (8) will determine the focus of depth for image system. As expected, the greater the numerical aperture of the system the shorter is focus of depth.
In inline digital holographic imaging system, the parameters λ and NA are known where NA can be calculated with the length L of CCD and recording distance z according to Eq. (5). Therefore, we can obtain the curve of the numerical relationship between $\Delta z'$ and $\lambda/(NA)^2$ under given conditions. The parameter N_F can be obtained by calculating the slope of the above line.
It is usual to regard a maximum loss of about 20% intensity as limit at the centre of the image of a point object, i.e. $I_{z'}(0,0) = 0.8I_z(0,0)$, is regarded as permissible for focused images, which is corresponding to $N_F = 0.5$. So the focus of depth can be written as

$$2\Delta z' = \frac{\lambda}{(NA)^2} \tag{13}$$

For visual detection the tolerance is more flexible. So the focus of depth will change with it.

2.2 The impact of depth of focus

From Eq. (8), the focus of depth in a in-line digital holographic system is mainly determined by the system parameters, the larger the numerical aperture of the system the shorter is focus of depth and the smaller the recording wavelength the shorter is focus of depth. Therefore, when the system parameters are fixed, we can shorten recording distance to increase the numerical aperture and reduce the DOF of system according to Eq. (5).
From the above analysis, we can see that the depth of focus significantly increases with recording distance. Generally, the depth of focus is tens of times or even hundreds of times more than particle diameter with tens of microns in in-line digital holographic system. So there are two adverse effects for the depth of focus to obtain precise axial location of particles. On the one hand, because within the depth of focus, all the reconstructed image of particle is clearer, it is very difficult to distinguish the clearest particle image. So the real axial location of particle is uncertain. It results lower measurement accuracy for the axial depth position of particle. Especially in DHPIV technology, the axial location of particle is key factors in analysis of velocity field. The impact of the DOF results in uncertainty and brings about larger errors. On the other hand, it is very important during the numerical reconstruction process. Because the reconstruction process is performed numerically, plane by plane to produce a region of the reconstructed volume, this is a very time-consuming process. With larger sampling interval, although the computational time can be reduced, if it is larger than the DOF, the real particle image may be “loss” and can not be reconstructed.

With smaller sampling interval, although real particle image can be reconstructed, the computational time will be increased significantly. Therefore, we must know the DOF to determine right sampling interval to locate all possible particles with less time before reconstructing particle fields.

3. General process for extracting 3D particle locations

Analyzing particle fields (such as velocity field computation from pairs of images and particle tracking in 3D turbulence flow) require a robust and accurate method for extracting particle coordinates from reconstructed 3D images. Our 3D particle field extraction process which consists of six different steps is shown in Fig.2. Here is a brief overview of the algorithm.

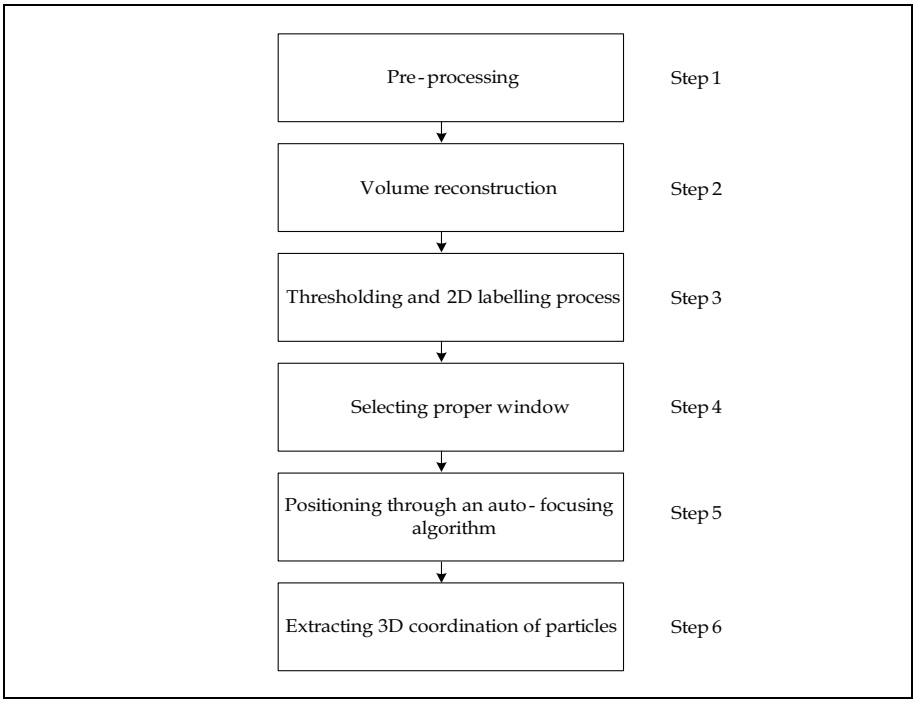


Fig. 2. General process for determining 3D particle locations

Step 1. The hologram of particles is pre-processed in this step to remove or suppress zero-order image from particle hologram by means of digital image processing technique, and to enhance the signal-to-noise ratio.

Step 2. The particle fields composed with amount of image slices are reconstructed from hologram at different depths. After reconstruction of each slice, median filtering is performed. Median filtering is effective in removing noise without disrupting edges, as long as the filter size is kept smaller than object feature sizes.

Step 3. This step is to identify the lateral position and size for particles in each depth by adaptive threshold. Then we will find all possible the centroids of particles. Note

that the lateral position and size of each window in reconstruction space varies depending on the reconstruction distance, growing larger and drifting farther from the optic axis as distance increases.

Step 4. Selecting proper window contains particle to be ready for digital image processing.

Step 5. An auto-focusing algorithm based on the most gradient for extracting particle axial position by selecting proper window in accordance with particle size is established. The focus measure is calculated for each window in each slice. For each window, the variation in the focus measure and irradiance over depth is analyzed. Hence, the axial depth of the particle in each window is determined. The variation of focus measure over depth is called the "focus curve." Note that the units are generally arbitrary, since the focus measure may not always have a direct physical significance.

Step 6. The final step is to extract 3D coordination of particles.

4. Auto-focusing algorithm for particle based on the most gradient

The particles that lie within the depth of focus of the imaging system are in focus (appearing sharp), while the particles that lie outside of the depth of field of the system are out of focus (appearing blurred). We now proceed to focus detection and demonstrate how, using a focus measure, we can determine the axial depth of particles in digital holographic reconstructions.

Focus measures are functions that attempt to determine the relative level of focus of sets of images. The accepted image property maximized by these functions is the high-spatial-frequency energy of the image. Hence, the edge of fully focused particle images has the larger gradient. So whether the particle is focused can be determined by calculating the edge gradient. Generally, the process for calculating the edge gradient of particles only calculates one or two gradient direction. However the real gradient direction for the edge of particle may not be the appointed direction, hence the gradients in all possible directions is calculated and chosen the largest gradient as the result.

To solve this problem, an auto-focusing algorithm based on the most gradient is proposed in this paper. First of all, the gradients in all possible directions is calculated in the focus window and chosen the largest gradient as the result. At the same time, taking into account the accuracy and the stability of algorithm, a threshold parameter which can effectively get rid of the background information and noise is introduced.

The edge gradient direction of a particle can only be in the horizontal direction, vertical direction, the 45 degree direction and negative 45 degree direction. In order to reduce the calculation error and obtain the real gradient direction, it is calculated as follows: the edge gradients in all directions are calculated in accordance with Formula (14) to (18) using any pixel including its eight surrounding pixels in focus window.

the edge of particle in horizontal direction (i=1)

$$\begin{cases} P1_i = f(x-1,y-1) + f(x-1,y) + f(x-1,y+1) \\ P2_i = f(x,y-1) + f(x,y) + f(x,y+1) \\ P3_i = f(x+1,y-1) + f(x+1,y) + f(x+1,y+1) \end{cases} \tag{14}$$

the edge of particle in vertical direction (i=2)

$$\begin{cases} P1_i = f(x-1,y+1)+f(x,y+1)+f(x+1,y+1) \\ P2_i = f(x-1,y)+f(x,y)+f(x+1,y) \\ P3_i = f(x-1,y-1)+f(x,y-1)+f(x+1,y) \end{cases} \tag{15}$$

the edge of particle in 45 degree direction (i=3)

$$\begin{cases} P1_i = f(x-1,y-1)+f(x-1,y)+f(x,y-1) \\ P2_i = f(x-1,y+1)+f(x,y)+f(x+1,y-1) \\ P3_i = f(x,y+1)+f(x+1,y)+f(x+1,y+1) \end{cases} \tag{16}$$

the edge of particle in -45 degree direction (i=4)

$$\begin{cases} P1_i = f(x-1,y)+f(x-1,y+1)+f(x,y+1) \\ P2_i = f(x-1,y-1)+f(x,y)+f(x+1,y+1) \\ P3_i = f(x,y-1)+f(x+1,y-1)+f(x+1,y) \end{cases} \tag{17}$$

$$\begin{cases} G1 = \left|P2_i - P1_i\right| \\ G2 = \left|P3_i - P2_i\right| \\ MAX_i = \max(G1,G2) \end{cases} \tag{18}$$

MAX1, MAX2, MAX3 and MAX4 are greater value in horizontal direction, vertical direction, the 45 degree direction and negative 45 degree direction. Because the real gradient direction can only be one of them, so a greater value in them can be selected as the real gradient direction. It is shown in Formula (19).

$$MAX = \max(MAX_1, MAX_2, MAX_3, MAX_4) \tag{19}$$

We can divide the focus window into several 3 × 3 windows to obtain real gradient value. However, due to the smaller 3 × 3 window, the result is noise-sensitive. In order to remove noise and reduce the impact of the background, a threshold having characterization of overall noise level distribution is defined as following

$$T = \sqrt{\frac{1}{M \times N} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (f(x,y) - F_{mean})^2} \tag{20}$$

where M, N and F_{mean} denote the number of row and column and the mean value in focus window.

After T is calculated, then we can calculate the variance σ^2 of all 3×3 windows. Only if σ^2 is greater than T, the corresponding window maintains the gradient value, otherwise the gradient value is set to zero. The specific formula is as following

$$Gradient = \begin{cases} MAX^2 & \text{if} \quad \sigma^2 \geq T \\ 0 & \text{else} \end{cases} \tag{21}$$

The purpose for square of most gradient is to further enhance the effect of target image gradient values because the values of the edge gradient are always larger than the values of

non-edge gradient. Finally, the gradient values of each 3×3 window can be summed as the final gradient focusing value.

5. Results and discussion

The above extraction method has been tested using standard particles which is approximately parallel to the recording plane. The light source used is a He-Ne laser. The wavelength of the generated laser beam is $\lambda = 632.8$nm . The diffraction patterns are recorded by means of a CCD camera with 2032×1520 pixels of 6.4 mm×4.8mm.

Fig. 3 shows the axial location result of auto-focusing algorithm based on the most gradient for a particle with 500μm diameter which is recorded at 100mm. There are three auto-focusing curves under different sampling interval at 100μm, 500μm and 1000μm. As can be seen from the Fig. 3, the best focus plane of the particle can be determined by the maximum value of auto-focusing evaluation function. The auto-focusing curves based on the most gradient have the characteristics with better single peak feature, even when the sampling interval is 100μm. Auto-focusing algorithm has the same pulse width the sampling interval. There are no location errors.

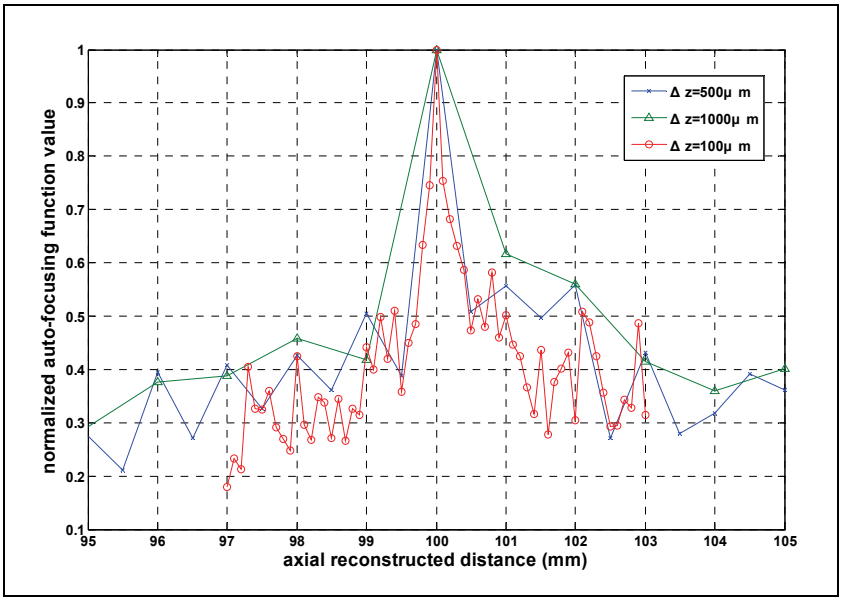


Fig. 3. Auto-focusing evaluation function based on the most gradient for a particle

Further two spherical particles with 500μm and 600μm diameter are selected as research objects. They are respectively located at 100mm and 105mm on sides of the optical glass. The results of auto-focusing function evaluation for two particles are shown in Fig.4. The axial sampling interval of auto-focusing curves is 500μm. As can be seen from the Fig.4, The auto-focusing curves based on the most gradient can accurately locate the axial depth position of two particles. Auto-focusing algorithm has the same pulse width the sampling interval. There are no location errors at 500μm sampling interval.

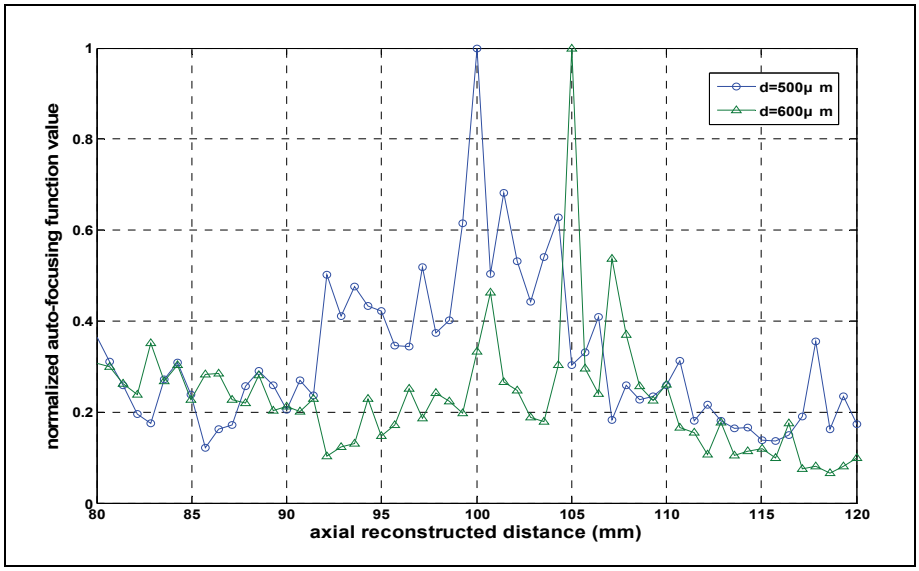


Fig. 4. Auto-focusing evaluation function based on the most gradient for two particles

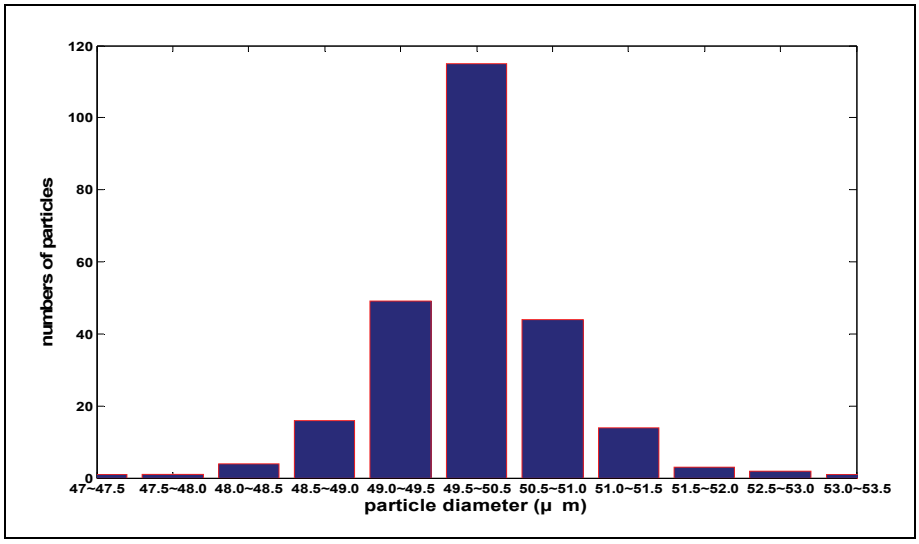


Fig. 5. The histogram of particles' diameter

To extract the particle field, we have built a sample volume corresponding to a set of reconstructed planes from 59.5mm to 60.5mm. The particle fields are consist of standard particles with 50μm diameter and 2 particles/mm3 concentration. Then, this volume has been processed according to the procedure discussed in Section 3. We have represented in Fig. 5 the histogram of the particles' diameter. This result is in good agreement with the standard particle distribution.

The above experimental study show that the axial locating errors mainly caused by larger depth of focus can be reduced by shortening the axial sampling interval and using auto-focusing algorithm based on the most gradient.

6. Conclusion

An auto-focusing algorithm based on the most gradient for extracting particle axial position by selecting proper window in accordance with particle size is established. This method provides a way of determining the depth of particles without prior knowledge of particle location. Using this algorithm, 3D coordination of particles can be extracted from the reconstructed volume. Taking into account the precision and stability of auto-focusing algorithm, another threshold parameter is introduced to remove noise and reduce the impact of the background. Experiment results show that the algorithm has the characteristics such as strong single peak feature and good stability.

7. Acknowledgment

This work was supported by Leading Academic Discipline Project of Shanghai Municipal Education Commission (Number: J51301) and the Program of Shanghai Municipal Education Commission under Grant (Number: 05LZ12)

8. References

Adams, M.; Kreis, T. & Jüptner, W. (2000). Particle analysis with digital holography. *Proceedings of SPIE*, pp. 314-320.

Dubois, F.; Schockaert, C.; Callens, N. & Yourassowsky, C. (2006). Focus plane detection criteria in digital holography microscopy. *Opt. Express* , Vol.14, 5895–5908.

Fournier, C.; Ducottet, C. & Fournel, T. (2004). Digital holographic particle image velocimetry: 3D velocity field extraction using correlation. Journal of Flow Visualization & Image Processing, Vol.11, pp. 53-72.

Ferraro, P.; Coppola, Nicola, G. Finizio, S. & Peirattini, G. (2003). Digital holographic microscope with automatic focus tracking by detecting sample displacement in real time. *Opt. Lett.* Vol.28, 1257–1259.

Gillespie, J.; King, R. (1989). The use of self-entropy as a focus measure in digital holography. *Pattern Recogn. Lett.* , Vol.9, 19–25.

GOODMAN, J. (2006). Introduction to Fourier optics. 3rd ed. Beijing: Publishing House of Electronics Industry, pp.141-143.

Kreis, T.; Jüptner, W. (1997). Suppression of DC term in digital holography. Optical Engineering, Vol. 36, No. 8, pp. 2357-2360.

Liebling, M.; Unser, M.. (2004). Autofocus for digital Fresnel holograms by use of a Fresnelet-sparsity criterion. *J. Opt. Soc. Am*. Vol. A21, 2424–2430.

Ma, L.; Wang, H.; Li,Y. & Jin, H. (2004). Numerical reconstruction of digital holograms for three-dimensional shape measurement. *J. Opt.*, Vol. A6, 396–400.

Meng, H.; Pan, G.; Pu, Y. et. al.. (2004). Holographic particle image velocimetry: from film to digital recording. *Meas. Sci. T*echnol., Vol. 15, pp.673-685.

Shen, G.; Wei, R. (2005). Digital holography particle image velocimetry for the measurement of 3Dt-3c flows. *Optics and Lasers in Engineering*, Vol. 43, pp. 1039-1055.

Takaki,Y.; Kawai, H. & Ohzu, H. (1999). Hybrid Holographic Microscopy Free of Conjugate and Zero-Order Images. *Applied Optics*, Vol. 38, No. 23, pp. 4990-4996.

15

Hybrid Numerical-Experimental Holographic Interferometry for Investigation of Nonlinearities in MEMS Dynamics

Minvydas Ragulskis[1], Arvydas Palevicius[2] and Loreta Saunoriene[3]
[1,3]*Research Group for Mathematical and Numerical Analysis of Dynamical Systems, Kaunas University of Technology, Studentu 50-222, 51638 Kaunas*
[2]*International Studies Center, Kaunas University of Technology, A. Mickeviciaus 37, 44244 Kaunas*
Lithuania

1. Introduction

Holographic interferometry is a powerful experimental technique for analysis of structural vibrations, especially if the amplitudes of those vibrations are in the range of micrometers (Caponero et al. (2000); Fagan et al. (1972); Fein (1997); Ganesan et al. (2000); Rastogi (2000); Vest (1979)). Recent advancements in optical measurement technology and development of hybrid numerical-experimental techniques require application of computational algorithms not only for post-processing applications like interpretation of experimental patterns of fringes, but embedding real time algorithms into the measurement process itself (Ragulskis & Saunoriene (2007)).

Computation and plotting of patterns of time average holographic fringes in virtual numerical environments involves such tasks as modelling of the optical measurement setup, geometrical and physical characteristics of the investigated structure and the dynamic response of the analysed system (Ragulskis et al. (2003)). Calculation of intensity of illumination at any point on the hologram plane requires computation of definite integrals over the exposure time. If the analysed structures perform harmonic oscillations that do not impose any complications – there exist even analytical relationships between the intensity of illumination, amplitude of oscillation, laser wavelength, etc. But if the oscillations of the investigated structures are non-harmonic (what is common when structures are non-linear) and the formation of patterns of fringes is implemented in the real time mode, the calculation of definite integrals becomes rather problematic. One of the objects of this study is to propose an order adaptive algorithm which could be effectively applicable for calculation of definite integrals in different real time holography applications.

Another goal of this study is to show that holographic interferometry, being a non-destructive whole field technique capable of registering micro oscillations of micro electromechanical systems (MEMS) components, cannot be exploited in a straightforward manner (Ostasevicius et al. (2005)). There exist numerous numerical methods and techniques for interpretation of patterns of fringes in the registered holograms of different oscillating objects and surfaces. Unfortunately, sometimes straightforward application of these motion reconstruction

methods (ordinary fringe counting technique, etc.) does not produce acceptable results. A typical example is holographic analysis of a MEMS switch dynamics which will be described in this chapter.

2. Hybrid numerical-experimental optical investigation of non-linearity in MEMS dynamics

2.1 Technological features of the MEMS cantilever

Investigation of dynamics of micro electromechanical systems (MEMS) is an important problem of engineering, technology and metrology. Specifically, recent interest in applying MEMS technology to miniaturization of relays, sensors, actuators for variety of applications requires design of appropriate testing and measurement tools for investigation of dynamic properties of those systems. Though MEMS technology offers great promise in addressing the need for smaller dynamical systems, the development of new types of MEMS structures is very costly and complicated procedure. On the other hand, straight application of principles of design of macro-mechanical systems is rather limited in MEMS applications. Therefore, application of measurement technologies capable of detecting the dynamic properties of micro scale systems may help to understand and evaluate the functionality of the systems.

Hybrid numerical-experimental optical techniques are applied for investigation of micro-mechanical relays, in particular their cantilevers. The apparent simplicity of the problem is misguiding due to non-linear interaction between the cantilever and the bottom electrode. Therefore the results of optical measurements of the cantilever dynamics are inaccurate due to the shift of the fringes in time average laser holographic interferograms. Numerical modelling helps to solve non-uniqueness of the inverse problem and to validate the interpretation of the pattern of fringes.

The fabrication sequence of the micro-electromechanical switch begins with the patterning and reactive ion etching of silicon using SF_6/N_2 gas chemistry in the cantilever source (support) area fabricating microstructures to increase the cantilever bond strength either durability of the device. After treatment of the substrate in the O_2/N_2 gases mixture plasma chrome layer of about 30 nm thickness and gold layer of about 200 nm thickness were deposited. Patterning of the source, gate and drain electrodes were performed using lift-off lithography. Electron beam evaporation was performed to deposit a sacrificial copper layer with thickness of about 3000 nm. Copper layer covered the whole area of the substrate. Patterning of the copper layer was performed in two steps. First of all, the copper layer was partially etched (etchant: H_2SO_4:CrO_3:H_2O) to define the contact tips for the cantilever and etching duration directly determined the spacing between tip's top and drain electrode. Next, the copper layer was etched away to uncover the source cantilever support area. The next step was photo resist patterning on the top of the sacrificial layer to define the mask for the cantilever sector and lift-off lithography of the evaporated gold layer with thickness of about 200 nm was performed. Afterwards, the photoresist was spun and patterned once again in the same sector and thick nickel layer was electroplated (sulfamate electrolyte: $Ni(NH_2SO_3)_2$:$4H_2O$) fabricating cantilever structure. Finally, the sacrificial layer was removed away using the same wet copper etchant to release the free-standing cantilever. The general view of MEMS cantilever is presented in Fig.1.

Figure 2 presents time average holograms of the previously described cantilever. The method used for registering optical holograms is described in Ostasevicius et al. (2004). Figure 2 (a) presents a hologram of acoustically excited cantilever. Figure 2(b) presents a hologram of cantilever excited by oscillating charge of the drain electrode.

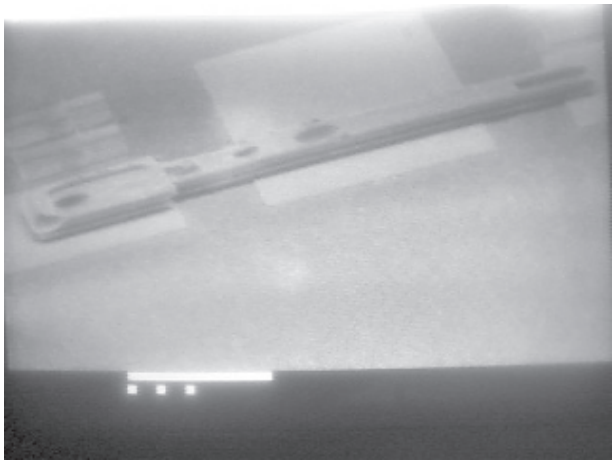
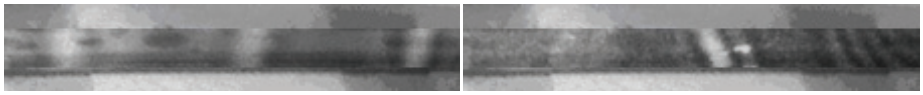


Fig. 1. Microscopic photo of MEMS cantilever



(a) (b)

Fig. 2. Holograms of cantilever: (a) acoustically excited cantilever; (b) excited by oscillating charge

Interpretation of hologram in Fig. 2(a) is straightforward – white fringes are regularly distributed on the surface of cantilever and ordinary fringe counting techniques can be applied for the reconstruction of the field of vibration amplitudes. The interpretation of pattern of fringes in Fig. 2(b) is much more complicated – one white and several dark fringes are distributed on the surface of cantilever and it is quite difficult to understand the dynamics of the cantilever. Moreover, longer exposure times produce dark images when the cantilever is excited by oscillating charge and the pattern of fringes is sensitive to exposure time. Such effects probably are caused by specific dynamical properties of the system.
Difficulties in interpretation of the generated patterns of fringes in holograms of MEMS cantilevers excited by oscillating charge originated the necessity for developing hybrid numerical-experimental models of analysed systems. Simulation of the dynamic as well as optical processes taking place in the analysed systems could help understanding experimental results.

2.2 Phenomenological model of MEMS cantilever

First step in analysing complex MEMS cantilever system is developing of simple one-dimensional phenomenological model which is presented in Fig. 3.
Governing equation of motion of the system presented in Fig. 3 takes the following form:

$$m\ddot{x} + h\dot{x} + kx = F_e(x), \tag{1}$$

where m, h, k – mass, viscous damping and stiffness coefficients; x – coordinate; F_e – electrostatic force; top dots denote full derivatives by time t. It is assumed that the charged

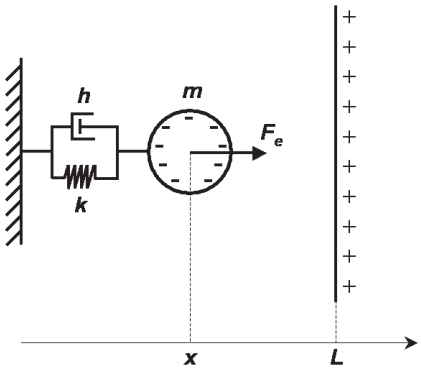


Fig. 3. One degree of freedom model of MEMS cantilever

contact plane is motionlessly fixed at coordinate L (Fig. 3). Mass m is negatively charged, while the charge Q of the contact plane varies harmonically in time:

$$Q = q \sin(\omega t), \tag{2}$$

where q – maximum charge of the contact plane; ω – frequency of charge oscillation. Then the electrostatic force F_e acting to mass m is

$$F_e = C\frac{Q}{L - x}, \tag{3}$$

where constant C depends from the charge of mass m, density of air, etc.

Harmonic oscillation of the charge of the contact plane can excite strongly non-linear response. That is a natural result as the governing equation of motion is non-linear. The non-linearity of the motion of the mass m is especially clear when the frequency of charge oscillation is around the natural frequency of the mechanical system (Fig. 4). In practical applications the frequency of charge oscillation very rarely reaches the fundamental frequency of the MEMS cantilever only due to the fact that it is very high. Usually the excitation frequencies are of magnitude lower than fundamental frequencies.

The presented one degree of freedom cantilever model is analysed when the excitation frequency is much lower that the fundamental frequency. It can be noted that the differential equation turns to be stiff and special care is required applying direct time marching integration techniques. First the system is integrated until the transient processes cease down. Then the attractor in phase plane $x - \dot{x}$ is drawn. Repetition of such procedure at different values of maximum charge q produces an array of attractors as shown in Fig. 5. It can be noted that the motion of the mass m is almost linear at small excitation. The form of the stable attractor gradually deforms at increasing excitation what is a natural result for a non-linear system.

One of the classical techniques used for experimental analysis of MEMS cantilevers is time average laser holography. It is a non-destructive whole field technique capable of detection microscopic dynamic displacements of the objects performing harmonic vibrations. How could the experimental results be interpreted if the analysed object would oscillate not harmonically, but as shown in Fig. 5.

One-dimensional system is considered for simplicity. Then the intensity of illumination I in the hologram plane will be:

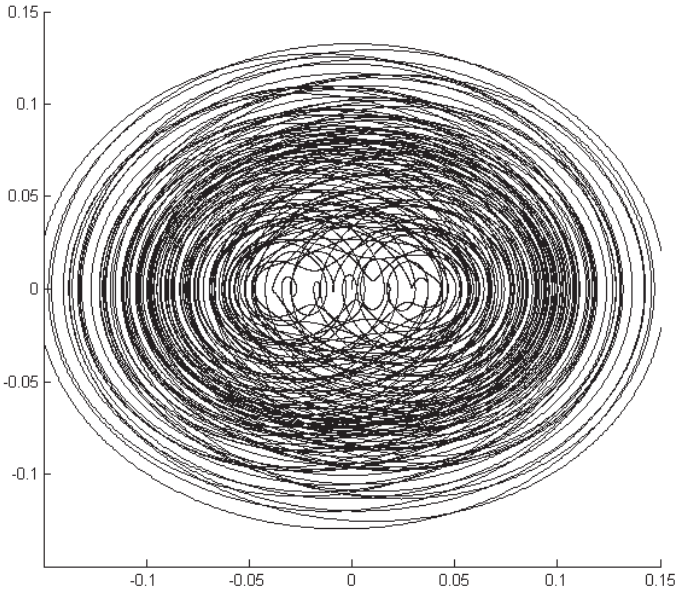


Fig. 4. Chaotic motion of mass m in phase plane $x - \dot{x}$ at $m = 1$; $k = 1$; $h = 0.003$; $\omega = 0.73$; $q = 0.1$; $L = 2$

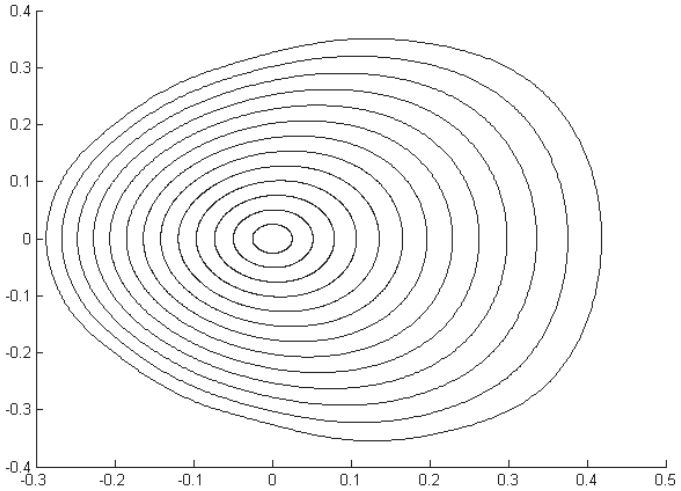


Fig. 5. Stable attractors at $m = 0.25$; $k = 20$; $h = 0.1$; $\omega = 1$; $L = 2$; $q = 1, 2, \ldots 13$

$$I = \lim_{T\to\infty} \frac{1}{T^2} \left| \int_0^T \exp\left(j\frac{2\pi}{\lambda} \zeta(t) \right) dt \right|^2, \tag{4}$$

where T – exposure time; λ – laser wavelength; ζ – scalar time process; j – imaginary unit. When $\zeta(t) = u\sin(\omega t + \varphi)$ where u, ω, φ – amplitude, angular frequency and phase of oscillations, the intensity of illumination takes the form:

$$I = \left|J_0\left(\frac{2\pi}{\lambda}u\right)\right|^2 = \lim_{T\to\infty}\frac{1}{T^2}\left(\int_0^T \cos\left(\frac{2\pi}{\lambda}u\sin(\omega t+\varphi)\right)dt\right)^2 \approx \left(\frac{1}{m}\sum_{i=1}^{m}\cos\left(\frac{2\pi}{\lambda}u\sin\left(\frac{2\pi}{m}(i-1)\right)\right)\right)^2, \tag{5}$$

where J_0 – zero order Bessel function of the first kind. It can be noted that $\lim_{T\to\infty}\int_0^T \sin\left(\frac{2\pi}{\lambda}u\sin(\omega t+\varphi)\right)dt = 0$ due to evenness of the sine function, and that the angular frequency and phase have no effect to the intensity of illumination. The second approximate equality builds the ground for numerical modelling of the relationships governing the formation of interference fringes.

If $\zeta(t)$ is not a harmonic process the intensity of illumination can be numerically reconstructed from Eq. (4), but the calculation is more complex than in Eq. (5) due to the fact the integral $\lim_{T\to\infty}\int_0^T \sin\left(\frac{2\pi}{\lambda}\zeta(t)\right)dt$ does not converge to zero. If $\zeta(t)$ is a periodic process and T_p is the time length of the period, the approximate numerical calculation scheme takes the following form:

$$I \approx \left(\frac{1}{m}\sum_{i=1}^{m}\cos\left(\frac{2\pi}{\lambda}\zeta\left(t_0+\frac{T_p}{m}(i-1)\right)\right)\right)^2 + \left(\frac{1}{m}\sum_{i=1}^{m}\sin\left(\frac{2\pi}{\lambda}\zeta\left(t_0+\frac{T_p}{m}(i-1)\right)\right)\right)^2, \tag{6}$$

where t_0 – arbitrary selected time moment. If $\zeta(t)$ is a process characterising time history of a dynamical system setting to a stable limit cycle type attractor, time moment t_0 must be selected large enough so that the transient processed are ceased. Such calculations are performed for an array of attractors shown in Fig. 6. One hundred separate solutions of Eq. (1) are analysed at intermittent values of q in the range from 0 to 13. The produced intensities of illumination are presented in Fig. 6 (x axis denotes 100 separate problems).

Remarkable is the fact that the relationship between the intensity of illumination and parameter q holds the same character as the square of Bessel function in Eq. (5). In other words, the inverse problem of the reconstruction of dynamic displacements does not have a unique solution. Interference fringes can be formed when the analysed object performs harmonic vibrations. Very similar interference fringes will be formed when the object will perform non-linear periodic oscillations. Thus though time average laser holography is a very attractive technique for analysis of MEMS cantilever vibrations, the interpretation of experimentally produced interference fringes is rather complex procedure if one cannot be sure if the vibrations are harmonic. This effect is illustrated in Fig. 7. The presented non-linear periodic oscillation and harmonic vibration will both generate the same intensity of illumination corresponding to the centre of the sixth interference fringe. It can be noted that time average laser holography is insensitive to static shifts of harmonic oscillations. That follows from the property of Bessel function:

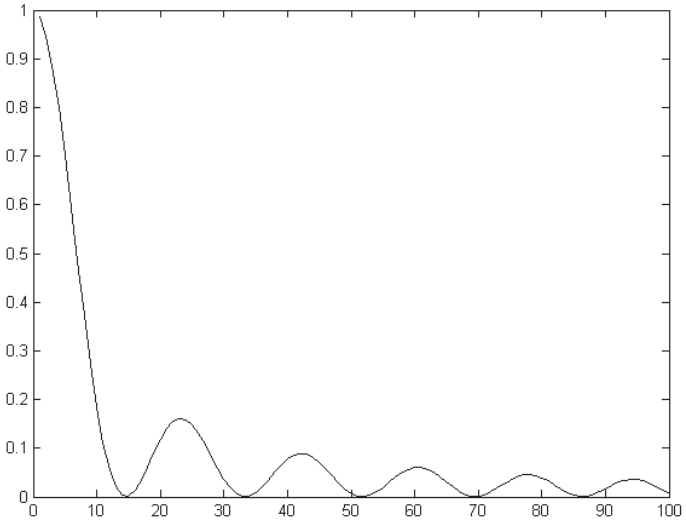


Fig. 6. Relationship between intensities of illumination and parameter q

$$\begin{aligned} &\lim_{T\to\infty}\frac{1}{T^2}\left|\int_0^T \exp\left(j\tfrac{2\pi}{\lambda}\left(u\sin\left(\omega t+\varphi\right)+C\right)\right)dt\right|^2 \\ &=\lim_{T\to\infty}\frac{1}{T^2}\left|\int_0^T \exp\left(j\tfrac{2\pi}{\lambda}\left(u\sin\left(\omega t+\varphi\right)\right)\right)dt\right|^2=\left(J_0\left(\tfrac{2\pi}{\lambda}u\right)\right), \end{aligned} \tag{7}$$

where C – constant. Results presented in Fig. 7 are remarkable not for the difference between the averages of non-linear and harmonic vibrations. Interesting is the fact that two different trajectories generate same intensity of illumination.

The problem of reconstruction of dynamic displacements from a pattern of interference fringes is even more complicated when the oscillations are chaotic. In that case the illumination of intensity can be calculated for certain stochastic time series approximating process $\zeta(t)$. If a time series ζ_i is normally distributed with variance σ^2 then, the decay of intensity of illumination can be calculated as follows:

$$\begin{aligned} I &\approx \lim_{m\to\infty}\left(\left(\frac{1}{m}\sum_{i=1}^{m}\cos\left(\tfrac{2\pi}{\lambda}\zeta_i\right)\right)^2+\left(\frac{1}{m}\sum_{i=1}^{m}\sin\left(\tfrac{2\pi}{\lambda}\zeta_i\right)\right)^2\right) \\ &=\lim_{m\to\infty}\left(\frac{1}{m}\sum_{i=1}^{m}\sum_{k=0}^{+\infty}\frac{(-1)^k\left(\frac{2\pi}{\lambda}\zeta_i\right)^{2k}}{(2k)!}\right)^2=\left(\sum_{k=0}^{+\infty}\frac{(-1)^k\left(\frac{2\pi}{\lambda}\right)^{2k}}{(2k)!}\lim_{m\to\infty}\sum_{i=1}^{m}\frac{(\zeta_i)^{2k}}{m}\right)^2 \\ &=\left(\sum_{k=0}^{+\infty}\frac{(-1)^k\left(\frac{2\pi}{\lambda}\right)^{2k}}{(2k)!}\cdot(2k-1)!!\cdot\sigma^{2k}\right)^2=\left(\sum_{k=0}^{+\infty}\frac{(-1)^k\left(\frac{2\pi}{\lambda}\sigma\right)^{2k}}{(2k)!!}\right)^2 \\ &=\left(\sum_{k=0}^{+\infty}\frac{(-1)^k\left(\frac{2\pi}{\lambda}\sigma\right)^{2k}}{2^k k!}\right)^2=\left(\sum_{k=0}^{+\infty}\frac{(-1)^k}{k!}\left(\frac{1}{2}\left(\frac{2\pi}{\lambda}\sigma\right)^2\right)^k\right)^2=\exp^2\left(-\frac{1}{2}\left(\frac{2\pi}{\lambda}\sigma\right)^2\right). \end{aligned} \tag{8}$$

The following identities are used in Eq. (8). If $\zeta\sim N(0,\sigma^2)$, then

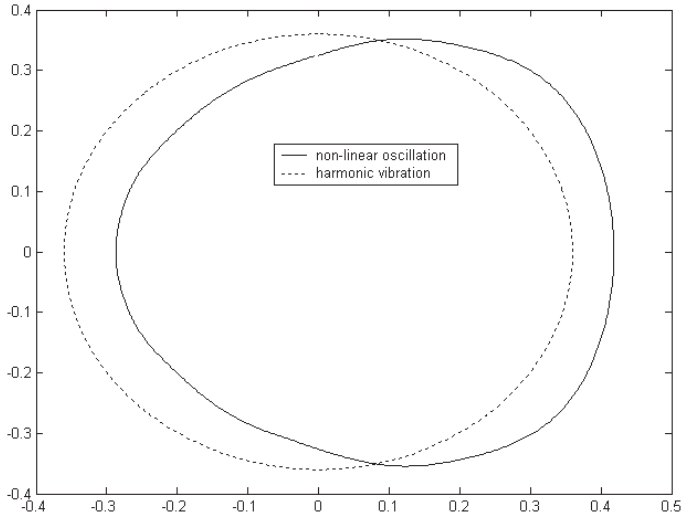


Fig. 7. Two different trajectories generating same intensity of illumination

$$E\zeta^{2k-1} \equiv 0, \quad k = 1, 2, 3, ...;$$
$$E\zeta^{2k} \equiv 1 \cdot 3 \cdot ... \cdot (2k-1)\,\sigma^{2k} = (2k-1)!!\sigma^{2k}, \quad k = 1, 2, 3, ... \tag{9}$$

It can be noted, that in this case no interference fringes will be formed at all – the intensity of illumination will gradually decrease at increasing variance σ^2.

This is a classical example of an ill-posed inverse problem. The problem of reconstruction of motion from a pattern of interference fringes has a unique solution only if the vibration of the analysed structure is harmonic. When the oscillations are non-linear (what is likely in MEMS cantilever dynamics) the interpretation of pattern of fringes can be complicated. Next sections address questions related to numerical construction of time averaged holographic interferograms and computational aspects of the calculation of definite integrals in real time applications.

3. Order adaptive quadrature rule for real time holography applications

As mentioned previously, computation of the intensity of illumination at any point on the hologram plane requires computation of definite integrals over the exposure time. But standard higher order Newton-Cotes quadrature formulas (Davis & Rabinowitz (1984)) require that the number of nodes must be a divisible numeral. For example the second order Newton-Cotes rule already requires that the number of nodes must be odd. Such conditions mean that a significant number of nodes at the end of an experimental time series must be deleted and the integration interval artificially shortened for higher order Newton-Cotes rule, if the number of nodes is not known at the beginning of the experiment. Therefore there exists a definite need for a high order integration rule with a constant time step without any requirement for the number of time steps. Such quadrature formula is proposed in Ragulskis & Saunoriene (2008):

$$\int_{t_0}^{t_0+(k-1)h} f(t)\,dt = \left(\sum_{i=1}^{m} a_i f_i + \sum_{i=1}^{k-2m} f_{m+i} + \sum_{i=1}^{m} a_{m-i+1} f_{k-m+i}\right) h, \tag{10}$$

where a_i are the weights and f_i are the discrete values of sampled function f at time moments $t_0 + (i-1)\cdot h$, $i = 1,\ldots,k$. It has been proved that this integration rule is exact when the integrated function is a polynomial of the m-th order, if only m is odd (Ragulskis & Saunoriene (2008)). The numerical values of the weights a_i are presented in the Table 1 at different values of m. The parameter p in this table denotes the maximum order of exactly integrated polynomials; l is the order of the error term expressed in the form $O(h^l)$.

m	2	3	4	5	6	7
a_1	0.5	0.37500000	0.33333333	0.32986111	0.31875000	0.30422454
a_2	1	1.1666667	1.2916667	1.3208333	1.3763889	1.4603836
a_3		0.95833333	0.83333333	0.76666667	0.65555556	0.45346396
a_4			1.0416667	1.1013889	1.2125000	1.4714286
a_5				0.98125000	0.92569444	0.73939319
a_6					1.0111111	1.0824735
a_7						0.98863261
p	1	3	3	5	5	7
l	2	4	4	6	6	8

Table 1. Nodal weights of the integration rule

It can be noted that finite element method was used for the derivation of the proposed quadrature rule which can be interpreted as a new variant of Gregory type formulas (Davis & Rabinowitz (1984)). Unfortunately, the proposed quadrature rule (also Gregory type rules) can be used only when the order is predefined before the experiment and does not change over the integration process. This paper proposes a multi-processor parallel algorithm with full order adaptability in real time calculation mode.

3.1 The basic real time integration rule

Let's suppose that function f is sampled starting from t_0 at equally spaced time steps; the length of a time step is h. Due to the real time process the number of nodes is not predefined before the experiment and process continues until the end of the sampling. Let's suppose that the terminal moment of the sampling occurs at $t_0 + 7h$ (8 function values f_i, $i = 1,\ldots,8$ are produced during the sampling process). Order of the integration rule is predetermined to be $m = 3$.

1. The first sum on the right side of Eq. (10) is computed:

$$Sum_1 = a_1 f_1 + a_2 f_2 + a_3 f_3, \tag{11}$$

   where $a_1 = 0.375, a_2 = 1.1666667, a_3 = 0.95833333$ (Table 1).

2. Starting from the fourth node, the following sum is computed until the end of the time series:

$$Sum_2 = f_4 + f_5 + f_6 + f_7 + f_8. \tag{12}$$

3. When the sampling is terminated, reverse computation of the third sum of eq. (10) is done:

$$Sum_3 = (a_3 - 1)\, f_6 + (a_2 - 1)\, f_7 + (a_1 - 1)\, f_8. \tag{13}$$

4. Finally, the definite integral $\int_{t_0}^{t_0+7h} f(t)\, dt$ is calculated according to eq. (10):

$$I = (Sum_1 + Sum_2 + Sum_3)\, h. \tag{14}$$

The process can terminate at any time step, if only $k \geq 2m$, but the last three values of the sampled function must be saved at every time moment in order to calculate Sum_3.
Now we will generalize the presented example for m-th order integration rule, if only the minimum number of nodes is $2m$. The algorithm is based on Master-Worker paradigm (Mattson et al. (2004)). Schematic graphical representation in Fig. 8 helps to interpret the computation process.

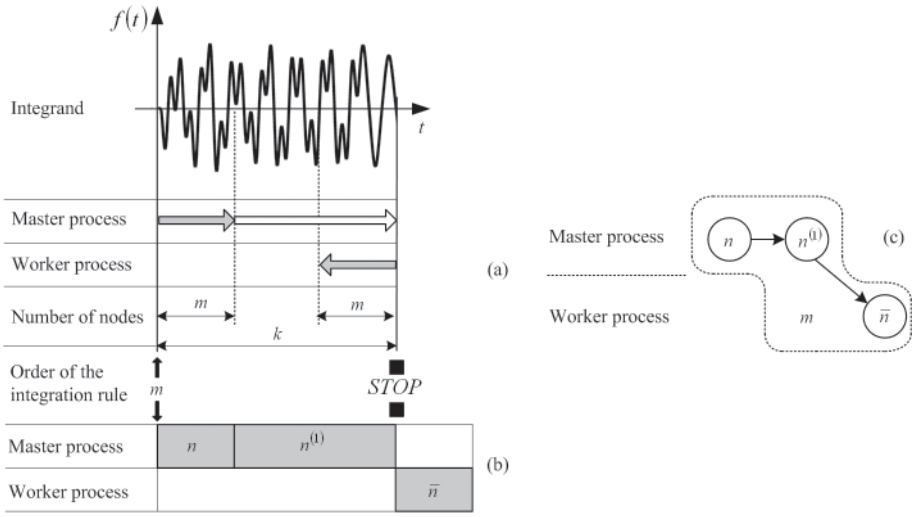


Fig. 8. Schematic representation of the basic model:(a) signal diagram; (b) time diagram; (c) flow chart diagram

Several notations used in Fig. 8 can be explained in more detail. Order of the integration rule m is predefined before the experiment. Calculation of Sum_1 is performed by Master processor (grey right arrow in signal diagram; block n in time diagram and node n in flow chart diagram). After m terms are included into Sum_1, the Master processor continues summation of nodal values of the integrand until the sampling process is terminated (white right arrow in signal diagram; block $n^{(1)}$ in time diagram and node $n^{(1)}$ in flow chart diagram). When the sampling is over, Worker processor performs reverse calculation of Sum_3 (grey left arrow in signal diagram; block $\bar{n}$ in time diagram and node $\bar{n}$ in flow chart diagram).
It can be noted that the last m values of the sampled function must be remembered at every time node in order to calculate Sum_3.

3.2 Order adaptive algorithm for real time applications

The presented basic real time integration rule copes well with integrands which can be approximated by a polynomial of a definite order in the domain of integration. But if the variation of the integrand is fast in some regions and slow in another regions, then order adaptability should be used to increase the accuracy of a definite integral. One can suggest to select very large m at the beginning of the experiment, but then we may face the risk that $k < 2m$.

We assume that there exists a detector which measures the values of the integrand and recommends the order of integration rule at any time moment in the domain of integration. Let's assume that the present order is m_1 and the detector recommends order m_2. Then two different situations may occur. If the number of sampled nodes since m_1 was declared is higher than or equal to $2m_1$, the transition to order m_2 can be performed fluently. The Master processor starts calculating Sum_1 for order m_2, while Worker processor takes care for reverse calculation of Sum_3 for terminated m_1.

But if the number of sampled nodes since m_1 was declared is less than $2m_1$, the Worker processor cannot start reverse calculation of Sum_3 without damaging Sum_1. Therefore a much more complex transition to order m_2 takes place in this situation. If m_2 is higher than m_1 the Worker processor must return to the point where order m_1 was declared and must recalculate Sum_1 with order m_2. But the simplicity is misleading – the Master processor has already summated Sum_1 with order m_1 to the total sum! Therefore the Worker processor must evaluate different weighting coefficients for orders m_1 and m_2. Moreover, the length of the queue where the last function values are stored must be already not m_i, but $2m_i$ (here m_i is the current order).

If m_2 is lower than m_1, but the number of sampled nodes since m_1 was declared is less than $2m_1$, the integration with order m_1 must be continued until the number of nodes is equal to $2m_1$, and only after that the order m_2 can be accepted.

Finally, we may comment what would happen if the sampling process is terminated and the number of sampled nodes since m_i was declared is lower than $2m_i$. Unfortunately, there will be no any possible techniques to preserve order m_i (time step is constant and reverse sampling with smaller time step is impossible in real time mode). The only solution is to select maximum possible order for the available number of time steps (floored half of the number of time steps).

We will illustrate the described situations with the following example (Fig. 9 and Fig. 10).

One Master processor and two Worker processors are necessary for full real time mode. Algorithm control, integrand sampling and summation of sums Sum_1 and Sum_2 is performed by Master processor. The Worker processors run only when the order is changed. Worker processors send back the results to the Master processor.

As an extreme situation we describe the transition from order m_3 to order m_4 (Fig. 9) where the second Worker processor is necessary for real time integration. Master processor starts calculating Sum_1 (with weights corresponding to order m_3) as soon as the order m_3 is declared. The number of discrete time nodes necessary for this procedure is m_3. As soon as Sum_1 is finished, Master processor starts summing non-weighted discrete function values. This process continues until order m_4 is declared. But the order detector has sensed a burst in the digital time series, so m_4 is much higher than m_3. In this particular situation we have that m_4 is even greater than k_3 (Fig. 9). Thus, the Worker processor must recalculate both the old Sum_1 and the rest non-weighted part (n_3 and $n_3^{(1)}$ in time diagram). Moreover, at the same time Sum_3 for order m_2 must be accomplished ($\bar{n}_2$ in time diagram). Thus Worker-2 processor is unavoidable for real time computation ($\bar{n}_2$ and $\bar{n}_3$ overlap in time diagram).

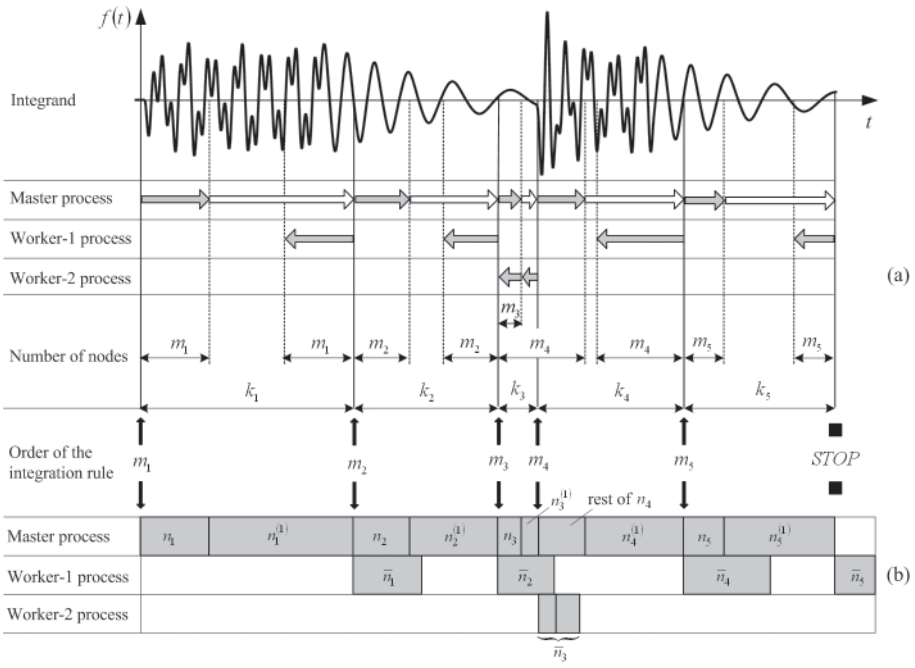


Fig. 9. Real time integration, general case: (a) signal diagram; (b) time diagram

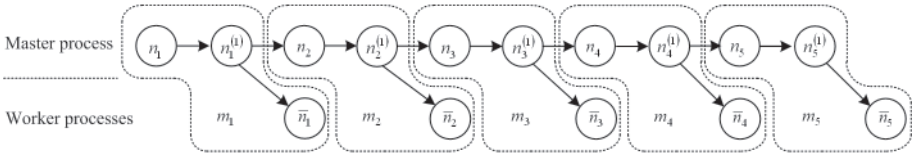


Fig. 10. Real time integration, general case: flow chart diagram

The presented procedure for real time calculation of definite integrals can be effectively applied in hybrid numerical-experimental techniques where time average intensities of illumination are reconstructed in virtual computational environment. Implementation of the proposed integration rule enables full real time computations with minimal data queue lengths and effective management of integration order.

4. Plotting holographic interferograms for visualisation of dynamic results from finite element calculations

The amount of data produced by finite element calculation places a particular challenge to scientific visualization. Meaningful and accurate visualization is an important task of computer modeling, especially when hybrid numerical-experimental techniques are used to construct digital holographic interferograms of such objects as non-linear MEMS cantilevers. Procedures plotting computer generated interferograms from the results of FEM analysis based on the principles of optical holography can provide realistic view of dynamic processes

taking place in such structures. Detailed description of the formation of digital holographic interferograms applicable for hybrid numerical-experimental techniques will be given in this chapter.

4.1 The coupling of FEM model with optical relationships

Let the planes A and B be defined in the 3 dimensional space (Fig. 11). Let the triangle defined by points 1, 2 and 3 be located in the plane A. This triangle is interpreted as an elementary body surface unit meshed using the Finite Element Method. Let the coordinates of the point r_1 define the observation point and the projection plane be defined by the points r_2, r_3 and r_4:

$$r_i = (x_i, y_i, z_i), \quad i = 1, 2, 3, 4, \tag{15}$$

where r_2 is the origin of the projection plane and the vectors $(r_3 - r_2)$, $(r_4 - r_2)$ form the ortho-normal base of the projection plane. The L – coordinates of the point of intersection of the line defined by the point of the structure (x, y, z) and the observation point with the projection plane can be obtained using the relationship in Appendix A, where the system matrix is substituted by:

$$\begin{bmatrix} x_2 & x_3 & x_4 & -x_1 & -x \\ y_2 & y_3 & y_4 & -y_1 & -y \\ z_2 & z_3 & z_4 & -z_1 & -z \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}. \tag{16}$$

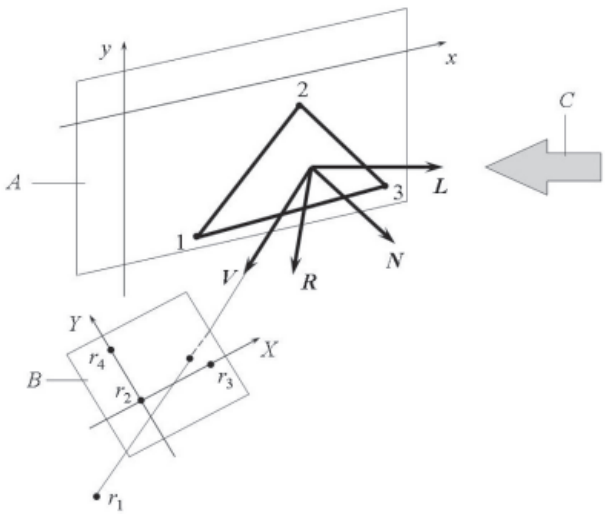


Fig. 11. Determination of the intensity of the surface: A – plane of the structure, B – projection plane, C – the incident laser beam

Then the spatial coordinates of the point of intersection $(\tilde{x}, \tilde{y}, \tilde{z})$ are:

$$\begin{aligned} \tilde{x} &= L_4 x_1 + L_5 x, \\ \tilde{y} &= L_4 y_1 + L_5 y, \\ \tilde{z} &= L_4 z_1 + L_5 z. \end{aligned} \tag{17}$$

The coordinates of the point (X, Y) in the projection plane can be obtained:

$$\begin{aligned} X &= (\tilde{x} - x_2)(x_3 - x_2) + (\tilde{y} - y_2)(y_3 - y_2) + (\tilde{z} - z_2)(z_3 - z_2), \\ Y &= (\tilde{x} - x_2)(x_4 - x_2) + (\tilde{y} - y_2)(y_4 - y_2) + (\tilde{z} - z_2)(z_4 - z_2). \end{aligned} \tag{18}$$

Assuming the projection plane corresponds to the computer graphical screen and contains n rows and m columns of pixels, the calculations are performed for every pixel of the mapped interferogram:

$$\begin{aligned} &X = X_{\min} + (j-1)(X_{\max} - X_{\min})/(m-1), \\ &Y = Y_{\max} - (i-1)(Y_{\max} - Y_{\min})/(n-1), \\ &i = 1, ..., n; \quad j = 1, ..., m, \end{aligned} \tag{19}$$

where the subscripts min and max indicate the minimum and maximum values of the corresponding quantities.
The unit normal vector of the projection plane is:

$$\begin{aligned} x_0 &= (y_3 - y_2)(z_4 - z_2) - (z_3 - z_2)(y_4 - y_2), \\ y_0 &= (z_3 - z_2)(x_4 - x_2) - (x_3 - x_2)(z_4 - z_2), \\ z_0 &= (x_3 - x_2)(y_4 - y_2) - (y_3 - y_2)(x_4 - x_2). \end{aligned} \tag{20}$$

Finally the spatial coordinates (x^*, y^*, z^*) of the point in the projection plane can be determined from:

$$\begin{aligned} &\begin{bmatrix} x_3 - x_2 & y_3 - y_2 & z_3 - z_2 \\ x_4 - x_2 & y_4 - y_2 & z_4 - z_2 \\ x_0 & y_0 & z_0 \end{bmatrix} \begin{Bmatrix} x^* \\ y^* \\ z^* \end{Bmatrix} \\ &= \begin{Bmatrix} X + x_2(x_3 - x_2) + y_2(y_3 - y_2) + z_2(z_3 - z_2) \\ Y + x_2(x_4 - x_2) + y_2(y_4 - y_2) + z_2(z_4 - z_2) \\ x_2 x_0 + y_2 y_0 + z_2 z_0 \end{Bmatrix}. \end{aligned} \tag{21}$$

It is assumed that the analysed structure is located in the plane $z = 0$ in the status of equilibrium, and $(\hat{x}_i, \hat{y}_i)$, $i = 1, 2, 3$ are nodal coordinates of a linear triangle element. Then the L-coordinates can be determined using the relationship in Appendix A, where the system matrix is substituted by:

$$\begin{bmatrix} \hat{x}_1 & \hat{x}_2 & \hat{x}_3 & -x_1 & -x^* \\ \hat{y}_1 & \hat{y}_2 & \hat{y}_3 & -y_1 & -y^* \\ 0 & 0 & 0 & -z_1 & -z^* \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}. \tag{22}$$

Thus the approximate coordinates of the intersection point $(\alpha, \beta, 0)$ between the line through the observation point and the pixel (i, j) and the surface of the analysed structure are obtained by interpolation.
It is assumed that the laser rays are coherent and parallel to each other.
The unit vector L from the point of the structure in the state of equilibrium to the isophase cross section of the laser beam in the direction of the beam is predefined.
The normal vector N coincides with the unit vector of the z axis:

$$N = [N_x, N_y, N_z]^T = [0, 0, 1]^T, \tag{23}$$

The relationship between the direction to the light source and the direction of reflection can be described by the following equation (Ivanov & Batrakov (1995)):

$$R = \begin{bmatrix} 2N_x^2 - 1 & 2N_xN_y & 2N_xN_z \\ 2N_xN_y & 2N_y^2 - 1 & 2N_yN_z \\ 2N_xN_z & 2N_yN_z & 2N_z^2 - 1 \end{bmatrix} \cdot L, \tag{24}$$

where R is the direction of reflection. The vectors L, N, R (Fig. 11) are complanar, the angle between L and N is equal to the angle between N and R.
Higher order elements can be subdivided into triangles and the approximate local coordinates of the element can be obtained by interpolation from the local coordinates of the nodes of the triangle.
For the three dimensional problem the calculations are performed for a sequential number of values of the local coordinates (ξ, η) in the current finite element:

$$\begin{aligned} \xi &= -1 + \tfrac{2}{n-1}(i-1), \quad i = 1, ..., n, \\ \eta &= -1 + \tfrac{2}{n-1}(j-1), \quad j = 1, ..., n. \end{aligned} \tag{25}$$

The spatial orthogonal Cartesian coordinates of these points can be calculated using the shape functions of the analysed finite element. Afterwards the corresponding values X and Y are obtained from Eq. (18).
The reconstructed digital image consists of the matrix of pixels where the columns are indexed from 0 to m_x and the rows – from 0 to m_y. Thus the point (X, Y) is mapped to the pixel (i_x, i_y):

$$\begin{aligned} i_x &= \text{round}\left(\tfrac{X - X_{\min}}{X_{\max} - X_{\min}} m_x\right), \\ i_y &= \text{round}\left(m_y - \tfrac{Y - Y_{\min}}{Y_{\max} - Y_{\min}} m_y\right). \end{aligned} \tag{26}$$

The rounding operation in Eq. (26) can distort the quality of the reconstructed image, especially when the resolution of the digital image is low, or the density of interference bands is high. Therefore, the shift operation to the coordinates of the center of the corresponding pixel (i_x, i_y) is introduced:

$$\begin{aligned} X &= i_x \tfrac{X_{\max} - X_{\min}}{m_x} + X_{\min}, \\ Y &= Y_{\min} - \left(i_y - m_y\right) \tfrac{Y_{\max} - Y_{\min}}{m_y}. \end{aligned} \tag{27}$$

Now, the spatial coordinates of the center of the pixel (x^*, y^*, z^*) can be calculated from Eq. (21). As the projection of the analysed point on the surface of the structure does not necessary coincide with the center of the pixel, the further calculations are dedicated for the location of a point on the surface corresponding to the center of that pixel.
The tangential vectors to the surface of the analysed finite element

$$\begin{bmatrix} x_\xi & y_\xi & z_\xi \\ x_\eta & y_\eta & z_\eta \end{bmatrix} \tag{28}$$

are determined by multiplying the derivatives of the shape functions (with respect to the local coordinates) by the nodal coordinates of the corresponding finite element. Subscripts in Eq. (28) denote partial derivatives.
The L–coordinates of intersection of the line (guided through the observation point and the center of the pixel) with the plane tangential to the analysed finite element are obtained from the relationship in Appendix A, where the system matrix is substituted by:

$$\begin{bmatrix} x & x+x_\xi & x+x_\eta & -x_1 & -x^* \\ y & y+y_\xi & y+y_\eta & -y_1 & -y^* \\ z & z+z_\xi & z+z_\eta & -z_1 & -z^* \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}. \tag{29}$$

Thus the calculations are performed for the corrected local coordinates $(\xi + L_2, \eta + L_3)$. It can be noted that such modifications make the procedure more precise and lowers the noise effects in the reconstructed interferograms.
The normal vector N for the analysed three-dimensional structure is obtained as a normalised vector product of the rows of Eq. (28). Other relationships used in the process of calculations coincide with the previously described ones for planar structures.

4.2 Determination of the interference bands in holographic interferograms

The distribution of intensity of the laser beam $a^2(x, y)$ is assumed (Vest (1979)):

$$a^2(x,y) = I_L\left(k_d\left(N \cdot L\right) + k_s\left(V \cdot R\right)^n\right), \tag{30}$$

where I_L is the intensity of the incident laser beam, k_d is the diffuse reflection coefficient, k_s – the specular reflection coefficient, n is a coefficient describing the smoothness of the surface, V is the direction to the viewer, dot in parenthesis denotes a scalar product.
When the surface of the plate performs harmonic oscillations according to an appropriate eigenform, the intensity can be calculated on the basis of Eq. (4):

$$I = \left(\overline{\cos\left(4\pi\left(u \cdot L\right)\sin\omega t/\lambda\right)}\right)^2 a^2(x,y), \tag{31}$$

where u is the amplitude of harmonic oscillations, λ is the laser wavelength, the top line denotes averaging by time.
The first term of Eq. (31) corresponds to the direct calculation of the real part of Eq. (5); the imaginary part is infinitesimal:

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T \sin\left(\frac{4\pi}{\lambda}\left(u \cdot L\right)\sin\omega t\right)dt = 0. \tag{32}$$

Numerical averaging of Eq. (31) can be performed using a certain number of intermediate state positions of the surface during one cycle of oscillations. The averaged term takes the form:

$$\overline{\cos\left(4\pi\left(u \cdot L\right)\sin\omega t/\lambda\right)} = \frac{1}{n}\sum_{i=1}^{n}\cos\left(\frac{4\pi}{\lambda}\left(u \cdot L\right)\sin\frac{2\pi\left(i-1\right)}{n}\right). \tag{33}$$

It can be noted that such averaging procedure enables to generalise the calculation of intensity for more complex (not only harmonic) dynamic processes. The convergence of the square of Eq. (33) to the square of Eq. (6) can be analysed numerically. It can be noted that when n = 16 the precision of calculations is acceptable for the first three interference bands, when $n = 32$ – for the first five bands, when $n = 64$ – for the first 8 bands.
The intensity levels decrease rapidly with the number of the intereference band due to the qualities of Bessel functions. As a limited number of intensity levels is used for the digital

representation of images for better visualisation of the results of calculations a sigmoidal mapping function is used to distort the intensity scale:

$$F(I) = \frac{\tanh(kI)}{\tanh(k)}, \tag{34}$$

where parameter k characterises the level of distortion, $0 < k < \infty$ (Fig. 12). It can be noted, that

$$\begin{aligned} &F(0) = 0, \\ &F(1) = 1, \\ &\lim_{k\to\infty} F(I) = \operatorname{sign}(I), \\ &\lim_{k\to 0} F(I) = I. \end{aligned} \tag{35}$$

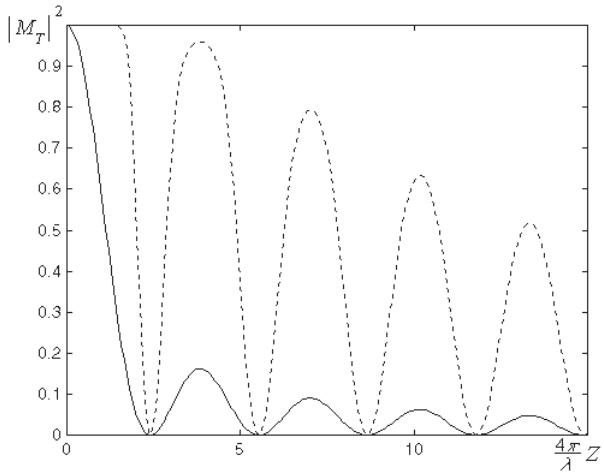


Fig. 12. The decay of intensity for the time averaged method without (solid line) and with intensity mapping at $k = 5$ (dashed line)

In fact,

$$\lim_{k\to 0} \frac{dF(I)}{dI} = \lim_{k\to 0} \frac{e^k + e^{-k}}{e^k - e^{-k}} \frac{d}{dI} \left(\frac{e^{kI} - e^{-kI}}{e^{kI} + e^{-kI}} \right) = \lim_{k\to 0} \frac{2k}{e^k - e^{-k}} = 1. \tag{36}$$

Optical stroboscopic method is used for better interpretation of micro vibrations (Vest (1979)). The structure is lightened two times per period of oscillations at the moments of extreme deflections.

The term $\overline{\cos(4\pi(u \cdot L)\sin\omega t/\lambda)}$ is interpreted as the average value of the two extreme deflections of the structure, corresponding to the moments $\omega t = \pi/2$ and $\omega t = 3\pi/2$, so the averaged term takes the form $\cos(4\pi(u \cdot L)/\lambda)$ due to the evenness of the cosine function. The Eq. (31) is transformed to:

$$I = \cos^2\left(4\pi\,(u \cdot L)/\lambda\right) I_L \left(k_d\,(N \cdot L) + k_s\,(V \cdot R)^n\right). \tag{37}$$

The decay of intensities for the time averaged and stroboscopic models are presented in Fig. 13. It can be noted that the locations of the centers of dark interference bands for the stroboscopic method (dashed line) do not coincide with the time averaged ones (solid line).

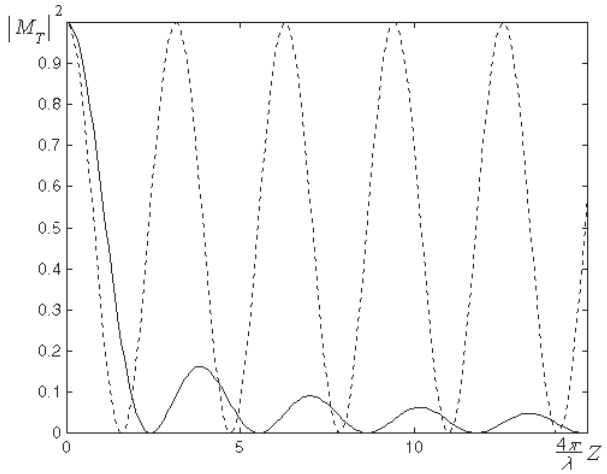


Fig. 13. The decay of intensity for the time averaged method (solid line) and for the stroboscopic method (dashed line)

The stroboscopic method does not produce a fast decay of the intensities and thus enables to interpret the higher interference bands from the interferogram. Thus the visualisation of higher amplitudes (higher interference bands) requires either the intensity mapping of the time averaged interferograms, or application of the stroboscopic method of analysis.
The problem of visualisation of microvibrations is important in engineering of precise mechanical systems. The plotting of interference bands from the results of finite element analysis has clear physical background compared to other visualisation techniques. Also that is important because of the ability of direct comparisons with the experimental results of holographic optical analysis. Presented methodology for plotting holographic interferograms is scalable in parallel computations and applicable to a wide variety of problems.

5. Computational visualization of time averaged holographic interferograms of the MEMS cantilever

The presented procedure for real time calculation of definite integrals can be effectively applied in hybrid numerical-experimental techniques where time average intensities of illumination in a holographic interferogram are reconstructed in virtual computational environment. Implementation of the proposed integration rule enables full real time computations with minimal data queue lengths and effective management of integration order.

These simulation results are illustrated for the dynamics of the MEMS cantilever under oscillating charge excitation what provides rather complex dynamic response of the tip of the cantilever. Figure 14 presents interferograms for the first eight eigenmodes. The displacement of the tip of the cantilever is presented in Fig. 15. Computations of the intensity of illumination for different time exposures for the MEMS cantilever are used to construct numerical approximations of holographic interferograms; the results are presented in Fig. 16.

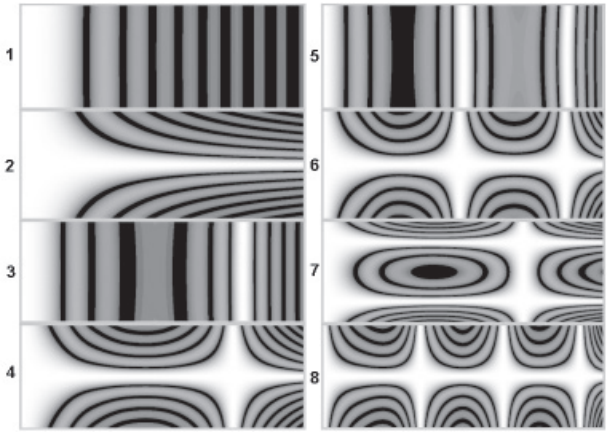


Fig. 14. Time average interferograms of the first eight eigenmodes of MEMS cantilever

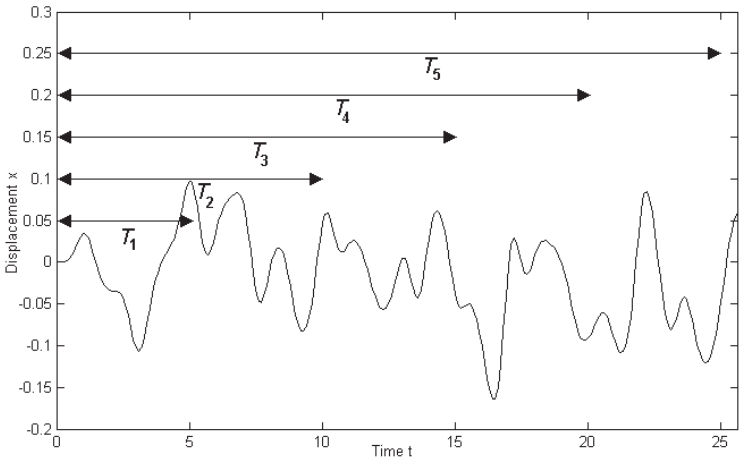


Fig. 15. Chaotic dynamics of the tip of MEMS cantilever and times of exposure $T_1 - T_5$

Numerical results validate the theoretical predictions and help to explain the complexity of the dynamical processes taking place in the analysed MEMS systems, especially when the analysed system oscillates chaotically.

Though time average holography is a powerful experimental technique for analysis of MEMS dynamics, the interpretation of produced patterns of fringes in holograms must be performed with care. Complex physical and non-linear dynamical processes taking place in MEMS

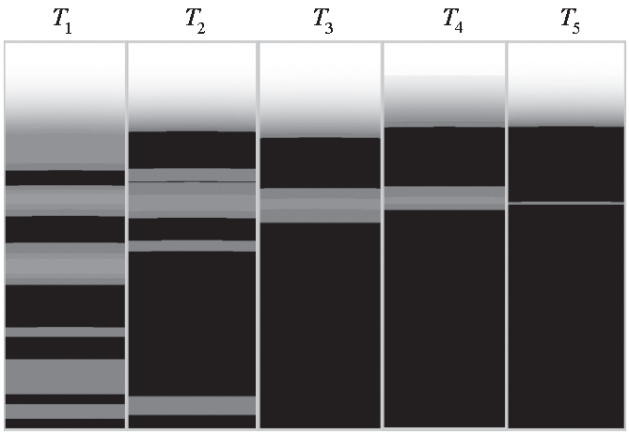


Fig. 16. Time avegare interferograms of the MEMS cantilever at different times of exposure

systems may influence the results produced by optical methods. This book chapter provides insight in the complexity of the MEMS dynamics and the problems of its optical analysis based on holographic interferometry.

6. Concluding remarks

The validation of experimental investigations versus numerical analysis provides the necessary background to analyse the dynamical characteristics of MEMS elements in virtual numerical environments. Direct application of fringe counting techniques for reconstruction of motion from time average holograms cannot be straightforward if the analysed MEMS system performs chaotic oscillations. Modifications of a classical time average holographic technique enable qualitative analysis of MEMS and can be applied for investigation of dynamical properties of much broader classes of MEMS systems.
Is well known that non-linear systems can exhibit periodic, quasiperiodic, and even chaotic responses under periodic forcing. Therefore it is important to understand what time averaged image would be produced if time average holographic interferometry is used to explore a chaotically oscillating elastic structure.
The ability to interpret such holographic interferograms would help to improve the uncertainty of the inverse problem and to distinguish malfunctions of the optical setup and also physical reasons causing specific optical effects.

7. Appendix A

Denoting shape functions L_1, L_2 and L_3 as L-coordinates of the triangle (Zienkiewicz & Morgan (2006)), and L_4 and L_5 as L-coordinates of the line, the following relationships are true:

$$\begin{aligned} x &= L_1 x_1 + L_2 x_2 + L_3 x_3, \\ y &= L_1 y_1 + L_2 y_2 + L_3 y_3, \\ z &= L_1 z_1 + L_2 z_2 + L_3 z_3, \\ L_1 &+ L_2 + L_3 = 1, \end{aligned} \tag{38}$$

and

$$\begin{aligned} x &= L_4\tilde{x}_1 + L_5\tilde{x}_2, \\ y &= L_4\tilde{y}_1 + L_5\tilde{y}_2, \\ z &= L_4\tilde{z}_1 + L_5\tilde{z}_2, \\ L_4 &+ L_5 = 1, \end{aligned} \tag{39}$$

where x_i, y_i, z_i, $i = 1,2,3$ – coordinates of three points in a plane, $\tilde{x}_1$, $\tilde{x}_2$, $\tilde{x}_3$ and $\tilde{y}_1$, $\tilde{y}_2$, $\tilde{y}_3$ – coordinates of two points lying on the line. Then the L-coordinates of the point of intersection between the line and the plane are given by the solution of the following system of equations:

$$\begin{bmatrix} x_1 & x_2 & x_3 & -\tilde{x}_1 & -\tilde{x}_2 \\ y_1 & y_2 & y_3 & -\tilde{y}_1 & -\tilde{y}_2 \\ z_1 & z_2 & z_3 & -\tilde{z}_1 & -\tilde{z}_2 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{Bmatrix} L_1 \\ L_2 \\ L_3 \\ L_4 \\ L_5 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{Bmatrix}. \tag{40}$$

The line intersects the plane inside the triangle when the conditions

$$L_i \in [0,1], \quad i = 1,2,3 \tag{41}$$

are satisfied.

8. References

Caponero, M.A.; Pasqua, P.; Paolozzi, A. & Peroni, I. (2000). Use of holographic interferometry and electronic speckle pattern interferometry for measurements of dynamic displacements, *Mechanical Systems and Signal Processing*, Vol.14, No.1, 49–62, ISSN 0888-3270.

Davis, P.J. & Rabinowitz, P. (1984). *Methods of Numerical Integration*, Academic Press, ISBN 978-0122063602, New York.

Fagan, W.F., Waddell, P. & McCracken, W. (1972). The study of vibration patterns using real-time hologram interferometry, *Optics & Laser Technology*, Vol.4, No.4, 167–172, ISSN 0030-3992.

Fein, H. (2000). Holographic interferometry: nondestructive tool, *The industrial Physicist*, Vol.3, No.3, 37–39, ISSN 1082-1848.

Ganesan, A.R.; Hinsch, K.D. & Meinlschmidt P. (2000). Transition between rationally and irrationally related vibration modes in time-average holography, *Optics Communications*, Vol.174, No.5-6, 347–353, ISSN 0030-4018.

Ivanov, V.P. & Batrakov, A.S. (1995). *Three Dimensional Computer Graphics*, Radio i Sviaz, Moscow.

Mattson, T.; Sanders, B. & Massingill, B. (2004). *Patterns for Parallel Programming*, Addison Wesley Professional, ISBN 0-321-22811-1, Reading, Massachusetts.

Ostasevicius, V.; Palevicius, A.; Daugela, A; Ragulskis, M. & Palevicius, R. (2004). Holographic imaging technique for characterization of MEMS switch dynamics, *SPIE Proceedings, 2004*, 73–84, ISSN 0277-786X.

Ostasevicius, V.; Tamulevicius, S.; Palevicius, A.; Ragulskis, M; Grigaliunas, V. & Minialga, V. (2005). Synergy of contact and non-contact techniques for design and characterization of vibrating MOEMS elements, *Journal of Microlithography, Microfabrication, and Microsystems*, Vol. 4, No. 4, 1–9, ISSN 1537-1646.

Ragulskis, M.; Palevicius, A. & Ragulskis, L. (2003). Plotting holographic interferograms for visualisation of dynamic results from finite element calculations, *International Journal for Numerical Methods in Engineering*, Vol. 56, 1647-1659, ISSN 0029-5981.

Ragulskis, M. & Saunoriene, L. (2007). Order adaptive quadrature rule for real time holography applications, *Lecture Notes in Computer Science*, Vol. 4310, 685–692, ISSN 0302-9743.

Ragulskis, M. & Saunoriene, L. (2008). Generalisations of the compound trapezoidal rule, *Applied Numerical Mathematics*, Vol.58 , No.1, 40–58, ISSN 0168-9274.

Rastogi, P.K. (2000). Principles of holographic interferometry and speckle metrology, *Photomechanics Topics in Applied Physics*, Vol. 77, 103–150, ISSN 0303-4216.

Vest, C.M. (1979). *Holographic Interferometry*, Wiley, ISBN 978-0471906834, New York.

Zienkiewicz, O.C. & Morgan, K. (2006). *Finite Elements and Approximation*, Dover Publications, ISBN 0486453014, New York.

16

Vibration Measurement by Speckle Interferometry between High Spatial and High Temporal Resolution

Dan Nicolae Borza
Institut National des Sciences Appliquées de Rouen
France

1. Introduction

Speckle interferometry is a widely known successor of holographic interferometry. It is usually based on the use of a continuous wave laser. The use of temporal phase stepping allows obtaining the full-field of vibration amplitudes at the surface of the object under study. The simplest method is that of real-time, time-averaged speckle interferometry. It has been described for the first time in (Butters & Leendertz, 1971). The vibration amplitude map is usually presented as a two-dimensional fringe pattern, whose intensity modulation is given by the square of the Bessel function of the first kind and zero order whose argument is proportional with the local vibration amplitude. Roughly, the fringes are loci of points having the same vibration amplitude. The difference between the vibration amplitudes of points situated on adjacent fringes is close to a quarter of the laser light wavelength, which is something between 0.12 μm and 0.16 μm.

The most important general characteristics of this technique are justifying the interest in its use. It is a non-contact technique, so the object is not disturbed during the measurement. This is very important since there is no added mass or forces applied to the tested object, so its dynamic characteristics are not affected by the measurement process. It is a full-field technique, so there is no need to use many sensors or scan the surface in order to measure the vibration amplitudes of different points. All amplitudes are measured simultaneously. Finally, the sensitivity is very high and only small vibration amplitudes, up to a few micrometers, are necessary to produce the fringe pattern.

Along with these very favourable characteristics, a certain number of limitations also exist. The most obvious limitation is the noisy aspect of interferograms. Both additive and multiplicative speckle noise are affecting the interferograms. Further processing of the Bessel-type fringe patterns is difficult because of the speckle noise and of the weak contrast of successively increasing order fringes. As will be shown in this chapter, this task may become simpler by reducing the high-frequency multiplicative noise in the orthogonal components of the interferogram. Another limitation is related to the fact that the fringe pattern is obtained by integration, during the 3- or 4-frame bucket, of several cycles of vibration. During this time the vibration must be stationary and the temporal phase information related to the vibration is lost. In this chapter several advances in this field will be shown, allowing to make full-field measurements and resolve temporally the vibration.

2. Real-time, time-average speckle interferometry

The typical setup of a speckle interferometry vibration measurement system is shown in Fig. 1.

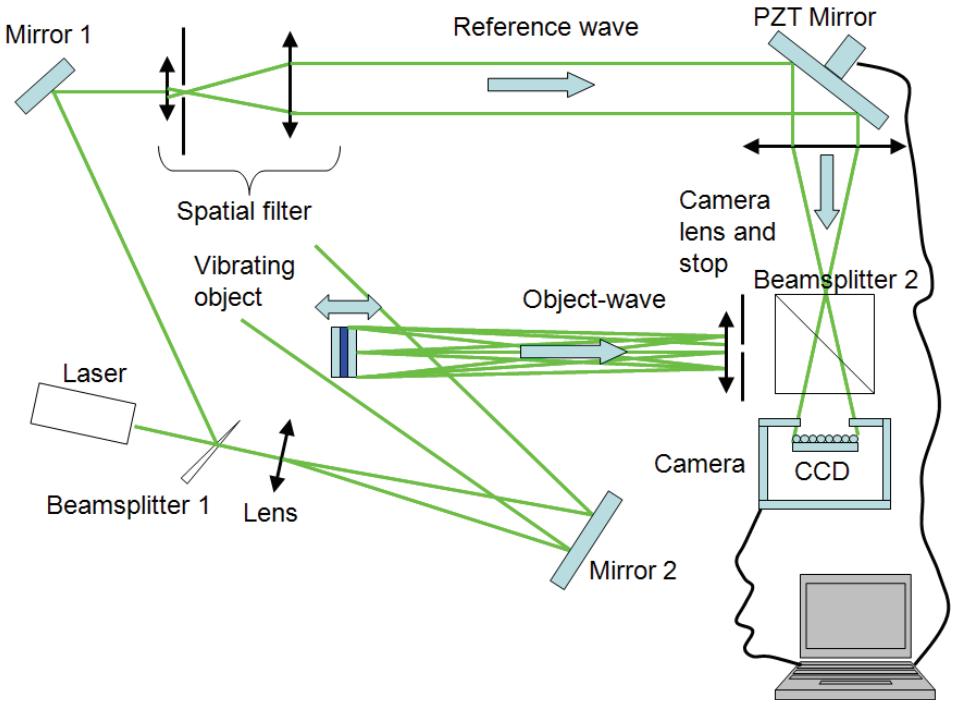


Fig. 1. Schematic lay-out of an out-of-plane sensitive time-average speckle interferometry system for vibrations

The vibrating object is illuminated by the CW laser beam transmitted by the beamsplitter, expanded by a lens and finally redirected by Mirror2 towards the vibrating object. The part of the laser beam reflected by the beamsplitter and by Mirror1 is expanded by a spatial filter and redirected, through the PZTMirror and the beamsplitter 2, towards the CCD detector. It is the reference beam. The PZT actuator placed behind PZTMirror is implementing the temporal phase stepping, producing a 4-step reference phase variation given by:

$$\Delta\varphi_i = (i-1)\frac{\pi}{2};\ i = 1,\ 2,\ 3,\ 4... \tag{1}$$

as first described in (Joenathan, 1991).
In phase-stepped speckle interferometry, the raw data consist in number (usually 3 or 4) of correlograms recorded by the CCD detector. They represent the spatial sampling of the interference field between a uniform complex reference wave (usually of constant real amplitude R):

$$\overline{R} = R \cdot \exp\left[j\left(\varphi_r - \Delta\varphi_i\right)\right] \tag{2}$$

and a speckled object-wave

$$\begin{aligned}\bar{O} &= O\cdot\exp\left(j\varphi_o\right) = O\cdot\exp\left[j\left(\varphi_r+\varphi_{o-r}+\varphi_v\cos\omega t\right)\right] = \\ &= O\cdot\exp\left[j\left(\varphi_r+\varphi_{o-r}\right)\right]\cdot\exp\left[j\left(\varphi_v\cos\omega t\right)\right] = \bar{O}_0\cdot\exp\left[j\left(\varphi_v\cos\omega t\right)\right]\end{aligned} \tag{3}$$

able to produce with the help of the camera lens a plane image of the harmonically vibrating object under study. The interference field of these waves is integrated by the camera during the frame acquisition time over an integer or large number of vibration periods. By putting $m = O^2 + R^2$ and $n = 2\cdot O\cdot R$, the result may be described by the equation:

$$I_i = m + n\cos\left[\varphi_{o-r}\left(x,y\right)+\alpha\right]\cdot J_0\left[\varphi_v\left(x,y\right)\right] \tag{4}$$

In Eq. (2), φ_{o-r} is the random phase difference between the uniform reference wave and the speckled object wave corresponding to a point *(x, y)* of the object in the equilibrium position. $J_0(\varphi_v)$ is the first kind, zero-th order Bessel function whose argument is the vibration-related phase given by the approximate relation:

$$\varphi_v\left(x,y\right) = \frac{4\pi}{\lambda}d\left(x,y\right) \tag{5}$$

where d is the vibration amplitude of the *(x, y)* point. In any speckle interferometry system based on 4-frame buckets (Creath, 1985), the four phase-stepped data fields given by Eq. (1) are used to calculate (either in real-time or during post-processing) the differences:

$$C = I_1 - I_3 = 2OR\cos\phi_{o-r}\cdot J_0\left(\varphi_v\right) \tag{6}$$

and:

$$S = I_4 - I_2 = 2OR\sin\varphi_{o-r}\cdot J_0\left(\varphi_v\right) \tag{7}$$

These two orthogonal data fields (Borza, 2004) may be used to compute and display the usual time-averaged fringe pattern:

$$I_{TAV} = \sqrt{S^2+C^2} = 2OR\left|J_0\left[\varphi_v\left(x,y\right)\right]\right| \tag{8}$$

They may also be considered as the real and imaginary components, in the detector plane, of a reconstructed complex wave (A is a constant):

$$\bar{O}_R = A\cdot O\cdot J_0\left[\varphi_v\left(x,y\right)\right]\exp\left(j\varphi_{o-r}\right) \tag{9}$$

One should be aware that any individual frame or the corresponding data fields or the interferogram itself are time-averaged, even if, as a particular case, the object is immobile during the integration time. In this case, Eq. (9) simply becomes:

$$\bar{O}_R = A\cdot O\cdot\exp\left(j\varphi_{o-r}\right) \tag{10}$$

The maxima (bright fringes) of Eq. (8) correspond to vibration amplitudes of 0 (zero) - for the brightest fringes showing the nodal lines, then to amplitudes of

$$d = (k + \frac{1}{2}) \cdot \frac{\lambda}{4} \quad (k=1, 2, 3 \ldots) \tag{11}$$

The minima (dark fringes) correspond to amplitudes:

$$d = k \cdot \frac{\lambda}{4} \quad (k = 1, 2, 3 \ldots) \tag{12}$$

The typical aspect of time-averaged fringes seen either in real-time on the monitor while looking at a vibrating object, or saved on the disk, is shown in Fig. 2 (Nistea & Borza, 2010a).

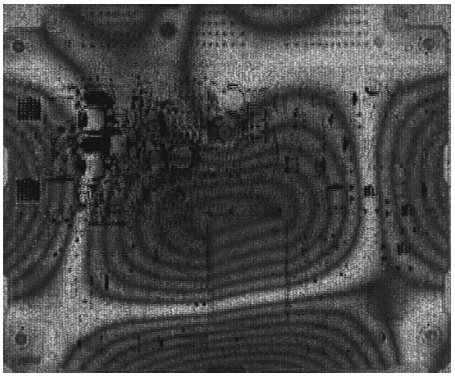


Fig. 2. Time-averaged speckle interferogram of an electronic assembled microcontroller card excited at one of its resonance frequencies at 841 Hz

It corresponds to a vibration during which the object reaches periodically the position of extreme deflection with regard to the equilibrium position, as shown in Fig. 3.

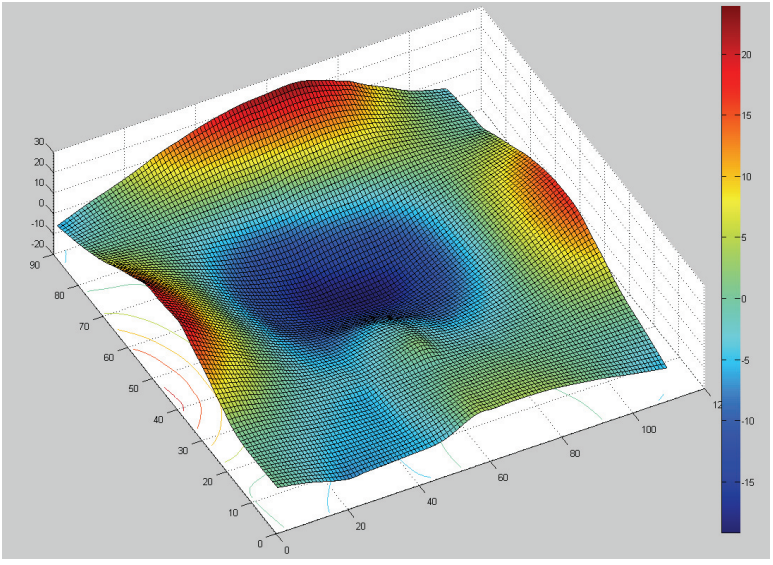


Fig. 3. Vibration amplitude-related phase map corresponding to interferogram in Fig. 2. Values are in radians.

The interferograms obtained by this method are also called *electronic holograms*, or, when it's no possibility of confusion with a "classical" hologram obtained through diffraction of the reference beam by the primary fringes recorded on photorecording media, simply *holograms*. To adress one of the limitations already mentioned for the time-average method, the speckle noise, one may record several time-average interferograms and eventually rotate the object illumination beam between recordings. By averaging a number of N interferograms, the speckle noise is reduced by $\sqrt{N}$, as shown in Fig. 4

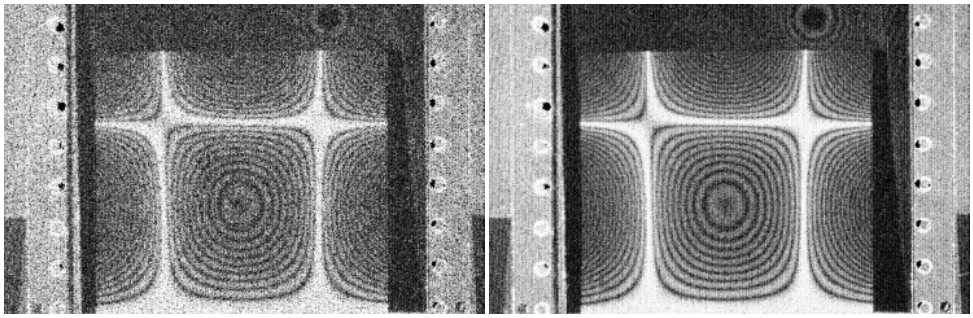


Fig. 4. Time-averaged holograms of a vibrating plane plate: at left a single hologram, at right the result of four hologram averaging.

Filtering a time-averaged hologram by convolution with an average or other low-pass filter may reduce the high-frequency noise, but strongly affects the fringe resolution and contrast.

3. Development of speckle interferometry vibration measurement methods

3.1 Reference wave modulation

One of the well-known drawbacks of time-average holograms is the impossibility to obtain by an automatic procedure the explicit full-field of vibration amplitudes. This problem was adressed (Pryputniewicz & Stetson, 1999) by using the reference wave homodyne modulation, in order to record several fringe patterns spatially shifted. Through the use of look-up tables, the final result is a wrapped amplitude-related phase fringe pattern. Such a fringe pattern is shown in Fig. 5.

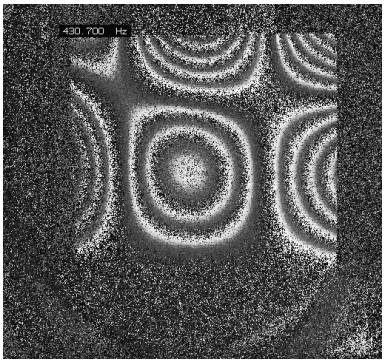


Fig. 5. Phase image of the vibration mode presented in Fig. 4.

The introduction of the homodyne modulation of the reference wave thus allows using slightly modified 2d unwrapping procedures in order to unwrap the fringe pattern and obtain the explicit amplitude map.
The method works well, but from the practical point of view is rather cumbersome as it implies changing several times the amplitude and the phase of the exciting voltages applied to the object and to the modulating mirror.
Other authors described the possible use of heterodyne phase modulation in time-average speckle interferometry. In (Høgmoen and Løkberg, 1977) this type of modulation is used to allow measurement of very small vibration amplitudes.

3.2 Stroboscopic principles

Apart from the methods based on the modulation of the reference wave, the only possibility to obtain phase fringe patterns for vibrating objects is to avoid the temporal averaging, or the integration during the recording of a frame. It means the CCD exposure should be much smaller than the period of vibration. This may be achieved by using stroboscopic principles or pulsed lasers. Several researchers studied the application of stroboscopic techniques to speckle interferometry. Stroboscopic techniques are "freezing" the vibration for the moments when data are recorded, like speckle interferometry with pulsed lasers. The correlograms recorded, which are the primary data fields from which interferograms are computed, are related only to certain well-defined positions. The temporal integration thus disappears, and so does the Bessel-type fringe profile. The simplest interferometric reconstructions are double-exposure holograms rather than time-averaged. The use of these techniques allows, as described in (Valera et al, 1997), the study of objects which are vibrating simultaneously at two frequencies whose ratio is an irrational number. The separation of the two vibration modes is made by stroboscopic heterodyned speckle interferometry. By choosing the frequency of the strobing pulses equal with one of the two frequencies of vibration, only the vibration of the second frequency was measured by the interferogram.
Stroboscopic techniques may also be used for measuring the response of a structure both in phase and in quadrature with the exciting force, allowing thus (Van der Auweraer et al, 2002) to estimate the real part and the imaginary part of the response of the object. This is an important step in experimental modal analysis.
Many other methods and two excellent reviews of the state-of-the-art were published by (Doval, 2000; Jacquot, 2008).

4. Vibration measurement with high spatial resolution

Several researchers proposed different experimental methods and algorithms based on the joint use of correlograms recorded with the object at rest and during vibration.
(Wang et al, 1996) made a thoroughful analysis of three methods based on the addition or subtraction of video signals recorded in such situations. For the particular case of the amplitude-fluctuation method, the subtraction between the signals corresponding to two different signals during vibration is shown to produce fringes dependent on the Bessel function $J_1[\varphi_v(x,y)]$ instead of $J_0[\varphi_v(x,y)]$. The subtraction of the speckled image of an object from that of the object at rest, as in (Nakadate, 1980), produces a fringe pattern given by:

$$I = 1 - 2 \cdot J_0(\varphi_v) \cdot [\cos(\varphi_{stat})] + J_0^2(\varphi_v) \tag{13}$$

If the speckle decorrelations are kept to a reasonable level, they might be approximated by:

$$I = [1 - J_0(\varphi_v)]^2 \tag{14}$$

As underlined in (Creath & Slettemoen, 1985) this type of fringes are equivalent to those appearing in real-time holographic interferometry. Their contrast is lower compared to time-average fringes, but they offer a valuable tool in studying vibrations of very small amplitudes.

In order to analyze the possibilities for reducing the speckle noise of the time-averaged speckle interferograms and advance towards high spatial resolution, one has to look at the expression of one of the correlograms recorded in the original 4-frame bucket. It is given by Eq. (4). In the expression $m + n\cos[\varphi_{o-r}(x,y) + \alpha] \cdot J_0[\varphi_v(x,y)]$, m (the bias) denotes the sum between the speckled image of the object and the uniform intensity of the reference beam, so it represents additive noise. The variable n (the modulation) denotes the square root of the product of the same terms (the speckled image of the object and the uniform intensity of the reference beam). This noisy term is multiplied by $\varphi_{o-r}(x,y)$ which is the random phase difference, of very high spatial frequency, between the object wave and the reference wave. The last factor of the product, $J_0(\varphi_v)$, is the only deterministic factor; the argument of the Bessel function is the spatial phase introduced by the object vibration. Unfortunately, with the exception of the nodal points where $\varphi_v = 0$ and $J_0(\varphi_v) = 1$, the product $n \cdot \varphi_{o-r}(x,y) \cdot J_0(\varphi_v)$ is dominated by the multiplicative speckle noise and the Bessel function is deeply covered by noise.

In the (usual) case of 4-frame correlogram buckets of vibrating objects, the additive noise is eliminated, as shown by Eq. (6) and (7), by making the differences between the frames phase-shifted by π. In the remaining expressions, $C = 2 \cdot O \cdot R \cdot \cos(\varphi_{o-r}) \cdot J_0(\varphi_v)$ and $S = 2 \cdot O \cdot R \cdot \sin(\varphi_{o-r}) \cdot J_0(\varphi_v)$ the multiplicative noise is still there. Any attempt to apply a low-pass filter to these two orthogonal components of the time-averaged hologram would "spread" the high-frequency noise and make impossible the interferogram reconstruction. To compensate the speckle noise we can use the similar terms from another 4-frame bucket, or, better, from a 4-frame bucket during which the object was at rest. Assuming that speckles didn't decorrelate between the acquisition of the frames with the object at rest and the acquisition with the vibrating object, but admitting a small phase difference $\Delta\varphi_{o-r}$, which may have as origine, for example, a small displacement or a thermal drift, these terms have the expressions:

$$J_i = m + n\cos\left[\varphi_{o-r}(x,y) + \Delta\varphi_{o-r} + \alpha\right] \tag{15}$$

and they are used to compute:

$$C' = J_1 - J_3 = 2OR\cos\left(\varphi_{o-r} + \Delta\varphi_{o-r}\right) \tag{16}$$

and

$$S' = J_4 - J_2 = 2OR\sin\left(\varphi_{o-r} + \Delta\varphi_{o-r}\right) \tag{17}$$

Eqs. (6), (7), (16) and (17) may be used to compute:

$$S_0 = S \cdot C' - S' \cdot C = 4O^2R^2 \cdot J_0(\varphi_v) \cdot \sin(\Delta\varphi_{o-r}) \tag{18}$$

and

$$C_0 = S \cdot S' + C' \cdot C = 4O^2R^2 \cdot J_0(\varphi_v) \cdot \cos(\Delta\varphi_{o-r}) \tag{19}$$

The data fields described by Eqs. (18) and (19) are important because they are almost noise-free; the high-frequency multiplicative speckle noise has been eliminated.
As described in (Borza, 2000; Borza, 2002; Borza, 2005) the two quadrature data fields may be used for hologram restoration by a procedure equivalent to the synchronous detection. Eqs. (18) and (19) allow computing the phase difference by using the arctangent function, $\tan^{-1}(S_0 / C_0)$.
If the 4-quadrant arctangent function is used instead, the sign of the numerator and of the denominator are evaluated separately and the result goes in the right quadrant. Then the identical factors in these expressions, $4 \cdot O^2 \cdot R^2 \cdot J_0(\varphi_v)$, will disappear, with one particular remark. The quadrant where the angle will be placed depends on the sign of the Bessel function $J_0(\varphi_v)$. Each time this factor will change its sign, the result

$$I_{ph} = a\tan 2\left(\frac{S_0}{C_0}\right) = \Delta\varphi_{o-r} \cdot \mathrm{sgn}\left[J_0(\Delta\varphi_{o-r})\right] \tag{20}$$

will also change sign. If the phase difference between the object at rest and the mean position of the vibrating object is close to zero, the fringe pattern is a quasi-binary one, as in fig. 6.

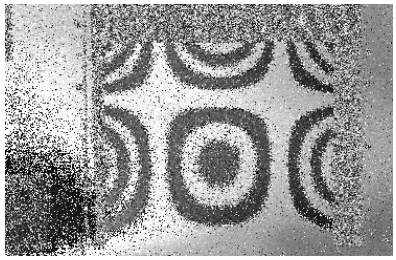


Fig. 6. Quasi-binary fringe pattern

This kind of "quasi-binary" fringes have also been reported by other authors (Picart et al, 2005; Singh et al, 2007).
Since the two orthogonal data fields are noise-free, the whole process may be described by the classical description of synchronous detection, as in fig. 7.
The two noise-free data fields may be low-pass filtered and the resulting quasi-binary hologram, showing the sign of the Bessel function (fig. 8a) may be the starting point for a procedure aiming at the obtention of the explicit vibration amplitude map.
Even more important is the reconstruction of the time-averaged hologram presented in fig. 8b. This becomes possible with the help of the filtered orthogonal components. It is a high spatial resolution hologram, allowing to determine the positions of the zeroes of the Bessel dark fringes in subpixel resolution.

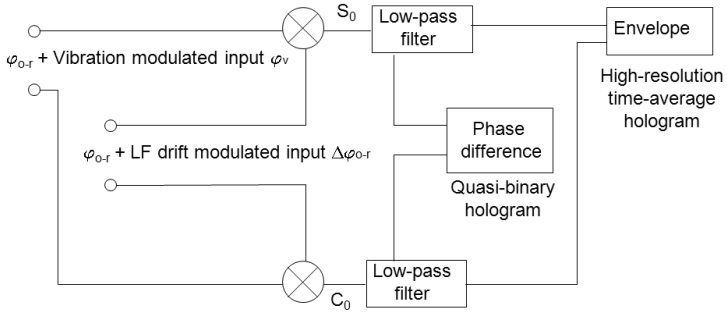
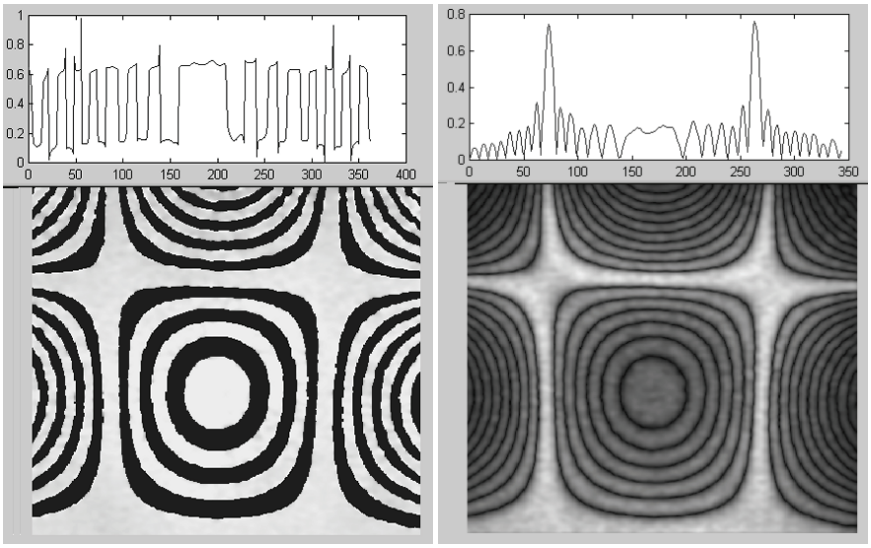


Fig. 7. Principle of synchronous detection.



a b

Fig. 8. (a) Quasi-binary hologram after filtering its orthogonal components; profile grey values are normalized between 0 and 1; (b) high spatial resolution time-average hologram

The paper (Borza, 2006) describes a method, having as starting point this kind of high-resolution time-average hologram, for automatically indexing in subpixel resolution all the dark and bright fringes and then obtaining the vibration amplitude in each point by simply inverting the Bessel function on its monotonicity intervals (fig. 9).

5. A unifying approach for speckle noise reduction in vibration measurement by high resolution time-average speckle interferometry

Since the methods presented in the last Section are using the phase difference between the states with the object at rest and with the object in vibration, and the eventual static deformation (described by $\Delta\varphi_{o-r}$) between these two instants has been taken into account in Eq. (15) and the following, it is reasonable to try doing a more systematic description in a unifying approach.

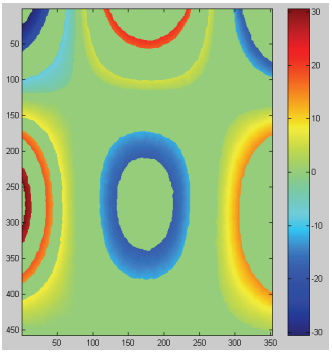


Fig. 9. Partially unwrapped vibration map by regional inverting of the Bessel function

In the most general situation, the two object states which are compared in the final interferogram may be characterized by:

- The first state: the object in vibration, described by the vibration-related phase φ_{v1};
- The second state: the object deformed (or displaced statically), as described by $\Delta\varphi_{o-r}$ and having also a second vibration, described by the vibration-related phase φ_{v2}.

In this case, for the first state the two orthogonal data fields computed from the 4-frame buckets acquired are:

$$\begin{aligned} C_{v1} &= A \cdot O \cdot J_0(\varphi_{v1}) \cdot \cos\varphi_{o-r} \\ S_{v1} &= A \cdot O \cdot J_0(\varphi_{v1}) \cdot \sin\varphi_{o-r} \end{aligned} \tag{21}$$

and:

$$\begin{aligned} C_{v2} &= A \cdot O \cdot \cos(\varphi_{o-r} + \Delta\varphi_{o-r}) \cdot J_0(\varphi_{v2}) \\ S_{v2} &= A \cdot O \cdot \sin(\varphi_{o-r} + \Delta\varphi_{o-r}) \cdot J_0(\varphi_{v2}) \end{aligned} \tag{22}$$

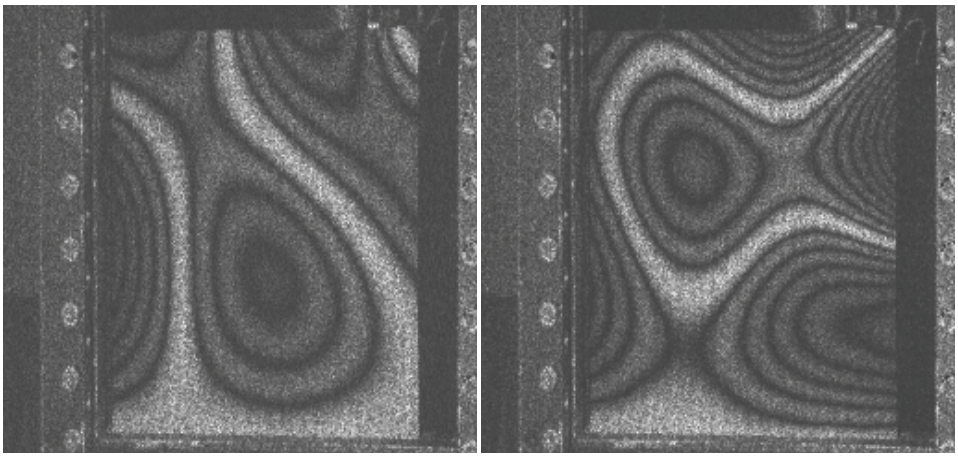


Fig. 10. Time-averaged holograms corresponding to the 2 vibration states.

In both cases, the time-averaged holograms obtained by eq. (8) are showing (fig. 10) Bessel-type fringe patterns which are related only to the vibration-related phase, either φ_{v1} or φ_{v2}.

$$I_{TAV1} = \sqrt{S^2 + C^2} = 2OR\left|J_0\left[\varphi_{v1}(x,y)\right]\right| \tag{23}$$

$$I_{TAV2} = \sqrt{S^2 + C^2} = 2OR\left|J_0\left[\varphi_{v2}(x,y)\right]\right| \tag{24}$$

Several situations may be imagined when combining the data fields from the two object states.

5.1 Phase difference of data fields - General case

The two orthogonal data fields which may be calculated from Eqs. (21) and (22) by using the formulæ:

$$\begin{aligned} S &= S_{v1}C_{v2} - C_{v1}S_{v2} \\ C &= C_{v1}C_{v2} + S_{v1}S_{v2} \end{aligned} \tag{25}$$

are given by the expressions:

$$\begin{aligned} S &= A^2O^2J_0(\varphi_{v1})J_0(\varphi_{v2})\sin\Delta\varphi_{o-r} \\ C &= A^2O^2J_0(\varphi_{v1})J_0(\varphi_{v2})\cos\Delta\varphi_{o-r} \end{aligned} \tag{26}$$

The square root of the sum of their squares is expressed by the absolute value of the product of the two Bessel functions:

$$I_{TAV12} = I_O \cdot \left|J_0(\varphi_{v1}) \cdot J_0(\varphi_{v2})\right| \tag{27}$$

and their relative phase is:

$$\Delta\varphi_{12} = \Delta\varphi_{o-r}\,\mathrm{sgn}\left[J_0(\varphi_{v1}) \cdot J_0(\varphi_{v2})\right] \tag{28}$$

They are illustrated in fig. 11.

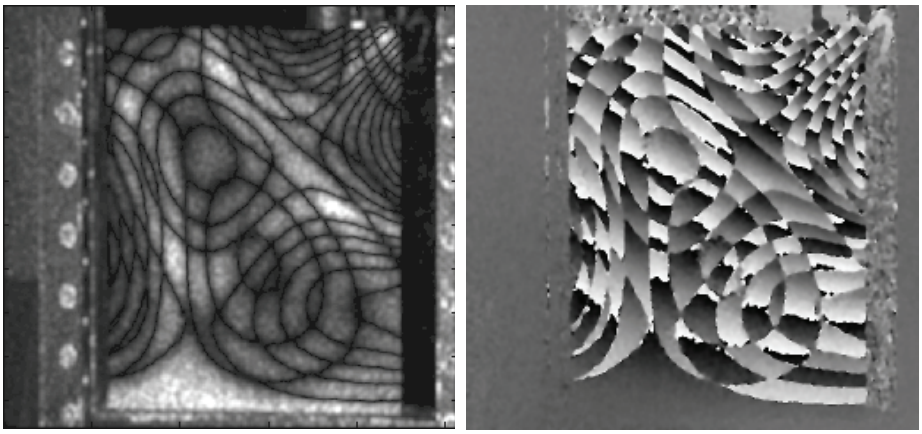


Fig. 11. Fringe patterns corresponding to Eqs. (27) and (28)

The static term, $\Delta\varphi_{o-r}$, may be caused by different phenomena: sometimes it appears because of air turbulences or thermal gradients between the two series of exposures, sometimes it reflects variations of the optical path in the object-wave due to a mechanical movement or the object deformation between the two recordings. If $\Delta\varphi_{o-r}$ is not too large, it may be easily compensated by post treatment of data (fig. 12).

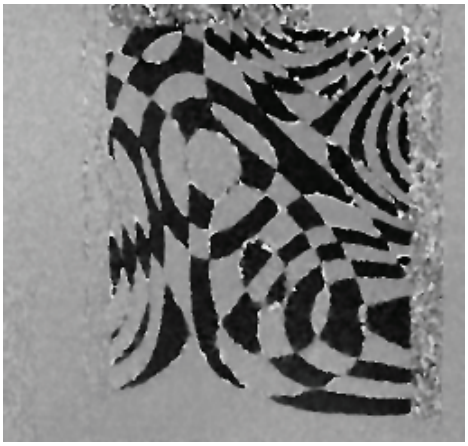


Fig. 12. Phase difference with compensated $\Delta\varphi_{o-r}$

5.2 Phase difference in a particular case: object at rest in the first state

This is the most current case; if during the second state there is no static phase difference with respect to the first state, one retrieves the situation already described in § 4. If the static phase difference is not negligible, then the phase difference is given by eq. (20) and shown in fig. 13..

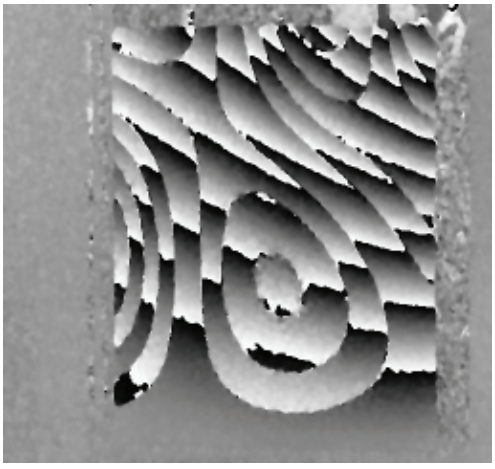


Fig. 13. Quasi-binary hologram with an important static phase difference and vibration of the object in one of the two states.

To fully appreciate the spatial resolution gain of time-average speckle interferometry when using the presented method, fig. 14 illustrates the usual time-average hologram and the high-resolution one, recorded during the same experience, along with the profiles across fringes. The values on the scales of grey levels are conventional values obtained by two different normalizations.

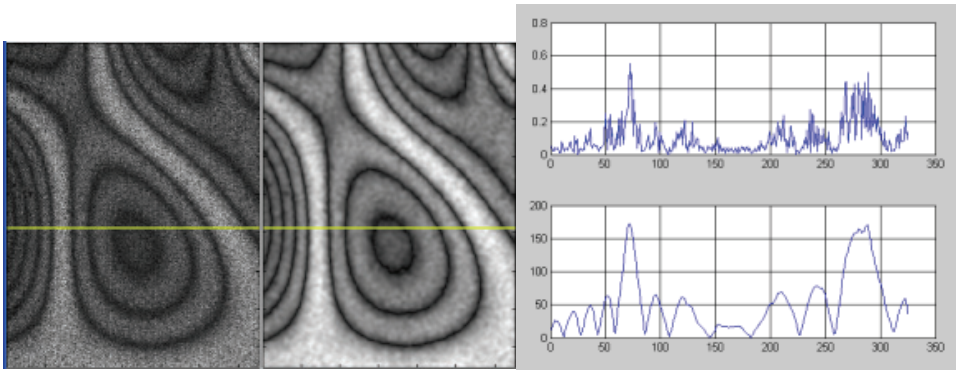


Fig. 14. Time-average (left), high resolution time-average (middle) interferograms and profiles across the fringe patterns (right)

5.3 Phase difference in a particular case: object vibrating identically in both states

The vibration amplitude and phase being identical in both states, $\varphi_{v1} = \varphi_{v2}$. In this case eq. (21) is replaced by:

$$\begin{aligned} C_{v1} &= C_{v2} = A \cdot O \cdot J_0(\varphi_{v1}) \cdot \cos\varphi_{o-r} \\ S_{v1} &= S_{v2} = A \cdot O \cdot J_0(\varphi_{v1}) \cdot \sin\varphi_{o-r} \end{aligned} \tag{29}$$

Eq. (25) become:

$$\begin{aligned} S &= S_{v1}C_{v1} - C_{v1}S_{v1} = A^2O^2J_0^2(\varphi_{v1}) \cdot \sin(\Delta\varphi_{o-r}) \\ C &= {C_{v1}}^2 + {S_{v1}}^2 = A^2O^2J_0^2(\varphi_{v1}) \cdot \cos(\Delta\varphi_{o-r}) \end{aligned} \tag{30}$$

Equation (28) becomes:

$$\Delta\varphi_{12} = \Delta\varphi_{o-r}\,\mathrm{sgn}\left[J_0{}^2(\varphi_{v1})\right] = \Delta\varphi_{o-r}\big|_{\mathrm{mod}\,2\pi} \tag{31}$$

This result is shown in fig.15.
The detailed explanations for this kind of fringe patterns are found in (Borza, 2008).

6. Speckle interferometry with high temporal resolution

The spatial resolution of speckle interferometry is in most cases high enough for the needs of the measurements, and the vibration amplitude maps found by this technique are used, as in (Moreau, 2008) to check and validate the results produced by other techniques (Laser Doppler Velocimetry, Near-Field Acoustical Holography).

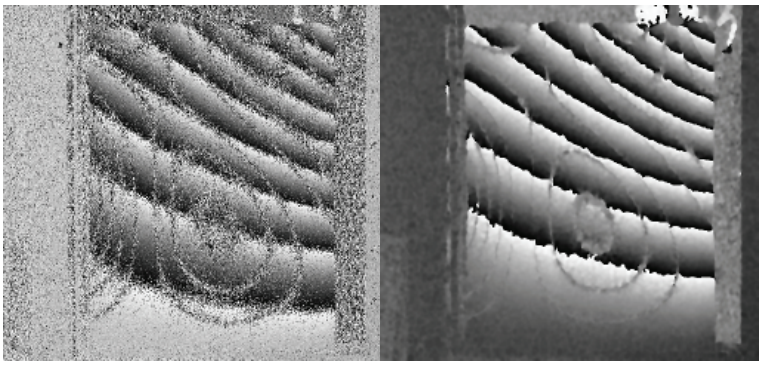


Fig. 15. Non-filtered (left) and filtered (right) patterns showing the static phase difference between two vibrating states of the object

The direction into which speckle interferometry should make the bigger steps is improving its temporal resolution, so as to be able to characterize the relative vibration phase between different points, the vibration waveform, or to cope with non-stationary phenomena.
The use of the pulsed laser is partially a solution, but the costs and technical difficulties are high, and the repetition rate of laser pulses rather limited.
The development of high-rate digital cameras and the increasing data transfer speed between such devices, memory and disk units in computers encouraged an increasing numer of researchers to work in speckle interferometry with high rate cameras.
Different systems and specific problems have been described in the litterature, allowing to appreciate that speckle interferometry is more and more able to study dynamic and non-periodic deformations with a good temporal resolution.
In (Aguanno et al, 2003) is presented the Single Pixel Carrier Based Demodulation approach with a digital CMOS-DSP camera for full-field heterodyne interferometry. The result is a full-field vibration measurement system able to operate, without scanning, like a classical scanning Laser Doppler Vibrometer (LDV). In (Asundi & Singh, 2006; Kaufmann, 2003) are presented the possibilities of dynamic digital holography, respectively dynamic speckle interferometry to measure amplitude and phase.
(Nistea & Borza, 2010b) presented a speckle interferometry system based on a use of a high-rate, low-cost CMOS camera and some of their work with this system. The measurement system uses a 300 mW CW Nd-YAG laser. It may be configured to work with or without temporal phase stepping. If used, phase stepping is implemented by an acousto-optic phase modulator. Some of the results obtained with this system are described in the next sections.

6.1 Object submitted to a mechanical shock: free oscillations, no phase stepping

The measurements of a small plate clamped along the left side (fig. 16) submitted to a shock were made by acquiring at the rate of 15600 frames/second, without phase stepping, correlograms having the dimension 118 x 39 pixels. The temporal history of any pixel contains 21840 points. For a randomly chosen pixel, marked with a cross in fig. 16, it is possible to determine the amplitude at any moment, the free oscillation frequency, the eventual delays between pixels and the damping.

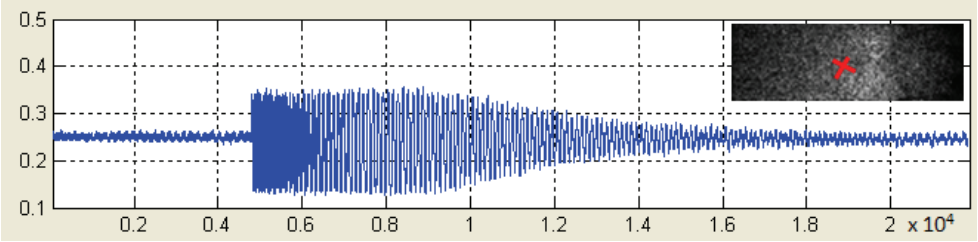


Fig. 16. Temporal history of a randomly chosen pixel, sampled at 15600 images/second

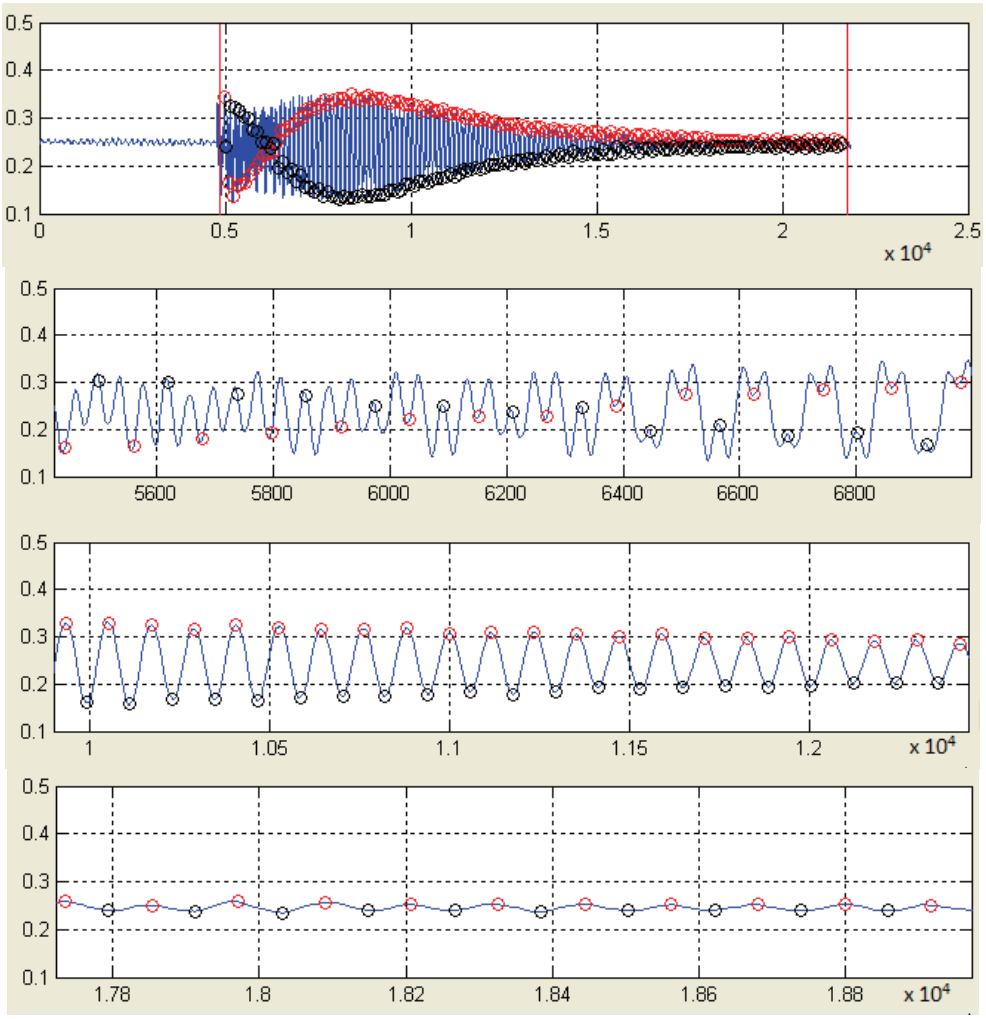


Fig. 17. Minima and maxima for each period of the damped temporal history for the chosen pixel

The frequency (or the period) is found after filtering the signal, fitting a sine function and finding, by an automatic procedure, its extrema (fig. 17) for any period considered. It may also be found by using the Fourier transform (fig. 18 a) of the signal corresponding to the free oscillations following the shock. The values for the period found by the two methods, expressed in number of images acquired at a rate of 15600 frames/second, are very close, 118.1 in the first case and 117.9 in the second. It corresponds to a frequency of about 132 Hz. The damping may be calculated either by Fourier transform followed by a minimum-search fit or by fitting exponentials to the envelopes of the signal, shown in fig. 18 b. The values found by the different methods are also very close, 0.00545 for the first method and between 0.0051 and 0.0057 for the second.

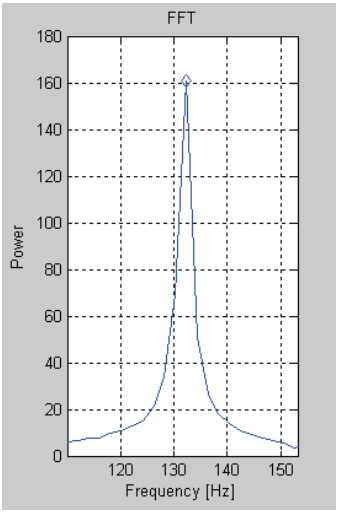


Fig. 18. (a) Frequency of the damped vibration found by FFT

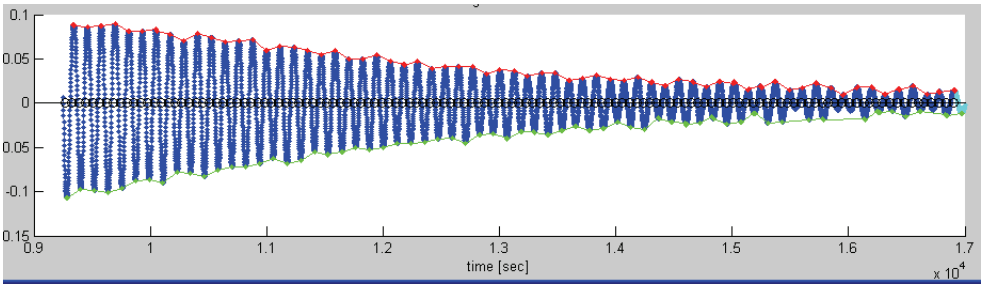


Fig. 18. (b) Envelopes of the free vibration for the considered pixel

6.2 Object in forced sinusoidal vibration: 2- and 4- step phase stepping

The measurements were made by acquiring 111 x 41 pixels wide correlograms at the rate of 26000 frames/second, with four-step phase stepping. The correlograms were then treated in two different ways, either as "classical" 4-step phase-shifted images (6500 interferograms/second), or as 2-step phase-shifted images (13000 interferograms/second).

In both cases, the calculated interferograms used two quadrature fields recorded either with the object at rest, or with the object in vibration.
For the 2-step case, one period of object vibration at 142 Hz was sampled by 93 full-field interferograms at equal time intervals.
The acquired data allow obtaining, at that rate, considering 2-step phase stepping, the following results:

- The wrapped phase differences between the reference state and the state corresponding to any of the 93 sampling instants (fig. 19);

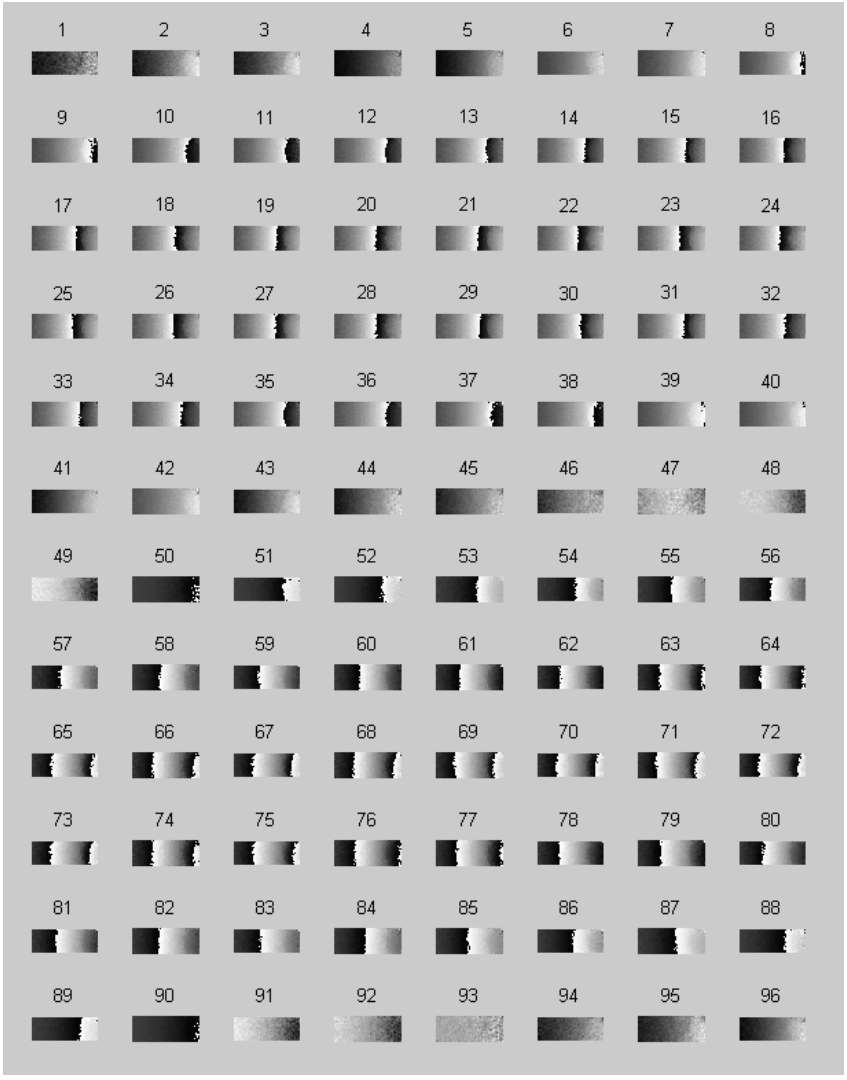


Fig. 19. Temporal history of full-field wrapped vibration amplitude fields

- The unwrapped phase differences between the reference state and the state corresponding to any of the 93 sampling instants (fig. 20);

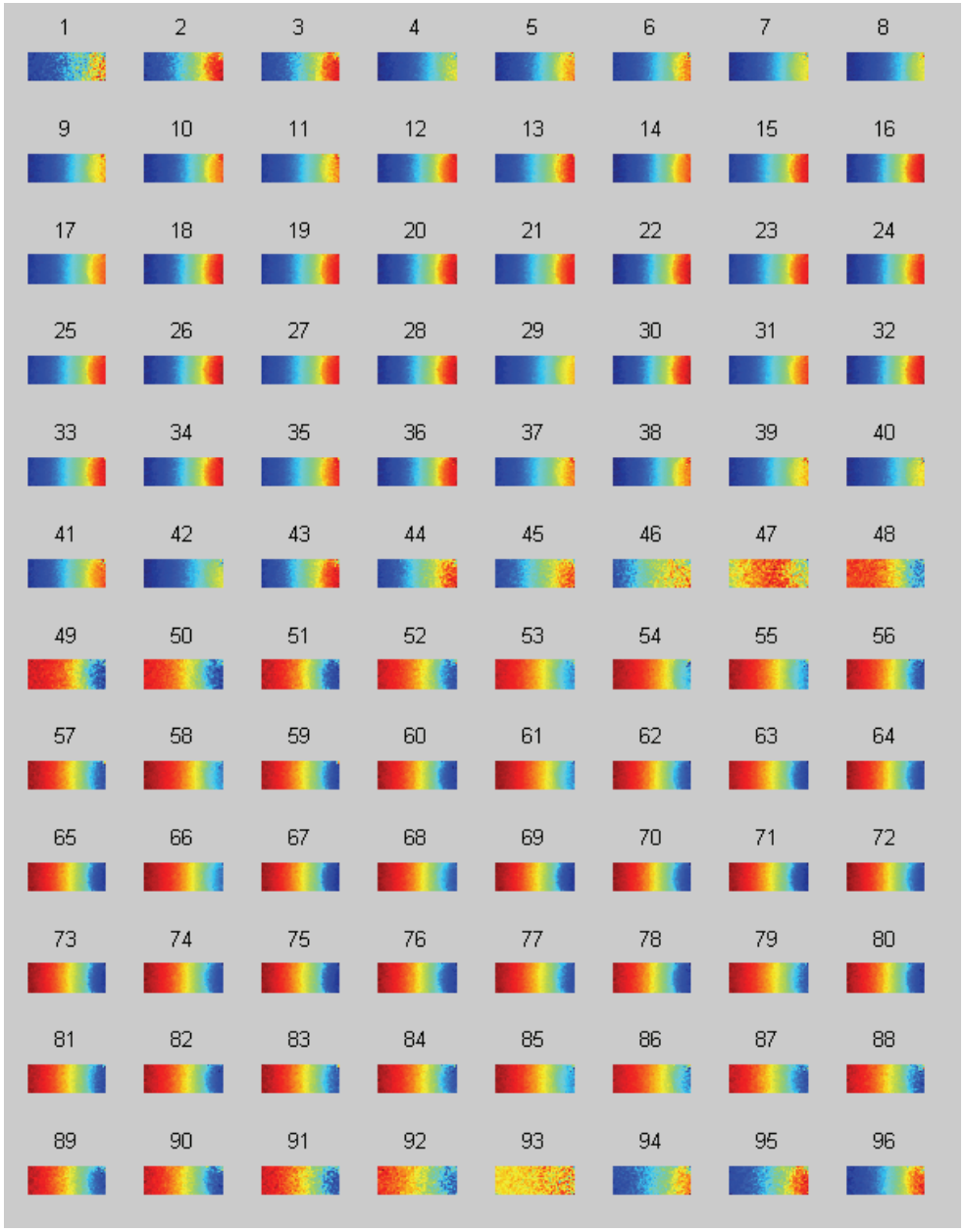


Fig. 20. Temporal history of full-field unwrapped vibration amplitude fields

- The unwrapped phase differences between any pair of phases corresponding to any of the 93 sampling instants (fig. 21);
- As a particular case, the phase differences around the zero-crossing and the extrema of a reference signal (the force signal), allowing to compute the complex frequency response, as mentioned at §3.2 and necessary for experimental modal analysis;

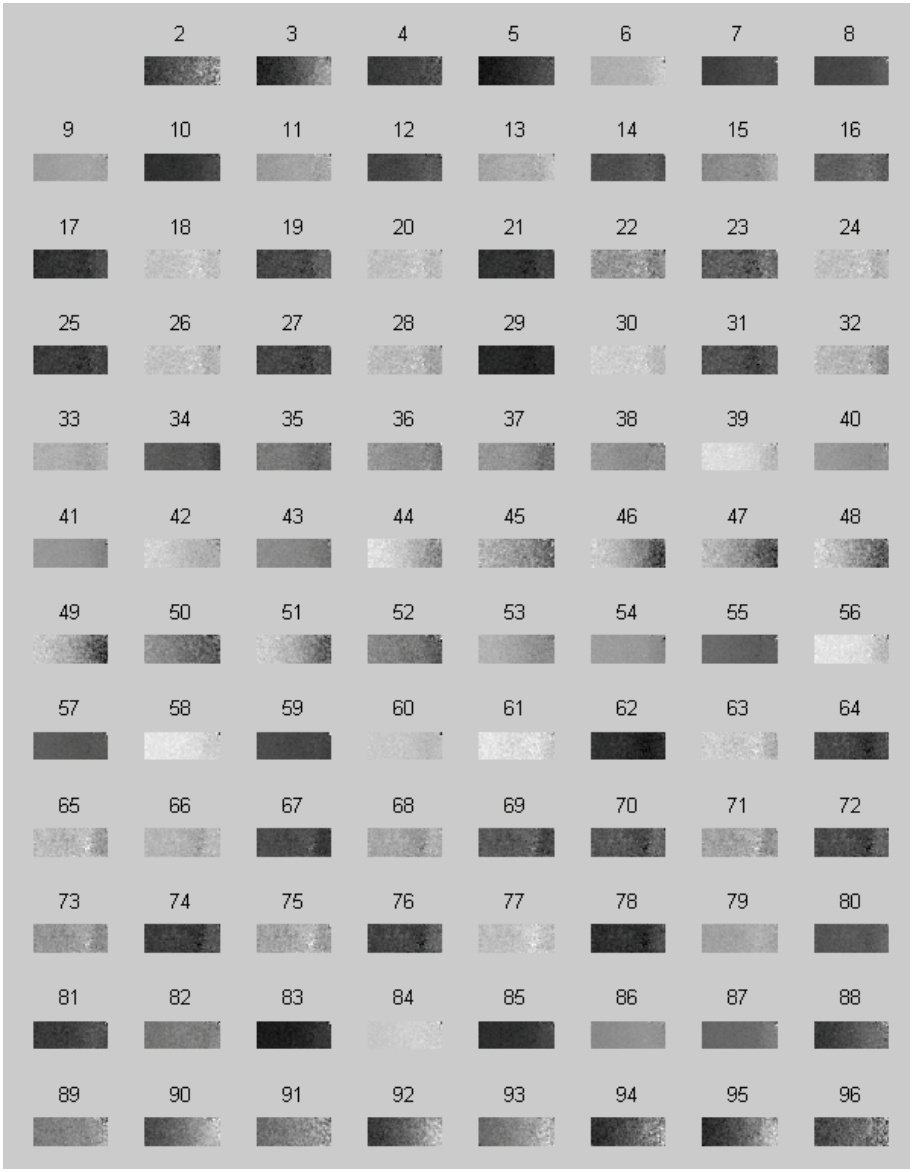


Fig. 21. Temporal history of full-field differences of unwrapped amplitude fields (or temporal history of object velocities)

- By integration during arbitrary time intervals, the full-field time-averaged holograms and the high-resolution time-average holograms (fig. 22).

Just as a remainder that these temporal histories have been recorded to resolve an harmonic vibration by a full-field interferometry system, by computing the sum of the interferograms noted from 1 to 93 in fig. 20, we may obtain either the "classical" time-average hologram in the upper part of fig. 22 or the high-spatial resolution time-average hologram in the lower part of fig. 22.

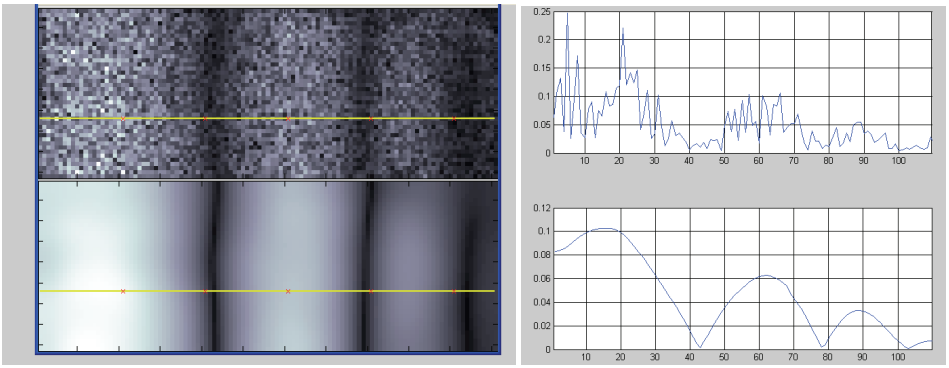


Fig. 22. Time-average and high-resolution time-average hologram during a vibration period

6.3 Object subjected to a variable force of arbitrary waveform; 2- and 4- step phase stepping

The object was perturbed with an arbitrary signal of which one period is shown in Fig. 23 (left) applied to a piezoelectric actuator. The measurements were made at a rate of 26000 images/second, in the same conditions as those in § 6.2. At right is presented the unwrapped (upper plot) and wrapped (lower plot) temporal history of a randomly chosen pixel from the series of 5000 interferograms. It represents the deflection of that point between the two moments indicated at left (START, END).

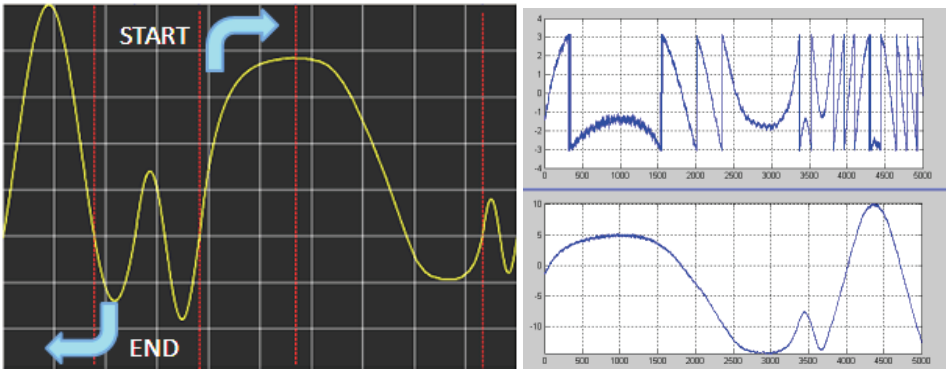


Fig. 23. One period of the force of arbitrary waveform applied by the piezoelectric actuator to the object (left) and the temporal history of a randomly selected pixel in the 5000 interferograms (right)

Such time histories as the one shown in fig. 23 (right) can be obtained for any of the 111 x 41 pixels of the 5000 interferograms. A few, randomly selected interferograms, are presented as wrapped phase patterns in fig. 24.

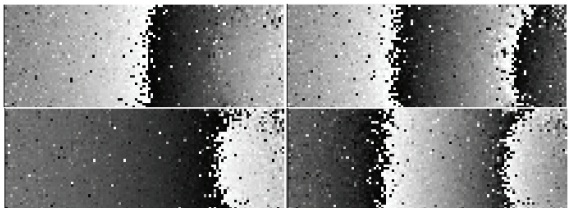


Fig. 24. A few randomly selected interferograms sampling the variable object deformation

7. Conclusions

Speckle interferometry has not stopped developing, in order to overcome its limitations. Important progress has been made towards higher temporal resolution measurements so as to give an adequate answer to the complex problems existing in vibration measurement and in vibracoustics. Higher spatial resolution has also been obtained in spite of the additive and multiplicative noise that covers the interferometric signal.
Many valuable results have been already obtained, and much work still have to be done in the algorithmics and data processing. New challenges appear, related to the manipulation and storage of massive data blocks characterizing high-rate multipoint temporal histories.

8. References

Aguanno, M.V., Lakestani, F., Whelan, M.P., Connelly, M.J. (2003). Single pixel carrier based approach for full field laser interferometry using a CMOS-DSP camera, Proc. SPIE 5251-5242, 304-312, ISSN 0277-786X

Asundi, A. & Singh, V. (2006). Amplitude and phase analysis in digital dynamic Holography, Optics Letters, Vol. 31, 16, 00-01, ISSN 0146-9592

Butters, J.N. & Leendertz, J. A. (1971). Holographic and video techniques applied to engineering measurement, *J. Meas. Control*, 4, 349–354, ISSN: 0142-3312

Borza, D.N. (2000). Stepped-amplitude modulation interferometry a new real-time mechanical vibrations measurement technique, in: Interferometry in Speckle Light, Ed. Jacquot, P., Fournier, J-M., 305-310, Springer, ISBN 3-540-67943, Berlin – Heidelberg, ISBN: 3-540-67943-X

Borza, D.N. (2002). A New Interferometric Method for Vibration Measurement by Electronic Holography, Experimental Mechanics, 42, 4, 1-7, ISSN: 0014-4851

Borza, D.N. (2004). High-resolution time-average electronic holography for vibration measurement, Opt. Lasers Eng., 41, 415-527, ISSN 0143-8166

Borza, D.N. (2005). Mechanical vibration measurement by high-resolution time-averaged digital holography, Meas. Sci. Technol. 16, 1853–1864, ISSN: 0957-0233

Borza, D.N. (2008). Speckle Noise Reduction in Vibration Measurement by Time-average Speckle Interferometry and Digital Holography – a Unifying Approach, Journal of Solid Mechanics and Materials Engineering, 2, 6, 695-706, ISSN:1880-9871

Borza, D.N. (2006). Full-field vibration amplitude recovery from high-resolution time-averaged speckle interferograms and digital holograms by regional inverting of the Bessel function, Opt. Lasers Eng., 44, 747-770, ISSN: 0143-8166

Creath, K. (1985). Phase-shifting speckle interferometry, Appl. Opt., Vol. 24, No. 18, 3053-3058, ISSN: 0003-6935

Creath, K. & Slettemoen, G. (1985). Vibration-observation techniques for digital speckle-pattern Interferometry, J. Opt. Soc. Am. A, Vol. 2, No. 10, 1629-1636, ISSN 1084-7529

Doval, A.F. (2000). A systematic approach to TV holography, Meas. Sci. Tech., 11, R1-R36, ISSN: 0957-0233

Høgmoen, K. & Løkberg, O.J. (1977). 1Detection and measurement of small vibrations using electronic speckle pattern interferometry, Appl. Opt. 16, 1869–75, ISSN: 0003-6935

Jacquot, P. (2008). Speckle interferometry: A review of the principal methods in use for experimental mechanics applications, Strain, 44, 57-69, ISSN 0039-2103

Joenathan, C. (1991). Vibration fringes by phase stepping on an electronic speckle pattern interferometer: an analysis, *Appl. Opt*. 30, 4658–4665, ISSN: 0003-6935

Kaufmann, G. (2003). Phase measurement in temporal speckle pattern interferometry using the Fourier transform method with and without a temporal carrier, *Opt. Comm*., 217, 141-149, ISSN: 0030-4018

Moreau, A., Borza, D.N. & Nistea, I. (2008). Full-field vibration measurement by time-average speckle interferometry and by Doppler vibrometry – A comparison, Strain, 44, 386-397, ISSN: 0039-2103

Nistea, I. (2010). *Développement des techniques optiques et acoustiques de mesure de champs orientées vers la vibroacoustique*, PhD Thesis, Institut National des Sciences Appliquées de Rouen, France

Nistea, I. & Borza, D.N. (2010a). Experimental Analysis Of Failure In Embedded Electronics And Mechatronical Systems Under Thermal Stress, *Proceedings of IEEE-TTTC International Conference on Automation, Quality and Testing, Robotics*, May 28-30, 2010 Cluj Napoca, Romania, ISBN: 978-1-4244-6722-8

Nistea, I. & Borza, D.N. (2010b). Vibration measurement by Speckle Interferometry between high spatial and high temporal resolution, *Proceedings of Colloque francophone "Mesures et Techniques Optiques pour l'Industrie"* – SFO, CMOI, Toulouse (accepted)

Nakadate, S., Yatagai, T., & Saito, H. (1980). Digital speckle-pattern shearing interferometry, Applied Optics 19, 24, 4241-4246, ISSN: 0003-6935

Picart, P., Leval, J., Mounier, D. & Gougeon, S. (2005). Some opportunities for vibration analysis with time averaging in digital Fresnel holography Appl. Opt. 44, 337–43, ISSN: 0003-6935

Pryputniewicz, R.J. & Stetson, K.A. (1989). Measurement of vibration patterns using electro-optic holography. Proc SPIE, 1162, 456–67, ISSN 0277-786X

Singh V., Miao J., Wang, Z., Hegde, G. & Asundi, A. (2007). Dynamic characterization of MEMS diaphragm using time averaged in-line digital holography, *Optics Communications* 280, 285–290, ISSN: 0030-4018

Valera, J., Jones, J., Løkberg, O., Buckberry, C. & Towers, D. (1997). Bi-modal vibration analysis with stroboscopic heterodyned ESPI, *Meas. Sci. Technol*., 8, 648–655, ISSN: 0957-0233

Van der Auweraer H., Steinbichler, H., Vanlanduit, S., Haberstok, C., Freymann, R., Storer, D. & Linet, V. (2002). Application of stroboscopic and pulsed-laser electronic speckle pattern interferometry (ESPI) to modal analysis problems, *Meas. Sci. Technol*., 13, 451–463, ISSN: 0957-0233

Part 5

Digital Algorithms in Holography

17

Reconstruction of Digital Hologram by use of the Wavelet Transform

Jingang Zhong[1,3] and Jiawen Weng[2,1]
[1]*Department of Optoelectronic Engineering, Jinan University,*
[2]*College of Science, South China Agricultural University,*
[3]*Key Laboratory of Optoelectronic Information and Sensing Technologies of Guangdong Higher Education Institutes, Jinan University,*
China

1. Introduction

In combination with charge coupled device (CCD) sensor technology and image processing systems, real time dynamic analysis becomes one of the significant advantages of digital holography. In order to achieve the dynamic analysis, lots of holograms should be recorded and every hologram should be numerical reconstructed. The most popular numerical reconstruction methods for digital holography include the well-known Fresnel diffraction integral method, the angular spectrum method and the convolution-based method. In order to filter out the zero-order term, the twin image term and the parasitic interferences, the process of the spatial filtering must be carried out in these methods. When some noises and parasitic interferences are introduced into the hologram, the spectrum of the virtual image would be disturbed by some other spectrum. It brings difficulties to define the spatial filter because of the blurry boundary and non-regular distribution of the spectrum. Manual spatial filters are often employed. However, defining different manual spatial filters would consume plenty of time for dynamic analysis. Therefore, the process of the spatial filtering limits the application of the digital holography for the dynamic and automatic analysis.
The wavelet transform, a tool excelling for its multi-resolution and localization in the space-frequency domains, is applied to the reconstruction of the object wave from the digital hologram in this chapter. Here the one-dimensional (1D) and two-dimensional (2D) Gabor wavelet transform (GWT) are employed for the analysis. By calculating the wavelet coefficients of the hologram at the ridge and the peak automatically, the object wave at the hologram plane can be reconstructed. At the same time the effect of the zero-order and the twin-image terms are eliminated without the process of the spatial filtering. The reconstruction approach for digital holography by use of the GWT is described in theory. And the experimental results are shown.

2. Principle of the reconstruction technique employing the 1D-GWT

2.1 1D-GWT

The continuous one-dimensional wavelet transform is defined as follows:

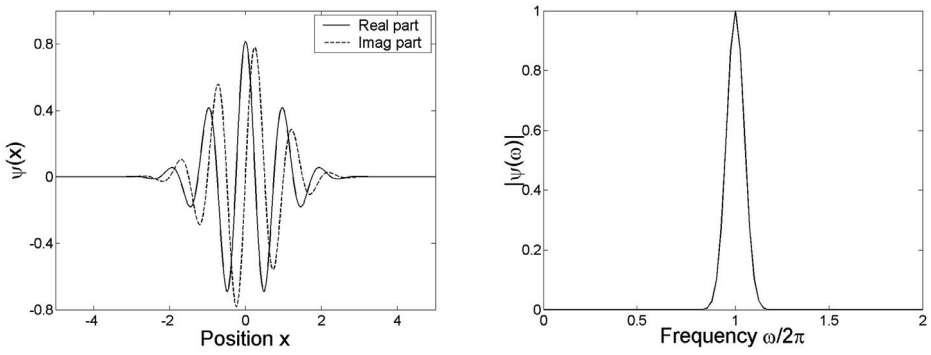
$$W(s,b) = \int_{-\infty}^{\infty} f(x)\psi_{s,b}^{*}(x)\mathrm{d}x = < f(x), \psi_{s,b}(x) > \tag{1}$$

$$\psi_{s,b}(x) = \frac{1}{s}\psi\left(\frac{x-b}{s}\right) \tag{2}$$

where $s > 0$ is the scale parameter related to the frequency; b is the shift parameter related to the position; $f(x)$ is the signal to be analyzed; $\psi_{s,b}(x)$ is the analyzing wavelet obtained by shifting and scaling the mother wavelet $\psi(x)$; * indicates the complex conjugate.
The Gabor wavelet is employed as the mother wavelet. The Gabor wavelet function and its Fourier transform are given as:

$$\begin{cases} \psi(x) = \dfrac{1}{\sqrt[4]{\pi}}\sqrt{\dfrac{2\pi}{\gamma}}\exp\left[-\dfrac{(2\pi/\gamma)^2 x^2}{2} + j2\pi x\right] \\ \psi(\omega) = \dfrac{\sqrt{2\pi}}{\sqrt[4]{\pi}}\sqrt{\dfrac{\gamma}{2\pi}}\exp\left[-\dfrac{(\gamma/2\pi)^2}{2}(\omega - 2\pi)^2\right] \end{cases} \tag{3}$$

where $\gamma = \pi\sqrt{2/\ln 2}$. The Gabor wavelet may be considered as a Gaussian window centered at position $x = 0$ and its Fourier transform centered at frequency $\omega = 2\pi$, as shown in Fig. 1. The function $\psi_{s,b}(x)$ is then centered around $x = b$ with a full width of half maximum $\Delta x = 2s$, and its Fourier transform is centered around $\omega = 2\pi / s$. The scale parameter s controls the spatial and frequency resolution of the wavelet decomposition: A large s corresponds to a stretched wave (high frequency resolution and low spatial resolution), whereas a small s corresponds to a compressed wave (low frequency resolution and high spatial resolution). The wavelet coefficients $W(s,b)$ using the Gabor wavelet describes the distribution of the signal $f(x)$ on the space-frequency plane. Namely, it can be considered that each of the $W(s,b)$ contains the information about the local frequency ($\omega = 2\pi / s$) component of the signal at the position $x = b$.



(a) (b)

Fig. 1. (a) the Gabor wavelet and (b) its Fourier transform.

The reasons for the choice of the Gabor wavelet are as follows: (1) The Gabor wavelet consists of a sine wave and a cosine wave that have been modulated by a Gaussian function. (2) It can represent a complex sinusoidal function that correlates well with the sinusoidal

characteristics of the interferogram. (3) The Gaussian function has the least spread in both domains of space and frequency.
Computing the 1D-GWT, the modulus of the wavelet coefficients can be obtained by:

$$\left|W(s,b)\right| = \sqrt{\left\{imag\left[W(s,b)\right]\right\}^2 + \left\{real\left[W(s,b)\right]\right\}^2} \tag{4}$$

where $imag\left[W(s,b)\right]$ and $real\left[W(s,b)\right]$ indicate the imaginary part and the real part of the $W(s,b)$. At the position b, the ridge of the wavelet transform is defined as the location where the modulus $\left|W(s,b)\right|$ reaches its local maximum along the scaling direction s.

2.2 Reconstruction principle by means of the 1D-GWT

The hologram is created by the interference, in off-axis geometry, between two coherent waves: on one side the wave of interest, called object wave $O(x,y)$, coming from the object, and on the other side a reference wave $R(x,y)$ being plane:

$$O(x,y) = o(x,y)\cdot\exp\left[j\phi(x,y)\right] \tag{5}$$

$$R(x,y) = R_0\cdot\exp\left[j2\pi\frac{\sin\theta}{\lambda}x\right] \tag{6}$$

where (x,y) are the coordinates of the hologram plane; $o(x,y)$ and R_0 are the amplitude of the object and the reference waves; $\phi(x,y)$ is the phase of the object wave; λ is the wavelength; θ is the angle between the propagation direction of the object and the reference waves. The intensity of the hologram $I(x,y)$ can be written as:

$$\begin{aligned} I(x,y) &= R\cdot R^* + O\cdot O^* + O\cdot R^* + O^*\cdot R \\ &= |R_0|^2 + |o(x,y)|^2 + R_0 o(x,y)\exp\left\{j\left[-2\pi\frac{\sin\theta}{\lambda}x + \phi(x,y)\right]\right\} \\ &\quad + R_0 o(x,y)\exp\left\{-j\left[-2\pi\frac{\sin\theta}{\lambda}x + \phi(x,y)\right]\right\} \end{aligned} \tag{7}$$

The first and the second terms form the zero-order term, the third and the fourth terms are the virtual and the real image terms respectively. Take one row of the hologram to analyze. The one-dimensional intensity distribution of the hologram can be written as:

$$I(x) = A(x) + R_0 o(x)\exp\left[j\varphi(x)\right] + R_0 o(x)\exp\left[-j\varphi(x)\right] \tag{8}$$

where $A(x) = |R_0|^2 + |o(x)|^2$ and $\varphi(x) = -2\pi\frac{\sin\theta}{\lambda}x + \phi(x)$. The phase $\varphi(x)$ can be expanded to a Taylor series around b:

$$\varphi(x) = \left[-2\pi\frac{\sin\theta}{\lambda}b + \phi(b)\right] + \left[-2\pi\frac{\sin\theta}{\lambda} + \phi'(b)\right](x-b) + \frac{\phi''(b)}{2!}(x-b)^2 + \cdots \tag{9}$$

When the higher order terms of $(x-b)$ are neglected, the phase $\varphi(x)$ can be expressed as:

$$\varphi(x) \approx \left[-2\pi\frac{\sin\theta}{\lambda}b+\phi(b)\right]+\left[-2\pi\frac{\sin\theta}{\lambda}+\phi'(b)\right](x-b) \tag{10}$$

Setting $A(x)\approx A$, $o(x)\approx o$, the wavelet transform of $I(x)$ becomes:

$$\begin{aligned} W(s,b) &= \int_{-\infty}^{\infty} I(x)\psi_{s,b}^{*}(x)\mathrm{d}x \\ &= \int_{-\infty}^{\infty}\left\{A+R_0 o\exp\left[j\varphi(x)\right]+R_0 o\exp\left[-j\varphi(x)\right]\right\}\psi_{s,b}^{*}(x)\mathrm{d}x \\ &= W_1(s,b)+W_2(s,b)+W_3(s,b) \end{aligned} \tag{11}$$

Perform the above calculation by employing the Gabor wavelet as follows:

$$\begin{aligned} W_1(s,b) &= \int_{-\infty}^{\infty} A\cdot\psi_{s,b}^{*}(x)\mathrm{d}x \\ &= \frac{1}{s}\int_{-\infty}^{\infty} A\cdot\frac{1}{\sqrt[4]{\pi}}\sqrt{\frac{2\pi}{\gamma}}\exp\left[-(2\pi/\gamma)^2\frac{(x-b)^2}{2s^2}\right]\exp\left(-j2\pi\frac{x-b}{s}\right)\mathrm{d}x \\ &\left(set\ \frac{x-b}{s}=t\right) \\ &= A\cdot\frac{1}{\sqrt[4]{\pi}}\sqrt{\frac{2\pi}{\gamma}}\int_{-\infty}^{\infty}\exp\left[-(2\pi/\gamma)^2\frac{t^2}{2}\right]\exp\left(-j2\pi t\right)\mathrm{d}t \\ &= A\cdot\frac{1}{\sqrt[4]{\pi}}\sqrt{\frac{2\pi}{\gamma}}\left\{\int_{-\infty}^{\infty}\exp\left[-(2\pi/\gamma)^2\frac{t^2}{2}\right]\cos\left(2\pi t\right)\mathrm{d}t-j\int_{-\infty}^{\infty}\exp\left[-(2\pi/\gamma)^2\frac{t^2}{2}\right]\sin\left(2\pi t\right)\mathrm{d}t\right\} \\ &= A\cdot\frac{1}{\sqrt[4]{\pi}}\sqrt{\frac{2\pi}{\gamma}}\left\{2\int_{0}^{\infty}\exp\left[-(2\pi/\gamma)^2\frac{t^2}{2}\right]\cos\left(2\pi t\right)\mathrm{d}t\right\} \\ &\left(consider\int_0^{\infty}\exp(-\xi^2x^2)\cos(\eta x)\mathrm{d}x=\frac{\exp\left[-\frac{\eta^2}{4\xi^2}\right]}{2\xi}\sqrt{\pi},\ (\xi>0)\right) \\ &= A\cdot\frac{1}{\sqrt[4]{\pi}}\sqrt{\frac{2\pi}{\gamma}}2\frac{\exp\left[-\frac{(2\pi)^2}{4\frac{(2\pi/\gamma)^2}{2}}\right]}{2\sqrt{\frac{(2\pi/\gamma)^2}{2}}}\sqrt{\pi} \\ &= \frac{\sqrt{\gamma}}{\sqrt[4]{\pi}}A\exp\left(-\frac{\gamma^2}{2}\right) \end{aligned} \tag{12}$$

$$
\begin{aligned}
W_2(s,b) &= \int_{-\infty}^{\infty} R_0 o \exp\left[j\varphi(x) \right] \psi_{s,b}^{*}(x) \mathrm{d}x \\
&= \int_{-\infty}^{\infty} R_0 o \exp\left\{ j\left[-2\pi \frac{\sin\theta}{\lambda} b + \phi(b) \right] \right\} \exp\left\{ j\left[-2\pi \frac{\sin\theta}{\lambda} + \phi'(b) \right](x-b) \right\} \psi_{s,b}^{*}(x) \mathrm{d}x
\end{aligned} \tag{13}
$$

where,

$$
\begin{aligned}
&\int_{-\infty}^{\infty} \exp\left\{ j\left[-2\pi \frac{\sin\theta}{\lambda} + \phi'(b) \right](x-b) \right\} \psi_{s,b}^{*}(x) \mathrm{d}x \\
&= \frac{1}{s} \int_{-\infty}^{\infty} \exp\left\{ j\left[-2\pi \frac{\sin\theta}{\lambda} + \phi'(b) \right](x-b) \right\} \cdot \frac{1}{\sqrt[4]{\pi}} \sqrt{\frac{2\pi}{\gamma}} \exp\left[-(2\pi/\gamma)^2 \frac{(x-b)^2}{2s^2} \right] \exp\left(-j2\pi \frac{x-b}{s} \right) \mathrm{d}x \\
&\left(set \frac{x-b}{s} = t \right) \\
&= \frac{1}{\sqrt[4]{\pi}} \sqrt{\frac{2\pi}{\gamma}} \int_{-\infty}^{\infty} \exp\left\{ j\left[-2\pi \frac{\sin\theta}{\lambda} + \phi'(b) \right] st \right\} \cdot \exp\left[-(2\pi/\gamma)^2 \frac{t^2}{2} \right] \exp(-j2\pi t) \mathrm{d}t \\
&= \frac{1}{\sqrt[4]{\pi}} \sqrt{\frac{2\pi}{\gamma}} \int_{-\infty}^{\infty} \exp\left[-(2\pi/\gamma)^2 \frac{t^2}{2} \right] \exp\left\{ j2\pi t \left[\left(-\frac{\sin\theta}{\lambda} + \frac{\phi'(b)}{2\pi} \right) s - 1 \right] \right\} \mathrm{d}t \\
&= \frac{1}{\sqrt[4]{\pi}} \sqrt{\frac{2\pi}{\gamma}} 2 \int_{0}^{\infty} \exp\left[-(2\pi/\gamma)^2 \frac{t^2}{2} \right] \cos\left\{ 2\pi t \left[\left(-\frac{\sin\theta}{\lambda} + \frac{\phi'(b)}{2\pi} \right) s - 1 \right] \right\} \mathrm{d}t \\
&= \frac{1}{\sqrt[4]{\pi}} \sqrt{\frac{2\pi}{\gamma}} 2 \frac{\exp\left\{ -\dfrac{(2\pi)^2 \left[\left(-\dfrac{\sin\theta}{\lambda} + \dfrac{\phi'(b)}{2\pi} \right) s - 1 \right]^2}{4\dfrac{(2\pi/\gamma)^2}{2}} \right\}}{2\sqrt{\dfrac{(2\pi/\gamma)^2}{2}}} \sqrt{\pi} \\
&= \frac{\sqrt{\gamma}}{\sqrt[4]{\pi}} \exp\left\{ -\frac{\gamma^2}{2} \left[\left(-\frac{\sin\theta}{\lambda} + \frac{\phi'(b)}{2\pi} \right) s - 1 \right]^2 \right\}
\end{aligned} \tag{14}
$$

There is,

$$
W_2(s,b) = \frac{\sqrt{\gamma}}{\sqrt[4]{\pi}} R_0 o \exp\left\{ -\frac{\gamma^2}{2} \left[\left(-\frac{\sin\theta}{\lambda} + \frac{\phi'(b)}{2\pi} \right) s - 1 \right]^2 \right\} \exp\left\{ j\left[-2\pi \frac{\sin\theta}{\lambda} b + \phi(b) \right] \right\} \tag{15}
$$

$$
\begin{aligned}
W_3(s,b) &= \int_{-\infty}^{\infty} R_0 o \exp\left[-j\varphi(x) \right] \psi_{s,b}^{*}(x) \mathrm{d}x \\
&= \int_{-\infty}^{\infty} R_0 o \exp\left\{ -j\left[-2\pi \frac{\sin\theta}{\lambda} b + \phi(b) \right] \right\} \exp\left\{ -j\left[-2\pi \frac{\sin\theta}{\lambda} + \phi'(b) \right](x-b) \right\} \psi_{s,b}^{*}(x) \mathrm{d}x
\end{aligned} \tag{16}
$$

where,

$$\int_{-\infty}^{\infty}\exp\left\{-j\left[-2\pi\frac{\sin\theta}{\lambda}+\phi'(b)\right](x-b)\right\}\psi_{s,b}^{*}(x)\mathrm{d}x=\frac{\sqrt{\gamma}}{\sqrt[4]{\pi}}\exp\left\{-\frac{\gamma^2}{2}\left[\left(-\frac{\sin\theta}{\lambda}+\frac{\phi'(b)}{2\pi}\right)s+1\right]^2\right\} \quad (17)$$

There is,

$$W_3(s,b)=\frac{\sqrt{\gamma}}{\sqrt[4]{\pi}}R_0o\exp\left\{-\frac{\gamma^2}{2}\left[\left(-\frac{\sin\theta}{\lambda}+\frac{\phi'(b)}{2\pi}\right)s+1\right]^2\right\}\exp\left\{-j\left[-2\pi\frac{\sin\theta}{\lambda}b+\phi(b)\right]\right\} \quad (18)$$

Therefore, it can be got as follows:

$$\begin{cases} W_1(s,b)=\dfrac{\sqrt{\gamma}}{\sqrt[4]{\pi}}A\exp\left(-\gamma^2/2\right) \\ W_2(s,b)=\dfrac{\sqrt{\gamma}}{\sqrt[4]{\pi}}R_0o\exp\left\{-\dfrac{\gamma^2}{2}\left[\left(-\dfrac{\sin\theta}{\lambda}+\dfrac{\phi'(b)}{2\pi}\right)s-1\right]^2\right\}\exp\left\{j\left[-2\pi\dfrac{\sin\theta}{\lambda}b+\phi(b)\right]\right\} \\ W_3(s,b)=\dfrac{\sqrt{\gamma}}{\sqrt[4]{\pi}}R_0o\exp\left\{-\dfrac{\gamma^2}{2}\left[\left(-\dfrac{\sin\theta}{\lambda}+\dfrac{\phi'(b)}{2\pi}\right)s+1\right]^2\right\}\exp\left\{-j\left[-2\pi\dfrac{\sin\theta}{\lambda}b+\phi(b)\right]\right\} \end{cases} \quad (19)$$

Because of the feature of $s>0$, the modulus $|W(s,b)|$ reaches its maximum at:

$$s=\frac{1}{-\dfrac{\sin\theta}{\lambda}+\dfrac{\phi'(b)}{2\pi}} \quad (20)$$

Therefore, the wavelet coefficients at the ridge of the 1D-GWT, which is described as $W_{ridge}(b)$, become:

$$\begin{aligned} W_{ridge}(b)=&\frac{\sqrt{\gamma}}{\sqrt[4]{\pi}}A\exp\left(-\gamma^2/2\right)+\frac{\sqrt{\gamma}}{\sqrt[4]{\pi}}R_0o\exp\left\{j\left[-2\pi\frac{\sin\theta}{\lambda}b+\phi(b)\right]\right\} \\ &+\frac{\sqrt{\gamma}}{\sqrt[4]{\pi}}R_0o\exp(-2\gamma^2)\exp\left\{-j\left[-2\pi\frac{\sin\theta}{\lambda}b+\phi(b)\right]\right\} \end{aligned} \quad (21)$$

Here $\exp\left(-\gamma^2/2\right)=\exp\left(-\pi^2/\ln 2\right)\approx 0$ and $\exp\left(-2\gamma^2\right)=\exp\left(-4\pi^2/\ln 2\right)\approx 0$. Note that the position parameter b relates to x, so that Eqs. (21) can be rewritten as:

$$W_{ridge}(x)=\frac{\sqrt{\gamma}}{\sqrt[4]{\pi}}R_0o\exp\left\{j\left[-2\pi\frac{\sin\theta}{\lambda}x+\phi(x)\right]\right\} \quad (22)$$

Multiplying the $W_{ridge}(x)$ by an ideal wave corresponding to a replica of the reference wave, the reconstructed wave $U_{ridge}(x)$ at the hologram plane is obtained as:

$$U_{ridge}(x)=W_{ridge}(x)R_0\exp\left[j2\pi\frac{\sin\theta}{\lambda}x\right]=\frac{\sqrt{\gamma}}{\sqrt[4]{\pi}}{R_0}^2o\exp\left[j\phi(x)\right] \quad (23)$$

A conclusion can be obtained from Eqs. (23) that the reconstructed wave at the ridge of the 1D-GWT is equal to the object wave at the hologram plane multiplied by a constant coefficient. Therefore the numerical reconstruction of the object wave can be carried out by means of the 1D-GWT. At the same time the effect of the zero-order and the twin-image terms are eliminated without the process of the spatial filtering.

It should be noted that the conclusion here is obtained by assuming the higher order terms of $(x-b)$ in Eqs. (9) being neglected. In the following two cases this approximation can be satisfied. First, if $\phi'(x)$ is of slow variation, it can be set that $\phi''(x) \approx \phi'''(x) \approx \cdots \approx 0$, subsequently the higher order terms of $(x-b)$ can be neglected. The phase of the object wave $\phi(x)$ corresponds to the optical path length (OPL) when a plane wave propagates through a phase object. Therefore, the slow variation of $\phi'(x)$ denotes that the first derivative of OPL is of slow variation, or the first derivative of refraction-index distribution of the phase object is of slow variation in physics. Second, because the function $\psi_{s,b}(x)$ is concentrated around $x=b$ with a full width of half maximum $\Delta x = 2s$, the higher order terms of $(x-b)$ can be neglected when $(x-b)<1$ in the full width.

The reconstructed wave at the object plane perpendicular to the propagation axis can be calculated by the Fresnel diffraction integral operation or by the angular spectrum method. Suppose that the wave front at the hologram plane is described as $E(x,y;0)$, and its angular spectrum $A(\xi,\eta;0)=\Im\{E(x,y;0)\}$ can be obtained by taking the Fourier transform. Here $\Im\{\}$ denotes the Fourier transform, and (ξ,η) are the corresponding spatial frequencies of (x,y). The angular spectrum of the reconstructed wave at the plane z perpendicular to the propagation axis, $A(\xi,\eta;z)$, can be calculated from $A(\xi,\eta;0)$ as:

$$A(\xi,\eta;z)=A(\xi,\eta;0)\cdot\exp\left[j\frac{2\pi z}{\lambda}\sqrt{1-(\lambda\xi)^2-(\lambda\eta)^2}\right] \tag{24}$$

Subsequently the reconstructed wave can be obtained by taking the inverse Fourier transform as $E'(x',y';z)=\Im^{-1}\{A(\xi,\eta;z)\}$, where $\Im^{-1}\{\}$ denotes the inverse Fourier transform.

3. Principle of the reconstruction technique employing the 2D-GWT

3.1 2D-GWT

There is a deficiency of the 1D-GWT method that the two-dimensional hologram must be divided into a series of one-dimensional signals along x direction when computing the 1D-GWT on x direction with y fixed. Thereby, the spatial frequency on y direction would be lost. It means that it would fail to analyze the fringes along x direction. Especially when there are some interference fringes of the hologram are along x direction but others along y direction proposed, or there are some parasitic interference signals along x direction, the two-dimensional wavelet transform is necessary.

The continuous two-dimensional wavelet transform is defined as follows:

$$W_f(s,\theta,a,b)=\int\int_{-\infty}^{\infty} f(x,y)\psi^*_{s,\theta}(x,y,a,b)\mathrm{d}x\mathrm{d}y \tag{25}$$

$$\psi_{s,\theta}(x,y,a,b)=\frac{1}{s^2}\psi\left(\frac{x-a}{s},\frac{x-b}{s},\theta\right) \tag{26}$$

where $s(>0)$ and θ are the scale and the rotation parameters related to the frequency; (a,b) is the shift parameter related to the position; $f(x,y)$ is a two-dimensional real signal to be analyzed; $\psi_{s,\theta}(x,y,a,b)$ is the analyzing wavelet obtained by shifting, scaling and rotating the mother wavelet $\psi(x,y)$, and it can be considered as a spatial window; * indicates the complex conjugate. Here the one-dimensional Gabor wavelet is extended to the two-dimensional situation to create a two-dimensional Gabor wavelet as a mother wavelet for its optimal joint space and frequency resolution for a signal. The two-dimensional Gabor wavelet function is given as:

$$\psi(x,y)=\frac{1}{\sqrt[4]{\pi}}\sqrt{\frac{2\pi}{\gamma}}\exp\left[-\frac{(2\pi/\gamma)^2\left(x^2+y^2\right)}{2}+j2\pi\left(x+y\right)\right] \tag{27}$$

where $\gamma=\pi\sqrt{2/\ln 2}$, and j is the imaginary unit. A series of analyzing wavelets, namely "daughter wavelets", are built up by shifting, scaling and rotating the two-dimensional Gabor wavelet, namely "mother wavelet", as follows:

$$\begin{aligned}\psi\left(\frac{x-a}{s},\frac{x-b}{s},\theta\right)=\frac{1}{\sqrt[4]{\pi}}\sqrt{\frac{2\pi}{\gamma}}\exp\left\{-\frac{(2\pi/\gamma)^2\left[\left(x-a\right)^2+\left(y-b\right)^2\right]}{2s^2}\right\}\\ \cdot\exp\left\{j2\pi\frac{\left(x-a\right)\cos\theta+\left(y-b\right)\sin\theta}{s}\right\}\end{aligned} \tag{28}$$

Computing the 2D-GWT of the signal $f(x,y)$, the modulus $\left|W_f(s,\theta,a,b)\right|$ and the phase $\psi(s,\theta,a,b)$ of the wavelet coefficients $W_f(s,\theta,a,b)$ are obtained as:

$$\left|W_f(s,\theta,a,b)\right|=\sqrt{\left\{imag\left[W_f(s,\theta,a,b)\right]\right\}^2+\left\{real\left[W_f(s,\theta,a,b)\right]\right\}^2} \tag{29}$$

$$\psi(s,\theta,a,b)=\arctan\left\{\frac{imag\left[W_f(s,\theta,a,b)\right]}{real\left[W_f(s,\theta,a,b)\right]}\right\} \tag{30}$$

where $imag[W_f(s,\theta,a,b)]$ and $real[W_f(s,\theta,a,b)]$ indicate the imaginary part and the real part of the $W_f(s,\theta,a,b)$. The result of the 2D-GWT is a four-dimensional array related to a parameter a, a parameter b, a scale parameter s and a rotation angle parameter θ. Fig. 2 shows an illustration of a four-dimensional map of size 512×512 pixel, in schematic form, for a specific example of a 2D-GWT at five different scale values $\{2,3,4,5,6\}$ and three different rotation angle values $\{0^0,30^0,60^0\}$.

At a position (a,b), the peak of the 2D-GWT is defined as the location where the modulus $\left|W_f(s,\theta,a,b)\right|$ reaches its local maximum along both the scaling direction s and the rotating direction θ. The determination of the peak of the 2D-GWT can be carried out by the following process. First, an array is constructed by picking up the modulus of the 2D-GWT

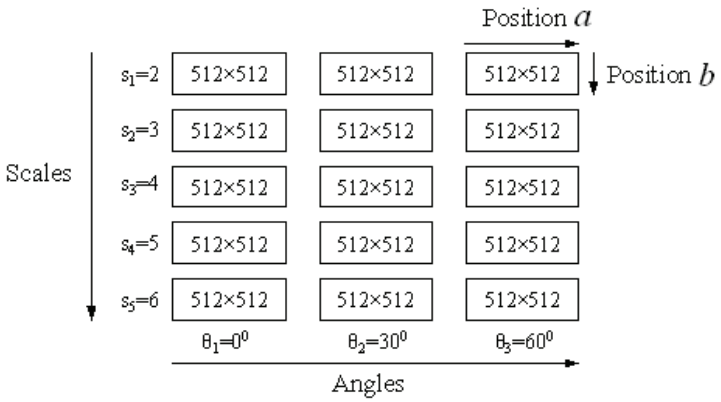


Fig. 2. Illustration of a four-dimensional array produced by the 2D-GWT.

coefficients corresponding to the position (a,b) at each scale and rotation angle value. For the above specific example including five different scale values and three different rotation angle values, a array of size 5×3 is obtained as shown in Fig. 3. Second, the maximum value of this array is found which is pointed out by an arrow in Fig. 3. Finally, the corresponding scale value and rotation angle value are determined, and the peak of the 2D-GWT at the position (a,b) is determined.

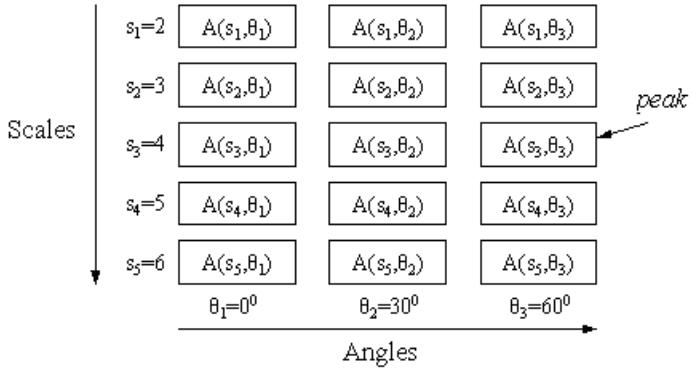


Fig. 3. Illustration of the peak of the 2D-GWT.

3.2 Reconstruction principle by means of the 2D-GWT

The object wave $O(x,y)$ and the reference wave $R(x,y)$ being plane can be written as:

$$O(x,y) = o(x,y)\cdot\exp\left[j\phi(x,y)\right] \tag{31}$$

$$R(x,y) = R_0\cdot\exp\left[j\frac{2\pi}{\lambda}\left(x\cos\alpha + y\cos\beta\right)\right] \tag{32}$$

α and β are the angles between the propagation direction of the object wave and the reference wave on x and y direction respectively. The intensity of the hologram $I(x,y)$ can be written as:

$$\begin{aligned} I(x,y) &= R\cdot R^{*} + O\cdot O^{*} + O\cdot R^{*} + O^{*}\cdot R \\ &= |R_0|^2 + |o(x,y)|^2 + R_0 o(x,y)\exp\left\{ j\left[-\frac{2\pi}{\lambda}\left(x\cos\alpha + y\cos\beta \right) + \phi(x,y) \right] \right\} \\ &\quad + R_0 o(x,y)\exp\left\{ -j\left[-\frac{2\pi}{\lambda}\left(x\cos\alpha + y\cos\beta \right) + \phi(x,y) \right] \right\} \end{aligned} \tag{33}$$

The first and the second terms form the zero-order term, namely the background intensity, which is written as $A(x,y) = |R_0|^2 + |o(x,y)|^2$. The third and the fourth terms are the virtual and the real image terms respectively. Defined the phase of the hologram as follows:

$$\varphi(x,y) = -\frac{2\pi}{\lambda}\left(x\cos\alpha + y\cos\beta \right) + \phi(x,y) \tag{34}$$

The intensity distribution of the hologram can be written as:

$$I(x,y) = A(x,y) + R_0 o(x,y)\exp\left[j\varphi(x,y) \right] + R_0 o(x,y)\exp\left[-j\varphi(x,y) \right] \tag{35}$$

The phase $\varphi(x,y)$ can be expanded to a two-dimensional Taylor series around (a,b):

$$\begin{aligned} \varphi(x,y) = \ & \varphi(a,b) + \left[(x-a)\frac{\partial\varphi(x,y)}{\partial x}\bigg|_{x=a,y=b} + (y-b)\frac{\partial\varphi(x,y)}{\partial y}\bigg|_{x=a,y=b} \right] \\ & + \frac{1}{2!}\left[(x-a)^2\frac{\partial^2\varphi(x,y)}{\partial x^2}\bigg|_{x=a,y=b} + (y-b)^2\frac{\partial^2\varphi(x,y)}{\partial x^2}\bigg|_{x=a,y=b} \right] + \ldots \end{aligned} \tag{36}$$

where $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$ represent the partial differential operation. Assuming $\frac{\partial\left[\varphi(x,y\right]}{\partial x}$ and $\frac{\partial\left[\varphi(x,y\right]}{\partial y}$ are of slow variation, so that the higher order terms of the Taylor series can be neglected. The phase $\varphi(x,y)$ can be expressed as:

$$\varphi(x,y) = \varphi(a,b) + \left[(x-a)\frac{\partial\varphi(x,y)}{\partial x}\bigg|_{x=a,y=b} + (y-b)\frac{\partial\varphi(x,y)}{\partial y}\bigg|_{x=a,y=b} \right] \tag{37}$$

The instantaneous frequency is defined as follows:

$$\begin{cases} f_x = \dfrac{1}{2\pi}\dfrac{\partial\left[\varphi(x,y)\right]}{\partial x} \\ f_y = \dfrac{1}{2\pi}\dfrac{\partial\left[\varphi(x,y)\right]}{\partial y} \end{cases} \tag{38}$$

And the spatial frequency is:

$$\begin{cases} f_x = \dfrac{\cos\alpha}{T} \\ f_y = \dfrac{\cos\beta}{T} = \dfrac{\sin\alpha}{T} \end{cases} \tag{39}$$

where T represents the spatial period of the hologram pattern around the position (a,b). Therefore, Eqs. (37) can be rewritten as:

$$\varphi(x,y) = \varphi(a,b) + 2\pi\frac{(x-a)\cos\alpha + (y-b)\sin\alpha}{T} \tag{40}$$

And the intensity of the hologram expressed by Eqs. (35) can be rewritten as:

$$\begin{aligned} I(x,y) = A(x,y) &+ R_0 o(x,y)\exp\left\{j\left[2\pi\frac{(x-a)\cos\alpha + (y-b)\sin\alpha}{T} + \varphi(a,b)\right]\right\} \\ &+ R_0 o(x,y)\exp\left\{-j\left[2\pi\frac{(x-a)\cos\alpha + (y-b)\sin\alpha}{T} + \varphi(a,b)\right]\right\} \end{aligned} \tag{41}$$

Because of the reference wave being plane, R_0 is constant. And the phase object can be considered to be nearly transparent, i.e. it does not modulate the intensity of the illuminated wave but the phase, therefore it can be set that $o(x,y) \approx o$ and $A(x,y) \approx A$. For the non-intensity-only object, when the amplitude of the object wave $o(x,y)$ is of slow variation compared with the phase term $\exp\left[j\phi(x,y)\right]$, we can set a slowly-varying-amplitude approximation $o(x,y) \approx o$. This approximation is widely adopted in nonlinear optics. And it can be satisfied for the non-intensity-only objects easily, because when the optical path difference δp changes one wavelength, i.e. $\delta p = \lambda$, the phase would changes $(2\pi/\lambda)\delta p = 2\pi$ and $\exp\left[j\phi(x,y)\right]$ would has one periodic variation. For the intensity-only object, it seems that this approximation cannot be satisfied. However, there is not strict intensity-only wave due to the diffraction effect. Therefore the mentioned approximation condition can be satisfied in general. Setting $o(x,y) \approx o$, $A(x,y) \approx A$, and computing the 2D-GWT of $I(x,y)$ becomes:

$$
\begin{aligned}
W(s,\theta,a,b) &= \iint_{-\infty}^{\infty} I(x,y)\psi_{s,\theta}^{*}(x,y,a,b)\mathrm{d}x\mathrm{d}y \\
&= \iint_{-\infty}^{\infty}\left\{A + R_0 o\exp\left[j\varphi(x,y)\right] + R_0 o\exp\left[-j\varphi(x,y)\right]\right\}\psi_{s,\theta}^{*}(x,y,a,b)\mathrm{d}x\mathrm{d}y \\
&= W_1(s,\theta,a,b) + W_2(s,\theta,a,b) + W_3(s,\theta,a,b)
\end{aligned} \tag{42}
$$

Performing the above calculation by employing the two-dimensional Gabor wavelet as follows:

$$
\begin{aligned}
&W_1(s,\theta,a,b) = \iint_{-\infty}^{\infty} A\cdot\psi_{s,\theta}^{*}(x,y,a,b)\mathrm{d}x\mathrm{d}y \\
&= \frac{1}{s^2}\iint_{-\infty}^{\infty} A\cdot\frac{1}{\sqrt[4]{\pi}}\sqrt{\frac{2\pi}{\gamma}}\exp\left\{-\frac{(2\pi/\gamma)^2\left[(x-a)^2+(y-b)^2\right]}{2s^2}\right\}\cdot\exp\left\{-j2\pi\frac{(x-a)\cos\theta+(y-b)\sin\theta}{s}\right\}\mathrm{d}x\mathrm{d}y \\
&= A\cdot\frac{1}{\sqrt[4]{\pi}}\sqrt{\frac{2\pi}{\gamma}}\int_{-\infty}^{\infty}\frac{1}{s}\exp\left[-\frac{(2\pi/\gamma)^2(x-a)^2}{2s^2}\right]\exp\left[-j2\pi\frac{(x-a)}{s}\cos\theta\right]\mathrm{d}x \\
&\qquad\cdot\int_{-\infty}^{\infty}\frac{1}{s}\exp\left[-\frac{(2\pi/\gamma)^2(y-b)^2}{2s^2}\right]\exp\left[-j2\pi\frac{(y-b)}{s}\sin\theta\right]\mathrm{d}y \\
&\left(set\ \frac{x-a}{s}=u,\ \frac{y-b}{s}=v\right) \\
&= A\cdot\frac{1}{\sqrt[4]{\pi}}\sqrt{\frac{2\pi}{\gamma}}\int_{-\infty}^{\infty}\exp\left[-\frac{(2\pi/\gamma)^2u^2}{2}\right]\exp\left[-j2\pi u\cos\theta\right]\mathrm{d}u\cdot\int_{-\infty}^{\infty}\exp\left[-\frac{(2\pi/\gamma)^2v^2}{2}\right]\exp\left[-j2\pi v\sin\theta\right]\mathrm{d}v \\
&= A\cdot\frac{1}{\sqrt[4]{\pi}}\sqrt{\frac{2\pi}{\gamma}}\left\{2\int_0^{\infty}\exp\left[-\frac{(2\pi/\gamma)^2u^2}{2}\right]\cos\left[2\pi u\cos\theta\right]\mathrm{d}u\right\}\cdot\left\{2\int_0^{\infty}\exp\left[-\frac{(2\pi/\gamma)^2v^2}{2}\right]\cos\left[2\pi v\sin\theta\right]\mathrm{d}v\right\} \\
&= A\cdot\frac{1}{\sqrt[4]{\pi}}\sqrt{\frac{2\pi}{\gamma}}\left\{2\frac{\exp\left[-\frac{(2\pi\cos\theta)^2}{4\frac{(2\pi/\gamma)^2}{2}}\right]}{2\sqrt{\frac{(2\pi/\gamma)^2}{2}}}\sqrt{\pi}\right\}\cdot\left\{2\frac{\exp\left[-\frac{(2\pi\sin\theta)^2}{4\frac{(2\pi/\gamma)^2}{2}}\right]}{2\sqrt{\frac{(2\pi/\gamma)^2}{2}}}\sqrt{\pi}\right\} \\
&= \frac{\sqrt{\gamma^3}}{\sqrt[4]{4\pi^3}}A\exp\left(-\frac{\gamma^2}{2}\right)
\end{aligned} \tag{43}
$$

$$
\begin{aligned}
W_2(s,\theta,a,b) &= \iint_{-\infty}^{\infty} R_0 o\exp\left[j\varphi(x)\right]\cdot\psi_{s,\theta}^{*}(x,y,a,b)\mathrm{d}x\mathrm{d}y \\
&= \iint_{-\infty}^{\infty} R_0 o\exp\left[j2\pi\frac{(x-a)\cos\alpha+(y-b)\sin\alpha}{T}\right]\exp\left[j\varphi(a,b)\right]\cdot\psi_{s,\theta}^{*}(x,y,a,b)\mathrm{d}x\mathrm{d}y
\end{aligned} \tag{44}
$$

where

$$
\begin{aligned}
&\iint_{-\infty}^{\infty}\exp\left[j2\pi\frac{(x-a)\cos\alpha+(y-b)\sin\alpha}{T}\right]\cdot\psi_{s,\theta}^{*}(x,y,a,b)\mathrm{d}x\mathrm{d}y\\
&=\frac{1}{s^2}\iint_{-\infty}^{\infty}\exp\left[j2\pi\frac{(x-a)\cos\alpha+(y-b)\sin\alpha}{T}\right]\\
&\qquad\cdot\frac{1}{\sqrt[4]{\pi}}\sqrt{\frac{2\pi}{\gamma}}\exp\left\{-\frac{(2\pi/\gamma)^2\left[(x-a)^2+(y-b)^2\right]}{2s^2}\right\}\exp\left\{-j2\pi\frac{(x-a)\cos\theta+(y-b)\sin\theta}{s}\right\}\mathrm{d}x\mathrm{d}y\\
&=\frac{1}{\sqrt[4]{\pi}}\sqrt{\frac{2\pi}{\gamma}}\int_{-\infty}^{\infty}\frac{1}{s}\exp\left[j2\pi\frac{(x-a)\cos\alpha}{T}\right]\exp\left[-\frac{(2\pi/\gamma)^2(x-a)^2}{2s^2}\right]\exp\left[-j2\pi\frac{(x-a)}{s}\cos\theta\right]\mathrm{d}x\\
&\qquad\cdot\int_{-\infty}^{\infty}\frac{1}{s}\exp\left[j2\pi\frac{(y-b)\sin\alpha}{T}\right]\exp\left[-\frac{(2\pi/\gamma)^2(y-b)^2}{2s^2}\right]\exp\left[-j2\pi\frac{(y-b)}{s}\sin\theta\right]\mathrm{d}y\\
&\left(set\ \frac{x-a}{s}=u,\frac{y-b}{s}=v\right)\\
&=\frac{1}{\sqrt[4]{\pi}}\sqrt{\frac{2\pi}{\gamma}}\int_{-\infty}^{\infty}\exp\left[-\frac{(2\pi/\gamma)^2u^2}{2}\right]\exp\left[j2\pi\frac{s}{T}u\cos\alpha\right]\exp\left[-j2\pi u\cos\theta\right]\mathrm{d}u\\
&\qquad\cdot\int_{-\infty}^{\infty}\exp\left[-\frac{(2\pi/\gamma)^2v^2}{2}\right]\exp\left[j2\pi\frac{s}{T}v\sin\alpha\right]\exp\left[-j2\pi v\sin\theta\right]\mathrm{d}v\\
&=\frac{1}{\sqrt[4]{\pi}}\sqrt{\frac{2\pi}{\gamma}}\left\{2\int_0^{\infty}\exp\left[-\frac{(2\pi/\gamma)^2u^2}{2}\right]\cos\left[2\pi\left(\frac{s}{T}\cos\alpha-\cos\theta\right)u\right]\mathrm{d}u\right\}\\
&\qquad\cdot\left\{2\int_0^{\infty}\exp\left[-\frac{(2\pi/\gamma)^2v^2}{2}\right]\cos\left[2\pi\left(\frac{s}{T}\sin\alpha-\sin\theta\right)v\right]\mathrm{d}v\right\}\\
&=\frac{1}{\sqrt[4]{\pi}}\sqrt{\frac{2\pi}{\gamma}}\left\{2\frac{\exp\left[-\dfrac{4\pi^2\left(\frac{s}{T}\cos\alpha-\cos\theta\right)^2}{4\frac{(2\pi/\gamma)^2}{2}}\right]}{2\sqrt{\frac{(2\pi/\gamma)^2}{2}}}\sqrt{\pi}\right\}\cdot\left\{2\frac{\exp\left[-\dfrac{4\pi^2\left(\frac{s}{T}\sin\alpha-\sin\theta\right)^2}{4\frac{(2\pi/\gamma)^2}{2}}\right]}{2\sqrt{\frac{(2\pi/\gamma)^2}{2}}}\sqrt{\pi}\right\}\\
&=\frac{\sqrt{\gamma^3}}{\sqrt[4]{4\pi^3}}\exp\left\{-\frac{\gamma^2}{2}\left[\left(\frac{s}{T}-1\right)^2+2\frac{s}{T}\left(1-\cos(\alpha-\theta)\right)\right]\right\}
\end{aligned}
\tag{45}
$$

There is,

$$W_2(s,\theta,a,b)=\frac{\sqrt{\gamma^3}}{\sqrt[4]{4\pi^3}}R_0 o\exp\left\{-\frac{\gamma^2}{2}\left[\left(\frac{s}{T}-1\right)^2+2\frac{s}{T}\left(1-\cos\left(\alpha-\theta\right)\right)\right]\right\}\exp\left[j\varphi(a,b)\right] \quad (46)$$

$$\begin{aligned} W_3(s,\theta,a,b)&=\iint_{-\infty}^{\infty}R_0 o\exp\left[-j\varphi\left(x\right)\right]\cdot\psi_{s,\theta}^{*}(x,y,a,b)\mathrm{d}x\mathrm{d}y \\ &=\iint_{-\infty}^{\infty}R_0 o\exp\left[-j2\pi\frac{(x-a)\cos\alpha+(y-b)\sin\alpha}{T}\right]\exp\left[-j\varphi(a,b)\right]\cdot\psi_{s,\theta}^{*}(x,y,a,b)\mathrm{d}x\mathrm{d}y \end{aligned} \quad (47)$$

where

$$\begin{aligned} &\iint_{-\infty}^{\infty}\exp\left[-j2\pi\frac{(x-a)\cos\alpha+(y-b)\sin\alpha}{T}\right]\cdot\psi_{s,\theta}^{*}(x,y,a,b)\mathrm{d}x\mathrm{d}y \\ &=\frac{\sqrt{\gamma^3}}{\sqrt[4]{4\pi^3}}\exp\left\{-\frac{\gamma^2}{2}\left[\left(\frac{s}{T}+1\right)^2-2\frac{s}{T}\left(1-\cos\left(\alpha-\theta\right)\right)\right]\right\} \end{aligned} \quad (48)$$

There is,

$$W_3(s,\theta,a,b)=\frac{\sqrt{\gamma^3}}{\sqrt[4]{4\pi^3}}R_0 o\exp\left\{-\frac{\gamma^2}{2}\left[\left(\frac{s}{T}+1\right)^2-2\frac{s}{T}\left(1-\cos\left(\alpha-\theta\right)\right)\right]\right\}\exp\left[-j\varphi(a,b)\right] \quad (49)$$

Therefore, it can be got as follows:

$$\begin{cases} W_1(s,\theta,a,b)=\dfrac{\sqrt{\gamma^3}}{\sqrt[4]{4\pi^3}}A\exp\left(-\dfrac{\gamma^2}{2}\right) \\ W_2(s,\theta,a,b)=\dfrac{\sqrt{\gamma^3}}{\sqrt[4]{4\pi^3}}R_0 o\exp\left\{-\dfrac{\gamma^2}{2}\left[\left(\dfrac{s}{T}-1\right)^2+2\dfrac{s}{T}\left(1-\cos\left(\alpha-\theta\right)\right)\right]\right\}\exp\left\{j\varphi(a,b)\right\} \\ W_3(s,\theta,a,b)=\dfrac{\sqrt{\gamma^3}}{\sqrt[4]{4\pi^3}}R_0 o\exp\left\{-\dfrac{\gamma^2}{2}\left[\left(\dfrac{s}{T}+1\right)^2-2\dfrac{s}{T}\left(1-\cos\left(\alpha-\theta\right)\right)\right]\right\}\exp\left\{-j\varphi(a,b)\right\} \end{cases} \quad (50)$$

Because of the feature of $s>0$, the modulus $\left|W_f(s,\theta,a,b)\right|$ reaches its maximum at $s=T$ and $\theta=\alpha$ in theory according to Eqs. (50). It means that not only the spatial frequency on x direction but also the spatial frequency on y direction, which cannot be obtained by the 1D-GWT, are determined according to Eqs. (39). Therefore, the wavelet coefficients at the peak of the 2D-GWT are obtained according to not only f_x but also f_y, which is described as $W_{peak}(a,b)$:

$$\begin{aligned} W_{peak}(a,b)=&\frac{\sqrt{\gamma^3}}{\sqrt[4]{4\pi^3}}A\exp\left(-\gamma^2/2\right)+\frac{\sqrt{\gamma^3}}{\sqrt[4]{4\pi^3}}R_0 o\exp\left\{j\varphi(a,b)\right\} \\ &+\frac{\sqrt{\gamma^3}}{\sqrt[4]{4\pi^3}}R_0 o\exp(-2\gamma^2)\exp\left\{-j\varphi(a,b)\right\} \end{aligned} \quad (51)$$

Here $\exp\left(-\gamma^2/2\right)=\exp\left(-\pi^2/\ln 2\right)\approx 0$ and $\exp\left(-2\gamma^2\right)=\exp\left(-4\pi^2/\ln 2\right)\approx 0$. Note that the position parameter (a,b) relates to (x,y), so that Eqs. (51) can be rewritten as:

$$W_{peak}(x,y)=\frac{\sqrt{\gamma^3}}{\sqrt[4]{4\pi^3}}R_0 o\exp\left\{j\left[-\frac{2\pi}{\lambda}\left(x\cos\alpha+y\cos\beta\right)+\phi(x,y)\right]\right\} \tag{52}$$

Multiplying the $W_{peak}(x,y)$ by an ideal wave corresponding to a replica of the reference wave, the reconstructed wave $U_{peak}(x,y)$ at the hologram plane is obtained as:

$$\begin{aligned} U_{peak}(x,y) &= W_{peak}(x,y)R_0\exp\left[j\frac{2\pi}{\lambda}\left(x\cos\alpha+y\cos\beta\right)\right] \\ &= \frac{\sqrt{\gamma^3}}{\sqrt[4]{4\pi^3}}{R_0}^2 o\exp\left[\phi(x,y)\right] \end{aligned} \tag{53}$$

A conclusion can be obtained from Eqs. (53) that the reconstructed wave at the peak of the 2D-GWT is equal to the object wave at the hologram plane multiplied by a constant coefficient. Therefore the numerical reconstruction of the object wave, including not only the amplitude but also the phase information, can be carried out by means of the 2D-GWT. At the same time the effect of the zero-order and the twin-image terms is eliminated without the process of the spatial filtering. The reconstructed wave at the object plane perpendicular to the propagation axis can be calculated by the Fresnel diffraction integral operation or by the angular spectrum method, which is described by Eqs. (24).

4. Experimental result

The system depicted in Fig. 4 is analogous to a Mach-Zehnder interferometer. A He-Ne laser beam, followed by a beam enlarger (BE), is separated into two beams by a beam splitter BS1. One beam serves as a reference wave, and another beam illuminates the biological specimen. A microscope objective MO1 collects the object wave transmitted by the specimen and produces a magnified image of the specimen at a distance behind the CCD camera.

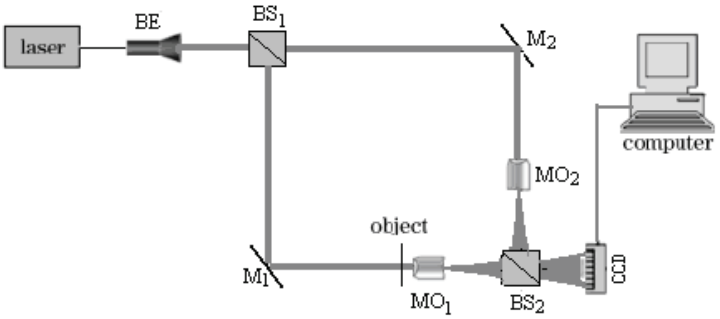


Fig. 4. Apparatus for digital holography. (BE: beam enlarger; BS1, BS2: beam splitter; M1, M2: mirror; MO1, MO2: microscope objective; CCD: charge coupled device.)

Another MO2 is placed in the reference arm to get the matching wave front curvatures. The beam splitter BS2 placed in front of the CCD camera combines the two beams, and the holograms are recorded by the CCD camera. The slight angle is introduced between the object and the reference beams by tilting the beam splitter BS2 for the off-axis holography.
An onion specimen placed between the slide and the cover slide is illuminated by the laser beam. Fig. 5(a) shows the image of the onion specimen recorded by the CCD camera when the reference wave is turn off. Fig. 5(b) is the hologram with the onion specimen recorded by the CCD camera of 512×512 pixels. In order to compensate for phase aberrations and image distortion, a reference hologram without the presence of any specimen is also recorded, as shown in Fig. 5(c). Fig. 5(d) is the spectrum of the hologram with the specimen on the logarithmic coordinates.

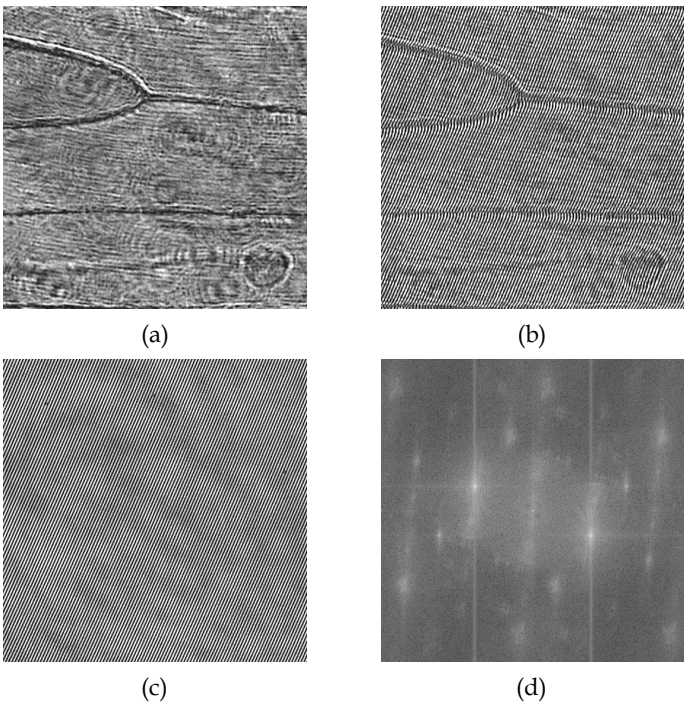
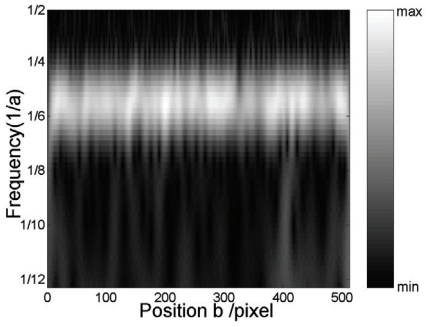
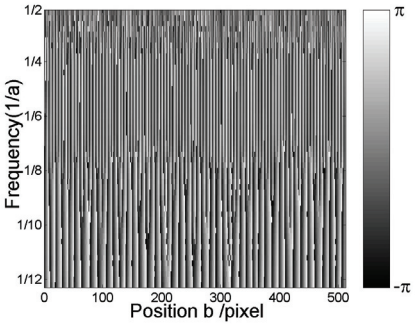


Fig. 5. Hologram. (a) the image of the onion specimen; (b), (c) are the respectively hologram with and without the onion specimen; (d) the spectrum of the hologram with the onion specimen on the logarithmic coordinates.

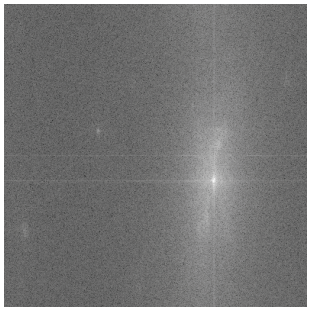
First, the 1D-GWT is employed for the analysis. Fig. 6(a) and (b) show the modulus and the phase of the wavelet coefficients at the 256th row of the hologram. By detecting the ridge of the 1D-GWT and multiplying the wavelet coefficients at the ridge of the 1D-GWT by the numerical reference wave, the object wave is reconstructed. Fig. 6(c) shows the spectrum of the reconstructed wave by employing the 1D-GWT. It shows that the high frequency information of the object wave is reserved, which might be filtered out by the process of the filtering. Fig. 6(d) and (e) show the amplitude and the phase modulo 2π, coded to 256 gray levels, of the reconstructed wave. The amplitude image shown in Fig. 6(d) seems



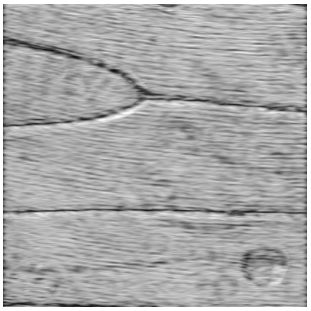
(a)



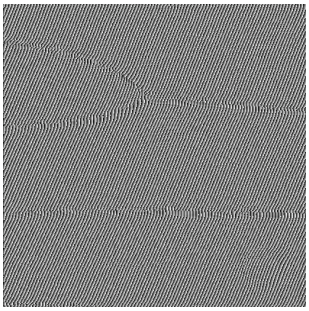
(b)



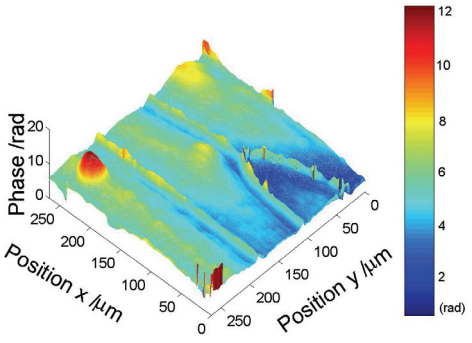
(c)



(d)



(e)



(f)

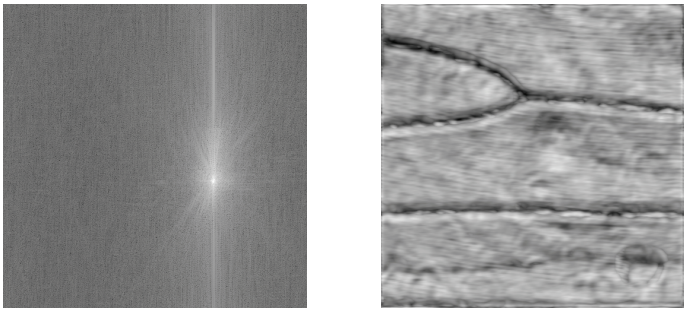
Fig. 6. Holography of an onion specimen by the analysis of the 1D-GWT. (a) the modulus and (b) the phase of the wavelet coefficients at the 256th row of the hologram; (c) the spectrum, (d) the amplitude, (e) the wrapped phase and (f) the unwrapped phase of the reconstructed wave.

$\lvert W_I(s,\theta,256,256)\rvert$		s								
		2	3	4	5	6	7	8	9	10
θ	0°	5.22E+00	2.38E+01	2.45E+01	5.37E+00	6.99E+00	5.83E+00	3.36E+00	3.33E+00	4.82E+00
	18°	3.91E+00	2.35E+01	1.13E+02	**1.49E+02**	5.57E+01	6.26E+00	2.69E+00	1.36E+00	7.02E-01
	36°	2.51E+00	1.58E+01	6.70E+01	7.37E+01	1.97E+01	6.95E+00	3.52E+00	9.38E-01	2.12E+00
	54°	3.16E+00	1.01E+01	1.19E+01	9.29E+00	8.86E+00	5.45E+00	2.36E+00	3.03E+00	4.36E+00
	72°	2.43E+00	5.24E+00	5.90E+00	4.79E+00	8.69E+00	7.49E+00	3.95E+00	1.11E+00	4.61E+00
	90°	6.22E+00	6.35E+00	1.16E+01	1.34E+01	1.85E+01	2.74E+01	2.67E+01	1.66E+01	6.25E+00
	108°	1.91E+00	2.01E+00	3.87E+00	3.61E+00	5.20E+00	6.66E+00	5.49E+00	2.69E+00	1.75E+00
	126°	7.92E-02	9.85E-02	1.32E-01	1.53E-01	1.83E-01	1.42E-01	6.80E-02	6.17E-02	7.50E-02
	144°	1.34E-03	1.91E-03	2.00E-03	1.70E-03	9.15E-04	7.14E-04	7.03E-04	4.03E-04	1.73E-04
	162°	2.93E-04	3.19E-04	3.48E-04	3.66E-04	3.71E-04	3.70E-04	3.71E-04	3.73E-04	3.77E-04

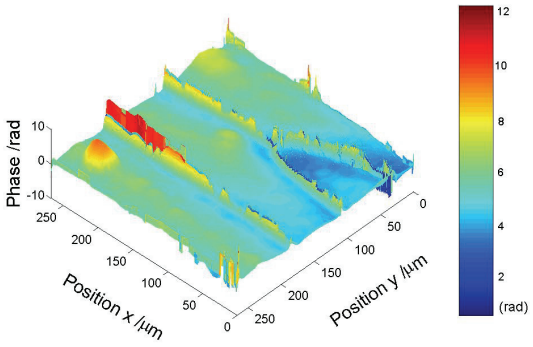
Table 1. The modulus of the wavelet coefficients at the position $(256,256)$.

to suffer from horizontal "ripples". This "ripples" is likely to be an additional interference signal produced by the reflected wave on the slides, which appear in the image of the onion specimen shown in Fig. 5(a). Because 1D-GWT is carried out on the horizontal direction, it fails to eliminate this parasitic interference signal. It might be eliminated by employing the 2D-GWT. After removal of the 2π ambiguity by a phase unwrapping process, the unwrapped phase is obtained, as shown in Fig. 6(f).

In order to eliminate the "ripples" as show in Fig. 6(d), the 2D-GWT is employed for the analysis. Table 1 shows the modulus of the wavelet coefficients with nine different scale values from 2 to 10 with interval of 1, and ten different rotation angle values from 0^0 to 162^0 with interval of 18^0 at the position $(256,256)$. It shows that the modulus reaches the maximum when $s=5$ and $\theta=18^0$, i.e. the peak of the 2D-GWT of the hologram at this position is determined. Fig. 7(a), (b) and (c) show the spectrum on the logarithmic coordinates, the amplitude and the unwrapped phase of the reconstructed wave by the analysis of the 2D-GWT. From the comparison of Fig. 5(d), Fig. 6(c) and Fig. 7(a), i.e. the spectrum distribution of the hologram, the reconstructed wave by the analysis of the 1D-GWT and the 2D-GWT, it can be seen that there is some unwanted spectrum by the analysis of the 1D-GWT, which results in failing to eliminate the parasitic interference signals along x direction like "ripples" as show in Fig. 6(d). While there is a more accurate spectrum by the analysis of the 2D-GWT as shown in Fig. 7(a), and the result is improved by employing the 2D-GWT as show in Fig. 7(b).



(a) (b)



(c)

Fig. 7. Holography of an onion specimen by the analysis of the 2D-GWT. (a) the spectrum, (d) the amplitude, (b) the wrapped phase and (c) the unwrapped phase of the reconstructed wave.

5. Conclusion

In this chapter, the theory of the 1D-GWT and 2D-GWT techniques for the analysis of the digital holography to obtain the quantitative phase especially for the phase object are demonstrated in detail. The results of the experiment of an onion specimen show the conclusion that the numerical reconstruction of the phase object can be carried out by the analysis of the GWT. At the same time the effect of the zero-order term and the twin-image term are eliminated without the spatial filtering. One of the significant advantages of the technique here we presented is that it can be applied to the dynamic live biological specimen analysis in the digital holographic phase-contrast microscopy.

6. References

Abid, A. Z.; Gdeisat, M. A.; Burton, D. R.; Lalor, M. J. & Lilley, F. (2007). Spatial fringe pattern analysis using the two-dimensional continuous wavelet transform employing a cost function. *Applied Optics*, Vol. 46, No. 24, August 2007 6120-6126, ISSN 0003-6935

Anderson, W. L. & Diao, H. (1995). Two-dimensional wavelet transform and application to holographic particle velocimetry. *Applied Optics*, Vol. 34, No. 2, January 1995 249-255, ISSN 0003-6935

Belaïd, S.; Lebrun, D. & Özkul, C. (1997). Application of two-dimensional wavelet transform to hologram analysis: visualization of glass fibers in a turbulent flame. *Optical Engineering*, Vol. 36, No. 7, July 1997 1947, ISSN 0091-3286

Carl, D.; Kemper B.; Wernicke, G. & Bally, G. von (2004). Parameter-optimized digital holographic microscope for highresolution living-cell analysis. *Applied Optics*, Vol. 43, No. 36, December 2004 6536-6544, ISSN 0003-6935

Cesar, A. & Taeeeui, K. (2003). Determination of strains from fringe patterns using space-frequency representations. *Optical Engineering*, Vol. 42, No. 11, November 2003 3182-3193, ISSN 0091-3286

Colomb, T.; Cuche, E.; Charrière, F.; Kühn, J.; Aspert, N.; Montfort, F.; Marquet, P. & Depeursinge, C. (2006). Automatic procedure for aberration compensation in digital holographic microscopy and applications to specimen shape compensation. *Applied Optics*, Vol. 45, No. 5, February 2006 851-863, ISSN 0003-6935

Colomb, T.; Kühn, J.; Charrière, F. & Depeursinge, C. (2006). Total aberrations compensation in digital holographic microscopy with a reference conjugated hologram. *Optics Express*, Vol. 14, No. 10, May 2006 4300-4306, ISSN 1094-4087

Cuche, E.; Marquet, P. & Depeursinge, C. (1999). Simultaneous amplitude-contrast and quantitative phase-contrast microscopy by numerical reconstruction of Fresnel off-axis holograms. *Applied Optics*, Vol. 38, No. 34, December 1999 6994-7001, ISSN 0003-6935

Gdeisat, M. A.; Burton, D. R. & Lalor, M. J. (2006). Spatial carrier fringe pattern demodulation by use of a two-dimensional continuous wavelet transform. *Applied Optics*, Vol. 45, No. 34, December 2006 8722-8732, ISSN 0003-6935

Goodman, J. W. (1996). *Introduction to Fourier Optics*, Mc Graw-Hill Companies, ISBN 0-07-024254-2, United States of America

Jeong, H. (2001). Analysis of plate wave propagation in anisotropic laminates using a wavelet transform. *NDT & E International*, Vol. 34, No. 3, April 2001 185-190, ISSN 0963-8695

Kadooka, K.; Kunoo, K.; Uda, N.; Ono, K. & Nagayasu, T. (2003). Strain analysis for moiré interferometry using the two-dimensional continuous wavelet transform. *Experimental Mechanics*, Vol. 43, No. 1, March 2003 45-51, ISSN 0014-4851

Kemper, B. & Bally, G. von (2008). Digital holographic microscopy for live cell applications and technical inspection. *Applied Optics*, Vol. 47, No. 4, February 2008 A52-A61, ISSN 0003-6935

Kühn, J.; Colomb, T.; Montfort, F.; Charrière, F.; Emery, Y.; Cuche, E.; Marquet, P. & Depeursinge, C. (2007). Real-time dual-wavelength digital holographic microscopy

with a single hologram acquisition. *Optics Express*, Vol. 15, No. 12, June 2007 7231-7242, ISSN 1094-4087

Langehanenberg, P.; Kemper, B.; Dirksen, D. & Bally, G. von (2008). Autofocusing in digital holographic phase contrast microscopy on pure phase objects for live cell imaging. *Applied Optics*, Vol. 47, No. 19, July 2008 D176-D182, ISSN 0003-6935

Lebrun, D.; Belaïd, S. & Özkul, C. (1999). Hologram reconstruction by use of optical wavelet transform. *Applied Optics*, Vol. 38, No. 17, June 1999 3730-3734, ISSN 0003-6935

Li, Zh.; Gu, B. & Yang, G. (1999). Slowly varying amplitude approximation appraised by transfer-matrix approach. *Physical Review B*, Vol. 60, No. 15, October 1999 10644 – 10647, ISSN 0163-1829

Liebling, M.; Blu, T. & Unser, M. (2003). Fresnelets: new multiresolution wavelet bases for digital holography. *IEEE Transactions on Image Processing*, Vol. 12, No. 1, January 2003 29-43, ISSN 1057-7149

Mann, C. J.; Yu, L.; Lo, C. & Kim, M. K. (2005). High-resolution quantitative phase-contrast microscopy by digital holography. *Optics Express*, Vol. 13, No. 22, October 2005 8693-8698, ISSN 1094-4087

Marquet, P.; Rappaz, B.; Magistretti, P. J.; Cuche, E.; Emery, Y.; Colomb, T.; & Depeursinge, C. (2005). Digital holographic microscopy: a noninvasive contrast imaging technique allowing quantitative visualization of living cells with subwavelength axial accuracy. *Optics Letters*, Vol. 30, No. 5, March 2005 468-470, ISSN 0146-9592

Montfort, F.; Charrière, F. & Colomb, T. (2006). Purely numerical compensation for microscope objective phase curvature in digital holographic microscopy: influence of digital phase mask position. *Journal of the Optical Society of America A*, Vol. 23, No. 11, November 2006 2944-2953, ISSN 1520-8532

Shimobaba, T.; Sato, Y.; Miura, J.; Takenouchi, M. & Ito, T. (2008). Real-time digital holographic microscopy using the graphic processing unit. *Optics Express*, Vol. 16, No. 16, August 2008 11776-11781, ISSN 1094-4087

Wang, Z. & Ma, H. (2006). Advanced continuous wavelet transform algorithm for digital interferogram and processing. *Optical Engineering*, Vol. 45, No. 4, April 2006 1-5, ISSN 0091-3286

Weng, J.; Zhong, J. & Hu, C. (2008). Digital reconstruction based on angular spectrum diffraction with the ridge of wavelet transform in holographic phase-contrast microscopy. *Optics Express*, Vol. 16, No. 26, December 2008 21971-21981, ISSN 1094-4087

Weng, J.; Zhong, J. & Hu, C. (2009). Phase reconstruction of digital holography with the peak of the two-dimensional Gabor wavelet transform. *Applied Optics*, Vol. 48, No. 18, June 2009 3308-3316, ISSN 0003-6935

Weng, J.; Zhong, J. & Hu, C. (2010). Automatic spatial filtering to obtain the virtual image term in digital holographic microscopy. *Applied Optics*, Vol. 49, No. 2, January 2010 189-195, ISSN 0003-6935

Yu, L. & Kim, M. K. (2005). Wavelength-scanning digital interference holography for tomographic three-dimensional imaging by use of the angular spectrum method. *Optics Letters*, Vol. 30, No. 16, August 2005 2092-2094, ISSN 0146-9592

Zhong, J. & Weng, J. (2004). Spatial carrier-fringe pattern analysis by means of wavelet transform: wavelet transform profilomatry. *Applied Optics*, Vol. 43, No. 26, September 2004 4993-4998, ISSN 0003-6935

Zhong, J. & Weng, J. (2005). Phase retrieval of optical fringe pattern form the ridge of a wavelet transform. *Optics Letters*, Vol. 30, No. 19, October 2005 2560-2562, ISSN 0146-9592

18

Iterative Noise Reduction in Digital Holographic Microscopy

Victor Arrizón[1], Ulises Ruiz[1] and Maria Luisa Cruz[2]
[1]*Instituto Nacional de Astrofísica, Optica y Electronica,*
[2]*Universidad del Istmo, Oaxaca,*
México

1. Introduction

Digital holographic microscopy (DHM) is an optical technique useful to retrieve quantitative information of microscopic objects with subwavelength axial accuracy (Haddad et al., 1992; Zhang-Yamaguchi, 1998; Pedrini et al., 1999; Ferraro et al., 2005). An important feature of DHM is its non-invasive character that allows the appropriate study of microorganisms in vivo (Car et al., 2004; Marquet et al., 2005; Rappaz et al., 2005, Colomb et al., 2006; Javidi et al., 2006; Charrière et al., 2006). Some DHM setups employ microscope objectives to form the image of the object (Ferraro et al., 2003; Mann et al., 2005; Colomb et al., 2006). However, it is also possible to implement DHM in a lensless setup, where the light field scattered from the object is used instead of its projected image (Wagner et al., 1999; Schnars-Juptner, 2002; Repetto et al., 2004; Cruz et al., 2008; Oh et al., 2010). A conventional setup in DHM is based on the use of an external reference wave (RW), which interferes with the field generated by the object, forming an intensity pattern which is known as hologram of the field (Takeda et al., 1982; Kreis, 1986; Yamaguchi-Zhang, 1997; Yamaguchi et al., 2001; Arrizón-Sanchez, 2004; Liebling, 2004; Quian, 2006; Meneses-Fabian, 2006; Guo, 2007; Cruz et al., 2009). Alternately, it is possible to omit the external RW, with a consequent simplification of the optical setup (Pedrini, 1998; Xu, 2002; Javidi, 2005; Morlens, 2006; Garcia-Sucerquia, 2006; Hwang-Han, 2007; Singh-Asundi, 2009). In any case, either the object hologram or the intensity of the field scattered by object are recorded with an electronic intensity meter, such as a charge-coupled device (CCD). As a final step of DHM, the recorded object hologram (and other intensity patterns) are digitized and processed in a computer, with the aim of reconstructing the microscopic features of the object.

In this chapter we first discuss the lensless setup that employs an off-axis RW, which is known as "lensless off axis digital holography" (LOADH) setup. We also consider a more simplified optical setup, which omits both the lens and the external RW. We refer to this system using the expression "referenceless on axis digital holography" (RLOADH) setup. In RLOADH it is still possible and convenient to split the field propagated from the object into a non scattered field and a scattered field. Even when the non scattered field is not added externally, it is often identified as a RW.

The LOADH and RLOADH approaches employ fewer optical components, and tend to be cheaper and more robust, than other DHM setups. In consequence, the necessity of special

considerations to compensate the influence of imperfections of multiple optical components is reduced in LOADH and RLOADH. In both simplified approaches the presence of the RW (external and internal respectively), generates a hologram which is formed by three terms, namely: the object term, its complex conjugate, and the zero order term. The goal of the digital reconstruction process is to obtain the complex object field from the hologram intensity pattern. This process is usually facilitated by using additional information, e. g. a constraint in the object plane. In the case of the LOADH it is also possible to measure, in addition to the hologram, the intensity patterns generated by the RW and by the object field. These two intensity patterns allow the elimination of the zero order in the hologram modulation.

If the spatial frequency bandwidth (BW) of the object under study is small enough, the isolation of the object field term can be performed (at least in the LOADH procedure) by employing the Fourier domain spatial filtering (FDSF) (Takeda, 1982). However, if DHM is used to study microscopic samples of large BW, the isolation of the object field component, in the hologram, cannot be realized by this simple spatial filtering procedure. In this case, one of the useful procedures is to perform the elimination of non desired hologram components by iterative procedures (Fienup, 1987; Cederquist et al., 1989; Wu, 2004; Denis et al., 2005; Hwang-Han, 2007; Nakamura, 2007).

In this chapter we describe an iterative method that effectively recovers the object field component from the recorded hologram, in both the LOADH and RLOADH setups. In both cases, it is required to implement multiple field propagations from the hologram plane to the object plane, and vice verse. In each one of these planes, specific amplitude constraints are applied. The constraint in the hologram plane is the recorded hologram amplitude itself. On the other hand, we assume that in the object plane the field is different to zero only within the area of a binary transmittance pupil.

In order to avoid stagnation during iterations, and to ensure that convergence provides the information of the true object, with low error, we generate an initial good approximation to the object field. To obtain an appropriate initial reconstruction in the LOADH setup, applied to the reconstruction of high BW objects, we measure the intensities of the fields generated by the RW and the object field, in addition to the hologram intensity. With this information we eliminate the zero order term in the hologram modulation. As a second step we generate the initial object field reconstruction using a modified FDSF procedure. In the case of the RLOADH approach we assume that the pupil (or support) at the object plane is known. Thus we can compute digitally the field that propagate to the CCD plane from the pupil alone. We combine this information with the intensity of the field propagated from the test object, to obtain digitally a modified hologram which essentially contains the object field and a noise field. When we back-propagate this modified hologram to the object plane, the object field appears well defined (focused) at this plane, while the noise term appears defocused. Thus we obtain, within the pupil area, the object field distorted with certain amount of noise.

The approximate object field reconstruction, obtained with the described procedures, is used as the input of an iterative error reduction algorithm, which is successful in reducing the reconstruction error. A key aspect of this algorithm is a gradual application of the amplitude constraints, at the object and CCD planes. We illustrate the proposed methods by means of numerical simulations. In particular we evaluate the convergence of the iterative procedure and the fidelity of the reconstructed object field.

2. Lensless off-axis digital holography with external reference wave

The LOADH optical setup, schematically depicted in Fig. 1, is a Mach-Zender type interferometer, which is convenient for the study of transparent objects. Lenses L1 and L2 are employed only to expand and collimate the laser beam. The interferometer itself is a lensless module. It is assumed that the object to be studied is limited by a small circular pupil whose transmittance is denoted p(x,y). The object, which appears at the lower branch of the interferometer, is illuminated by an expanded and collimated laser gaussian beam, of wavelength λ. If the width of this beam is large in comparison to the pupil size, it is considered as a plane wave inside this pupil.

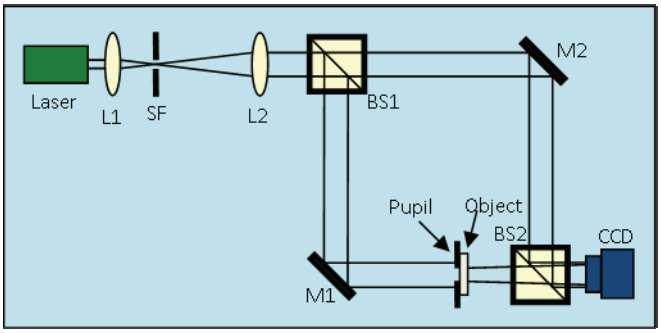


Fig. 1. LOADH optical setup. The components are: Laser, Lenses (L_1,L_2), spatial filter (SF), beam splitters (BS_1, BS_2), mirrors (M_1, M_2) and CCD.

For simplicity we assume that the object only modulates the phase of the illuminating beam. Thus, considering that the beam amplitude is normalized, the field transmitted at the object plane is

$$f(x,y) = p(x,y)\exp[i\phi(x,y)]. \tag{1}$$

This object field, freely propagated to the CCD plane, or hologram plane, is expressed as

$$g(x,y) = |g(x,y)|\exp[i\psi(x,y)]. \tag{2}$$

This field, obtained with the exact angular spectrum (AS) approach (Goodman, 1996), is mathematically represented as

$$g(x,y) = \Im^{-1}[F(u,v)H(u,v)]. \tag{3}$$

In Eq. (3), F(u,v) is the Fourier transform of f(x,y), $\Im^{-1}$ denotes the inverse Fourier transformation, and H(u,v) is the propagation factor

$$H(u,v) = \exp\left[ikz\sqrt{1-\lambda^2(u^2+v^2)}\right], \tag{4}$$

where z is the propagation distance, k=2π/λ, and $i = \sqrt{-1}$.
The propagated field computed with the AS approach is exact in the context of the scalar wave equation. Another advantage of the AS method is that when the operations of Eq. (3)

are performed with a fast Fourier transform (FFT) algorithm, the sampling resolution of functions f(x,y) and g(x,y) are identical. It must be noted that the propagation factor H(u,v) can be appropriately sampled more easily if the propagation distance z tends to be small. Thus, the AS approach is convenient to study high bandwidth objects (i. e. microscopic objects), for which the required distance from the CCD to the object plane is relatively short.
The reference beam, which propagates in the upper branch of the setup, is redirected to the CCD by the beam splitter BS_2. We assume that this beam is a plane wave that arrives with a slight tilt to the CCD. This RW is given by

$$r(x,y)=|r(x,y)|\exp[i2\pi(u_0x+v_0y)], \tag{5}$$

where the tilt is determined by the spatial frequencies (u_0,v_0). In this representation of the RW, the amplitude $|r(x,y)|$ can be spatially variable in the CCD area.
The superposition of the object field f(x,y) with the RW r(x,y) at the CCD plane, generates an intensity pattern, or hologram, which is given by

$$h(x,y)=|r(x,y)|^2+|g(x,y)|^2+g^*(x,y)\,r(x,y)+g(x,y)\,r^*(x,y), \tag{6}$$

where the asterisk denotes complex conjugation. The fourth term in h(x,y) contains the information of the object field, modulated by the plane wave r*(x,y). The task of DHM is to obtain the information of the object field, g(x,y), from the measured intensity h(x,y). To facilitate this purpose it is convenient to measure the intensities $I_r=|r(x,y)|^2$ and $I_g=|g(x,y)|^2$. Each one of these intensities is measured by blocking one of the branches of the interferometer. Considering these intensities together with h(x,y) we can compute the modified hologram

$$h_m(x,y)=\frac{h(x,y)-I_r-I_g}{I_r^{1/2}}. \tag{7}$$

The division by $(I_r)^{1/2}$ in Eq. (7) can be performed for the intensities measured by the CCD, making the reasonable assumption that $I_r\neq 0$ in the CCD area. Performing the operations in Eq. (7), the modulation of the modified hologram can be expressed as

$$h_m(x,y)=g(x,y)\exp[-i2\pi(u_0x+v_0y)]+g^*(x,y)\exp[i2\pi(u_0x+v_0y)]. \tag{8}$$

As noted in Eq. (8), the modulation of the modified hologram, $h_m(x,y)$, which is computed only from intensity measurements in the CCD, is formed by the object field g(x,y) and its conjugate pair, which are modulated by linear phase factors. The Fourier transform of this modified hologram, given by

$$H_m(u,v)=G(u-u_0,v-u_0)+G^*(-u-u_0,-v-u_0), \tag{9}$$

presents the function G, which is the Fourier spectrum of the object field, and its conjugated G*, placed at different positions in the Fourier domain of the hologram. It is remarkable that if these functions G and G* are not overlapped, it is possible to apply a spatial filtering that eliminates the conjugate term leaving alone the term $G(u-u_0, v-v_0)$. By centering this function it is obtained G(u,v), from which it is possible to recover the field, at the object plane, performing the inversion of Eq. (3), by means of the operation

$$f(x,y)=\Im^{-1}\left[G(u,v)H^{-1}(u,v)\right]. \tag{10}$$

The object field reconstruction, with the FDSF method described above, generates an accurate reconstructed field, only if the object bandwidth BW_{OB} is smaller than $2(u_0^2+v_0^2)^{1/2}$, otherwise the spectra bands of G and G* may appear overlapped. On the other hand, the frequency bandwidth of the CCD, given by $BW_{CCD}=p^{-1}$, being p the pixel pitch of the CCD, also restricts the object BW. For simplicity we assume that the pixel pitch is the same in the horizontal and the vertical axes of the CCD. The Fourier transform of the sampled version of the hologram $h_m(x,y)$, recorded by the CCD, is formed by multiple shifted replicas of the Fourier spectrum $H_m(u,v)$ that are repeated with periodicity BW_{CCD}. Considering the structure of $H_m(u,v)$ in Eq. (9), a convenient distribution of the multiple spectra terms G and G* in the Fourier domain of the sampled hologram is obtained adopting the carrier spatial frequencies $u_0=v_0=BW_{CCD}/4$. In this case the multiple spectra terms G and G* appear uniformly distributed in the hologram Fourier domain (Fig. 2), propitiating a reduced noise in the main object field spectrum term G, which is centered at frequency coordinates (u_0,v_0).

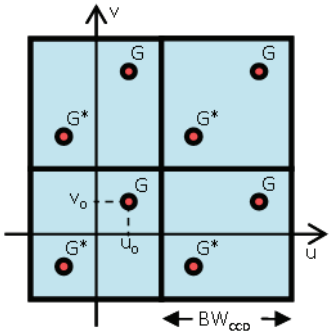


Fig. 2. Distribution of multiple spectra terms G and G* in the Fourier transform of the sampled version of the hologram $h_m(x,y)$, recorded by the CCD. This spatial distribution is uniform by using a RW with spatial frequencies $u_0=v_0=BW_{CCD}/4$.

Considering the spatial frequencies $u_0=v_0=BW_{CCD}/4$, the restriction previously established for the object BW, takes the form

$$BW_{OB}<BW_{CCD}/\sqrt{2}\ . \tag{11}$$

On the other hand, the spatial frequency bandwidth of the field must be limited so that the propagated field is captured by the CCD. This condition can be expressed as

$$BW_{OB} < 2\,\lambda^{-1}\,\sin(\theta_{max})\ , \tag{12}$$

where θ_{max} (shown in Fig. 3) is the angle subtended from the border of the CCD to the center of the object pupil.

The object reconstruction using the FDSF method will show small error if BW_{OB} is much smaller than the limits established in Eqs. (11) and (12). It is expected that the reconstruction error increases significantly when BW_{OB} is equal or greater than these limits. In the next subsections we perform numerical simulations corresponding to both cases.

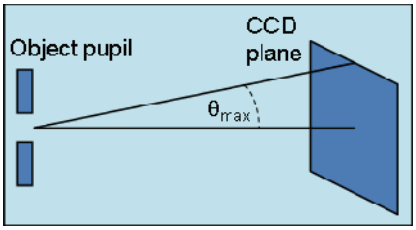


Fig. 3. The maximum angle θ_{max}, subtended from the border of the CCD to the center of the object pupil, also restricts the object bandwidth.

2.1 Reconstruction of a low bandwidth object

For the numerical simulations on DHM employing the LOADH setup we assume that the CCD has 1024x1024 pixels and pixel pitch p=5μm, which corresponds to a CCD band width BW_{CCD}=0.2 μm^{-1}. In addition, it is assumed that the distance from the CCD to the object plane is z=45 mm and that the wavelength is λ=500 nm.

We will consider first the computational reconstruction of an object with low frequency BW. The test object is an array of micro lenses, limited by a circular pupil of radius R=1mm. Each micro lens has a focal length f=30 mm and is contained in a square of width w_L=155 μm. The estimated BW of the array of lenses is obtained by the relation $BW_{OBJ}=(2)^{1/2}\ w_L\ (\lambda f)^{-1}$ (Arrizón et al., 2000). Considering the above specified parameters, one obtains $BW_{OBJ}\cong 0.015\mu m^{-1}$. On the other hand, for the determination of the RW defined in Eq. (5), we assume that the modulus is $|r(x,y)|=1$, and that the spatial frequencies are $u_0=v_0=BW_{CCD}/4$. For these parameters the object BW is much smaller than the limits given by the inequalities in Eqs. (11) and (12). For computational purposes, the object is sampled with a resolution of 5μm, which is equal to the CCD pixel pitch. In addition, the object matrix is centered in an extended (zero-padded) matrix of 1024x1024 pixels. The phase modulation within the object pupil is shown in Fig. 4 (a), and the amplitude (modulus) of the object field g(x,y), propagated to the CCD plane, is displayed in Fig. 4 (b). The amplitude values in the gray-levels bar of Fig. 4 (b) are normalized to the unitary amplitude of the beam that illuminates the test object. This type of normalization is also employed in subsequent sections.

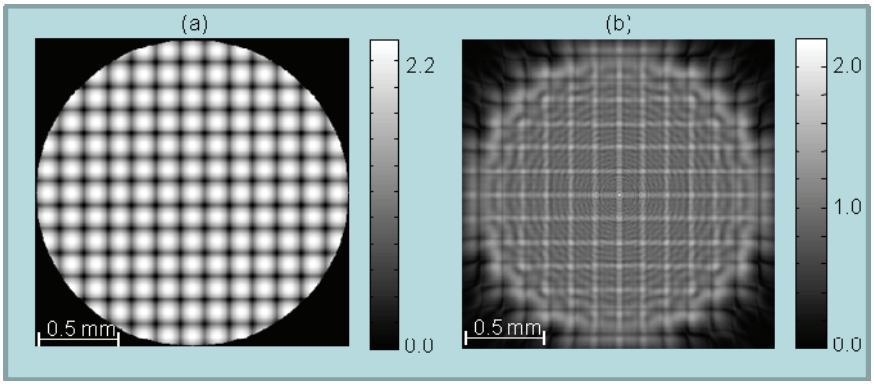


Fig. 4. (a) Phase modulation (radians) of the test object, formed by an array of micro-lenses, and (b) amplitude of the field propagated to the CCD plane.

Considering the RW, computed with Eq. (5), and the object field we obtain the hologram intensity h(x,y) and the modified hologram $h_m(x,y)$ [Eq. (8)]. In Fig. 5 we display the real valued modulation of $h_m(x,y)$ and the modulus of its Fourier spectrum $H_m(u,v)$. The object field term and its conjugate appear clearly separated in Fig. 5 (b). The object field spectrum $G(u-u_0,v-v_0)$ is placed at the left-top corner in Fig. 5 (b).

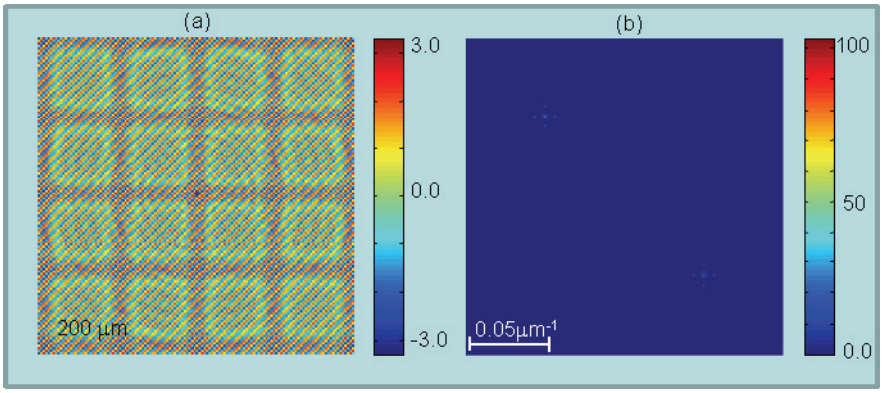


Fig. 5. (a) Close view of the real valued modulation of modified hologram $h_m(x,y)$, and (b) modulus of its Fourier spectrum $H_m(u,v)$.

The next steps are the digital elimination of the conjugate term $G^*(-u-u_0,-v-v_0)$, the centering of the field spectrum $G(u-u_0,v-v_0)$, and the inverse Fourier transformation of the centered spectrum. The result is the reconstructed object field, at the hologram plane, which we denote as $g_R(x,y)$. The reconstructed field at the object plane, denoted $f_R(x,y)$, is computed by the operations in Eq. (10), replacing $g_R(x,y)$ instead of g(x,y). The unwrapped phase of the recovered field $f_R(x,y)$ is displayed in Fig. 6(a). The reduced phase error $\arg\{f_R(x,y)\}-\arg\{f(x,y)\}$, shown in Fig 6(b), is an indicator of the high reconstruction fidelity. The discussed reconstruction process, based on the use of the FDSF technique, was presented to put in context the reconstruction of high BW objects that will be discussed in following sections.

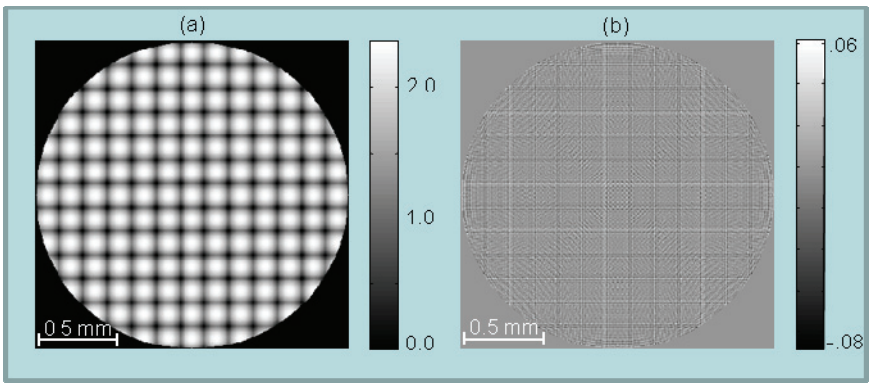


Fig. 6. (a) Unwrapped phase modulation (radians) of the reconstructed object $f_R(x,y)$, and (b) phase error $\arg\{f_R(x,y)\}-\arg\{f(x,y)\}$.

2.2 Reconstruction of high bandwidth objects

In the example of section 2.1 the FDSF is employed successfully to reconstruct a low BW object field using the LOADH setup. The use of an iterative error reduction for the reconstruction of a high BW object in the LOADH setup, is next studied. Let us consider a second test object formed by an array of micro-lenses, whose only difference with the array considered in subsection 2.1, is a new focal length f=3mm. In the LOADH setup we assume the CCD distance z=40 mm. The new estimated object BW is 0.15μm^{-1} which is 25% smaller than the CCD BW. The wrapped phase modulation of the array of lenses is shown in Fig. 7.

For the considered object, the modified hologram $h_m(x,y)$ has a wider Fourier spectrum $H_m(u,v)$, in comparison to the case discussed in section 2.1. The modulus of $H_m(u,v)$, displayed in Fig. 8 (a), shows the object field contribution (top-left corner) partially overlapped with the conjugate term (bottom-right corner). In addition, the object Fourier spectrum term overflows the limits of the spectrum window obtained with the FFT algorithm, making difficult the implementation of the spatial filtering. A solution to this problem is obtained by introducing an additional modification into the hologram $h_m(x,y)$. This change, obtained by multiplying the original modified hologram with the phase modulation of the RW, leads to the new modified hologram

$$h_m(x,y) = g(x,y) + g^*(x,y)\exp[i4\pi(u_0x + v_0y)]\,. \tag{13}$$

To obtain this new hologram modulation it is only required the knowledge of the reference wave. The spatial frequencies (u_0,v_0) of the RW can be determined, for example, by measuring the center of the object field spectrum term in the hologram Fourier domain (Cruz et al., 2009). The function $|H_m(u,v)|$, corresponding to the new hologram $h_m(x,y)$, is displayed in Fig. 8 (b), where the object spectrum term appears centred at the spectrum window. Of course, the overlapping with the conjugate term is still present. This conjugate term now has been splitted, and appears at the corners of the spectrum window. The diameter of the circle in Fig. 8 (b) is equal to the nominal object BW. For the implementation of the FDSF the Fourier spectrum field outside this circle is eliminated. This spatial filtering is a little deficient since the spectrum domain within the circle in Fig. 8 (b) includes some contributions of the conjugate spectrum term, and excludes high frequency information of the object field.

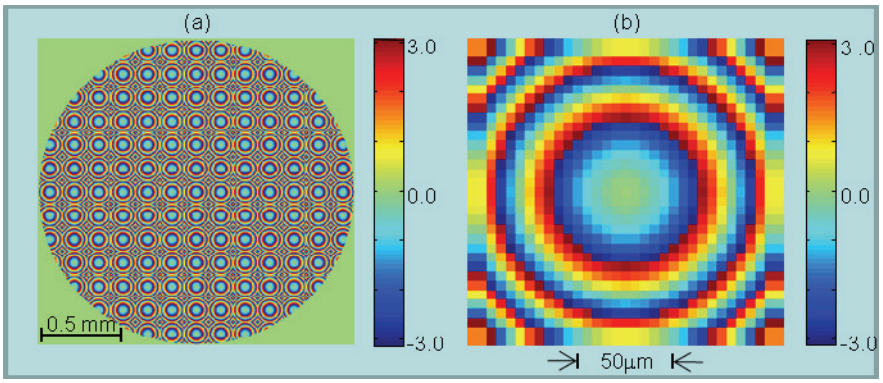


Fig. 7. (a) Wrapped phase modulation of the second test object formed by an array of micro lenses with focal length f=3mm. (b) Phase of a single lens in the array.

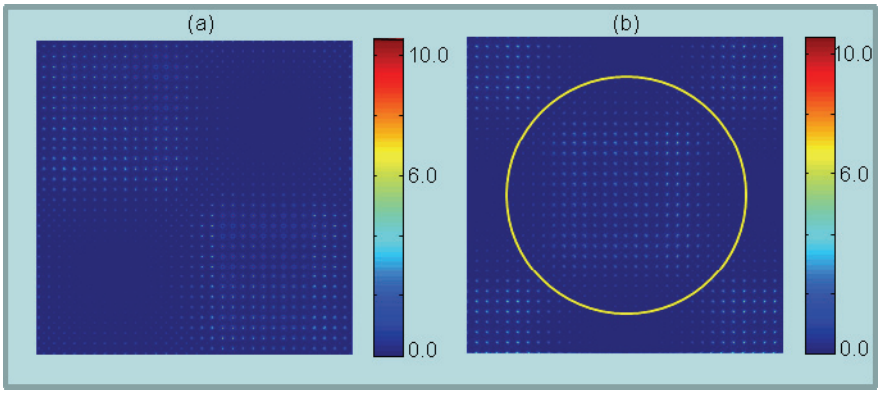


Fig. 8. Modulus of the hologram Fourier spectrum $H_m(u,v)$: (a) for the first modified hologram [Eq. (8)], and (b) for the second modified hologram [Eq. (13)].

To complete the FDSF process, we first obtain a filtered version, $h_f(x,y)$, of the second modified hologram $h_m(x,y)$, implementing the inverse Fourier transform of the spatially filtered hologram Fourier spectrum. The hologram $h_f(x,y)$ is formed essentially by the object term g(x,y), distorted with the part of the conjugate term which is not eliminated. We propagate the field $h_f(x,y)$ to the object plane, recovering an approximation of the object field. The wrapped phase of the recovered object field is shown in Fig. 9 (a), and a close view of the phase modulation of a single lens is displayed in Fig. 9 (b). In order to visualize more clearly the phase errors of this reconstruction, we display in Fig. 10 the central pixels of (a) the original object and (b) the reconstructed object, representing a reduced phase domain in the colour bar.

A quantitative evaluation of the reconstructed object field $f_R(x,y)$ can be made using the root mean square (RMS) error relation

$$\mathrm{Errf} = \sqrt{N_P^{-1} \sum_{\Omega_P} \frac{\left|f - f_R\right|^2}{\left|f\right|^2}}\,, \tag{14}$$

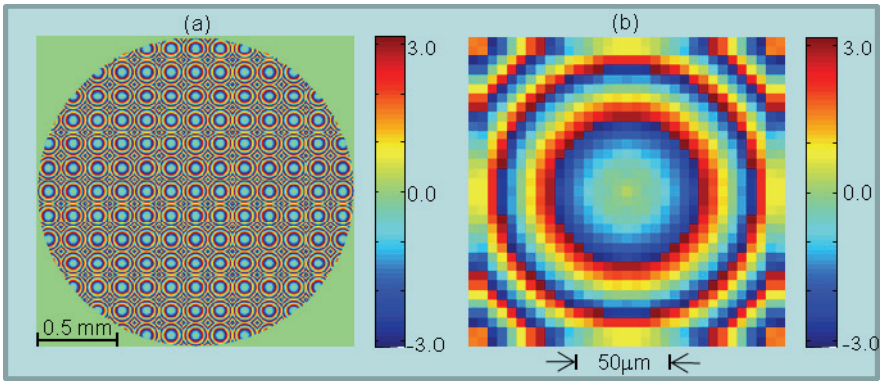


Fig. 9. (a) Wrapped phase modulation in the reconstructed object field, and (b) phase modulation at the central lens of the reconstructed object.

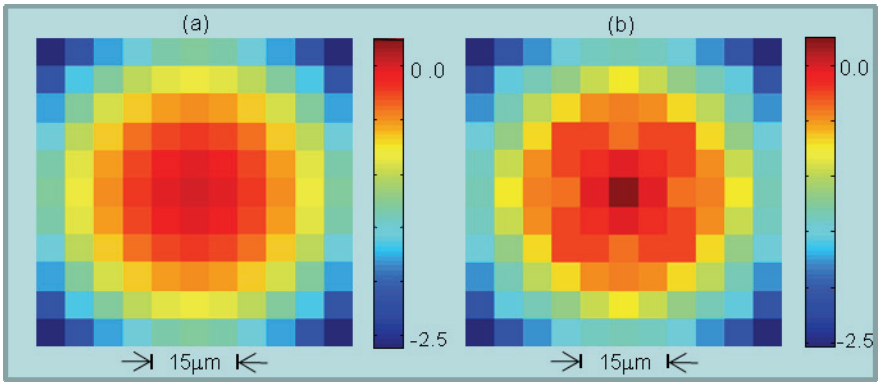


Fig. 10. Phase modulation at the central pixels of the original object (a), and at the corresponding pixels of the reconstructed object (b).

where f(x,y) is the true object field, N_P is the number of pixels in the object pupil Ω_P, and the summation is over the pixels in Ω_P. Eq. (14) provides a simultaneous evaluation of the error for both the modulus and the phase of the recovered field f_R. Although this error is dimensionless, Eq. (14) approximates a phase modulation error (in radians) if the modulus error $|f_R| - |f|$ is negligible. The error in the reconstruction of the second test object, computed with Eq. (14), is approximately 0.124.

Unfortunately, the error in Eq. (14) cannot be computed in an experimental reconstruction, for which the object complex field modulation f(x,y) is unknown. A convenient alternative to evaluate the reconstruction is to measure the RMS error of the field amplitude $|g_R|$ obtained at the CCD plane. This error is expressed as

$$\mathrm{Errg} = \sqrt{N_g^{-1} \sum_{\Omega g} \frac{\left[|g| - |g_R|\right]^2}{|g|^2}} \,. \tag{15}$$

where $|g| = (I_g)^{1/2}$, I_g is the measured intensity in the LOADH setup when the RW is blocked out, Ω_g is the set of pixels in the CCD for which $|g| \neq 0$, and N_g is the number of pixels in the set Ω_g. Since the involved functions in Eq. (15) can be measured by the CCD, the error Errg is useful as an indirect measure of the reconstruction quality in a experimental situation.

2.2.1 Iterative error reduction algorithm in object reconstruction

The error in the reconstructed object field, obtained with the FDSF procedure applied in section 2.2, is partially due to the contributions of the conjugate term, contained within the area of the applied spatial filter [see Fig. 8 (b)]. Another source of error is the loss of high spatial frequencies of the object field due to the low-pass filtering. Next we describe a simple iterative algorithm that substantially reduces the reconstruction error.

A key procedure of the algorithm, which is referred to as iterative error reduction algorithm (ITERA), is the application of amplitude constraints in both the object and CCD planes. In the object plane the constraint is given by the support or pupil function p(x,y). Assuming that the transmittance of the object under study has constant modulus (=1), the amplitude of

the reconstructed field f_q (at the stage q of the process) is replaced by $(1+M_1|f_q|)/(M_1+1)$, which represents a weighted average of 1 and $|f_q|$, determined by a positive parameter M_1.
In the CCD plane the amplitude constraint is the modulus of the field propagated from the object alone, i. e. $|g|=(I_g)^{1/2}$, where I_g is the measured intensity in the LOADH setup when the RW is blocked out. At the q-th stage of the iterative process, the amplitude of the field g_q propagated to the CCD plane is replaced by $(|g|+M_2|g_q|)/(M_2+1)$, which is a weighted average of $|g|$ and $|g_q|$, determined by the positive parameter M_2.
The algorithm steps are schematically represented in Fig. 11. The input f_0 is the approximated reconstructed field, obtained with the FDSF procedure described in section 2.2. The positive numbers M_1 and M_2 determine the speed of adjustment of the field amplitude to the constraint amplitude. For example, for $M_1=0$ amplitude a_{q-1} at the object plane is replaced by the constraint amplitude, i.e. $a_q=1$. On the other hand for M=1, the amplitude a_{q-1} is replaced by $a_q=(1+a_{q-1})/2$. It is noted that large values of M_1 and M_2 tend to reduce the convergence speed, but they can be necessary to avoid fast stagnation. In the implemented ITERA we employ $M_1=20\alpha$ and $M_2=5\alpha$, where α takes the initial value of 1. During iterations, the value of α is made proportional to the evaluated error Errg (right block in Fig. 11).
In the block at the right side of Fig. 11, the operator PR represents the propagation to the CCD plane, and in the lower block, the operator PR^{-1} represents the back-propagation to the object plane.

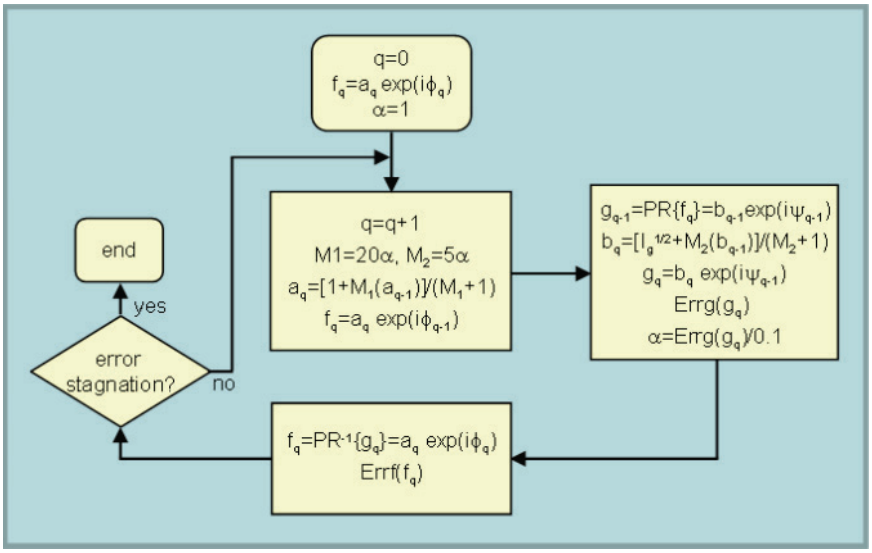


Fig. 11. Iterative error reduction algorithm (ITERA) for reducing the error in the first object field reconstruction.

We executed this algorithm using as input the reconstructed field obtained in section 2.2. The decrease of both errors Errf and Errg [Eqs. (14) and (15)], as a function of the iteration number, is shown in Fig. 12. The minimum errors obtained after 50 iterations are $Errf_{min}=0.001$ and $Errf_{min}=3.7\text{x}10^{-4}$, which represent a reduction of at least two magnitude orders in relation to the initial errors. The phase modulation at the central pixels of the

reconstructed field at the object plane, obtained for the numbers of iterations 8 and 50, are displayed in Fig. 13. The error Errf for these reconstructions corresponds to (a) the intermediate value 0.6 and (b) the minimum value 0.001. The quality of the final reconstruction can be visualized comparing Fig. 13 (b) with Fig. 10 (a).

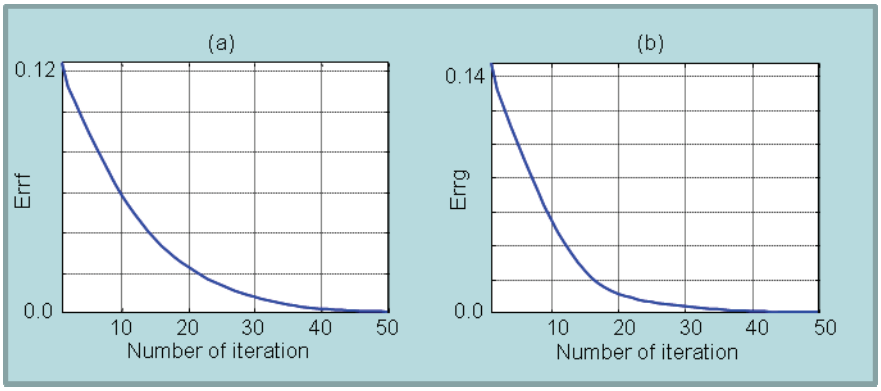


Fig. 12. Decrease of reconstruction errors (a) Errf and (b) Errg, as a function of the iteration number. The input is the first reconstructed object in section 2.2.

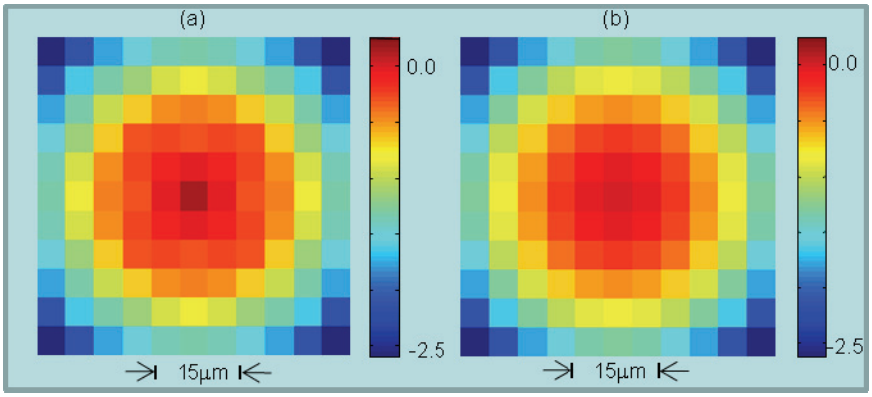


Fig. 13. Phase modulation at the central pixels of the reconstructed object field, obtained for the numbers of iterations (a) 8 and (b) 50.

2.2.2 Error reduction in an extremely noisy object reconstruction

To enhance the power of the ITERA algorithm, we consider a third test object whose BW is larger than the limit imposed by Eq. (11). This object is an array of micro lenses of focal distance f=1.5 mm, and a width W_L=125 µm. The pupil of the array of lenses has radius R=1 mm. The object bandwidth, obtained with the relation $BW_{OBJ}=(2)^{1/2}\ w_L\ (\lambda f)^{-1}$ is approximately 0.24, which exceeds the CCD BW in 20%. We employ the same setup and CCD parameters previously considered, except a new propagation distance z=31.25 mm. Fig. 14 shows the wrapped phase modulation (a) within the object pupil and (b) for the central lens of the array. We employ again a RW with spatial frequencies $u_0=v_0=BW_{CCD}/4$.

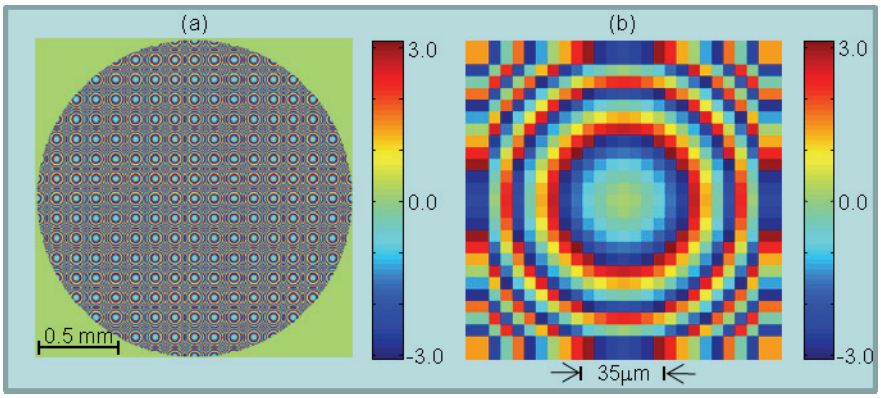


Fig. 14. (a) Wrapped phase modulation of the third test object formed by an array of micro lenses of focal length fL=1.5 mm. (b) Close view of the phase modulation in a single lens of the array.

After computing the propagated field, with Eq. (3), we obtained the second modified hologram $h_m(x,y)$ [Eq. (13)] and its Fourier spectrum $H_m(u,v)$. In the image of $|H_m(u,v)|$, displayed in Fig. (15), the object field term and the conjugate contribution appear highly overlapped. To make evident how these terms share the hologram spectrum space, we display the Fourier spectra for the separated terms $g(x,y)$ and $g^*(x,y)\exp[i4\pi(u_0x+v_0y)]$ that form $h_m(x,y)$. The modules of these Fourier spectra components are displayed in Fig. 16.
In spite of the severe overlapping of the object and conjugate terms in the Fourier spectrum of the hologram $h_m(x,y)$ [Fig. 15], we perform a first reconstruction using the FDSF method. The spatial filtering digitally eliminates the spectra information outside the circle shown in Fig. 15, whose diameter (in spatial frequency units) is DW_{CCD}=0.2 μm^{-1}. This spatial filtering is deficient by two reasons: it maintains a significant part of the conjugate contribution and it eliminates high frequency data of the object field. After performing the required steps of generating the spatially filtered hologram $h_m(x,y)$ and the propagation of this filtered

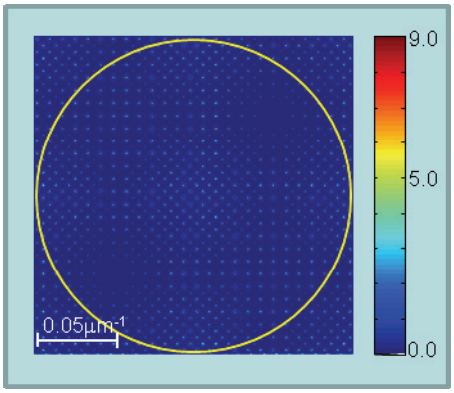


Fig. 15. Modulus of $H_m(u,v)$, for the second modified hologram [Eq. (13)], corresponding to the object field in Fig. 14. The spectrum contributions outside the yellow circle will be eliminated for the first reconstruction.

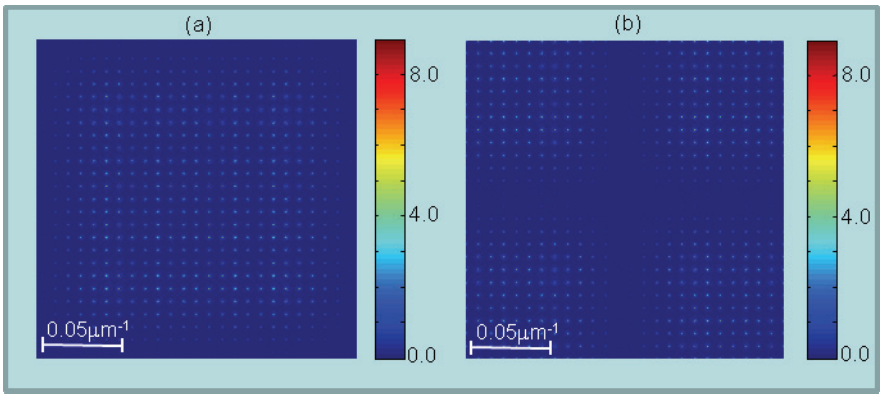


Fig. 16. Modules for the Fourier spectra of the separated terms (a) g(x,y) and (b) $g^*(x,y)\exp[i4\pi(u_0x+v_0y)]$, which form the modified hologram $h_m(x,y)$ of the test object.

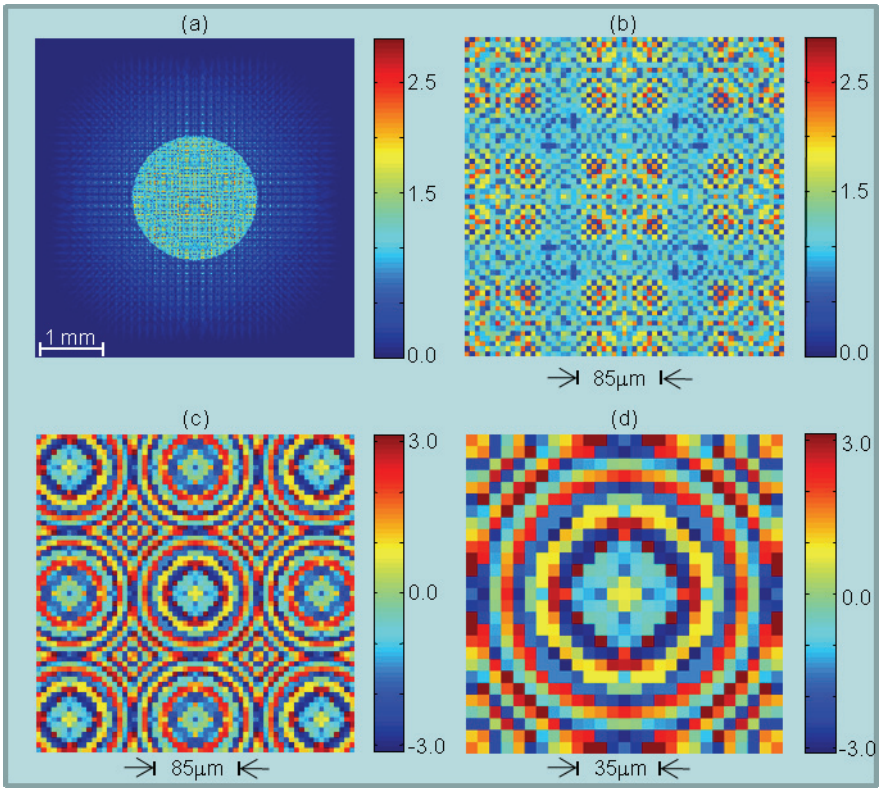


Fig. 17. First reconstruction of the third test object. Modulus of the field in (a) an extended and in (b) a close view. (c) Phase modulation of the central 9 micro lenses and (d) close view of the central lens phase modulation.

hologram to the object plane, we obtain the first reconstruction field. Fig. 17 (a,b) displays the amplitude of this reconstructed field, (c) the reconstructed phase modulation for the 9 central lenses of the array, and (d) a close view of the phase modulation in the central reconstructed lens. The object pupil, which is clearly seen in Fig. 17 (a), is polluted with a significant amount of light from the conjugate field, which fortunately appears out of focus. The detailed effect of the conjugate is observed in the close views of the amplitude (b) and phase modulation (c,d). Due to the presence of the conjugate field, the amplitude in Fig. 17 (a,b) takes values in the interval [0,2.9], which represents a large variation around the true object field amplitude A=1, within the pupil. On the other hand, the presence of the conjugate field also introduces a significant distortion in the phase of the reconstructed micro lenses (c,d). This distortion is evident when we compare the reconstructed lenses with the true lens phase modulation, shown in Fig. 14 (b).

We executed the ITERA algorithm with the same parameters employed in section 2.2.2, using as input the initial reconstructed field, described in the previous paragraph. The reduction of both errors Errf and Errg, as a function of the iteration number, is evidenced in Fig. 18. The minimum errors, obtained with 300 iterations, are $Errf_{min}$=0.01 and $Errf_{min}$= $6x10^{-4}$. Fig. 19 displays the reconstruction field amplitude and the phase modulation at the central reconstructed lens, at the end of the execution of ITERA. The negligible amplitude error within the reconstructed object pupil is consistent with the computed error $Errf_{min}$=0.01.

The low reconstruction error obtained in the present example makes evident that most of the noise introduced by the conjugated field is successfully eliminated by the application of ITERA. Now we wonder if the other error source, in the first reconstruction, namely the low-pass filtering of the object field, has been corrected or not. In other words we want to know if ITERA allows the recovery of the high frequency object features lost in the spatial filtering. Our answer to this significant question is yes. Fig. 20 (a) shows the object Fourier spectrum modulus $|F(u,v)|$, subject to the same low-pass filtering applied to the modified hologram (illustrated in Fig. 15). On the other hand, Fig. 20 (b) shows the modulus of the Fourier spectrum, $F_R(u,v)$, of the recovered field. It is remarkable that high frequency information which was eliminated in the object spectrum (Fig. 20 (a)) reappeared in the

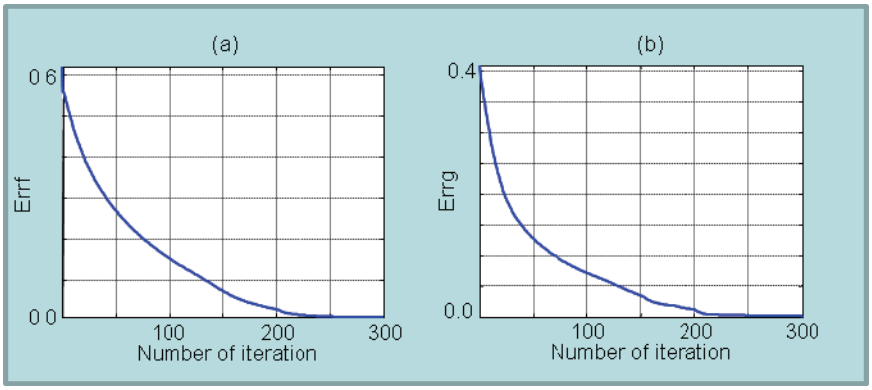


Fig. 18. Reduction of errors (a) Errf and (b) Errg, as a function of the iteration number, when the input is the first reconstructed object field, presented in Fig. 17.

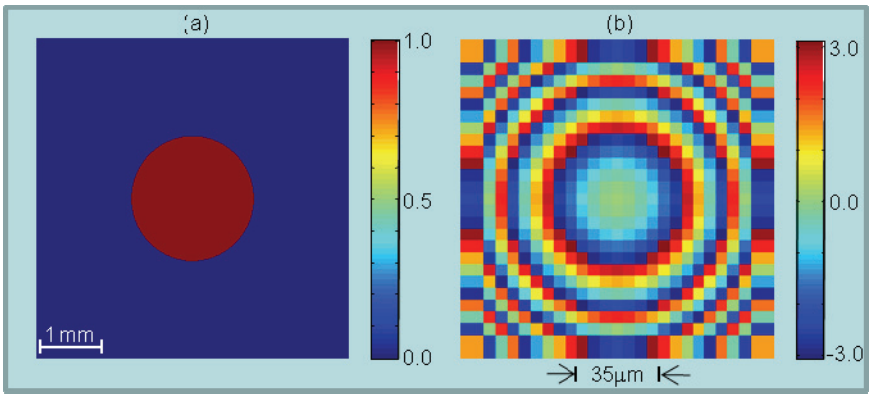


Fig. 19. (a) Modulus of the reconstructed object field, and (b) phase modulation at the central reconstructed micro lens, at the iteration 300 of the reducing error algorithm.

Fourier spectrum of the recovered object field (Fig. 20 (b)). This can be observed comparing the corners of both images in Fig. 20. We have verified the high fidelity of the recovered spatial frequency information $F_R(u,v)$ by a quantitative comparison with the Fourier spectrum F(u,v) of the test object.

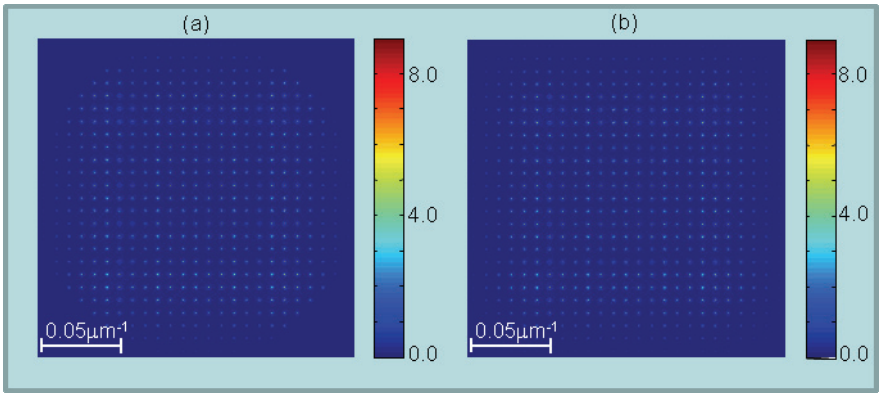


Fig. 20. Square root of the modules of (a) the spatially filtered object spectrum F(u,v), and (b) the Fourier spectrum $F_R(u,v)$, of the reconstructed object field $f_R(x,y)$.

3. Lensless on-axis digital holography without external reference wave

The error reduction algorithm (ITERA) described in section 2.2.1, has been successfully tested in the error reduction of large BW test objects, initially reconstructed by means of a FDSF technique. In the discussed cases it is assumed that the intensity information used for the reconstruction process is obtained in a LOADH optical setup. In this section we test the performance of ITERA applied to object field reconstruction in a RLOADH setup, schematically depicted in Fig. 21. In this setup, it is assumed that the object plane is illuminated by a plane wave and that the field scattered by the object freely propagates to the CCD plane.

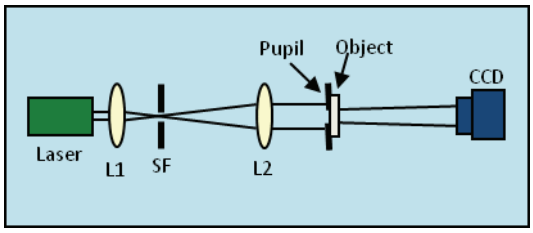


Fig. 21. RLOADH optical setup to implement DHM without an external reference wave.

We assume again that the object is limited by a pupil of binary transmittance p(x,y), and that the field transmitted at the object plane is given by Eq. (1). A common procedure used in DH, without an external RW, is to represent the object field by the expression

$$f(x,y) = p(x,y) + d(x,y) \tag{16}$$

where

$$d(x,y) = p(x,y)\{\exp[i\phi(x,y)] - 1\}\,. \tag{17}$$

The field propagated to the CCD plane, computed by means of the AS approach [Eq. (3)], can be expressed as

$$g(x,y) = g_p(x,y) + g_d(x,y)\,, \tag{18}$$

where $g_p(x,y)$ and $g_d(x,y)$ are the fields contributions from the object terms p(x,y) and d(x,y) respectively. Assuming a complete knowledge of the pupil function p(x,y), we can obtain, at least using numerical computation, the complex field $g_p(x,y)$. The intensity of the field g(x,y), or in-line hologram, can be measured by the CCD in the RLOADH setup. This intensity can be expressed as

$$h(x,y) = g_p g_p^* + g_d g_d^* + g_p g_d^* + g_p^* g_d\,. \tag{19}$$

Since we assume a complete knowledge of $g_p(x,y)$, we can obtain the modified hologram modulation

$$h_m(x,y) = \frac{h(x,y) - g_p g_p^*}{g_p^*} = g_d(x,y) + n(x,y)\,, \tag{20}$$

where

$$n(x,y) = \frac{g_d g_d^*}{g_p^*} + \frac{g_p g_d^*}{g_p^*}\,. \tag{21}$$

The division by $g_p^*(x,y)$ in Eqs. (20) and (21) is possible only in the region of the hologram where $g_p^*(x,y) \neq 0$. This is not an important restriction when the non-zero scattered field $g_d(x,y)$ occupies an area for which $g_p^*(x,y) \neq 0$. It is remarkable that the modified hologram

$h_m(x,y)$ is formed by the complex diffracted field $g_d(x,y)$ and the function $n(x,y)$ that represents a noise term.

When the hologram modulation $h_m(x,y)$ is back-propagated to the object plane, the term $g_d(x,y)$ generates the function $d(x,y)$, in a well defined region within the pupil. On the other hand, the structure of $n(x,y)$ [Eq. (21)] makes unlikely that this term will produce well defined or focussed patterns within the pupil region. Thus we expect that the hologram back-propagation will allow the reconstruction of a good approximation to the field $d(x,y)$. With this approximated version of $d(x,y)$ we can obtain a first reconstruction of $f(x,y)$ by means of Eq. (16). In the next section, we will discuss the reconstruction of a specific object, obtained from the modified in-line hologram given by Eq. (20). The error in this approximated reconstruction will be reduced by using the program ITERA, presented in section 2.2.1.

3.1 Reconstruction and error reduction in the RLOADH setup

To illustrate the reconstruction process using the modified in-line hologram presented in Eq. (20), we consider that the test object to be processed has the phase modulation of Fig. 22 (a). In this figure the large circle of radius R=1 mm corresponds to the object pupil $p(x,y)$. The object field is zero outside this pupil. The phase modulation within the object pupil is zero except at the central circular region (whose radius is 200 μm.) that contains the phase modulation of a small array of micro lenses. A close view of this array of lenses is displayed in Fig. 22 (b). These micro lenses are identical to the ones employed in the example of section 2.2.2. The modulus of the field propagated to the CCD plane, when the object pupil is illuminated with a plane wave (of unitary amplitude), is displayed in Fig. 23 (a). In this case, the distance from the CCD to the object plane is z=20 mm. This distance was chosen to allow that most of the diffracted field $g_d(x,y)$ is covered by the field $g_P(x,y)$.

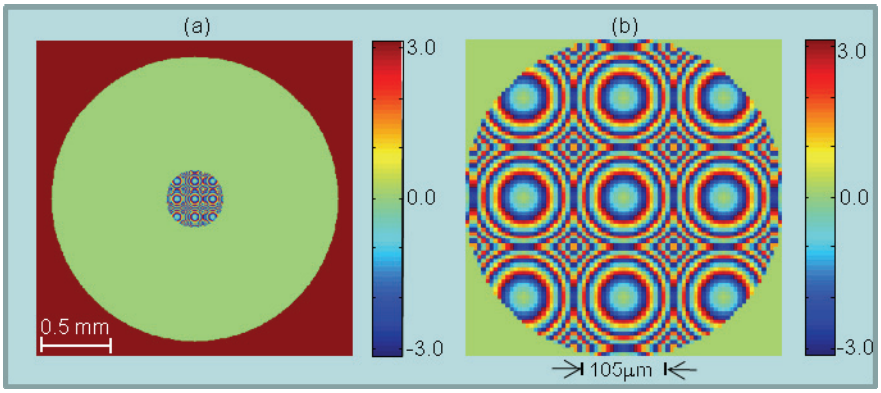


Fig. 22. (a) Phase modulation within the pupil of the fourth test object. (b) Close view of the small array of micro lenses at the center of the pupil.

For the generation of the modified in-line hologram [Eq. (20)] we only considered the region, $\Re_h$, where the amplitude of $g_P(x,y)$ is larger than 5% of its peak value. For the first reconstruction we omit the information outside this region. This is a practical method to avoid difficulties of division by g_P^* in Eq. (20). The modulus of the modified in-line hologram, displayed in Fig. 23 (b), shows the region $\Re_h$ (a circular domain) where this

hologram has been defined. A drawback of this method is that the information of g_d outside the region $\Re_h$, does not appear in the hologram. This missed information in function g_d corresponds to high spatial frequencies of the function d(x,y), in the object plane. An interesting feature of the reducing error algorithm ITERA, to be applied below, is that allow the recovery of the high frequency information of d(x,y), missed in the definition of the in-line hologram. Fig. 24 (a) shows the modules of functions g_p and g_d, respectively. It is noted (Fig. 24 (b)) that g_d contains information outside the region $\Re_h$ where the in-line hologram [Fig. 23 (b)] is defined.

After computing the modified in-line hologram $h_m(x,y)$ [Eq. (20)] we obtained the first reconstruction of function d(x,y). This is obtained by performing the back-propagation of the hologram to the object plane. The first reconstruction of the object field f(x,y), is obtained by adding the known function p(x,y) to the reconstructed function d(x,y). The amplitude and the phase of this reconstructed field, within the pupil domain, are shown in Fig. 25. The influence of noise is evident in the high amplitude variation shown in Fig. 25 (a). The

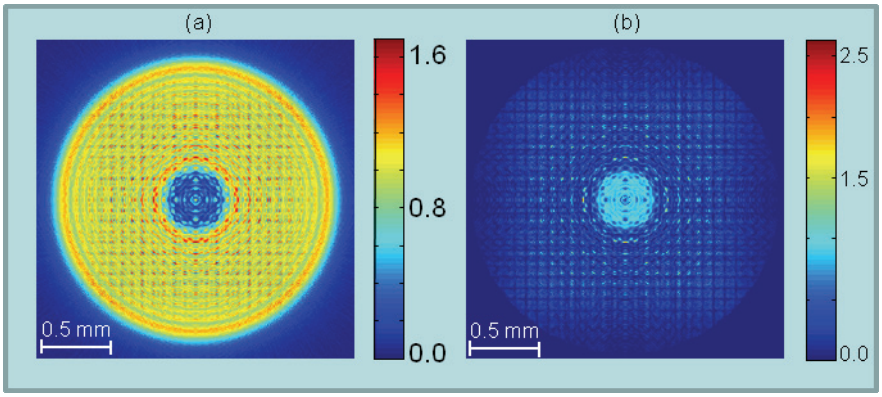


Fig. 23. (a) Amplitude of the object field, g(x,y), propagated to the CCD plane, and (b) modulus of the modified in-line hologram $h_m(x,y)$.

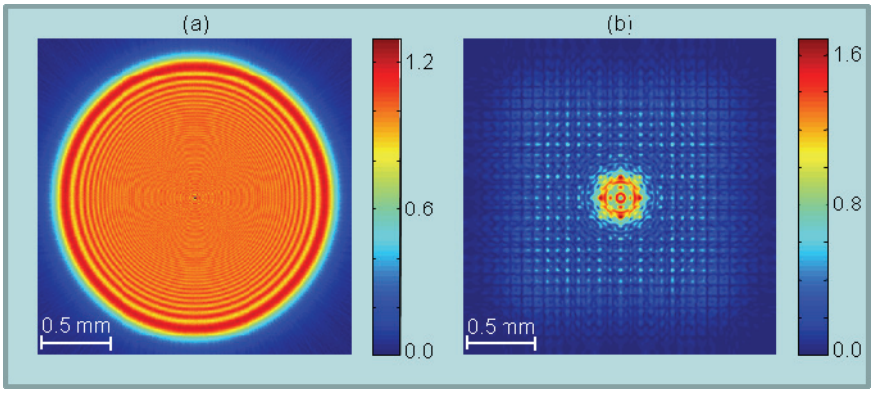


Fig. 24. Modules of the fields (a) $g_p(x,y)$ and (b) $g_d(x,y)$ that form the propagated object field g(x,y).

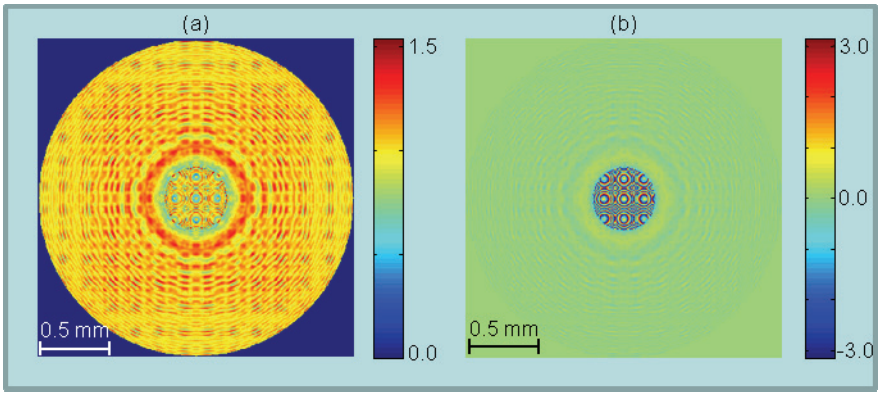


Fig. 25. Amplitude (a) and (b) phase of the first reconstruction of the object field (displayed in Fig. 22), obtained from the modified in-line hologram.

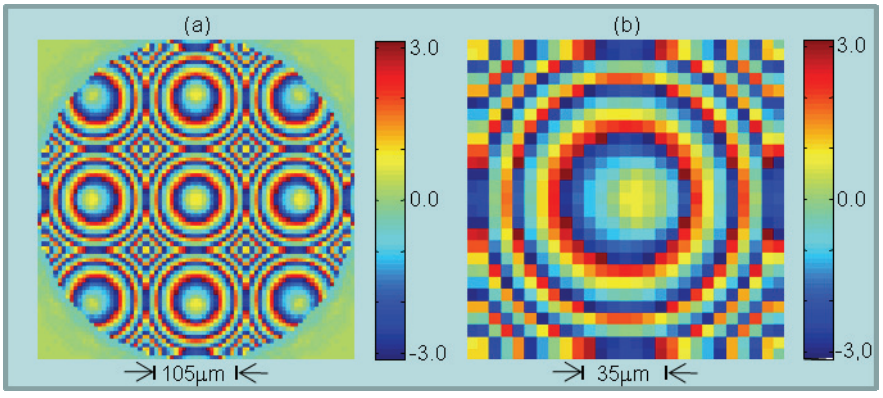


Fig. 26. (a) Phase modulation in the reconstructed array of lenses, and (b) close view of the phase modulation in the lens placed at the right side of the array, along the horizontal axis.

expected modulus within the pupil is A=1. A close view of the phase modulation in the array of lenses and is shown in Fig. 26 (a). The detailed image of a single lens, displayed in Fig. 26 (b), shows a clear asymmetry in the phase modulation.

We executed the ITERA program, with the parameters and constraints specified in section 2.2.1, using as input the first reconstructed object field mentioned in the previous paragraph. The reduction of errors Errf and Errg, as a function of the iteration number is displayed in Fig. 27. The error values (obtained at the iteration 185) are $Errf_{min}$=0.01 and $Errg_{min}$=0.0014 . We also computed the RMS error in the phase modulation obtaining 0.01 radians (at the iteration 185). As expected, this value coincides with the minimum value of Errf. The amplitude of the corrected object field, obtained at the iteration 185 of ITERA, is displayed in Fig. 28 (a). The corrected phase modulation of the asymmetrical lens in Fig. 27 (b) is displayed in Fig. 28 (b).

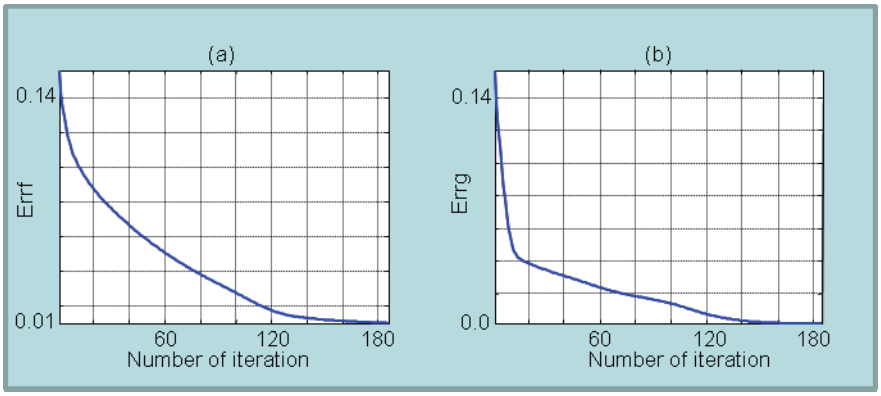


Fig. 27. (a) Reduction of errors (a) Errf and (b) Errg, as a function of the iteration number, when the input is the first reconstructed object field, presented in Fig. 25.

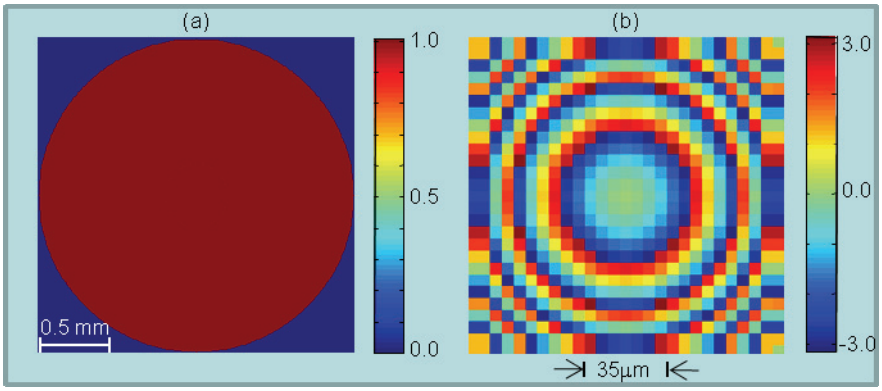


Fig. 28. (a) Amplitude of the reconstructed object field after the iteration 185, and (b) close view of the corrected version of the asymmetrical lens displayed in Fig. 26 (b).

The algorithm ITERA, applied to the object field reconstructions performed in the present and previous sections, have employed the adjusting parameters $M_1 = 20\alpha$ and $M_2 = 5\alpha$ (see Fig. 11). We note that the convergence speed of ITERA can be optimized by different definitions of these parameters. For example, in the application of ITERA to the reconstruction in this section, the error reduction obtained in Fig. 27, is also obtained with only 105 iterations, employing the parameters $M_1 = 6\alpha$ and $M_2 = 2\alpha$. The evolution of errors Errf and Errg, with these new parameters M_1 and M_2, is shown in Fig. 29.

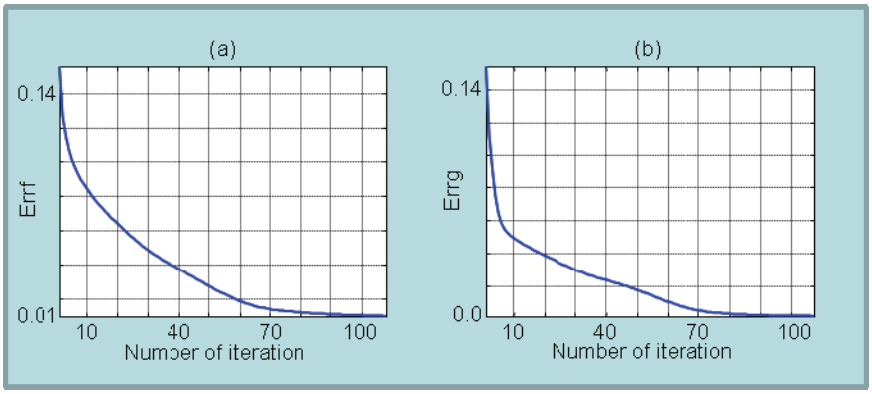


Fig. 29. (a) Reduction of errors (a) Errf and (b) Errg, as a function of the iteration number, using modified adjustment parameters M_1 and M_2.

4. Conclusion

We have implemented and tested an algorithm (ITERA) for reducing the error in the reconstruction of large BW object fields performed with intensity data from the LOADH and RLOADH optical setups. The initial step is the generation of a first reconstruction which is a good approximation of the true object field.

In the case of the LOADH setup, the first reconstruction is obtained by a FDSF method, based on the use of a modified hologram that contains the object field contribution on-axis, and the conjugate of this field off-axis. The contribution of the conjugate term in the hologram is reduced by spatial filtering in the hologram Fourier domain. On the other hand, when the modified filtered hologram is back-propagated to the object plane, the conjugated term appears defocused in this plane.

For the first reconstruction in the RLOADH setup we proposed a modified in-line hologram that essentially contains the object field information (in the CCD plane) and a noise term. An advantageous feature of the modified in-line hologram is that the term corresponding to the object field appears well focussed during propagation to the object plane, while the noise term appears defocused. This feature propitiates a dilution of the noise in the first reconstruction, within the object pupil.

The ITERA program allows a significant error reduction in both initial reconstructions. The good performance of ITERA is due in part to the relatively low noise of initial reconstructions. Another factor that also propitiates a fast convergence without stagnation is the gradual applications of constraints. Specifically, during a given iteration the obtained amplitude field a(x,y) (either at the object plane or the CCD plane) is replaced by a weighted average of the amplitude constraint with the field a(x,y).

A remarkable feature of ITERA is that it allows the recovery of high frequency information of the object field, which is lost in the first reconstruction. In the considered examples, it was possible to make negligible the reconstruction error at the end of application of ITERA, with a moderate number of iterations. The convergence speed of the algorithm can be optimized by the appropriate choice of the adjustment parameters M_1 and M_2.

5. References

Arrizón V. & Sánchez-de-la-Llave D., (2004). Common-path interferometry with one-dimensional periodic filters. *Opt. Lett.,* Vol. 29, No. 2, (Jan. 2004) 141-143, ISSN 0146-9592, 1539-4794.

Arrizón V., Testorf M., Sinzinger S., & Jahns J., (2000). Iterative optimization of phase-only diffractive optical elements based on a lenslet array. *J. Opt. Soc. Am. A,* Vol. 17, No. 12, (Dec. 2000) 2157-2164, ISSN 1084-7529, 1520-8532.

Car D.l, Kemper B., Wernicke G., & von Bally G. (2004). Parameter-Optimized Digital Holographic Microscope for High-Resolution Living-Cell Analysis. *Appl. Opt.,* Vol. 43, No. 36, (Dec. 2004) 6536-6544, ISSN 1559-128X , 2155-3165.

Cederquist J. N., Fienup J.R., Wackerman C. C., Robinson S. R., & Kryskowski D. (1989). Wave-front phase estimation from Fourier intensity measurements. *J. Opt. Soc. Am. A*, Vol. 6, No. 7, (Jul. 1989) 1020-1026, ISSN 1084-7529, 1520-8532.

Charrière F., Pavillon N., Colomb T., Depeursinge C., Heger T. J., Mitchell E. A. D., Marquet P., & Rappaz B. (2006). Living specimen tomography by digital holographic microscopy: morphometry of testate amoeba. *Opt. Express,* Vol. 14, No.16, (Aug. 2006) 7005-7013, ISSN 1094-4087.

Colomb T., Cuche E., Charrière F., Kühn J., Aspert N., Montfort F., Marquet P., & Depeursinge C. (2006). Automatic procedure for aberration compensation in digital holographic microscopy and applications to specimen shape compensation. *Appl. Opt.,* Vol. 45, No. 5, (Feb. 2006) 851-863, ISSN 1559-128X.

Cruz, M.L.; Castro, A. & Arrizón, V. (2008). Phase retrieval in digital holographic microscopy using a Gerchberg-Saxton algorithm, *Proceedings of Optics and Photonics for Information Processing II,* pp. 70721C, ISBN 0277-786X, USA, Sept. 2008, SPIE, San Diego, CA..

Cruz M.L., Castro A., & Arrizón A. (2009) Phase shifting digital holography implemented with a twisted-nematic liquid-crystal display. *Appl. Opt.,* Vol. 48, No. 36, (Dec. 2009) 6907-6912, ISSN 1559-128X.

Denis L., Fournier C., Fournel T., Ducottet C. (2005). Twin-image noise reduction by phase retrieval in in-line digital holography, *Proceedings of Wavelets XI* , Vol. 5914, 59140J, ISBN 9780819459190, USA, (Jul., 2005), SPIE , San Diego, CA..

Ferraro P., De Nicola S., Finizio A., Coppola G., Grilli S., Magro C., & Pierattini G. (2003). Compensation of the Inherent Wave Front Curvature in Digital Holographic Coherent Microscopy for Quantitative Phase-Contrast Imaging. *Appl. Opt.,* 42, Vol. 42, No. 11, (Apr. 2003) 1938-1946 , ISSN 1559-128X.

Ferraro P., Grilli S., Alfieri D., De Nicola S., Finizio A., Pierattini G., Javidi B., Coppola G., & Striano V. (2005). Extended focused image in microscopy by digital Holography. *Opt. Express* , Vol. 13, No. 18, (Sep., 2005) 6738-6749, ISSN 1094-4087.

Fienup J. R., (1987). Reconstruction of a complex-valued object from the modulus of its Fourier transform using a support constraint. *J. Opt. Soc. Am. A,* Vol. 4, No. 1, (Jan., 1987) 118-123, ISSN 1084-7529.

Garcia-Sucerquia J., Xu W., Jericho K. Klages P., Jericho M. H., & Kreuzer J. (2006). Digital in-line holographic microscopy. *Appl. Opt.*, Vol. 45, No. 5, (Feb. 2006) 836-850, ISSN 1559-128X.

Goodman J. W., (1996), *Introduction to Fourier Optics*, Roberts McGraw Hill Higher Education, 2nd edition, ISBN-10 0071142576, USA.

Guo H., Yu Y. & Chen M. (2007). Blind phase shift estimation in phase-shifting interferometry. *J. Opt. Soc. Am. A,* Vol. 24, No. 1, (Jan. 2007) 25-33, ISSN 1084-7529.

Haddad, W. S., Cullen, D., Solem, J. C, Longworth, J. W., McPherson, A., Boyer, K., & Rhodes, C. K. (1992). Fourier-Transform Holographic Microscope. *Appl. Opt.*, Vol. 31, No. 24, (Aug. 1992) 4973-4978, ISSN 1559-128X.

Hwang H., & Han P. (2007). Signal reconstruction algorithm based on a single intensity in the Fresnel domain. *Opt. Express,* Vol. 15, No. 7, (Apr. 2007) 3766-3776, ISSN 1094-4087.

Javidi B., Inkyu M, Yeom S., & Carapezza E. (2005). Three-dimensional imaging and recognition of microorganism using single-exposure on-line (SEOL) digital holography. *Opt. Express,* Vol. 13, No. 12. (Jun. , 2005) 4492-4506, ISSN 1094-4087.

Javidi B., Yeom S., Moon I., & Daneshpanah M. (2006). Real-time automated 3D sensing, detection, and recognition of dynamic biological micro-organic events. *Opt. Express,* Vol. 14, No. 9, (May 2006) 3806-3829, ISSN 1094-4087.

Kreis T. (1986). Digital holographic interference-phase measurement using the Fourier-transform method. *J. Opt. Soc. Am. A*, Vol. 3, No. 6, (Jun. 1986) 847-855, ISSN 1084-7529.

Liebling M., Blu T., & Unser M. (2004). Complex-wave retrieval from a single off-axis hologram. *J. Opt. Soc. Am. A,* Vol. 21, No. 3, (Mar. 2004) 367-377 , ISSN 1084-7529.

Mann C., Yu L., Lo C., & Kim M. (2005). High-resolution quantitative phase-contrast microscopy by digital holography. *Opt. Express,* Vol. 13, No. 22, (Oct. 2005) 8693-8698, ISSN 1094-4087.

Marquet P., Rappaz B., Magistretti P. J., Cuche E., Emery Y., Colomb T., & Depeursinge C. (2005). Digital holographic microscopy: a noninvasive contrast imaging technique allowing quantitative visualization of living cells with subwavelength axial accuracy. *Opt. Lett.,* Vol. 30, No. 5, (Mar, 2005) 468-470, ISSN 0146-9592.

Meneses-Fabian C., Rodriguez-Zurita G., & Arrizón V. (2006). Optical tomography of transparent objects with phase-shifting interferometry and stepwise-shifted Ronchi ruling. *J. Opt. Soc. Am. A,* Vol. 23, No. 2, (Feb. 2006) 298-305, ISSN 1084-7529, 1520-8532.

Morlens A.S., Gautier J., Rey G. , Zeitoun P. , Caumes J.P., Kos-Rosset M. , Merdji H. , Kazamias S. , Cassou K., & Fajardo M. (2006). Submicrometer digital in-line holographic microscopy at 32 nm with high-order harmonics. *Opt. Express,* Vol. 31, NO. 21, (Nov. 2006) 3095-3097, ISSN 1094-4087.

Nakamura T., Nitta K., & Matoba O. (2007). Iterative algorithm of phase determination in digital holography for real-time recording of real objects. *Appl. Opt.*, Vol. 46, No. 28, (Oct. 2007) 6849-6853, ISSN 1559-128X , 2155-3165.

Oh C., Isikman S. O., Khademhosseinieh B., & Ozcan A. (2010). On-chip differential interference contrast microscopy using lensless digital holography. *Opt. Express,* Vol. 18, No. 5, (Mar. 2010) 4717-4726, ISSN 1094-4087.

Pedrini G., Fröning P., Fessler H., & Tiziani H.J. (1998). In-Line Digital Holographic Interferometry. *Appl. Opt.,* Vol. 37, No. 26, (Sept. 1998) 6262-6269, ISSN 1559-128X, 2155-3165.

Pedrini G., Fröning P., Tiziani H.J., & Mendoza Santoyo F. (1999). Shape measurement of microscopic structures using digital holograms. *Optics Commun.*, Vol. 164, No. 4-6, (Jun. 1999) 257-268, ISSN 0030-4018.

Qian K., Fu Y., Liu Q., Seah H. S., & Asundi A. (2006). Generalized three-dimensional windowed Fourier transform for fringe analysis. *Opt. Lett.*, Vol. 31, No. 14, (Jul. 2006) 2121-2123, ISSN 0146-9592, 1539-4794.

Rappaz B., Marquet P., Cuche E., Emery Y., Depeursinge C., & Magistretti P. (2005). Measurement of the integral refractive index and dynamic cell morphometry of living cells with digital holographic microscopy. *Opt. Express*, Vol. 13, No. 23, (Nov. 2005) 9361-9373, ISSN 1094-4087.

Repetto L. , Piano E., & Pontiggia C., (2004). Lensless digital holographic microscope with light-emitting diode illumination. *Opt. Lett.*, Vol. 29, No. 10, (May. 2004) 1132-1134, ISSN 0146-9592, 1539-4794.

Schnars U., & Jüptner W. (2002). Digital recording and numerical reconstruction of holograms. *Meas. Sci. Technol.,* Vol. 13, No. 9, (Sept. 2002), R85, ISSN 0957-0233, 1361-6501.

Singh V. R. & Asundi A. (2009). In-line digital holography for dynamic metrology of MEMS. *Chin. Opt. Lett.,* Vol. 7, No. 12, (Dec. 2009) 1117-1122.

Takeda M., Ina H., & Kobayashi S. (1982). Fourier-transform method of fringe-pattern analysis for computer-based topography and interferometry. *J. Opt. Soc. Am.,* Vol. 72, No. 1, (Jan. 1982) 156-160, ISSN 1084-7529, 1520-8532.

Wagner C., Seebacher S., Osten W., & Jüptner W. (1999). Digital Recording and Numerical Reconstruction of Lensless Fourier Holograms in Optical Metrology. *Appl. Op*t. 38, No. 22, (Aug. 1999) 4812-4820, ISSN, 1559-128X , 2155-3165.

Wu J. S., Weierstall U., Spence J. C. H., & Koch C. T. (2004). Iterative phase retrieval without support. *Opt. Lett.,* Vol. 29, No. 23, (Dec., 2004) 2737-2739, ISSN 0146-9592, 1539-4794.

Xu W., Jericho M. H., Meinertzhagen I. A., & Kreuzer H. J. Digital In-Line Holography of Microspheres. *Appl. Opt.,* Vol. 41, No. 25, (Sept. 2002) 5367-5375, ISSN 1559-128X , 2155-3165.

Yamaguchi I. & Zhang T. (1997). Phase-shifting digital holography. *Opt. Lett.,* Vol. 22, No. 16, (Aug. 1997) 1268-1270, ISSN 0146-9592, 1539-4794.

Yamaguchi I., Kato J., Ohta S., & Mizuno J. (2001). Image formation in phase-shifting digital holography and applications to microscopy. *Appl. Opt.*, Vol. 40, No. 34, (Dec. 2001) 6177-6186, ISSN 1559-128X , 2155-3165.

Zhang T. & Yamaguchi I. (1998). Three-dimensional microscopy with phase-shifting digital holography. *Opt. Lett.*, Vol. 23, No. 15, (Aug, 1998) 1221-1223, ISSN 0146-9592, 1539-4794.

19

Image Quality Improvement of Digital Holography by Multiple Wavelengths or Multiple Holograms

Takanori Nomura
Wakayama University
Japan

1. Introduction

Digital holography (Yamaguchi & Zhang, 1997; Osten et al., 2002; Kreis, 2005; Ferraro et al., 2005; Frauel et al., 2006; Nomura & Imbe, 2010) is a useful technique for recording the fully complex field of a wavefront. In line with advances in imaging devices such as CCDs, digital holography is accessible. The digital holography has been used for lots of applications including encryption (Javidi & Nomura, 2000; Tajauerce & Javidi, 2000; Nomura et al., 2004) and three-dimensional object recognition (Poon & Kim, 1999; Tajauerce et al., 2001; Javidi & Kim, 2005). However, the quality of a reconstructed image of a digital hologram is not good, because it is suffering from speckles. Comparing a conventional film hologram, the extent of a digital hologram is small and the resolution of the hologram is low in single figure. This is why the reconstructed image has noticeable speckles. To reduce the speckles, some signal processing approaches have been proposed (Javidi et al., 2005; Do et al., 2005; Maycock et al., 2005; Kreis & Schlüter, 2007). By use of wavelet decomposition, three-dimensional image fusion has been proposed and some speckles reduction has been achieved. Independent component analysis has been also applied to reduce speckles. Aperture synthesis approach is also proposed to enhance the resolution. Somewhat speckle noises have been reduced by the approaches.
In this paper, two methods to reduce speckle of the reconstructed image are presented. One is the method based on the superposition of the reconstructed images with different wavelengths. Another is the method based on the multiple holograms. In Section 2, the characteristics of speckles are reviewed briefly. In Section 3, the former method is described. Some experimental results are shown to confirm the proposed method. In Section 4, the latter method is described with some experimental results for confirmation.

2. Speckle size in digital holography

Let $g(x_0, y_0)$ denote a complex field of an object wave at the object plane (x_0, y_0). At a distance of d from an object, the complex field is given by (Nomura et al., 2008)

$$u(X, Y) = \exp\left(i\frac{2\pi}{\lambda}\frac{X^2 + Y^2}{2d}\right)\mathcal{F}\left[g(x_0, y_0)\exp\left(i\frac{2\pi}{\lambda}\frac{x_0^2 + y_0^2}{2d}\right)\right], \tag{1}$$

where λ denotes a wavelength. The notation $\mathcal{F}$ denotes a Fourier transform operation. The digital hologram has a finite extent due to the extent of an imaging device. Let us assume that

the extent is $a \times b$. Therefore the digital hologram can be written as

$$u'(X,Y) = u(X,Y)\mathrm{rect}\left(\frac{X}{a}\right)\mathrm{rect}\left(\frac{Y}{b}\right). \tag{2}$$

The compelx field of the reconstructed object $U(x,y)$ can be calculated by

$$\begin{aligned} U(x,y) &= \exp\left(-i\frac{2\pi}{\lambda}\frac{x^2+y^2}{2d}\right)\mathcal{F}\left[u'(X,Y)\exp\left(-i\frac{2\pi}{\lambda}\frac{X^2+Y^2}{2d}\right)\right] \\ &= \exp\left(-i\frac{2\pi}{\lambda}\frac{x^2+y^2}{2d}\right)\left\{g(x_0,y_0)\exp\left(i\frac{2\pi}{\lambda}\frac{{x_0}^2+y_0^2}{2d}\right) * \mathrm{sinc}\left(\frac{ax_0}{\lambda d}\right)\mathrm{sinc}\left(\frac{by_0}{\lambda d}\right)\right\}, \end{aligned} \tag{3}$$

where the notation $*$ denotes a convolution operation. From Eq. (3), the reconstructed object is given by a convolution between the original complex field and a sinc function determined by the extent of the imaging device. If the speckle size is defined as a full width at half maximum of the main lobe of the sinc function, the size of the speckle ΔS_x and ΔS_y are approximately given by

$$\Delta S_x = \frac{\lambda d}{a}, \tag{4}$$

$$\Delta S_y = \frac{\lambda d}{b}, \tag{5}$$

respectively. These relations are also shown in the literature (Kreis, 2005). The speckle spoils the quality of the reconstructed image. To improve the quality, two approaches are considered. One approach is to make the effect weaken. Another approach is to reduce the size of the speckle using large $a \times b$ based on a synthetic aperture. In the following Sections, both approaches are described.

3. Speckles reduction by multiple wavelengths

3.1 Speckle model and reconstructed image with multiple wavelengths

If digital holograms with different wavelengths are recorded, the shape and the position of speckles are different from each other, because the complex field of the diffracted wave depends on the wavelength based on Eq. (1). Here it is assumed that the shape and the position of the speckles are randomly changed in relation to a wavelength. Namely the intensity of the reconstructed object $I'(x,y)$ is assumed to be given by

$$I'(x,y) = I(x,y) + s_i(x,y), \tag{6}$$

where $I(x,y)$ and $s_i(x,y)$ denote the intensity of the reconstructed object without the speckle and the speckle recorded by the wavelengh λ_i. By the summation of $I'(x,y)$ over different wavelengths,

$$\begin{aligned} \sum_{i=1}^{N} I'(x,y) &= \sum_{i=1}^{N} I(x,y) + \sum_{i=1}^{N} s_i(x,y), \\ NI'(x,y) &= NI(x,y) + S, \end{aligned} \tag{7}$$

is obtained, where S is considered as a constant value given by

$$\sum_{i=1}^{N} s_i(x,y) = S. \tag{8}$$

This equation is based on the above-mentioned assumption. If the constant value S is negligible smaller than $NI'(x,y)$, the relation written by

$$I'(x,y) = I(x,y), \tag{9}$$

is can be obtained. Namely, by superposing the multiple reconstructed image intensity with different wavelengths, the blight and dark spots, namely speckles, are smoothed so that the image quality is improved.

However, a difficulty to superpose the reconstructed images is come up. The pixel size Δx and Δy of the reconstructed image are given by

$$\Delta x = \frac{\lambda d}{N_X \Delta X}, \tag{10}$$

$$\Delta y = \frac{\lambda d}{N_Y \Delta Y}, \tag{11}$$

respectively, where, N_X and N_Y denote the number of pixels of the digital holograms, respectively, and ΔX and ΔY denote the pixels sizes of the imaging devices, respectively (Kreis, 2005; Javidi et al., 2005). The relations mean that the pixel size in the reconstructed image varies according to the wavelength. The pixel size in the reconstructed image should be equal to superpose. From here the case Δx is explained for simplicity. To equalize Δx, one of the parameters d, N_x, and ΔX have to be changed. It is unrealistic to change distance d according to the change of the wavelength, because it is not easy to implement. One approach is to change N_X for a different wavelength by padding zeroes (Javidi et al., 2005). This is a simple and smart method, because it needs neither experimental manipulation nor signal processing. However the Fast Fourier Transform (FFT) algorithm that the sampling number is equal to the Nth power of two can not be adopted.

Here, the method to change the pixel size (sampling interval) of the digital hologram by compensation is proposed. Let λ_c and λ_i $(i = 1, \ldots, N)$ denote a criterion wavelength and a wavelength used for recording. Using Eq. (6), the compensated sampling interval $\Delta x'$ can be written as

$$\Delta x' = \frac{\lambda_i}{\lambda_c} \Delta x, \tag{12}$$

where Δx denote the sampling interval before compensation. Base on the compensation, the pixel value (intensity) of the hologram should be changed. Linear interpolation between neighbor pixel values is adopted to determine the value of the compensated pixel. It is assumed here that the the change of the pixel value of the interferogram (digital hologram) is not sharp. Let $I(n)$ and $I'(m)$ denote the intensity of the digital hologram at the nth pixel and the intensity of the compensated digital hologram at the mth pixel, respectively. The schematic diagram of compensation of a digital hologram for a different wavelength is shown in Fig. 1. The compensated pixel value denoted by $I'(m+1)$ is given by

$$I'(m+1) = \frac{I(n+1) - I(n)}{\Delta x} \left\{ (m+1)\Delta x' - n\Delta x \right\} + I(n). \tag{13}$$

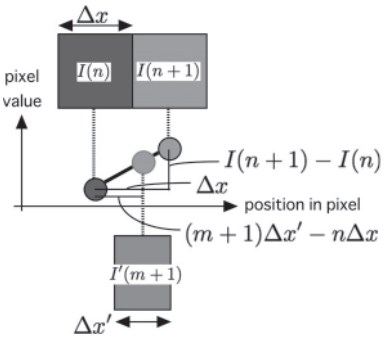


Fig. 1. Schematic diagram of compensation of the pixel value based on a linear interpolation.

Using the compensated digital holograms, reconstructed objects $U_i(x,y)$ $(i = 1,\ldots,N)$ are obtained. Finally the synthesized reconstructed object $|U_s(x,y)|^2$ is obtained by

$$|U_s(x,y)|^2 = \frac{1}{N}\sum_{i=1}^{N}|U_i(x,y)|^2 . \tag{14}$$

3.2 Optical experiments

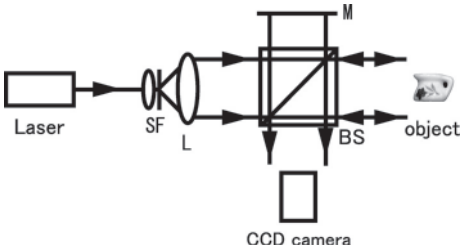


Fig. 2. Exemerimental setup: L, lens; SF, spatial filter; BS, beam splitter; M, mirror.

To confirm the proposed method, the optical experimental results are shown. The experimental setup is shown in Fig. 2. As a wavelength-tunable coherent light source, the dye (Rhodamine 6G) with the solid state diode-pumped, frequency-double Nd:Vanadate laser that provides 532 nm wavelength is used. For an object, a miniature pitcher with a height of $\sim$2 cm shown in Fig. 3 is used. A CCD camera with 1024 $\times$ 768 pixels and 8 bits gray levels is used. The pixel size of the CCD is 4.65 μm $\times$ 4.65 μm. The distance from the object to the CCD is 520 mm. An off-axis digital holography configuration is adopted.

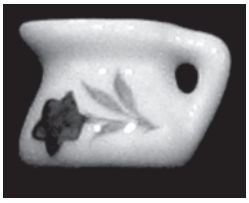
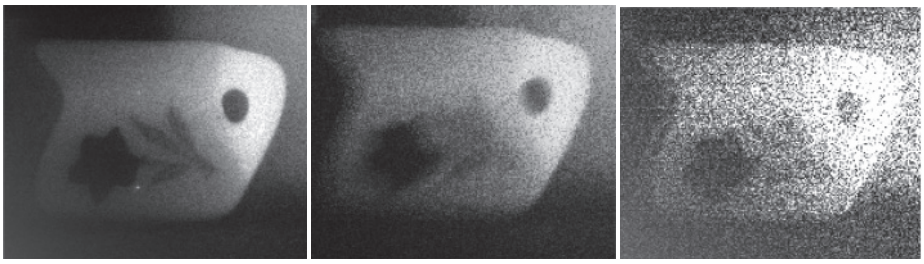


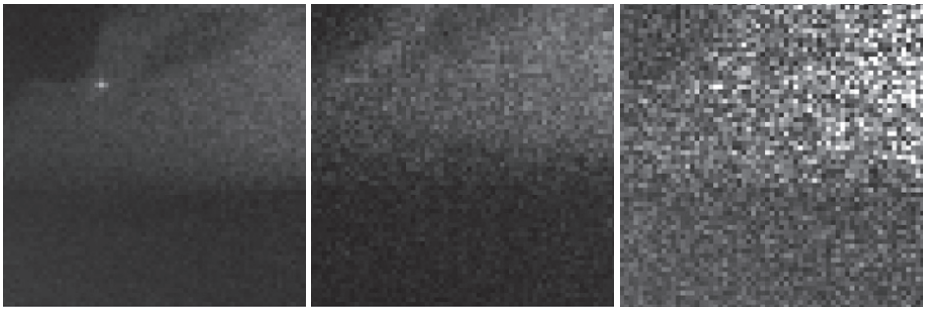
Fig. 3. A miniature pitcher as a three-dimensional object.

Thirty-eight digital holograms are recorded with the wavelength from 567 to 606 nm at an interval of ∼1 nm. The superposed image of the thirty-eight reconstructed images by the proposed method is shown in Fig. 4(a). The superposed image without sampling period compensation is shown in Fig. 4(b). The conventional reconstructed image by a single hologram (wavelength: 567 nm) is shown in Fig. 4(c). The brightness of the upper right corner in each image is due to the dc term of the off-axis digital holography. The reconstructed images with improved quality can be seen by the proposed method. Note that the image shown in Fig. 4(a) has the bright spot on the pitcher. In Fig. 4(b), the spot is weakened by disagreement of the resolution in the reconstructed plane. In Fig. 4(c), the spot is buried in speckles . To emphasize the difference, the magnified portions of the reconstructed images are shown in Fig. 5.



(a) The proposed superposition method. (b) The superposition without sampling period compensation. (c) A conventional single hologram.

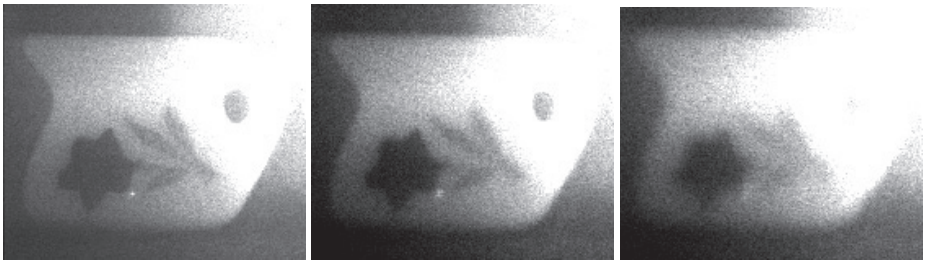
Fig. 4. The comparison of the reconstructed images.



(a) The proposed superposition method. (b) The superposition without sampling period compensation. (c) A conventional single hologram.

Fig. 5. The comparison of the magnified portions of the reconstructed images.

In Fig. 6, the reconstructed images by the proposed method at various distances from the CCD are also given. In Fig. 7, the images by the conventional method are also given. From the images, the proposed method keeps the characteristics of the digital holography that has the flexibility of reconstruction distance.

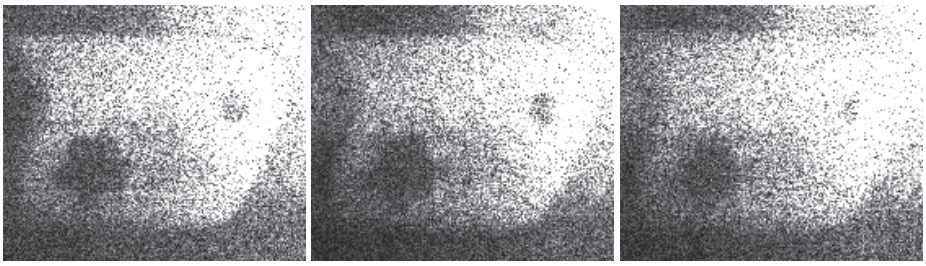


(a) A reconstructed image at a distance of 520 mm (in focus).

(b) A reconstructed image at a distance of 420 mm (out of focus).

(c) A reconstructed image at a distance of 320 mm (out of focus).

Fig. 6. Reconstructed images by the proposed superposition method.



(a) A reconstructed image at a distance of 520 mm (in focus).

(b) A reconstructed image at a distance of 420 mm (out of focus).

(c) A reconstructed image at a distance of 320 mm (out of focus).

Fig. 7. Reconstructed images from a single hologram (conventional).

In the approach, the reconstructed image loses its phase distribution because of intensity averaging. If the phase distribution is necessary in some cases such as instrumentation, this approach is not effective. In the cases the following synthetic aperture approach will wok. The method of speckle suppression should be chosen depending on the application.

4. Speckles reduction by multiple holograms

The quality of the reconstructed image is very poor, because it is suffering from the speckles due to the small extent of the hologram as mentioned in Section 2. To reduce the speckles, it is effective to use a hologram with a large extent (Binet et al., 2002). For example, if we record M holograms to be synthesized into a large extent hologram for each hologram, a good quality image will be reconstructed. In this case, the camera, which is used for digital hologram recording, has to be moved in plane for M times to obtain one large hologram. This is not practical. Here, the method to superpose digital holograms recorded from various viewing-zone angles on cyclic position.

4.1 Recording and synthesis of digital holograms

Consider the superposition of the holograms recorded from various viewing-zone angles on concyclic positions. Digital holograms of the object, which is placed on the rotational stage, from different angles are recorded by an image sensor concyclic positions as shown in

Fig. 8. In the recording process, the concerned hologram H_n and neighboring holograms H_{n-1} and H_{n+1} as shown in Fig. 9 are obtained. Figure 9 denotes two neighbors. If neighboring holograms are used to be synthesized, a large-extent digital hologram is obtained. However,

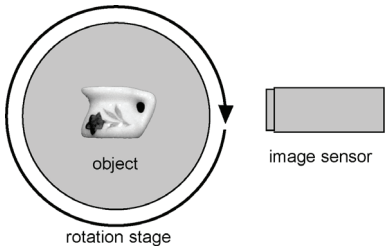


Fig. 8. Recording digital holograms by an image sensor on the concyclic positions.

simply synthesis between neighbor holograms is not available, because each hologram is in the different plane. To synthesize the holograms, holograms have to be brought on the same plane like H'_{n-1} and H'_{n+1} as shown in Fig. 9. To obtain the hologram on the same plane

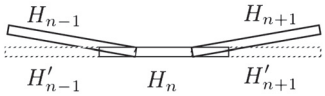


Fig. 9. A digital hologram and its neighbor holograms.

where the concerned hologram lays, the following procedure is introduced. It is assumed that the neighbor hologram $H_{n-1}(x', y')$ can be written as

$$\begin{aligned} H_{n-1}(x', y') &= |U_r(x', y') + U_o(x', y')|^2 \\ &= |U_r(x', y')|^2 + |U_o(x', y')|^2 \\ &\quad + U_o(x', y')U_r^*(x', y') + U_o^*(x', y')U_r(x', y'), \end{aligned} \tag{15}$$

where $U_o(x', y')$ and $U_r(x', y')$ denotes the complex amplitude distributions of the object wave and the reference wave on the hologram plane (x', y'). If the plane wave which incidents into the hologram at an angle is used, the third term on the right side $U_o(x', y')U_r^*(x', y')$ can be extracted using a spatial frequency filtering. After extracting the object wave information $U_o(x', y')U_r^*(x', y')$, the object wave information on the plane (x, y) is obtained numerically. For numerical calculation, it is assumed that wave propagates straightforward and only phase is changed according to the propagation distance as shown in Fig. 10. A hologram can be made

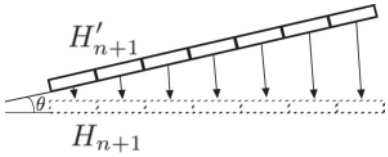


Fig. 10. Numerical propagation of the wavefront of the object wave.

using interference between the numerically obtained object wave on the plane (x, y) and the reference wave. After realignment of neighbor digital holograms, they can be synthesized by a correlational technique.

4.2 Experimental results

Experimental results are shown to confirm the feasibility of the proposed method. An optical experimental setup is shown in Fig. 11. A He-Ne laser with a wavelength of 632.8 nm is used as a coherent light source. A CCD camera is used to record the hologram. The number of pixels of the CCD camera is 1024×768, and its pixel size is $4.65\mu\text{m} \times 4.65\mu\text{m}$. The off-axis configuration with tilting the mirror is used for the spatially separation of the zero-order and the conjugate image.

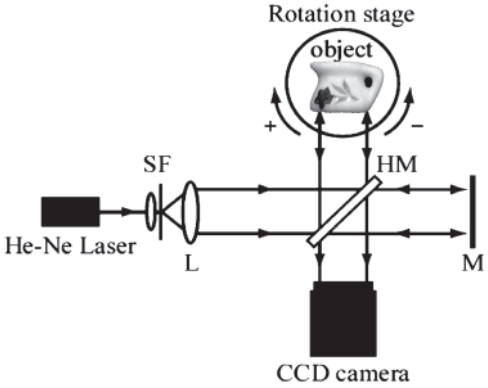
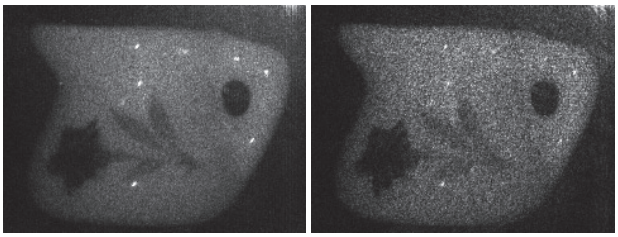


Fig. 11. An optical setup for digital hologram recording. SF: spatial filter, HM: half mirror, M: mirror, L: lens.

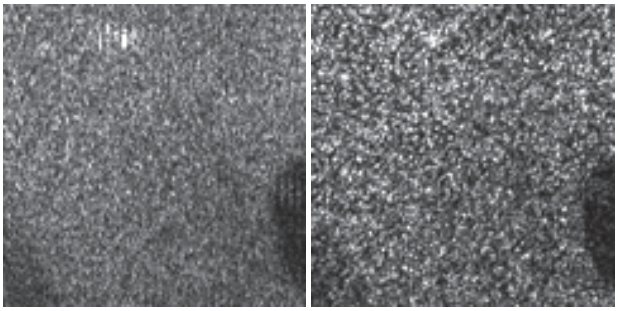
The distance from the object (a miniature pitcher as shown in Fig. 3) to the CCD camera is 400 mm. It is necessary to adjust the rotation angle to 0.4 deg or less to synthesize holograms, because two adjacent holograms are overlapped each other. In the experiments, the object was rotated from -180 deg to $+180$ deg at an interval of 0.2 deg and 1801 holograms were recorded. Then, some of recorded holograms were processed and superposed by the proposed method.

To confirm the feasibility of superposition, the reconstruction images from a single hologram and a superposed hologram are compared. The reconstructed images are shown in Fig. 12.



(a) (b)

Fig. 12. Reconstructed images from (a) a synthesized hologram and (b) a single hologram.



(a) (b)

Fig. 13. Magnified portions of the reconstructed images from (a) a synthesized hologram and (b) a single hologram.

The magnified part of the each reconstructed images are shown in Fig. 13. The horizontal size of speckles in Fig. 13(a) are smaller than that in Fig. 13(b), because the size of speckles is in inverse proportion to the size of holograms. Furthermore, as neither blurred nor double images can be seen in Fig. 12(a), it is shown that the synthesis is successful.

5. Conclusion

Two methods to reduce speckle of the reconstructed image have been presented. One is the method based on the superposition of the reconstructed images with different wavelengths. Another is the method based on the multiple holograms. After a brief review of the speckle size of the digital holography, the concepts of the two methods have been described. Some experimental results have been shown to confirm the proposed methods. Some experimental results are shown to confirm the proposed methods.

6. References

Yamaguchi, I. & Zhang, T. (1997). Phase-shifting digital holography, *Opt. Lett.* Vol. 22, 1268-1270.

Osten, W.; Baumbach, T. & Jüptner, W. (2002). Comparative digital holography, *Opt. Lett.*Vol. 27, 1764-1766.

Kreis, T. (2005) *Handbook of Holographic Interferometry* Wiley VCH, Weinheim.

Ferraro, P,; Grilli, S.; Alfieri, D.; Nicola, S. D.; Finizio, A.; Pierattini, G.; Javidi, B.; Coppola, G. & Striano, V. (2005). Extended focused image in microscopy by digital Holography, *Opt. Express* Vol. 13, 6738-6749.

Frauel, Y., Naughton, T.; Matoba, O.; Tahajuerce, E. & Javidi, B. (2006). Three-dimensional imaging and processing using computational holographic imaging, *Proc. IEEE* Vol. 94, 636-653.

Nomura, T. & Imbe, M. (2010). Single-exposure phase-shifting digital holography using a random-phase reference wave, *Opt. Lett.* Vol. 35, No. 13, 2281-2283.

Javidi, B. & Nomura, T. (2000). Securing information by use of digital holography, *Opt. Lett.* Vol. 25, No.1, 28-30.

Tajahuerce E. & Javidi, B. (2000). Encrypting three-dimensional information with digital holography, *Appl. Opt.* Vol. 39, 6595-6601.

Nomura, T.; Uota, K. & Morimoto, Y. (2004). Hybrid optical encryption of a 3-D object using a digital holographic technique, *Opt. Eng.* Vol. 43, 2228-2232.

Poon, T.-C. & Kim, T. (1999). Optical Image Recognition of Three-Dimensional Objects, *Appl. Opt.* Vol. 38, 370-381.

Tajahuerce, E.; Matoba, O. & Javidi, B. (2001). Shift-invariant three-dimensional object recognition by means of digital holography, *Appl. Opt.* Vol. 40, 3877-3888.

Javidi, B. & Kim, D. (2005). Three-dimensional-object recognition by use of single-exposure on-axis digital holography, *Opt. Lett.* Vol. 30, 236-238.

Javidi,B.; Ferraro, P.; Hong, S.-H.;De Nicola, S.; Finizio, A; Alfieri, D. & Pierattini, G. (2005) Three-dimensional image fusion by use of multiwavelength digital holography, *Opt. Lett.* Vol. 30, 144-146.

Do, C. M.; Hong, S.-H.; Nomura, T. & Javidi, B. (2005) Multi-wavelength holographic image fusions using discrete wavelet transform, *Proc. SPIE* Vol. 6016, 60160Z-1-6.

Maycock, J.; Mc Elhinney, C. P.; McDonald, J. B.; Naughton, T. & Javidi, B. (2005). Independent component analysis applied to digital holograms of three-dimensional objects, *Proc. SPIE* Vol. 5908, 590806-1-9.

Kreis, T. & Schlüter, K. (2007). Resolution enhancement by aperture synthesis in digital holography, *Opt. Eng.* Vol. 46, 055803.

Nomura, T., Okamura, M., Nitanai, E. & Numata, T. (2008). Image quality improvement of digital holography by superposition of reconstructed images obtained by multiple wavelengths, *Appl. Opt.* Vol. 47, No. 19, D38-D43.

Binet, R.; Colineau, J. & Lehureau, J. -C. (2002). Short-range synthetic aperture imaging at 633 nm by digital holography, *Appl. Opt.* Vol. 41, 4775-4782.

20

Digital Holography and Phase Retrieval

Hamootal Duadi[1], Ofer Margalit[1], Vicente Mico[2],
José A. Rodrigo[3], Tatiana Alieva[4], Javier Garcia[2] and Zeev Zalevsky[1]
[1]*School of Engineering, Bar-Ilan University, Ramat-Gan 52900,* [2]*Departamento de Óptica, Universitat de València, c/Dr. Moliner, 50, 46100 Burjassot,*
[3]*Instituto de Óptica (CSIC), Imaging and Vision Department. Serrano 121, Madrid 28006,*
[4]*Universidad Complutense de Madrid, Facultad de Ciencias Fisicas, Ciudad Universitaria s/n, Madrid 28040,*
[1]*Israel*
[2,3,4]*Spain*

1. Introduction

In imaging systems, digital as well as analogue, only the intensity is captured while the phase, containing the depth information and/or refractive index changes, is lost. Phase retrieval is the art of acquiring this information which is crucial in surface reconstruction, microscopy, location detection and depth measurements.
In this chapter we will first introduce holographic based methods, or the digital holography approach, as a solution to the phase retrieval problem. These methods use an optical setup, called an interferometer, which records the interference pattern between a reference beam and the analyzed wavefront. The interferometer image depends on the phase difference between the two beams; hence it holds the phase information. We will present a work of ours combining an interferometer setup with partial coherence and with specimen rotation. Note that instead of inspecting the contrast of the interference fringes one may use the contrast of the projected/reflected speckle patterns (self interference patterns) to map the axial domain.
We will then present phase retrieval methods based on wavefront propagation. This involves solving mathematical equations, related to physical conditions on the wavefront. Several numerical computational methods used for this task will be considered.
Further, we will overview the iterative approaches to the phase retrieval problem, discussing their abilities and limitations.
It will be shown that the iterative method based on multiple measurements allows more accurate and less noisy phase reconstruction. As an example we demonstrate the results of phase retrieval from the multiple images acquired in the gyrator domain.
Finally, some conclusions will be made.

2. Digital holography

One common method for acquiring the phase of an object is by digital holography. In regular holography, the interferometer image is written on film. Then the phase information

is reconstructed by illuminating the recording film with the correct reference beam. In digital holography the interferometer image is typically captured by a CCD camera and the reconstruction step is done using digital processing techniques. Holography based setups, named interferometer setups, are used to record the interference between a reference beam and the desired wavefront (Gabor, 1948; Leith & Upatnieks, 1962; Leith & Upatnieks, 1963; Leith & Upatnieks, 1964). This interference pattern contains the phase information, which can later be reconstructed. For the purpose of phase retrieval digital holography is used, since the idea is to use a CCD camera and digitally compute the missing phase information. Usage of in-line setups, i.e. phase shifting interferometer (Yamaguchi & Zhang, 1997; Bruning et. al., 1974; Xu et. al., 2002), permits the measurement of both phase and amplitude of an object. This is possible since the setup is an on axis setup, meaning that all beams are aligned to the optical axis (Fig. 1). The phase is recovered from the change in the diffraction pattern.

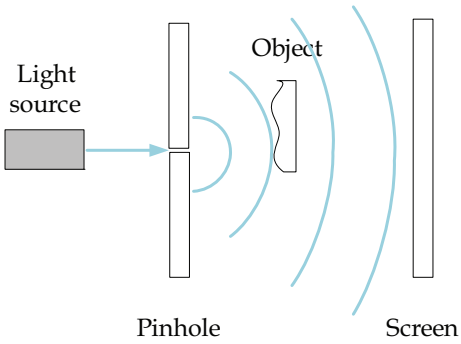


Fig. 1. Optical setup for in-line holography. A coherent wave propagates through a pinhole to create a point source. Partially scatters off an object, creating a highly magnified hologram on a screen.

The problem with holography is the phase wrapping ambiguity, i.e. the mapping range should not exceed one wavelength. Some methods are based upon the coherency of the light source (Rosen & Yariv, 1996), since interference fringes will only be generated within the coherence length. Various works propose tomographies with rotation of specimen or modification of illumination (Charriere et.al, 2006; Choi et. al, 2008). Other works propose using speckle properties as a unique axial characteristic (Karo, 1977; Cederquist et. al., 1988; Almoro et. al., 2009).

By combining an interferometer setup with partial coherence of a light source or specimen rotation (Zalevsky et. al., 2008), one may overcome the phase wrapping limitation. In the partial coherence method, a Michelson interferometer was used (Fig. 2). This interferometer splits the incoming beam to two optical paths. One path reaches a mirror and the other a reflective specimen. The two beams are combined together by a beam splitter (BS) to create an interferometer image relative to the surface profile of the specimen.

First the source's interference fringe contrast was mapped to create a reference lookup table (Table 1). This table includes the axial location corresponding to different fringe contrast values. One may see that in the very near field (up to 4mm) the fringe contrast is periodic, and in the farther field there is a linear decrease in the contras (maintaining the same

periodicity). Then, the 3D information is extracted from a single image using the lookup table. In the experimental results shown in Fig. 3, the depth difference between the two wafers placed on the mirror is approximately in the order of 20-30μm (Fig. 3. (a)) and 10-15μm (Fig. 3. (b)). One may see that the fringe contrast indeed changes for different depth positions. The fringe contrast, marked by C in the figures, is an indicator of the depth location, since the farther the object is, the lower the contrast. This method is a real time method, with tunable axial accuracy and without triangulation angle dependency.

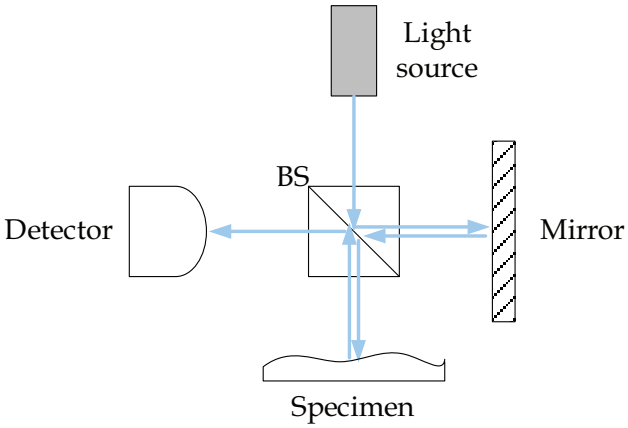
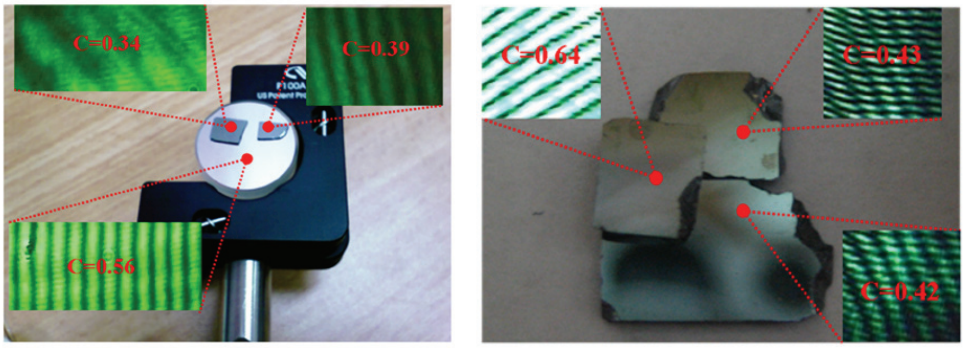


Fig. 2. Optical setup for Michelson interferometer. A coherent wave from the light source is spitted by the BS to two optical paths. One path leads to a mirror that reflects the ray back to the BS. The other path leads to a reflective specimen and back to the BS. The difference between the two optical paths depends on the profile of the specimen. Hence, by combining together the two rays, one receives an interferometer image relative to the profile of the specimen.



(a) (b)

Fig. 3. Image of the object used for the partial coherence experiment and the obtained interference pattern with the extracted local contrasts. (a) Two silicon wafers positioned on a mirror. (b). Two silicon wafers positioned on top of a third silicon wafer.

Path difference (mm)	Contrast
0	0.650
0.1	0.580
0.2	0.576
0.3	0.535
0.4	0.410
0.5	0.280
0.6	0.255
0.7	0.385
0.8	0.306
0.9	0.275
1.0	0.260
1.5	0.430
2.0	0.500
2.5	0.380
3.0	0.266
3.5	0.425
4.0	0.502
54.0	0.509
54.1	0.467
54.2	0.405
54.3	0.340
54.4	0.340
54.5	0.329
54.6	0.328
54.7	0.286
54.8	0.267
54.9	0.250
55.0	0.237

Table 1. Experimental lookup table of fringe contrast changes versus the difference betwean the two paths of the interferometer.

In the multiple angle setups, a Mach-Zehnder based interferometer was used, where the specimen was placed on a high precision rotation stage (Fig. 4). In this interferometer the input beam is split into two optical paths and the investigated sample is placed in one of these paths. The two beams are then combined by a BS to create an interferometer image relative to the depth profile of the sample.

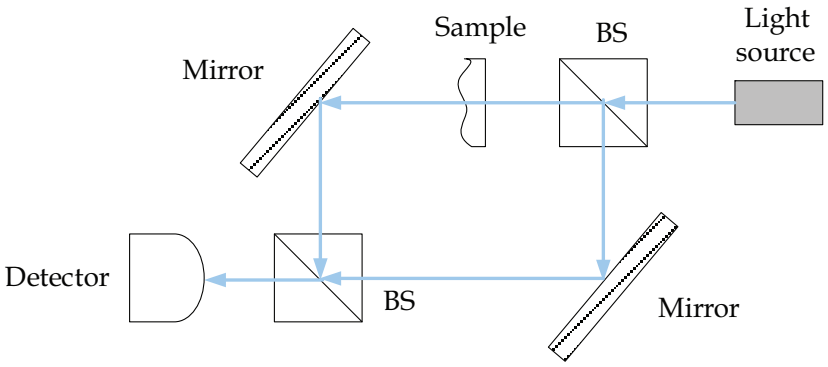


Fig. 4. Optical setup for Mach-Zender interferometer. A coherent wave from the light source is split by the BS to two optical paths. Both paths have the same optical distance and lead to a mirror that reflects the beam towards the second BS. However, a transparent sample is placed in one of the paths, which introduces a phase difference between the paths. The interferometer output of the BS is received by the detector.

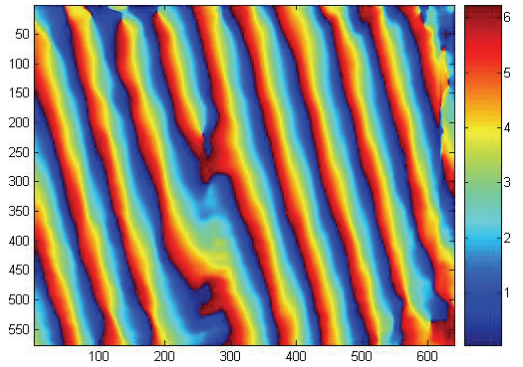
In this setup, two images were captured while the angular orientation of the specimen was changed by the rotation stage and a 3D image was reconstructed using the following equations for the phase:

$$\left\{\left[2\pi nd(x)\right]/\left[\lambda\cos\theta_1\right]\right\}\bmod\{2\pi\} = a(x) \quad ;\left\{\left[2\pi nd(x)\right]/\left[\lambda\cos\theta_2\right]\right\}\bmod\{2\pi\} = b(x) \tag{1}$$

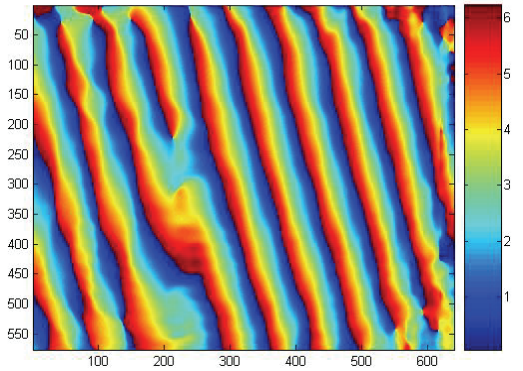
where λ is the wavelength, n is the refractive index of the sample, and d(x) is the true (unwrapped) profile that we wish to extract. Mod is the mathematical operation of modulo. The required angular change between images is less than one pixel in the camera.

The results of this setup are presented in Fig. 5. The first two images (Fig. 5. (a),(b)) are the phase reconstructions separated by a 1 degree specimen rotation. The lowest image (Fig. 5. (c)) is the reconstructed 3D profile computed from Eq. (1). This method allows imaging of phase only transparent objects and, in the case when the profile is known, allows refractive index estimation.

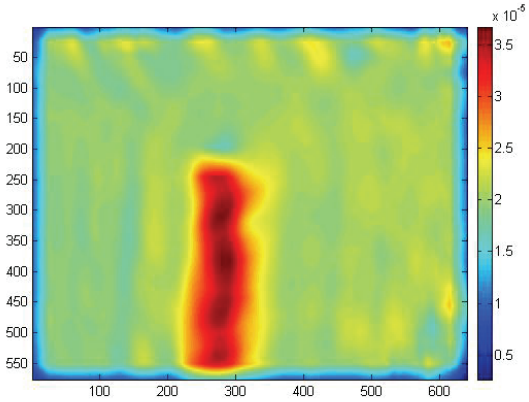
It is also possible to use speckle contrast instead of fringe contrast. Speckles are a random pattern created by a beam scattered from a rough surface. Each optical system has a unique speckle pattern that varies in the propagation direction (Goodman, 1984). This means that it is possible to generate for each axial position its own distinguishable random illumination pattern. Furthermore, the random patterns that are related to different planes have low correlation with each other. Hence, by recognizing the speckle pattern one may determine the axial location (Almoro & Hanson, 2008; Rastogi & Jacquot, 1987; Sjödahl, 1995). However, this approach is mainly used to measure distortion of an object and is usually combined with other methods. Furthermore, the speckle contrast decreases as the beam propagates along the axial direction. This allows phase retrieval through computation of speckle contrast (Irvine et. al., 2010). In both of these methods the speckle distributions are first studied. Then, the sample is placed in the setup and the depth is recovered from the reference speckle information.



(a)



(b)



(c)

Fig. 5. Experimental reconstruction using rotation of specimen. (a),(b) Phase reconstruction for two projection angles that were separated by 1 degree. (c) Profile reconstruction. Each pixel represents a 6.7μm camera pixel. The alleviation thickness is in meters.

3. Numerical computations

The numerical computational methods allow phase reconstruction with or without an interferometer setup. However, they are needed for more than just phase retrieval. The digital computation of the wavefront propagation is also needed in any numerical algorithm or simulation of an electromagnetic wave. The different kinds of numerical reconstruction also offer a solution for difficulties in real space processing, such as spurious noise, fringes, optical system aberrations etc. There are numerous computational methods, including Kirchhoff–Helmholtz transform (Gabor, 1949), Fresnel-Kirchhoff (Cuche et. al., 1999), transforming Guigay equations (Guigay, 1977; Teague, 1983), solving transport of intensity equations (Gureyev & Nugent, 1997), convolution (Colomb et. al., 2006), angular spectrum (Mann et. al., 2005) etc. Each method corresponds to a different mathematical description of the optical space and has its advantages for a specific circumstance. Some descriptions are differential while others are integral.

One of the most common equations, which describe the propagation of an input wave u(x,0) through a free space distance z, is the Fresnel integral:

$$u(x,y,z) = \exp(i2\pi z/\lambda)\left[u(x,y,0) * \exp\left(i2\pi\left(x^2+y^2\right)/z\lambda\right)\right]/iz\lambda \tag{2}$$

where λ is the wavelength and * denotes the mathematical convolution operation. However, since numerical simulations are discrete, according to the Nyquist theorem this is correct only if $z>Z_c$, while Z_c is defined as:

$$Z_c = Ndx^2/\lambda \tag{3}$$

where N is the number of pixels and dx is the pixel size.

If one would like to represent the very near field, then there is need for a different approximation.

The angular spectrum calculation of the output field is

$$u(x,y,z) = \iint U(\xi,\mu)\exp\left(-i\pi\left(\xi^2+\mu^2\right)\lambda z\right)\exp\left(2\pi i(\xi x+\mu y)\right)d\xi d\mu \tag{4}$$

where U(ξ) is the Fourier transform of the input field, defined by

$$U(\xi,\mu) = \iint u(x,y,0)\exp\left(-2\pi i(\xi x+\mu y)\right)dxdy \tag{5}$$

The meaning of this representation is to compute the product between the Fourier transform of the input field and the Fourier transform of the Kernel $\exp(i\pi(x^2+y^2)/\lambda z)$. In this case the Nyquist theorem require $z<Z_c$ (Mendlovic et. al., 1997).

Mathematically, Eq. (3) and (4) are the same, since it is a known Fourier property that the convolution operation becomes a product in the Fourier domain. However, converting these equations from the continuous domain to the discrete one reveals a contradiction.

The transport of intensity equation indicates what is the input phase φ(x,y) that solves a known intensity I(x,y,z) variation in the propagation direction.

$$-\nabla\cdot(I(x,y,z)\nabla\varphi(x,y)) = 2\pi/\lambda\partial_z I(x,y,z) \tag{6}$$

where $\nabla=(\partial_x,\partial_y)$ is the gradient operator. By decomposing the problem to the Fourier harmonics, it is posible to convert the differential problem into a matrix formulae.

There are other advantages in choosing certain computational method. For example, the angular spectrum computation allows control of spurious noise components (Mann et. al., 2005).

4. Iterative phase retrieval

One famous iterative method is the Gerchberg-Saxton (GS) algorithm (Gerchberg & Saxton, 1972). This method is widely used to retrieve the input phase, from known amplitude distributions of the input and output fields of an optical setup. Later on this algorithm was used for beam shaping applications to compute the required phase element for producing a specific output intensity, given the input illumination. The idea is to choose an initial random phase estimation and compute a new estimation in each new iteration (Fig. 6).

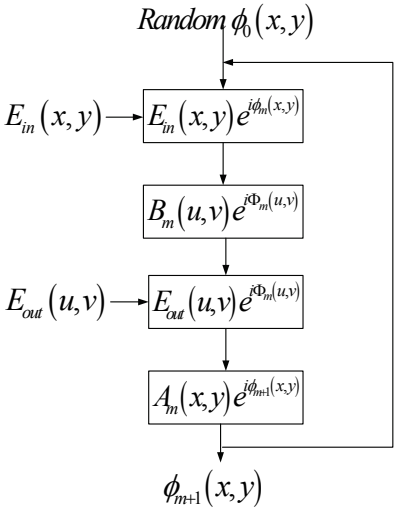


Fig. 6. A scheme of the GS algorithm.

In each iteration the contribution of the current phase estimation and the amplitude of the input illumination are computed, denoted by the input plane expression $E_{in}(x,y)\exp\{i\phi_m(x,y)\}$. Then, the resulting expression after passing through the optical domain is computed by transforming this input plane expression. This is denoted by the output plane expression $B_m(u,v)\exp\{i\Phi_{m+1}(u,v)\}$. In order to fulfill the apriori constrain for the output plane intensity, the magnitude of the output plane is replaced by $E_{out}(u,v)$. The output plane expression is now $E_{out}(u,v)\exp\{i\Phi_{m+1}(u,v)\}$. The new output plane expression, $E_{out}(u,v)\exp\{i\Phi_{m+1}(u,v)\}$, is then inverse transformed to the input plane, denoted by $A_m(x,y)\exp\{i\phi_{m+1}(x,y)\}$, and the apriori constrain for the input amplitude $E_{in}(x,y)$ is imposed on the input expression. The phase $\phi_{m+1}(x,y)$ is the new phase estimation. This process is repeated until convergence is obtained.

Note that it was proven that the GS algorithm indeed converges (Fienup, 1982), which means that the next phase estimation is at least as good as the previous one.

The original GS algorithm was used in a Fourier based optical setup. However, other optical transforms, such as the Fresnel and the fractional Fourier transform, were later on proven to be more efficient in terms of convergence rate (Zalevsky et. al., 1996).

Another iterative approach is the gradient search method (Fienup, 1993). This method has a better considered step size towards the desired phase, which allows faster convergence for multiple intensity constrains.
However, the problem with iterative methods is the phase ambiguity, resulting with converging to a local minima solution, which leads to an incorrect phase distribution. In order to overcome this problem, statistical methods (Nieto-Vesperinas et. al., 1988), stimulated annealing (Kirkpatrick et. al., 1983; Zhou et. al., 1999) and de-convolution methods (Seldin & Fienup, 1990) were proposed.

5. Iterative phase retrieval approach based upon multiple measurements

The concept of using multiple images was first suggested regarding the phase ambiguity in electron microscopy (Misell, 1973). Misell proposed that by taking several defocused images, one may later reconstruct a better phase using the iterative GS algorithm. This idea was implemented in later works as well (Redding et. al., 1998; Acton et. al., 2004; Brady & Fienup, 2006; Dean et. al., 2006). In order to capture such multiple plane images, one can change the axial position of the sample using a stepper motor (Barty et. al., 1998), change the position of the recording imaging device by using mechanical motion (Zhang et. al. 2003) or using defocusing implemented in a spatial light modulator in a motionless configuration (Camacho et. al., 2010).

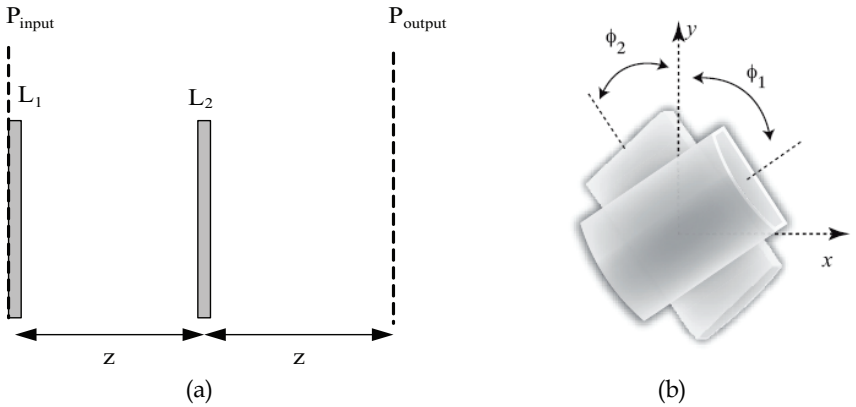


Fig. 7. (a) GT optical setup. Each generalized lens $L_{1,2}$ is an assembled set of two cylindrical lenses (b). The transformation is only reached by the proper rotation (with angles ϕ_1 and ϕ_2) of the lenses.

Recently the efficiency of the multiple constrains iterative method has been demonstrated using images obtained in the gyrator domains (Rodrigo et. al., 2010). The gyrator transform (GT) is a linear canonical integral transform, which produces the twisted rotation in position−spatial frequency planes of the phase space (Rodrigo et. al., 2007). Thus, the GT operation of a two dimensional function $f_i(r_i)$ (complex field amplitude) with a parameter α, known as the transformation angle, is given by

$$F^{\alpha}(r_o) = \frac{1}{2\lambda z|\sin\alpha|}\iint f_i(x_i, y_i)\exp\left(i2\pi\frac{(x_o y_o + x_i y_i)\cos\alpha - (x_i y_o + x_o y_i)}{2\lambda z\sin\alpha}\right)dx_i dy_i \quad (7)$$

where $\mathbf{r}_{i,o} = (x_{i,o}, y_{i,o})$ are the input and output coordinates, respectively, λz is a normalization parameter that for the case of the optical implementation of the GT correspond to the wavelength λ and to fixed axial interval between the lenses z. The gyrator projections $\left|F^{\alpha}(r_O)\right|^2$ can be measured using the optical setup displayed in Fig. 7 (a). The optical setup consists of two generalized lenses (Fig. 7 (a)). Each generalized lens is assembled from two cylindrical lenses (Fig. 7 (b)), properly rotated with angles ϕ_1 and ϕ_2 that fulfill the following relations with the transformation angle α:

$$\sin(2\varphi_1) = \cot(\alpha/2); \quad \sin(2\varphi_2) = (\sin\alpha)/2 \tag{8}$$

In order to improve phase retrieval in a gyrator base optical system it was proposed to use multiple rotation angles. As in the original GS algorithm, a random phase estimation is chosen, and in each iteration a new estimation is computed (Fig. 8). Note that this serial approach may also be used in different transformation domains than the gyrator e.g. in the Fresnel or the Fourier optical setups. Nevertheless the astigmatism of the GT allows to resolve some phase ambiguities (for example the direction of the rotation of the helicoidal phase) that is impossible to do using the isotropic transforms.

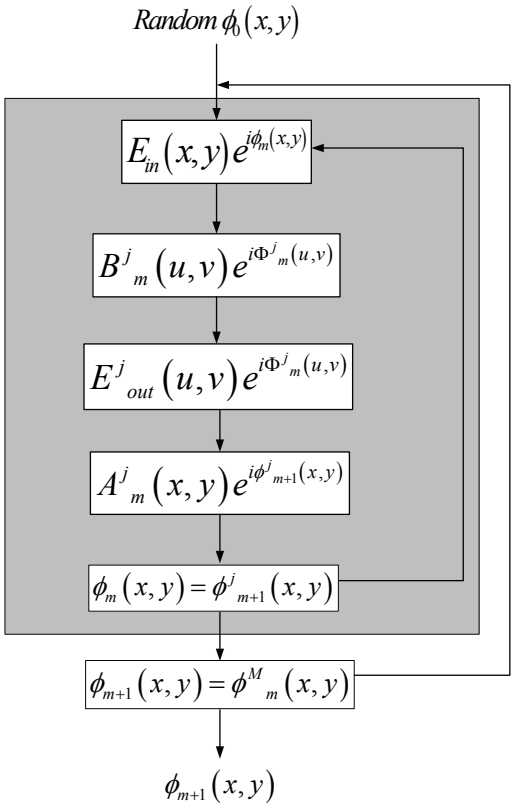
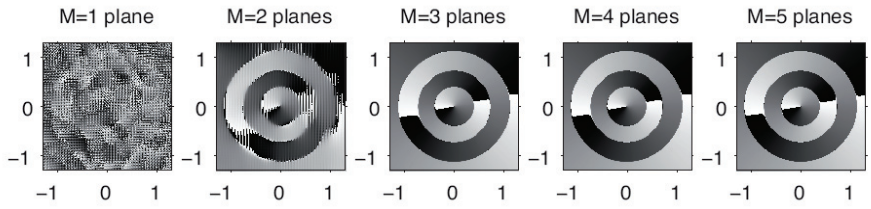
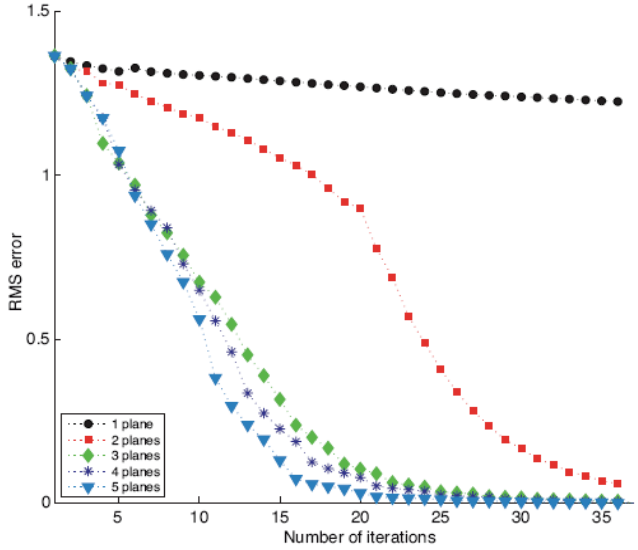


Fig. 8. Scheme of the iterative algorithm based upon multiple measurements.



(a)



(b)

Fig. 9. Numerical results for the multi-stage phase retrieval algorithm using GT. M is the number of constraint images (measurements). (a) Retrieved phase distribution after 23 iterations for the case of $LG_{+3,1}$ mode. (b) The phase retrieval is calculated by using constraint images involving angles: α = 45°, 55°, 65°, 70° and 80°. The evolution of RMS error is also displayed as a function of the number of constraint images as well as the number of utilized iterations.

In order to estimate the phase retrieval capabilities of a multiple plane method in comparison to a single plane method, a quality measuring tool is need. A common image processing tool for this is the relative root mean square (RMS) error, also known as Abserr, which is given by

$$\mathrm{RMS} = \left(\sum \left|E_{in}\exp(i\phi) - E_{in}\exp(i\phi_{rt})\right|^2\right)^{\frac{1}{2}} \times \left(\sum \left|E_{in}\exp(i\phi)\right|^2\right)^{-\frac{1}{2}} \tag{9}$$

where $\phi(x,y)$ is the phase to retrieve and $\phi_{rt}(x,y)$ is the estimated phase distribution at each iteration.

The reconstructed phase distribution for a different numbers of planes is shown in Fig. 9 (a). One may see that for a fixed number of iterations, the reconstructed phase improves as the

number of planes increases. The RMS error versus the number of iterations is shown in Fig. 9 (b). Adding planes decreases the convergence time. By defining successful retrieval as one having RMS error of less than 0.02, it was also demonstrated that the phase distribution is successfully retrieved (even in a noisy case) by using at least three constraint planes and after applying 36 iterations (Rodrigo et. al., 2010).
By working with the gyrator optical setup, instead of other optical domains, one may use images from different transform rotation angles (rather than using different axial positions requiring generation of axial movement) which makes the system more flexible and robust.

6. Conclusion

In this chapter we have presented an overview on the topic of the phase retrieval problem. We have presented some interesting perspectives in three different approaches: digital holography, iterative phase retrieval and multiple measurement methods.
First we have presented the interferometer as the optical setup needed for digital holography. We have discussed the advantages of in-phase setups as well. We have presented the problem of phase ambiguity in holography based methods. Our work of using partial coherence or specimen rotation was able to overcome this phase wrapping problem.
We have presented different computational methods needed in phase retrieval. The use of computational methods in digital holography is preferable over regular holography in terms of dealing with spurious noise, fringes, optical system aberrations etc.
Then we have discussed the abilities of various iterative methods having local minima solution problem. The idea of multiple measurements to create a multiple plane iterative method was presented as a possible solution. Out work implementing this approach in the gyrator domain was presented and was proven to allow phase retrieval with high accuracy and low complexity.
The phase retrieval methods, mentioned in this chapter, are commonly used in different three dimensional applications. Space telescopes, such as the James Webb space telescope (Acton et. al., 2004), use the multiple measurement iterative approach. Electron microscopy uses the multiple measurement method as well. Iterative phase retrieval algorithms are also used for planning diffractive optical elements (DOE) used as phase only filters for beam shaping (Zhou et. al., 1999). Digital holography microscopes (DHM) use different interferometer based setups (Charriere et. al., 2006).

7. References

Acton, D. S.; Atcheson, P. D.; Cermak, M.; Kingsbury, L. K.; Shi, F. and Redding, D. C. (2004). James Webb Space Telescope wavefront sensing and control algorithms, *Proceedings of SPIE 5487*, 887, Glasgow Scotland United Kingdom, June 2004, SPIE .

Almoro, P. F. and Hanson, S. G. (2008). Wavefront sensing using speckles with fringe compensation, *Opt. Exp.*, Vol. 16, No. 11, (May 2008) pp. 7608-7618.

Almoro, P. F.; Maallo, A. M. S. and Hanson, S. G. (2009) . Fast-convergent algorithm for speckle-based phase retrieval and a design for dynamic wavefront sensing, *Appl.* Opt., Vol. 48, No. 8, (Mar. 2009) pp. 1485-1493.

Barty, A.; Nugent, K. A.; Paganin,D. and Roberts, A. Quantitative optical phase microscopy, *Opt. Lett.*, Vol. 23, No. 11, (Jun. 1998) pp. 817-819.

Brady, G. R. and Fienup, J. R.(2006). Nonlinear optimization algorithm for retrieving the full complex pupil function, *Opt. Express*, Vol. 14, No. 2, (Jan. 2006) 474–486 .

Bruning, J. H.; Herriott, D. R.; Gallagher, J. E.; Rosenfeld, D. P.; White, A. D. and Brangaccio, D. J. (1974). Digital wavefront measuring interferometer for testing optical surfaces and lenses, *Appl. Opt.*, Vol. 13, No. 11, (Nov. 1974) pp. 2693–2703.

Camacho, L.; Micó, V.; Zalevsky, Z. and García, J. Quantitative phase microscopy using defocusing by means of a spatial light modulator, *Opt. Express,* Vol. 18, No. 7 (Mar. 2010) pp. 6755-6766.

Cederquist, J N; Fienup, J R; Marron, J C; Paxman, R G (1988). Phase retrieval from experimental far-field speckle data, *Opt. Lett.*, Vol. 13, No. 8, (Aug. 1988) pp.619-621

Charriere, F.; Marian, A.; Montfort, F.; Kuehn, J. and Colomb, T. (2006). Cell refractive index tomography by digital holographic microscopy, *Opt. Lett.*, Vol. 31, No. 2, (Jan. 2006) pp. 178-180.

Choi, W.; Fang-Yen, C; Badizadegan, K.; Dasari, R. R. And Feld, M. S. (2008). Extended depth of focus in tomographic phase microscopy using a propagation algorithm, *Opt. Lett.*, Vol. 33, No. 2, (Jan. 2008) pp. 171-173.

Colomb, T.; Montfort, F.; Kuhn,J; Aspert, N.; Cuche, E.; Marian, A.; Charrier, F.; Bourquin, S.; Marquet, P. and Depeursinge, C. (2006). Numerical parametric lens for shifting, magnification, and complete aberration compensation in digital holographic microscopy, *JOSA A.*, Vol. 23, No. 12, (Dec. 2006) pp. 3177-3190.

Cuche, E.; Bevilacqua, F. and Depeursinge, C. (1999). Digital holography for quantitative phase-contrast imaging, *Opt. Lett.*, Vol. 24, No. 5, (Mar. 1999) pp. 291–293.

Dean, B. H.; Aronstein, D. L.; Smith, J. S.; R. Shiri, and Acton, D. S. (2006). Phase retrieval algorithm for JWST Flight and Testbed Telescope, *Proceedings of SPIE 6265*, 626511, Orlando FL USA, May 2006, SPIE.

Fienup, J. R. (1982). Phase retrieval algorithms: a comparison, *Appl. Opt.*, Vol. 21, No. 15, (Aug. 1982) pp. 2758-2769.

Fienup, J. R. (1993). Phase-retrieval algorithms for a complicated optical system, *Appl. Opt.*, Vol. 32, No. 10, (Apr. 1993) pp. 1737–1746 .

Gabor, D. (1948). A new microscopic principle, *Nature*, Vol. 161, No. 4098, (May 1948) pp. 777–778.

Gabor, D. (1949). Microscopy by reconstructed wavefronts, *Proc. R. Soc. London Ser. A*, Vol. 197, No. 1051, (Jul. 1949) pp. 454–487.

Gerchberg ,R. W. and Saxton, W. O. (1972). A practical algorithm for the determination of phase from image and diffraction plane pictures, *Optik*, Vol. 35, No. 2 237–246.

Goodman, J. W. (2000). *Statistical optics,* Wiley Interscince, ISBN 0471399167, New York.

Guigay, J. P. (1977). Fourier-transform analysis of Fresnel diffraction patterns and in-line holograms, *Optik*, Vol. 49, pp. 121–125.

Gureyev, T. E. and Nugent, K. A. (1997). Rapid quantitative phase imaging using the transport of intensity equation, *Opt. Commun.*, Vol. 133, No.1, (Jan. 1997) pp.339 – 346.

Irvine, S. C.; Paganin, D. M.; Jamison, R. A.; Dubsky, S. and Fouras, A. (2010). Vector tomographic X-ray phase contrast velocimetry utilizing dynamic blood speckle, *Opt. Exp.*, Vol. 18, No. 3, (Jan. 2010) pp. 2368-2379.

Karo, D. P. and Schneiderman, A. M.(1977). Transfer functions, correlation scales, and phase retrieval in speckle interferometry, *JOSA*, Vol. 67, No, 11, (Nov. 1977) pp. 1583-1587.

Kirkpatrick, S.; Gelatt, C. D. and Vecchi, M. P. (1983). Optimization by simulated annealing, *Scienc,e* Vol. 220, No. 4598 , (May 1983) pp. 671–680.

Leith, E. N. and Upatnieks, J. (1962). Reconstructed wavefronts and communication theory, *JOSA,* Vol. 52, No. 10, (Oct. 1962) pp. 1123– 1128.

Leith, E. N. and Upatnieks, J. (1963). Wavefront reconstruction with continuous-tone objects, *JOSA,* Vol. 53, No. 12, (Dec. 1963) pp. 1377–1381.

Leith, E. N. and Upatnieks,J. (1964). Wavefront reconstruction with diffused illumination and three-dimensional objects, *JOSA*, Vol. 54, No. 11, (Nov. 1964) 1295–1301.

Mann, C.; Yu, L.; Lo, C. M. and Kim,M. (2005). High-resolution quantitative phase-contrast microscopy by digital holography, *Opt. Express*, Vol. 13, No. 22, (Oct. 2005) pp. 8693–8698.

Mendlovic, D.; Zalevsky, Z. and Konforti, N. (1997). Computation considerations and fast algorithms for calculating the diffraction integral, *J. Mod. Opt.*, Vol. 44, No. 2, (Feb. 1997) pp. 407-414.

Misell, D. L. (1973). A method for the solution of the phase problem in electron microscopy, *J. Phys. D: App. Phy.*, Vol. 6, No. 1,(Jan. 1973) pp. L6–L9.

Nieto-Vesperinas, M.; Navarro, R. and Fuentes, F. J. (1088). Performance of a simulated-annealing algorithm for phase retrieval, *JOSA A*, Vol. 5, No. 1, (Jan. 1988) pp. 30–38.

Rastogi, P. K and Jacquot, P. (1987). Measurement of difference deformation using speckle interferometry, *Opt. Lett.*, Vol. 12, No. 8, (Aug. 1987) pp. 596-598.

Redding, D. C.; Basinger, S.; Lowman, A.; Kissil, A.; Bely, P.; Burg, R.; Lyon, R.; Mosier, G.; Femiano,M.; Wilson, M.; Schunk, R. G.; Craig, L.; Jacobson, D.; Rakoczy, J. and Hadaway ,J. (1998). Wavefront Sensing and Control for a Next- Generation Space Telescope, *Proceedings of SPIE 3356*, 758,Liege Belguim, Jun. 1998.

Rodrigo, J. A.; Alieva, T. and Calvo, M. L. (2007). Experimental implementation of the gyrator transform, *JOSA A* , Vol. 24, No. 10,(Oct. 2007) pp. 3135–3139.

Rodrigo, J. A.; Duadi, H.; Alieva, T. and Zalevsky, Z. (2010). Multi-stage phase retrieval algorithm based upon the gyrator transform, *Opt. Exp.*, Vol. 18, No. 2, (Jan. 2010) pp. 1510–1520 .

Rosen, J. and Yariv, A. (1996). General theorem of spatial coherence: application to three-dimensional imaging, *JOSA A*, Vol. 13, No. 10, (Oct. 1996) 2091–2095.

Seldin, J. H. and Fienup, J. R. (1990). Iterative blind deconvolution algorithm applied to phase retrieval, *JOSA A*, Vol. 7, No. 3, (Mar. 1990) pp. 428–433 .

Sjödahl, M (1995). Calculation of speckle displacement, decorrelation, and object-point location in imaging systems, *Appl. Opt.* Vol. 34, No. 34,(Dec. 1995) pp. 7998-8010.

Teague, M. R. (1983). Deterministic phase retrieval: a Green's function solution, *JOSA,* Vol. 73, No. 11, (Nov. 1983) pp. 1434–1441.

Xu, W.; Jericho, M. H.; Meinertzhagen, I. A. and Kreuzer, H. J. (2002). Digital in-line holography of microspheres, *Appl. Opt.*,Vol. 41, No. 25, (Sep. 2002) 5367–5375 .

Yamaguchi, I. and Zhang, T. (1997). Phase-shifting digital holography, *Opt. Lett.*, Vol. 22, No. 16, (Aug. 1997) pp. 1268-1270.

Zalevsky, Z.; Mendlovic, D. and Dorsch, R. G. (1996). Gerchberg-Saxton algorithm applied in the fractional Fourier or the Fresnel domain, *Opt. Lett.*, Vol. 21, No. 12, (Jun. 1996) pp. 842-844.

Zalevsky, Z.; Margalit, O.; Vexberg, E.; Pearl, R. and Garcia, J. (2008). Suppression of phase ambiguity in digital holography by using partial coherence or specimen rotation, *App. Opt.*, Vol. 47, No. 19, (Apr. 2008) pp. D154-D163.

Zhang, Y.; Pedrini, G.; Osten, W. and Tiziani, H. Whole optical wave field reconstruction from double or multi in-line holograms by phase retrieval algorithm, *Opt. Express*, Vol. 11, No. 24 (Dec. 2003) pp. 3234–3241.

Zhou, G.; Chen, Y.; Wang, Z. and Song, H. (1999). Genetic Local Search Algorithm for Optimization Design of Diffractive Optical Elements, *Appl. Opt.*, Vol. 38, No. 20, (Jul. 1999) pp. 4281-4290.

Part 6

Non-optical Holography

21

In-line Hard X-ray Holography for Biomedical Imaging

Andrzej Krol
Department of Radiology, SUNY Upstate Medical University
Syracuse, NY
USA

1. Introduction

X-ray imaging for biomedical applications is based on detecting interactions of electromagnetic (EM) radiation with energy in the 15–200 keV range (corresponding to wavelength in the 0.8– 0.06 A range) with biological tissues. Such ultra short EM waves called x-rays can be generated through many different mechanisms (e.g. Bremsstrahlung in solid targets, characteristic x-ray emission, synchrotron radiation or in free-electron lasers) and may interact with biological objects being imaged via the coherent Rayleigh scattering, photoelectric absorption or Compton scattering. During x-ray imaging x-ray beam is directed on the object of interest and specially designed detectors detect the transmitted or scattered x-rays. Some x-ray detectors can count photons that interacted with the detector element while other simply record the average x-ray fluence over exposure time in a given detector element. Consequently, even though x-rays are EM radiation conventional detectors can only measure its amplitude but not phase. For biological tissues the complex refractive index is used to describe their interaction with x-rays. It is very close to unity and usually it is written as: $n = 1 - \delta - i\beta$, where the $\delta > 0$ (called the decrement of the real part of the refractive index) and $\beta > 0$ describe the phase and amplitude change of x-ray wave traversing tissue, respectively (Als-Nielsen & McMorrow, 2001; Lewis, 2004; Zhou & Brahme, 2008). In conventional x-ray imaging only information on attenuation (i.e. on the imaginary part β) is obtained. For water and soft tissue the ratio δ/β increases as a square of the x-ray energy up to E~40 keV and δ/β is in the 100–1200 range. The ratio δ/β is higher for lower effective Z compounds. Therefore, it is expected that effect of phase shift will be more pronounced in soft tissue imaging, as compared to bone. It follows that two x-ray waves traversing different biological tissues may exit the object with large phase difference even though the absorption they suffered was similar. Consequently, the ability to detect the x-ray phase shift during x-ray imaging could provide additional important information on the structure of the imaged object in addition to information on electron density that could be elucidated from conventional x-ray absorption imaging. The optimum energy for such phase-contrast imaging is higher than in absorption imaging possibly resulting in the radiation dose savings. We note that absorption–based x-ray imaging suffers from limited low-contrast resolution, i.e. the ability to differentiate two adjacent regions with similar electron density is rather poor. This can be remedied by use of imaging probes– contrast agents that provide higher or lower electron density around or in the structures of interest.

As an example, consider x-ray imaging of blood vessels (angiography). They cannot be detected while embedded in soft tissue by means of conventional absorption-based imaging unless iodinated contrast agent (e.g. Iohexol or Diatrizoate) is used. Typically only vessels with diameter larger than 200 μm could be detected. Unfortunately, use of contrast agents carries risk to some patients. Because x-ray phase shift is based on x-ray refraction rather than on absorption, blood vessels could be detected in phase-contrast imaging without any contrast agent or by using a new class of safer contrast agents. For example, Zhang et al. (2008) reported using in-line holography setup for imaging of blood vessels with 40 μm in diameter without any contrast agent and 30 μm vessels using physiological saline as contrast agent.

An important parameter that determines the character of phase contrast imaging is the Fresnel number $N_F=a^2/(\lambda z)$, where a is the size of the structure in the object being imaged, and z is the object-to-detector distance (Nugent et al., 2001). The far-field *(Frauenhofer region)*, intermediate-field Fresnel region and near-field Fresnel region can be defined as corresponding to $N_F<<1$ $(z>> a^2/\lambda)$, $N_F\approx 1$ $(z\approx a^2/\lambda)$, and $N_F>1$ $(z< a^2/\lambda)$, respectively. The near-field region Fresnel region is especially important for biomedical imaging. In this regime the phase shift depends linearly on the object-to-detector distance and in most cases the phase variation can be expected to be continuous allowing to obtain unique solutions in well-defined domain and correctly-defined boundary conditions (Gureyev & Wilkins, 1998).

2. Common approaches to phase-contrast imaging

2.1 Triple crystal x-ray interferometry

This method relies on precisely aligned three lamellae (thin perfect Si crystals) cut from a single large Si monocrystal forming Bonse–Hart interferometer (Momose, 1995; Momose et al., 1996; Levis, 2004; Zhou & Brahme, 2008). The incoming x-ray beam is split by the first lamellae into two beams via Laue diffraction, the second lamellae diffracts the two beams onto the same spot on the third crystal where they interfere. The object of interest is placed in one of the beams (sampling beam) between the second and the third lamellae while the other beam is used as a reference and a phase shifter is inserted in it. The detector is placed in the optical path of the sampling beam emerging from the third lamellae. Because the phase shifter introduces known phase shift this device allows direct measurement of the phase map ($\cos\Phi(x,y)$) in the object. The spatial (lateral) coherence width and temporal coherence length that are necessary for this method are $\xi_s = \lambda\,(\Delta\alpha/\alpha)^{-1}$ and $\xi_t = \lambda\,(\Delta E/E)^{-1}$, respectively, where λ is the x-ray wavelength, $\Delta\alpha/\alpha$ is the device angular acceptance and $\Delta E/E$ is the x-ray energy bandpass of the device. For the Bonse–Hart interferometer the typical values are $\Delta\alpha/\alpha < 10^{-4}$ and $\Delta E/E< 10^{-4}$, therefore $\xi_s >10^{-6}$ m and $\xi_t >10^{-6}$ m resulting in coherence volume $V_c= \xi_s \times\xi_s \times\xi_t = 10^{-18}$ m^3. Such stringent coherence conditions can be only met by high brilliance x-ray source, e.g. synchrotron (Bilderback et al., 2005) or free-electron laser (Grubel et al., 2007) working in sufficiently short pulses, equipped with a crystal monochromator and by an interferometer made of a large single silicon crystal. The spatial resolution of a few micrometers was reported (Momose, 2002). The practical limitation for the wide spread of this technology is imposed by the need for a synchrotron or free-electron laser x-ray source, and rather small field-of-view (5×5 cm^2) created. This method is not suitable for imaging objects with very sharp interfaces (e.g. bone/soft tissue) because the resulting Fresnel fringes might become too narrow to be resolved by conventional detectors. However, this approach was successfully applied to 3D imaging of

small samples of soft tissues (e.g. rat brain, or excised breast tissue sample) resulting in 3D maps of the x-ray refractive index in the object reconstructed from multiple projections (Momose et al., 2003; Takeda et al., 2007).

2.2 Talbot interferometry

Talbot interferometry (Yokozeki & Suzuki, 1971; Lohmann & Silva, 1971) is based on the Talbot effect (Weitkamp et al., 2005; Momose et al., 2006). The imaging system consists of a phase grating G1 and an analyzer absorption grating G2 very close to the detector. If an x-ray source with low spatial coherence is used, an additional source grating G0 is inserted between the x-ray source and the object located in front of G1 (Weitkamp et al., 2006; Pfeiffer et al. 2006). If the distance between G1 and G2 is equal to the Talbot distance, the phase grating creates periodic fringe (moiré) pattern on G2 via Talbot effect. The period of G2 needs to match the period of the moiré pattern. For weakly absorbing objects the recorded image is directly related to gradient of phase, $d\Phi(x,y)/dx$, where x and y are perpendicular to the x-ray propagation vector and x is in the transverse direction perpendicular to gratings' slits. The phase shift in the object can be estimated by integration along the x-axis. Better approach is to translate G1 or G2 in the x direction and to obtain at least four images. In such way one can obtain information on the absorption component and on the phase-shift component of the image. For Talbot interferometry the required spatial coherence is $\xi_s = \lambda l/w \sim 10^{-8}$, where l is the G0–G1 distance and w is the source size. Because $\Delta E/E < 10^{-1}$, therefore temporal coherence is $\xi_t > 10^{-9}$, thus offering most relax coherence conditions ($V_c = 10^{-25}$ m^3) among all phase contrast imaging methods. The spatial resolution is defined by the source size and magnification via $\sim wd/l$, where d is the G1–G2 distance. The advantage of this method is that it could be used with conventional medical x-ray tubes and conventional large field-of-view x-ray detectors. The disadvantage is the limited grating size (presently not exceeding 10 cm×10 cm) thus resulting in the limited field-of-view, limited angular acceptance that may extend the required imaging time while working with the cone-beam source (such as x-ray tube). Presence of three gratings inserted between the x-ray source and the detector may result in dose increase to the patient (Olivo & Speller, 2007).

2.3 X-ray diffraction-enhanced imaging and multi-image radiography technique

X-ray diffraction-enhanced imaging (DEI) system consists of a crystal monochromator and a silicon single crystal analyzer. The analyzer is aligned with the monochromator in such a way that the x-rays with angular incidence within the rocking curve width relative to the Bragg angle of analyzer are accepted. The object is inserted between the monochromator and the analyser crystal. If the rocking angle is set at the half of the maximum reflectivity (larger or smaller than the Bragg angle) than the images recorded by the detector will contain information about gradient of phase shift $d\Phi(x,y)/dx$ and also about absorption of the object. If the rocking angle is set to be equal to the Bragg angle only information about absorption of the object will be provided (Chapman et al., 1997). By obtaining multiple images at low and high angles about the Bragg angle one can retrieve information on the imaginary part and on the gradient of the real part of the x-ray refracting index of the object. Because in this imaging setup under some conditions one can also obtain images due to ultra-small-angle scattering, therefore this method was expanded to multi-image radiography (MIR) technique (Wernick et al., 2003; Rigon et al., 2007). MIR requires acquisition of multiple images at different angular settings and appropriate reconstruction

methods. The estimated coherence volume for DEI is about 10^{-18} m^3, therefore this method can be only implemented with very high brilliance x-ray source such as synchrotron or free-electron laser x-rays source. The reported spatial resolution was in the range of a few micrometers (Nesterets et al., 2004). Another limitation of this method stems from the fact that the x-ray beam emerging from the analyzer is vary narrow (typically no more than 1 mm height) with its width not exceeding 10 cm. Therefore only relatively small objects can be imaged via sample scan. Such approach might result in motion artifacts and is technically challenging.

2.4 In-line hard x-ray holography

The in-line holography also called propagation-based imaging does not require any diffracting crystals or gratings. It is the simplest approach to phase contrast imaging. It was first investigated by Snigirev et al. (1995), and Wilkins et al. (1996). This method requires sufficiently small x-ray source to fulfil the condition for spatial coherence $\xi_s > 10^{-6}$ m. However, a broad x-ray spectrum $\Delta E/E < 10^{-1}$ is adequate leading to relaxed condition for temporal coherence $\xi_t > 10^{-9}$ m (Wilkins et al., 1996). The geometry used is shown in Fig. 1.

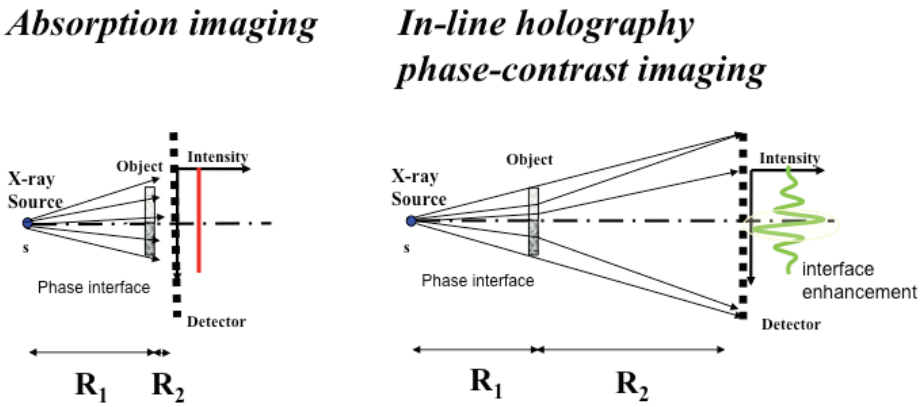


Fig. 1. Left panel: imaging geometry for absorption imaging. Right panel: imaging geometry for in-line holography setup. Absorption imaging setup does not allow detecting a phase interface in the object. In-line holography setup under some conditions might allow detection of a phase interface in the object.

The R_1 and R_2 denote the source-to-object and the object-to-detector distance, respectively. For the in-line holography the spatial coherence width condition can be expressed as

$$\xi_s = \lambda\,(\Delta\alpha/\alpha)^{-1} = \lambda(s/R_1)^{-1} = \lambda(R_1/s), \tag{1}$$

where s is the x-ray effective source size. Wu and Liu (2007) proposed a measure for estimation of the degree of coherence required to create detectable phase-contrast. Based on Wigner distribution function they introduced concept of the shearing length $L_{shear} = \lambda R_2 |u|/M$, where u is the spatial frequency of the object. They pointed out that the ratio

$$L_{shear}/\xi_s = R_2|u|/MR_1 = (M-1)s|u|/M \quad (2)$$

is a better measure of coherence of the x-ray wavefield within the shearing length for a given spatial frequency u. For $L_{shear}/\xi_s << 1$, the wavefield is highly coherent over the shearing length. It follows, that the shortest spatial wavelength (λ_o) of the object component that could be imaged via phase contrast is defined by

$$\lambda_o >> s(M-1)/M = s'/M, \quad (3)$$

where s' is the size of projected source on the detector plane (geometrical unsharpness of the system). If $L_{shear}/\xi_s \geq 1$ or $\lambda_o \leq s'/M$ than phase contrast cannot be detected. It is interesting that the above conditions do not depend on x-ray energy.

Wilikins et al. (1996) and Pogany et al. (1997) showed that for in-line holography of weakly absorbing objects in the near-field Fresnel region the observed image contrast is proportional to the Laplacian of the phase-shift ($d^2\Phi(x,y,z)/dxdy$) thus it is proportional to the Laplacian of projected electron density in the object. Consequently, only large variations in projected electron density i.e. large phase gradients ("sharp" interfaces) will significantly contribute to phase-shift-related image contrast. It this case the phase contrast is practically independent of energy.

3. Figure-of-merit for in-line holography phase-contrast imaging with cone-beam x-ray source

For laboratory-based applications of in-line holography only compact size x-ray sources need to be considered. They include microfocal x-ray tube and ultrafast-laser based plasma x-ray source (Kieffer et al., 2002). Such sources work in cone-beam geometry (Fig. 1). Hereafter, we will discuss optimization of geometrical parameters of such systems.

In order to optimize image quality in phase-contrast imaging we need to maximize the signal-to-noise ratio (SNR_A) for a given feature (A) in the object at a fixed absorbed dose (Krol et al., 2007). We introduce the following figure–of–merit that has to be maximized:

$$FOM_A = SNR_A|_{Dose=const} \quad (4)$$

One can define the image signal (S_A) as an absolute difference between the measured x-ray intensities in the image plane with and without the feature of interest, integrated over a selected region (A) in the image plane.

$$S_A = \int_A dxdy \left| \int_0^\infty dv \left[F_{obj}(x,y;v) - F_{bg}(x,y;v) \right] \eta(v) \right| \quad (5)$$

where A is the region in the image plane for which the signal is measured, F_{obj} and F_{bg} are the x-ray fluences (photons per unit area) per unit bandwidth in the image plane with and without the feature of interest, respectively, (x,y) are the Cartesian coordinates in the image plane, ν is the x-ray frequency ($\nu = c/\lambda$), where c is the speed of light and λ is the x-ray wavelength) and $\eta(\nu)$ is the quantum efficiency of the detector. We can rewrite (5) in a form

$$S_A = \iint_A dxdy \left| I_{obj}(x,y) - I_{bg}(x,y) \right| \quad (6)$$

where $I_{\mathrm{obj}}(x,y)=\int_0^\infty \mathrm{d}\nu F_{\mathrm{obj}}(x,y;\nu)\eta$ is the time-integrated intensity of the image and $I_{\mathrm{bg}}(x,y)=\int_0^\infty \mathrm{d}\nu F_{\mathrm{bg}}(x,y;\nu)\eta(\nu)$ is the background intensity. They are defined as detector counts per unit area. Therefore, it is a dimensionless quantity.
Because x-ray generation follows Poisson statistics, the noise can be expressed as

$$N_A = T_A^{1/2} \tag{7}$$

where $T_A=\int_A dxdy\left[I_{obj}(x,y)+I_{bg}(x,y)\right]$ is the total number of detector counts acquired in region *A*.
Consequently, the recorded signal-to-noise ratio is

$$SNR_A \equiv \frac{S_A}{N_A}=\frac{S_A}{T_A^{1/2}} \tag{8}$$

The region *A* over which the signal is measured is defined differently for the phase-shift- than for the absorption-related contrast. In the case of signal due to absorption, S_A^a, the natural choice of the region *A* is the whole image of the feature, *A*=Ω, while in the case of the near-field Fresnel region in-line phase-contrast signal, S_A^φ, the natural choice is a strip Γ around the image of the boundary of the feature *A* with its width proportional to the width of the first Fresnel fringe (Fig. 2). For this reason, we modify definition of the SNR_A to make it suitable to describe in-line phase-contrast imaging:

$$SNR_A=\sqrt{\left(\frac{S_\Gamma^\phi}{N_\Gamma}\right)^2+\left(\frac{S_\Omega^a}{N_\Omega}\right)^2} \tag{9}$$

Our goal is to maximize the SNR_A for a given feature *A* in the object at a fixed absorbed radiation dose. The absorbed dose is uniquely determined by the object radiological properties and the spectral distribution of the incident x-ray beam. Therefore, the incident fluence $F_{\mathrm{in}}(x',y';\nu)\equiv F_{\mathrm{in}}(\nu)$ is uniquely fixed by the dose. For a given x- ray spectrum we can only vary geometrical parameters of the imaging setup for the purpose of *SNR* maximization. In order to keep the incident fluence constant when the source-to-object distance R_1 is varied, the exposure time or the x-ray source intensity needs to be varied appropriately. Under assumption that in the vicinity of the feature of interest the incident x-ray fluence $F_{\mathrm{in}}(x',y';\nu)$ and the projected x-ray absorption coefficient in the object $E_{\mathrm{obj}}(x',y';\nu)$ are slowly varying as functions of transverse coordinates (x',y'). and the x-ray frequency ν, and at a fixed incident fluence, the signal and noise corresponding to x-ray absorption will be practically independent of imaging geometry as the decrease in photon fluence in the image plane as a function of the object-to-detector distance R_2 will be compensated by the corresponding change in system magnification resulting in increase of the projected feature area. Consequently, only SNR_A^φ i.e. SNR_A corresponding to the phase –shift contribution to image contrast needs to be maximized,

$$SNR_A^\varphi=\frac{S_\Gamma^\phi}{N_\Gamma}, \tag{10}$$

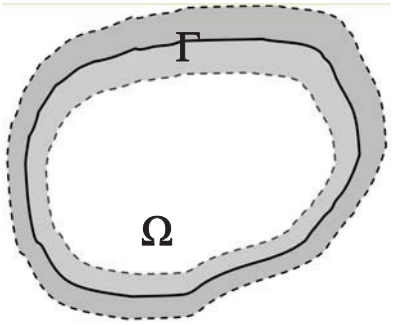


Fig. 2. Schematic depiction of the feature of interest *A:* Ω is the region inside the solid line (image of the feature of interest); Γ is the region between the dashed and solid lines (a strip around the boundary of the feature of interest).

We will examine SNR_A^{φ} characteristics as a function of geometrical parameters of the imaging setup and we will show that it exhibits more complex behavior, as compared to the absorption SNR_A^{a}. Following the assumption that the incident fluence and the object absorption properties are slowly varying in the vicinity of the feature of interest, one can also assume that the background intensity is also as slowly varying function of transverse coordinates in the vicinity of the image of the feature. Therefore, $I_{\mathrm{bg}}(x,y) \equiv I_{\mathrm{bg}}$ and $\iint_{\Gamma} \mathrm{d}x\mathrm{d}y I_{bg}(x,y) = |\Gamma| I_{bg}$, where $|\Gamma|$ is the area of the Fresnel fringe strip Γ. Because the phase shift does not change the integrated intensity of the image, one can make use of the approximation $\iint_{\Gamma} \mathrm{d}x\mathrm{d}y I_{\mathrm{obj}}(x,y) \cong |\Gamma| I_{\mathrm{bg}}$. We recall that the background intensity I_{bg} decreases as the inverse of the square of magnification, $I_{\mathrm{bg}} \sim M^{-2} I_{\mathrm{in}}$, while the intensity incident on the object (I_{in}) is fixed by the dose. Substituting these expressions into (4), we obtain an expression for the Figure of Merit (FOM) that has to be maximized:

$$FOM = \frac{S_{\Gamma}^{\varphi} M}{\sqrt{|\Gamma| I_{in}}} = \gamma S_{\Gamma}^{\varphi} |\Gamma|^{-1/2} M, \tag{11}$$

where $\gamma \sim I_{\mathrm{in}}^{-1/2}$ is a constant that depends on the object and the x-ray spectrum (because for different incident x-ray spectra the same dose may correspond to different incident intensities), but not on the geometric characteristics of the imaging setup. For the in-line holography imaging system

$$\Gamma \cong 2M\sigma_M, \text{ and } \sigma_M^2 \equiv \sigma_{\mathrm{obj}}^2 + \sigma_{\mathrm{sys}}^2(M) = \sigma_{\mathrm{obj}}^2 + (M-1)^2 M^{-2}\sigma_{\mathrm{src}}^2 + M^{-2}\sigma_{\mathrm{det}}^2, \tag{12}$$

where σ_{obj} describes the "unsharpness" of the object boundary, σ_{src} and σ_{det} are the standard deviations of the source intensity distribution and the detector point-spread-function (PSF), respectively, and *M* is the geometric magnification. Because the strip Γ is defined by the first Fresnel fringes at both sides of the boundary, its area is

$$|\Gamma| \cong 4M^2\sigma_M L = 4M^2 L\left[\sigma_{\mathrm{obj}}^2 + \sigma_{\mathrm{sys}}^2(M) = \sigma_{\mathrm{obj}}^2 + (M-1)^2 M^{-2}\sigma_{\mathrm{src}}^2 + M^{-2}\sigma_{\mathrm{det}}^2\right], \tag{13}$$

where L is the length of the boundary of the feature of interest. Let us introduce the incident intensity-independent average phase-shift image contrast $\overline{C}_\Gamma^\varphi$ over Γ

$$\overline{C}_\Gamma^\varphi \equiv S_\Gamma^\varphi / T_\Gamma, \tag{14}$$

where $T_\Gamma \equiv T_{A=\Gamma}$ is the total number of detector counts in the strip Γ.
By substituting (13) and (14) into (11) one obtains:

$$FOM = \gamma' \, \sigma_M^{-1/2} \, S_\Gamma^\varphi = \gamma'' \, \sigma_M^{1/2} \, \overline{C}_\Gamma^\varphi, \tag{15}$$

where $\gamma' \sim (LI_{\text{in}})^{-1/2}$ is a constant that depends on the object feature and the x-ray spectrum, but not on the geometric parameters of the imaging system, and $\gamma'' \equiv \gamma' T_\Gamma \sim (LI_{\text{in}})^{1/2}$ is another constant. Equation (8) allows practical optimization of geometrical parameters for the in-line holography with cone-beam x-ray source.
A universal analytical expression for the phase-contrast signal, S_Γ^φ and for the phase-shift contrast $\overline{C}_\Gamma^\varphi$, is difficult to derive because it depends on a number of factors including the shape and sharpness of the boundary of Ω. In general, these two parameters depend on of geometric parameters of the imaging setup.
Hereafter, we will optimize the FOM defined by (11) for some simple objects.

4. Examples of applications

4.1 In-line holography imaging of nylon fibers: system geometry optimization

An example of optimization of in-line holography setup for specific application – nylon fibers imaging – can be found in Krol et al., (2007). An image of Nylon fibers (10–330 μm diameter) obtained with ultrafast laser operating at 100 Hz, 28 fs and 40 mJ perpulse and x-ray focal spot size 10 μm (Advanced Laser Light Source [ALLS] Laboratory, Institut National de la Recherché Scientifique, University of Quebec, Canada), with Mo targets and Be filter are shown in Fig. 3 Projection images of nylon fibers of 10–330 μm diameter were obtained in in-line holography setup using an ultrafast laser-based x-ray source. [Krol et. al, 2007]

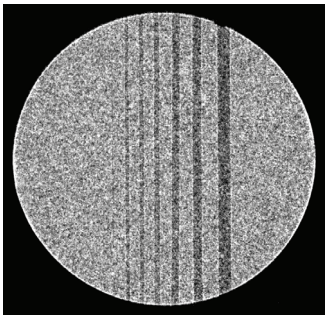
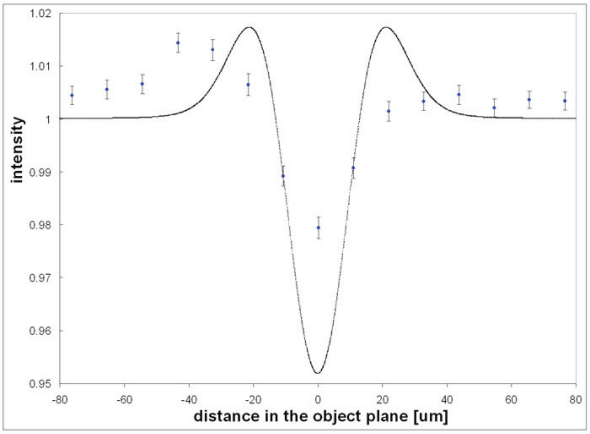
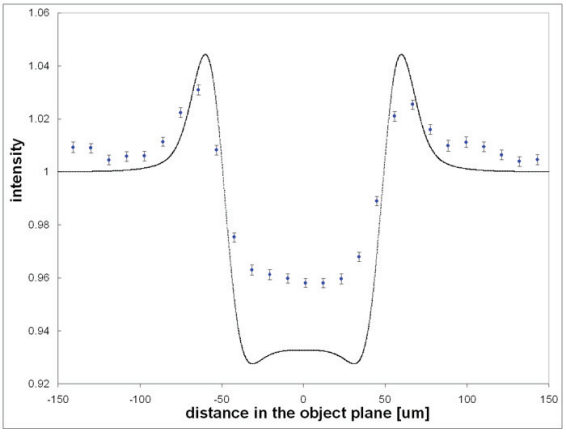


Fig. 3. Normalized image of nylon fiber phantom acquired at ALLS laboratory (INRS) with 100 Hz ultrafast laser-based x-ray source with Mo target and Be filter. Source-to-object distance, R_1 = 40.7 cm; source-to-detector distance R_2 = 48.9 cm; magnification M= 2.201. A cooled CCD camera (24 μm pitch) was used with a Gd_2OS_2 screen coupled via 1:1 optical taper.

Comparison of experimental line profiles through images of Nylon fibers shown in Fig. 3 with simulations are shown in Fig. 4 (Nesterets et al., 2005). We observe satisfactory agreement between experiment and theory. It should be stressed that no curve fitting was performed, nor was any "fudge factor" used. These results confirmed applicability of the theoretical model for Nylon fibers in-line holography phase-contrast imaging simulations.



(a)



(b)

Fig. 4. Comparison of experimental line profiles through the images of Nylon fibers obtained with 100 Hz ultrafast laser at 50% power level, at ALLS-INRS lab. The solid lines represent the theoretical calculations. The discrete data points are taken from experiments using a CCD detector (24 μm pitch) with Gd_2OS_2 screen. (a) Nylon fiber with 20 μm diameter, Mo target, Be filter, R_1 = 40.7 cm, R_2 = 48.9 cm. (b) Nylon fiber with 100 μm diameter, Mo target, Be filter, R_1 = 40.7 cm, R_2 = 48.9 cm.

The rigorous wave-optical formalism was used to estimate S_Γ^φ and $\overline{C}_\Gamma^\varphi$ in (8). The following four parameters were estimated:

Contrast:

$$Contrast = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \tag{16}$$

where I_{max} and I_{min} are the adjacent minimum and maximum of the intensity due to phase shift in the first Fresnel fringe image of the object.
Signal-to-noise ratio:

$$SNR = \frac{\int_{\Omega} \left| I_{obj} - I_{bg} \right| dS}{\left(\int_{\Omega} \left[I_{obj} + I_{bg} \right] dS \right)^{1/2}} \tag{17}$$

where the region Ω in the image is chosen such that it contains 90% of the total signal $\int_{\Omega} \left| I_{obj} - I_{bg} \right| dS$ and I_{obj} and I_{bg} are fluencies trough the object and the background at the image plane, respectively.
Resolution is defined as the distance between the adjacent minimum and maximum of the intensity in the image of the object due to phase shift in the first Fresnel fringe and divided by the magnification *M*.
Sampling is defined as the distance (d_F, in pixels of the detector) between the adjacent minimum and maximum of the intensity due to phase shift in the first Fresnel fringe image of the object.
We have investigated the influence of the geometrical parameters of the imaging setup on the characteristics of the phase-contrast images of Nylon fibers (Krol et al., 2007). We assumed Gaussian spatial intensity distribution of the source with $FWHM_S$ = 5 μm and Gaussian resolution function of the detector with $FWHM_D$ = 50 μm. We calculated dependences of the *Contrast*, *SNR*, *Resolution* and *Sampling* on the magnification *M* assuming the total source-to-detector distance to be fixed at *R* = 2 m (Figs. 5 and 6).

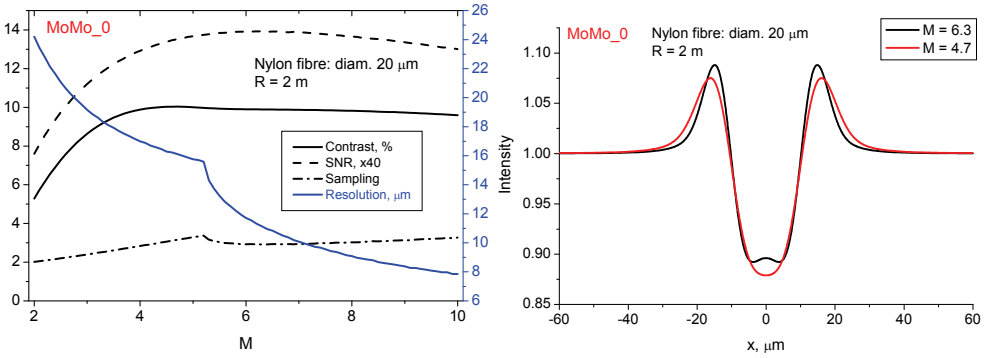


Fig. 5. Left panel: Calculated dependences of the Contrast, SNR, Resolution and Sampling on the magnification *M* at the total source-to-detector distance fixed at R = 2 m. Right panel: Calculated line profiles at magnification M=6.3 and M=4.7. ULX spectrum is obtained with Mo target and Be filter only.

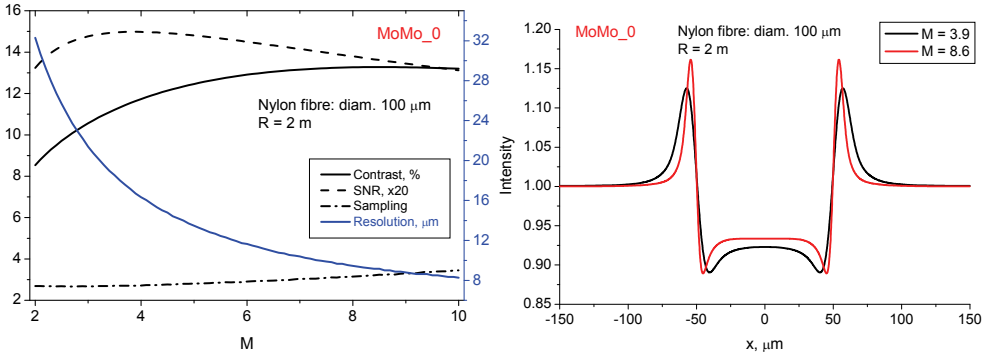


Fig. 6. Left panel: Calculated dependences of the Contrast, SNR, Resolution and Sampling on the magnification M at the total source-to-detector distance fixed at R = 2 m. Right panel: Calculated line profiles at magnification M=3.9 and M=8.6. Ultrafast laser-based x-ray source spectrum is obtained with Mo target and Be filter.

Analysis of Figs. 5 and 6 indicates that contrast and *SNR* vs. magnification exhibit broad maxima and that variation of M_C with fiber diameter is relatively small (7.7 – 8.6 for 10 –100 μm fiber diameter, respectively) while M_{SNR} changes significantly (7.7 – 3.9 for 10 –100 μm fiber diameter, respectively). The maximum contrast is almost independent of the fiber diameter (10.7% –13.2% for the 10 – 100 μm fiber diameter, respectively), while the maximum of *SNR* rapidly increases with increase of the fiber diameter (0.2461 –0.7492 for the 10 – 100 μm fiber diameter, respectively).

Let $R_1 = R/M$ be the source-to-object distance with R_1 and R_2 variable and $R=R_1+R_2=const$. We wish to optimize FOM defined in (4), In order to keep the dose constant when R_1 is varied, the exposure time or the x-ray source intensity has to be varied accordingly.

We conclude that:

i. The optimum magnification (which maximizes the *FOM*) is almost independent of the total source-to-detector distance, R, and depends only on the properties of the imaged object, the source size and the detector resolution.
ii. The corresponding maximum values of the contrast and *SNR* are almost linear with respect to R
iii. The optimum magnification decreases with fiber diameter;
iv. The minimum fiber diameter defines the minimum source-to-object distance R_1, if R is fixed and the object is moved.

These conclusions allow us to propose the following strategy for choosing the optimum geometry for phase-contrast imaging: First, for a given object feature type one should calculate the optimum magnification M^{opt} that maximizes the FOM. Second, assuming that R is fixed, calculate the optimum source-to-object distance as $R_1^{opt} = R/M^{opt}$.

Similar calculations were performed for in-line holography system with fixed R_1 and variable R_2 ($R=R_1+R_2$), where R is the total source-to-detector distance. The results show that M_C is almost independent of the R while M_{SNR} rapidly increases with R. Both parameters strongly depend on the radius of a fiber. For fixed R_1 and R_2 they increase with the fiber diameter.

Investigation of the influence of x-ray spectrum mean energy on *SNR* and contrast indicate that these two parameters decrease with increasing mean x-ray energy.

4.2 Soft tissue small avascular tumor imaging with x-ray phase-contrast micro-CT in-line holography setup

In-line holography phase-contrast imaging might allow detection and quantification of soft tissue tumors that cannot be imaged using conventional absorption radiography. We analyzed feasibility of small soft tissue avascular tumor in mouse model imaging with x-ray phase-contrast using in-line holography setup and compared its performance with absorption-based imaging (Nesterets et al., 2008). We simulated imaging of a small spheroidal avascular invading tumor in mouse model in an early stage of development, i.e. not containing a necrotic core (Jiang et al., 2005, Stein et al., 2007) with quiescent cell core (smaller than 250 μm) and a range of distributions of the proliferating cell density forming the outer shell.

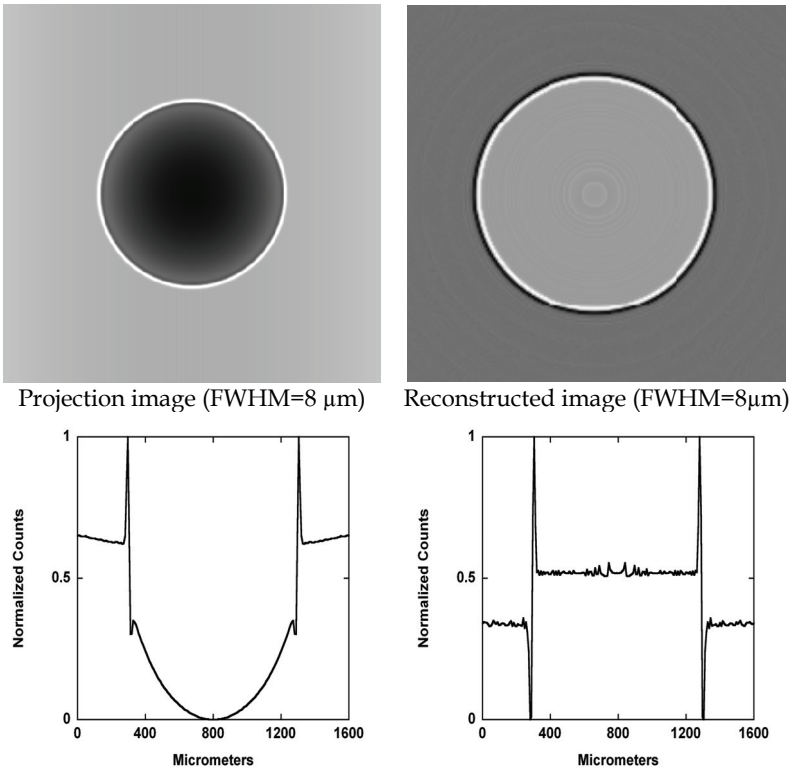


Fig. 7. Simulated projection images (M=2) of tumor model using in-line holography geometry. Proliferating cell density distribution FWHM=8μm.

This outer layer was modeled by various degree of Gaussian blurring of density profile across its interface with the surrounding tissue (FWHM in the 8 – 100 μm range). A mouse was approximated by a cylinder 28 mm in diameter and 30 mm in length. The tumor was centrally located and its core had the same radiological properties as the soft tissue and it was embedded in the adipose tissue. The projection approximation was used in simulations i.e. $E_{out}(x,y)$ of the wave transmitted through the object was related to the amplitude $E_{in}(x,y)$ of the wave incident on the object by $E_{out}(x,y) = \exp(i\varphi(x,y) - a(x,y))\, E_{in}(x,y)$, where $\varphi(x,y)$ and

$a(x,y)$ are the phase and the attenuation induced by the object, respectively. The free-space propagation from the object plane to the detector was modeled using the rigorous wave-optical formalism (Nesterets et al., 2005). For simplicity, a point source and a perfect detector ($\sigma_S = \sigma_D = 0$) were assumed with $M = 1.1$ and $M = M_{opt} = 2$ (the optimum magnification corresponds to the point source and perfect detector), $R = 2$ m (total source-to-detector distance), and 20 keV x-ray energy. We simulated sets of micro-CT projection images, each consisting of 1440 images and reconstructed them using a modified Feldkamp algorithm (Figs. 7-9).

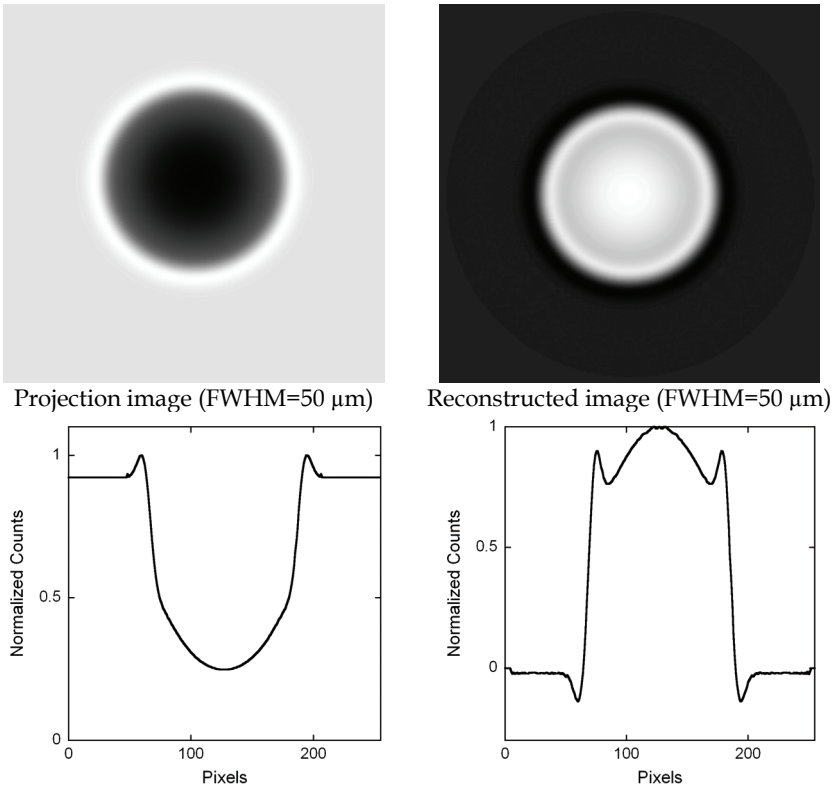


Fig. 8. Simulated projection images (M=2) of tumor model using in-line holography geometry. Proliferating cell density distribution FWHM=50 μm.

In addition, we estimated absorption and phase-shift image contrast in the projection images of simulated tumor vs. their diameter for different density profiles across its interface with the surrounding tissue (Fig. 10).

We observe phase-shift enhancement of the image contrast at the boundaries of the tumor in the projection images and even stronger enhancement in the reconstructed transaxial images. The enhancement strongly depends on the projected real part of the refractive index gradient at the interface between the tumor and the surrounding tissue. It is very pronounced at FWHM=8 μm, moderate at FWHM=50 μm and difficult to detect at

FWHM=100 μm. At large values of the tumor radii the absorption contrast dominates the phase contrast and the image contrast is therefore independent of the tumor/normal tissue interface blurring. For small tumors (r <1 mm) the phase contrast dominates the absorption contrast but it rapidly decreases with increase of the tumor/normal tissue interface blurring.

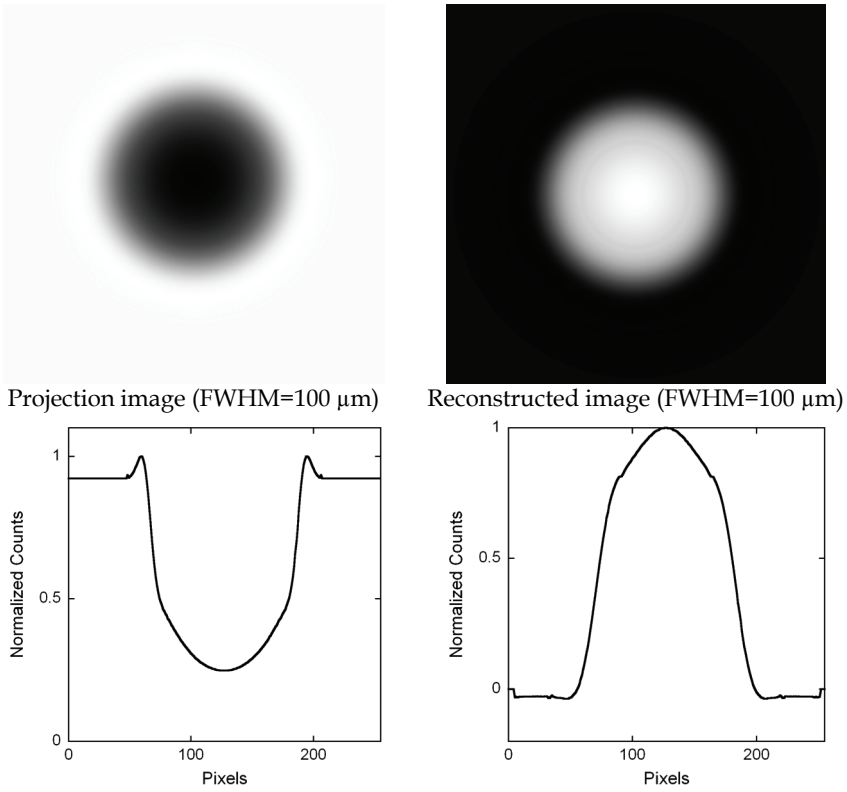


Fig. 9. Simulated projection images (M=2) of tumor model using in-line holography geometry. Proliferating cell density distribution FWHM=100 μm.

5. Conclusions

The in-line holography setup for phase-contrast imaging is the simplest and the easiest to implement in the context of biomedical imaging. It does not require high-brilliance x-ray source (e.g. synchrotron) or sophisticated x-ray diffraction crystals or gratings. It does not necessitate narrow x-ray energy band. The field-of-view is limited only by the x-ray detector size but not by the aperture of x-ray optics. The in-line holography is very well suited to detect steep gradients (interfaces) of the real part of the x-ray refraction index inside the object, as well as variation of the imaginary part i.e. x-ray absorption. Consequently, it might allow visualization of internal interfaces, e.g. between normal tissue and tumor (providing that they create sufficient x-ray phase shift) that otherwise do not produce sufficient absorption contrast to be detected in conventional radiological imaging.

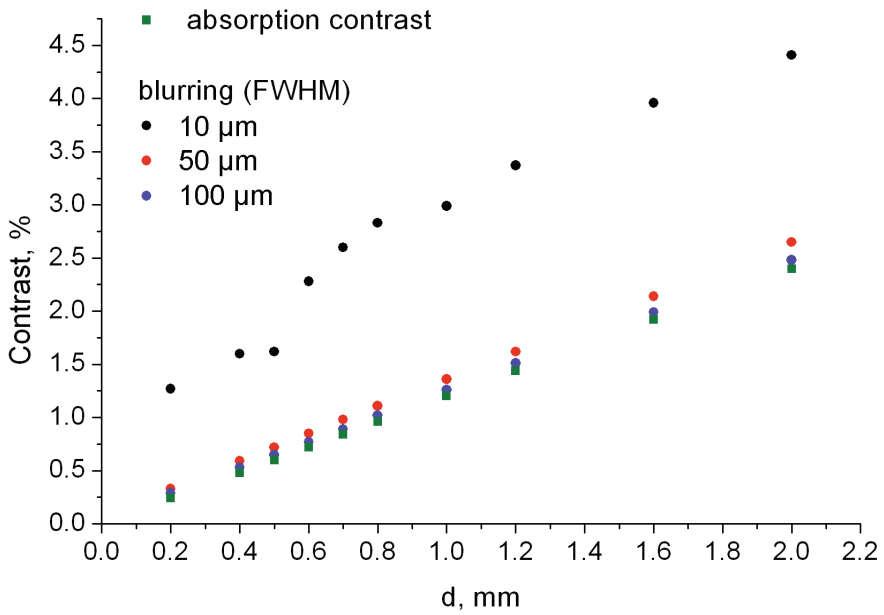


Fig. 10. In-line holography image contrast vs. tumor diameter d for proliferating cells for different density profiles across its interface with the surrounding tissue FWHM = 10 μm, 50 μm and 100 μm.

The in-line holography setup requirements include:

i. Sufficiently small x-ray source size to provide required spatial coherence to image the object's component of interest. This condition is met if the shortest spatial wavelength (λ_o) of the feature of interest fulfils the following inequality: $\lambda_o >> s(M-1)/M = s'/M$, where s' is the size of projected source (s) on the detector plane geometrical unsharpness of the system) and M is system magnification.

ii. The x-ray source intensity should match the imaging task, i.e. the x-ray fluence on the sample needs to create minimum acceptable *SNR* and contrast for the imaging task within acceptable exposure time.

iii. The detector sampling (in pixels of the detector) needs to be commensurate with the distance (d_F) between the adjacent minimum and maximum of the intensity due to phase shift in the first Fresnel fringe image of the object. Consequently, one needs at minimum 5 detector pixels to adequately sample the distance d_F.

The above requirements translate into the focal spot size in the 5 –50 μm range depending on the geometrical parameters of the system. The necessary average power of x-ray source depends strongly on the application: 20–100 W for micro-CT, 500–1000 W for breast or chest imaging, and 10–40 kW for whole body CT. Presently, compact x-ray sources that could fulfill some of these conditions include microfocal x-ray tube and ultrafast laser-based x-ray source. However, micro-focal x-ray tube suffers from maximum power limit imposed by the required small focal spot size and could be used for micro imaging only. It also produces rather broad x-ray spectra that are difficult to tune. Ultrafast lasers on the other hand, are

fast evolving technology and it is expected that they will reach kilowatt average power level within the next ten years. Presently, they could be used for micro (preclinical) imaging. The x-ray spectra produced by ultrafast laser-based x-ray source can be easily tuned by change of the target elemental composition and could be optimized for the specific imaging task. The required pixel size of the detector is in the 5 – 30 μm range, depending on the system parameters. The ongoing progress in the field of radiological x-ray detectors will lead to creation of large-area high-resolution, low-noise, photon-counting detectors that would match requirements of clinical in-line holography x-ray imaging systems.

6. References

Als-Nielsen, J. and McMorrow, D. (2001). *Elements of Modern X-Ray Physics*, Chichester: Wiley.

Bilderback, D.H,; Elleaume, P.; Weckert E. (2005). Review of third and next generation synchrotron light sources. *J Phys B At Mol Opt Phys*, 38, pp S773-97.

Chapman, D.; Thomlinson, W.; Johnston, R. E., Washburn, D.; Pisano, E.; Gmur, N.; Zhong, Z.; Menk, R.; Arfelli, F.; Sayers, D. (1997). *Phys. Med. Biol.,* 42, pp 2015–25.

Grubel, G.; Stephenson, G.B.; Gutt, C.; Sinn, H.; Tschentscher, T. (2007). XPCS at the European X-ray free electron laser facility, *Nucl Instrum Methods Phys Res,* B262, pp. 357-67.

Gureyev, T.E. & Wilkins, S.W. (1998). On X-ray phase retrieval from polychromatic images, *Optics Communications*, 147, pp. 229-232.

Kieffer, J.C.; Krol, A.; Jiang, Z.; Chamberlain, C.C.; Scalzetti, E.; Ichalalene, Z. (2002). Future of Laser-Based X-Ray Sources for Medical Imaging", *J. Appl. Phys. B*, 74, pp. S75-S81.

Jiang, Y.; Pjesivac-Grbovic, J.; Cantrell, C.; Freyer, J.P. (2005). A Multiscale Model for Avascular Tumor Growth, *Biophysical Journal*, 89, pp. 3884–3894.

Krol, A.; Kincaid, R.; Servol, M.; Kieffer, J.C.; Nesterets, Y.; Gureyev, T.; Stevenson, A.; Wilkins, S.; Ye, H.; Lipson, E.; Toth, R.; Pogany, A.; Coman, I. (2007). Initial experimentation with in-line holography x-ray phase-contrast imaging with an ultrafast laser-based x-ray source, *Proc SPIE,* 6510, pp. 65100L.

Lewis, R.A. (2004). Medical phase contrast X-ray imaging: current status and future prospects. *Phys Med Biol,* 49, pp. 3573-83.

Lohmann, A. W; & Silva, D. E. (1971). Opt. Commun. 2, pp. 413.

Momose, A. (1995). Demonstration of phase-contrast computed tomography using an X-ray interferometer, *Nucl Instrum Methods* A352:622-8.

Momose, A.; Takeda, T.; Itai, Y.; Hirano, K. (1996). Phase-contrast x-ray computed tomography for observing biological soft tissues, *Nat. Med.*, 2, pp. 473–5.

Momose, A. & Fukuda, J. (1995). *Med. Phys.*, 22, pp. 375.

Momose, A.; Takeda, T,; Itai, Y. et al. (1998). *J. Synchrotron Radiat.*, 5, pp. 309–14.

Momose, A. (2002). Phase-contrast X-ray imaging based on interferometry, *J Synchrotron Rad* 9, pp. 136-42.

Momose, A. (2003). Phase-sensitive imaging and phase tomography using X-ray interferometers, *Opt Expr*, 11(19) pp. 2303-14.

Momose, A.; Yashiro, W.; Takeda, Y.; Suzuki, Y.; Hattori, T. (2006). Phase tomography by x-ray Talbot interferometry for biological Imaging, *Japan. J. Appl. Phys*., 45, pp. 5254–62.

Nesterets, Y.I.; Gureyev, T.E.; Paganin, D.; Pavlov, K.M.; Wilkins, S.W. (2004). Quantitative diffraction-enhanced X-ray imaging of weak objects, *J Phys D Appl Phys*, 37, pp. 1262-74.

Nesterets, Y.I.; Wilkins, S.W.; Gureyev, T.E.; Pogany, A.; Stevenson, A.W. (2005). On the optimization of experimental parameters for x-ray in- phase-contrast imaging, *Rev. Sci. Instr.,* 76, 093706.

Nesterets,Y.; Gureyev, T.; Stevenson, A.; Pogany,A.; Wilkins, S.; Kincaid, R.; Ye, H.; Vogelsang, L.; Lipson, E.; Coman, I.; Fourmaux, S.; Kieffer, J.C.; Krol, A. (2008). Soft tissue small avascular tumor imaging with x-ray phase-contrast micro-CT in in-line holography setup, *Proceedings of SPIE*, 6913.

Nugent, K.A.; Paganin, D.; Gureyev, T. E. (2001). The phase odysse, *Physics Today*, pp. 27-32.

Olivo, A. & Speller, R. (2007). A coded-aperture technique allowing X-ray phase contrast imaging with conventional sources, *Appl Phys Lett*, 91, pp. 074106

Pfeiffer, F.; Weitkamp, T.; Bunk, O.; David, C. (2006). Phase retrieval and differential phase-contrast imaging with low-brilliance x-ray sources, *Nat. Phys.,* 2, 258–61.

Stein, M.A.; Demuth, T.; Mobley, D. et al.(2007). A Mathematical Model of Glioblastoma Tumor Spheroid Invasion in a Three-Dimensional In Vitro Experiment, *Biophysical Journal,* 92, pp. 356–365.

Pogany, A.; Gao, D.; Wilkins, S.W. (1997). Contrast and resolution in imaging with a microfocus x-ray source, *Rev. Sci. Inst*. 68, pp. 2774-2782.

Rigon, L.; Arfelli, F.; Menk, R.H. (2007). Three-image diffraction enhanced imaging algorithm to extract absorption, refraction, and ultrasmall-angle scattering. *Appl Phys Lett*, 90, pp. 114102.

Snigirev A, Snigireva I, Kohn V, Kuznetsov S, Schelokov I. On the possibilities of X-ray phase contrast microimaging by coherent high-energy synchrotron radiation. Rev Sci Instrum 1995;66:5486-92.

Takeda, T.; Wu, J.; Lwin, T. T.; et al. (2007). Interferometer-based phase-contrast X-ray computed of colon cancer specimens: comparative study with 4.74-T magnetic resonance imaging and optical microscopy, *J Comput Assist Tomogr*, 31, pp. 214-7.

Weitkamp, T.; Diaz, A.; David, C.; et al. (2005). X-ray phase imaging with a grating interferometer, *Opt. Exp*., 13, pp. 6296–304.

Weitkamp, T; David, C.; Kottlerb, C.; Bunkc, O.; Pfeiffer, F. (2006). Tomography with grating interferometers at low-brilliance sources, *Proc. SPIE,* 6318.

Wernick, M.N.; Wirjadi, O., Chapman D, et al.(2003). Multiple-image radiography. *Phys Med Biol,* 48, pp. 3875-95.

Wilkins, S.W.; Gureyev, T.E.; Gao, D.; Pogany, A.; Stevenson, A.W. (1996). Phase-contrast imaging using polychromatic hard X-rays. *Nature*, 384, pp. 335-8.

Yokozeki, S. & Suzuki, T. (1971). *Appl. Opt.*, 10, pp. 1575.

Wu, X.; & Liu, H. (2007). Clarification of aspects in in-line phase-sensitive, X-ray imaging., *Med Phys*, 34(2), pp. 737-43.

Zhang, X.; Liu, X.S.; Yang, X.R. et al. (2008). *Phys. Med. Biol.*, 53, pp. 5735–5743

Zhou, S.; & Brahme, A., (2008). Development of phase-contrast X-ray imaging techniques and potential medical applications, *Physica Medica,* 24, pp. 129-148.

22

Fundamentals and Applications of Electron Holography

Akira Tonomura
Advanced Research Laboratory, Hitachi, Ltd.
Advanced Science Institute, RIKEN
Japan

1. Introduction

Electron holography is a two-step imaging method devised by D. Gabor (Gabor, 1949): an electron hologram is formed as an interference pattern between an object wave and a reference wave, and the object electron wave is reconstructed as an optical wave by illuminating a laser beam onto hologram film. The reconstruction has been facilitated by advances in digital reconstruction techniques. The reconstructed optical wave is similar to the original object electron wave except for the scale ratio of the electron and light wavelengths, and it represents the "wavefunction" of the object electron wave.
Gabor intended to optically compensate for the aberrations in the electron lens system, a factor limiting the resolution of electron microscopes. Thanks to the development of bright and yet coherent field-emission electron beams, electron holography proved to be useful not only for high resolution microscopy but also for phase microscopy at the level of atoms and molecules (Tonomura, 2005). The wavefronts of electrons transmitted through an object contain information about the object's shape and size and also about the electromagnetic fields inside and outside the object. Furthermore, since the wavelengths of fast electrons are extremely short, such as 0.009 Å for 1-MV electrons, the wavefronts can be observed with high spatial resolution.
The interaction of an electron wave with electromagnetic fields is described by the Aharonov-Bohm (AB) effect (Aharonov & Bohm, 1959). The AB effect principle has been used to observe not only quantum phenomena but also the microscopic distributions of electromagnetic fields by measuring the phases of electrons. The behavior of vortices in superconductors has also been observed by Lorentz microscopy as well as by interference electron microscopy.
This chapter describes fundamental experiments in quantum mechanics and the application results of electromagnetic field observation using electron holography. These experiments include the observation of magnetic heads for perpendicular recording, of Josephson vortices in YBCO thin film, and of the nucleation of the ferromagnetic phase assisted by magnetic fields in colossal magnetoresistance.

2. Interaction of electrons with electromagnetic fields – the Aharonov-Bhom effect –

In electron holography, electron wavefronts, or "electron wavefuntions", are reconstructed. Therefore, we first discuss the interaction of incident electron waves with electromagnetic

fields. In classical mechanics, electrons passing through electromagnetic fields are deflected by Lorentz forces. In quantum mechanics, electrons propagate as waves, and their behavior is explained by the Schrödinger equation. In that realm, the concept of force is no longer relevant: electric field $\boldsymbol{E}$ and magnetic field $\boldsymbol{B}$, defined as forces acting on a unit charge, take on a secondary meaning. Instead of forces, "phase shifts" come into play. The primary physical entities are neither $\boldsymbol{E}$ nor $\boldsymbol{B}$ but electrostatic potential V and vector potentials $\boldsymbol{A}$ since phase shifts are produced by these potential, as described by the Schrödinger equation. In 1959, Y. Aharonov and D. Bohm theoretically predicted that a relative phase shift is produced by the potentials and can be observed as a displacement of interference fringes even when electrons pass through spaces free of $\boldsymbol{E}$ and $\boldsymbol{B}$. This effect was later dubbed "the Aharonov-Bohm (AB) effect".

The AB effect increased in significance in the late 1970s in relation to the theories attempting to unify all fundamental interactions in nature, in which potentials are extended to "gauge fields" and regarded as the most fundamental physical quantity. T. T. Wu and C. N. Yang (Wu & Yang, 1975) stressed the significance of the AB effect in relation to the physical reality of gauge fields.

Soon after the theoretical prediction of the AB effect, experimental testing was carried out. The first experiment was conducted by R. G. Chambers (Chambers, 1960) in 1960 (Fig. 1(a)). He placed a tapered iron whisker in the shadow of a biprism filament and detected a phase difference between two electron waves passing on both sides of the biprism though magnetic fields leaking out from the whisker and thus influencing incident electrons. The value of the phase difference was determined by the magnetic flux enclosed by the two waves, exactly as predicted by the AB effect. Other experiments followed: Fowler *et al*. (Fowler *et al*., 1961) used a uniform iron whisker as a biprism filament (Fig. 1(b)). Boersch *et al*. (Boersch *et al*., 1962) used a permalloy thin film evaporated on the bottom of a biprism filament (Fig. 1(c)). Möllenstedt and Bayh (Möllenstedt & Bayh, 1962) made a beautiful experiment using a tiny solenoid with a diameter as small as 1 μm.

They all detected the phase difference predicted by the AB effect. However, vector potentials have long been regarded as mathematical auxiliaries, so some people questioned the existence of the AB effect, thus causing a long controversy (Bocchieri & Loinger, 1978; Peshkin & Tonomura, 1989). The reported results were attributed by the opponents of the AB effect to the effect of magnetic fields leaking out from both ends of the finite solenoids or ferromagnets used in the experiments (Bocchieri & Loinger, 1981; Roy, 1980).

3. Conclusive experiments for the AB effect

An infinitely long magnet cannot be obtained experimentally, but an ideal geometry can be achieved in a finite size by using a toroidal magnet.

We first fabricated a leakage-free tiny magnet made of permalloy with a square-toroid geometry (Tonomura *et al*., 1982) (Fig. 2) and detected a non-zero phase difference between two electron waves passing inside the hole and outside the toroid (Fig. 2(b)). However, this finding was not accepted by everyone, and papers were published disagreeing with our conclusion (Bocchieri *et al*., 1982; Miyazawa, 1985). They argued that, to prove the AB effect, the electrons should not touch the magnetic field. They claimed that since the detected phase difference could not be due to 'vector potentials' , magnetic fields (although slight) leaked out from even our 'square' toroidal magnets, mainly due to the existence of corners in the square sample (Tonomura, 1983) (Fig. 2(c)), so that part of the electron waves touched and even penetrated the magnet (Fig. 2(a)).

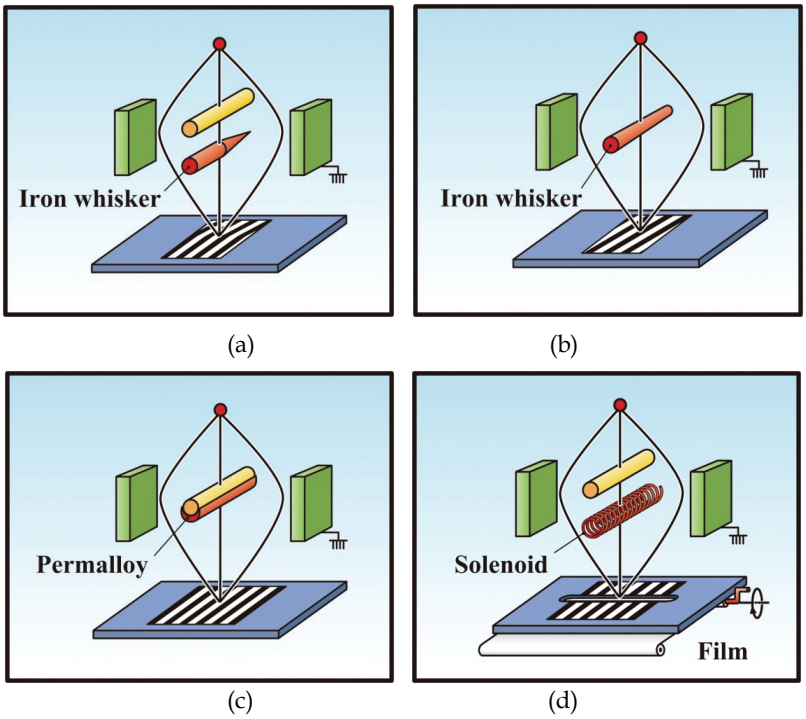


Fig. 1. Experimental arrangements for testing the AB effect in 1960s. (a) Chambers placed a tapered iron whisker in the shadow of a biprism filament, (b) Fowler *at al*. used a uniform iron whisker as a biprism filament, (c) Boersch *et al*. evaporated ferromagnetic permalloy on the bottom of a biprism filament, (d) Möllenstedt and Bayh placed a tiny solenoid in the shadow of a biprism filament. Tiny ferromagnets or solenoids with finite lengths were used for testing in these experiments. Later in the '80s, the existence of the AB effect was questioned because magnetic fluxes leaked out from both ends of the finite solenoids or magnets.

We thus carried out further experiments using a 'circular' toroidal magnet having even smaller leakage flux (Tonomura et al., 1983) (Fig. 3(a)) and again detected the AB phase difference. To present the overlap of incident electrons with the magnetic field inside the sample, we next conducted experiments using a 'circular' toroidal magnet on top of which a thick gold film was evaporated (Tonomura *et al.*, 1984), which prevented electron penetration into the magnet (Fig. 3(b)). In this experiment, however, we could not determine the absolute value of the phase difference between the two electron beams, only their fractions in units of 2π. The obtained interference pattern (Fig. 3(b)) indicates that the interference fringes inside the hole and outside the toroid were not aligned, thus providing evidence for the AB effect even when incident electrons do not penetrate the magnet.
Skeptics, however, asserted that the toroidal magnet should not only be covered on its top surface but also be fully covered on all its surfaces since the tail of the electron wavefunction curls around the sample and touches the magnetic field.

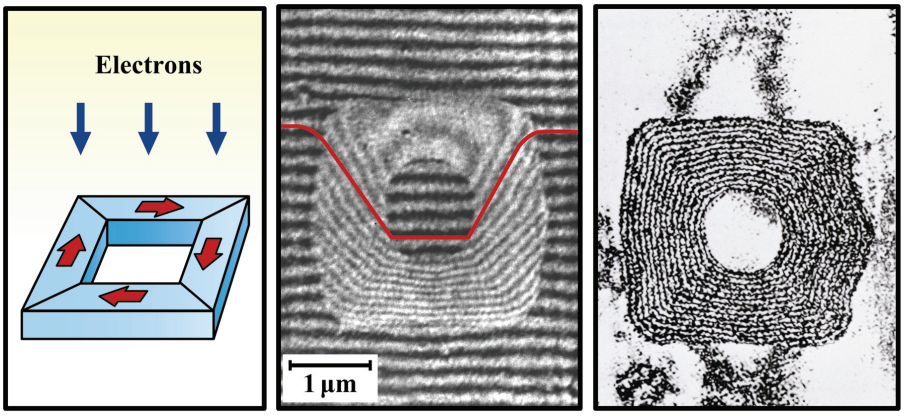


Fig. 2. Experiments using square toroidal magnet. (a) Schematic. (b) Interference pattern showing relative phase difference of $7\times(2\pi)$ between inside and outside toroid. (c) Flux leaking from toroidal magnet. Toroidal ferromagnetic samples were used to minimize leakage flux. Contour lines of transmitted electron wavefronts indicate magnetic lines in magnetic flux units of h/e, confirming that the amount of leakage flux was approximately 1/15 of total flux, $h/2e$, since the phase difference was amplified by a factor of 2.

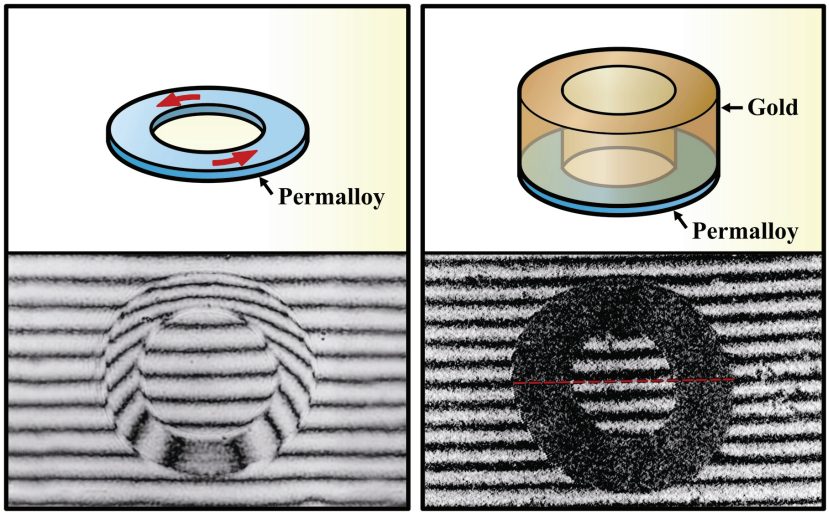


Fig. 3. Circular toroidal magnets made of permalloy. (a) Transparent toroidal magnet (permalloy film 10 nm thick). (b) Opaque toroidal magnet covered with gold film (permalloy film 10 nm thick; gold film 300 nm thick).

According this challenge, we fabricated tiny toroidal magnets completely surrounded by superconductors (Tonomura *et al.,* 1986), thereby removing any ambiguities about the leakage flux and also about the overlap of incident electrons with the magnetic field in the solenoid or magnet. A permalloy toroidal magnet was completely covered with a niobium superconductor to confine the magnetic flux within the sample (Figs. 4(a) and (b)). The AB phase difference was again detected (Fig. 4(c)) even though magnetic field leakage was prevented by the Meissner effect, and the electrons did not touch the magnetization inside the sample.

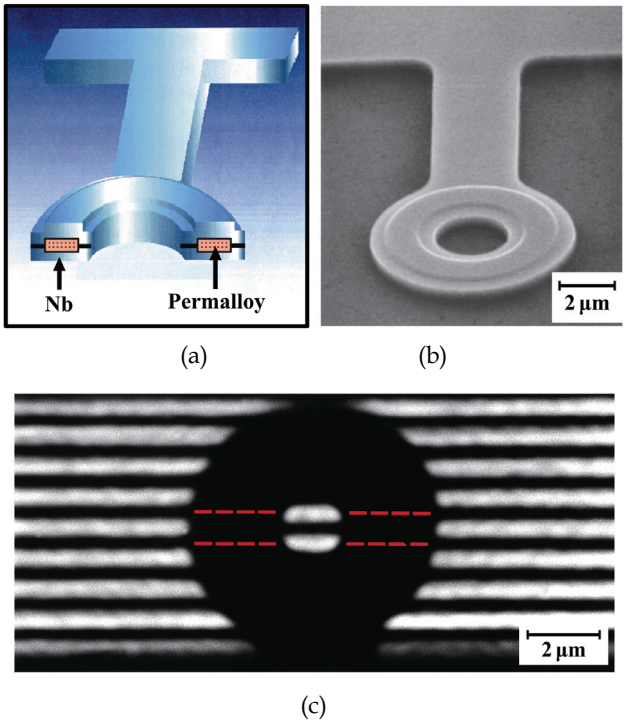


Fig. 4. Conclusive experiment proving existence of the AB effect using toroidal magnet completely covered with superconductor. (a) Schematic. (b) Scanning electron micrograph of sample. (c) Interference pattern at T = 5 K. Toroidal ferromagnetic permalloy sample was covered with Nb layers. Leakage magnetic fluxes were confirmed by electron holography to be less than 1/20 of total magnetic flux passing through permalloy magnet. When sample temperature was decreased below superconductive critical temperature of Nb 9.2 K, the phase difference between two electron beams passing through inside the hole and outside the sample was quantized in π units, either zero or π. This magnetic flux quantization in $h/2e$ units proves that the magnetic fluxes were completely confined within the sample due to the Meissner effect. The production of a phase difference of π under the condition of no overlap of electrons and magnetic fields confirmed the AB effect.

We fabricated many toroids, more than 100,000 of different sizes having different magnetic flux values thanks to advances in lithography (Fig. 5), but obtained only two phase-shift values. The detected phase difference value was either π or 0. The phase-shift quantization in π units, indicating the magnetic flux quantization in superconductors, confirmed the complete covering of the magnetic flux within the superconductor, so flux leakage was prevented by the Meissner effect. In addition, the toroidal samples were covered with additional Cu layers so that incident electrons could not penetrate them and touch the magnetic field inside. Therefore, the results shown in Fig. 4(c) decisively confirmed the AB effect.

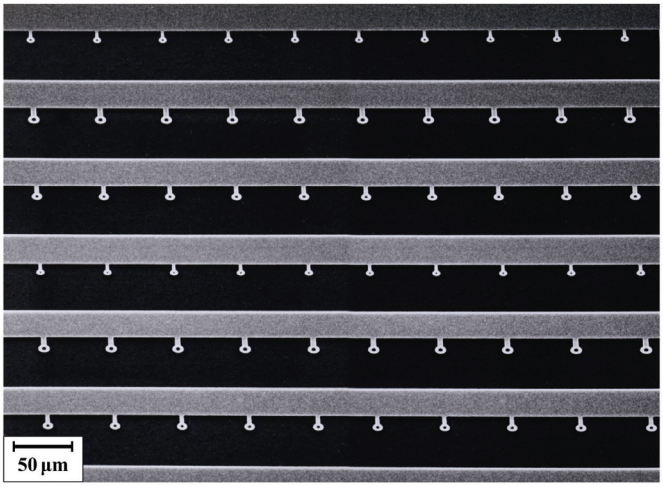


Fig. 5. Scanning electron micrograph of array of toridal magnets covered with superconductors. More than 100,000 various samples were fabricated photolithographically.

4. Applications of electron holographic interferometry

We found in 1980 that quantitative magnetic lines of force can be visualized using the AB effect principle (Tonomura *et al*., 1980). An example of a hexagonal cobalt particle is shown in Fig. 6. With electron microscopy, only its outline can be observed (Fig. 6(b)). With interference microscopy, however, circular phase contour fringes can be observed inside the particle (Fig. 6(c)). These fringes in the interference micrograph, phase-amplified two times, directly indicate quantitative magnetic lines in $h/2e$ flux units irrespective of the velocity of incident electrons. This results from the AB effect. The phase difference between two electron beams passing through two points in the specimen plane is determined by the magnetic flux enclosed by the two beam paths: a phase difference of 2π is produced by a magnetic flux of $h/e = 4.1 \times 10^{-19}$ Wb. Therefore, when two points in the specimen plane lie on a magnetic line, no phase difference is produced. That is, phase contour lines indicate magnetic lines of force. When the two electron beams enclose a magnetic flux of h/e, a phase difference of 2π is produced. That means phase contour fringes appear every h/e flux units. This observation principle is the same as that of a SQUID.

Not only magnetic fields but also electric fields can phase-shift incoming electrons. In fact, electrons entering a non-magnetic sample are accelerated by inner potential V_0 of the

materials, so their wavelength is shortened, as shown in Fig. 7(a). Therefore, the wavefronts inside the specimen are delayed proportional to the specimen thickness. The wavefront displacements cannot be observed in an electron micrograph (Fig. 7(b)), but can be observed in an interference micrograph (Fig. 7(c)) as displacements from regular fringes. The displacements indicate thickness contours in 20 nm thickness units.

An example of a triangular cobalt particle affected by both magnetic lines and sample thickness is shown in Fig. 8. A thickness contour map forms on the periphery (Fig. 8(b)) while circular magnetic lines form in the inner region since the sample thickness is uniform.

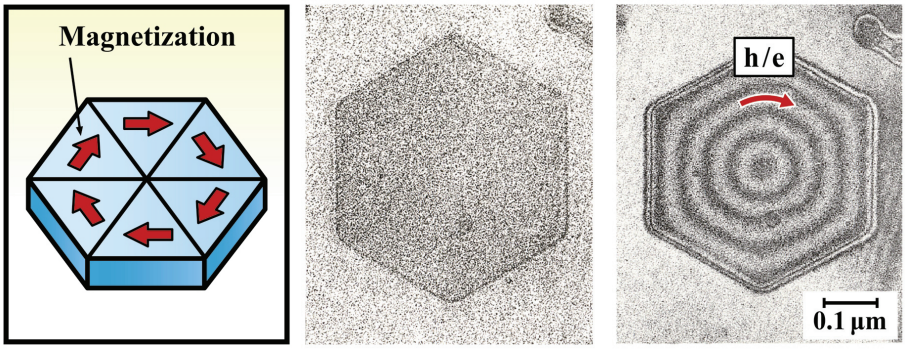


Fig. 6. Hexagonal cobalt particle. (a) Schematic. (b) Electron micrograph. (c) Interference micrograph (phase amplification × 2). Phase contours in interference micrograph (c) indicate magnetic lines in $h/2e$ flux units. Magnetic lines are circular inside the particle.

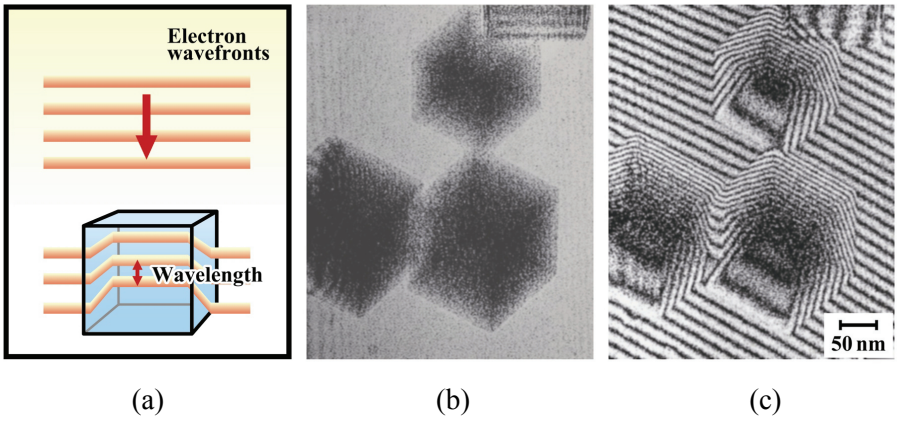


Fig. 7. Magnesium-oxide particle. (a) Schematic of wavefronts penetrating non-magnetic sample. (b) Electron micrograph. (c) Interference micrograph (phase amplification × 2). Electrons are accelerated by inner potential V_0, and their wavelength shortens as illustrated in (a). Although only the outlines of the particles are observed in the electron micrograph (b), thickness contours in units of 20 nm are observed as deviations from regular interference fringes in the interference micrograph (c).

A more practical application of electron holographic interferometry to "in-plane" magnetic recording (Osakabe *et al.*, 1983) is shown in Fig. 9. A magnetic head slides across a metal tape, and the bit patterns are recorded. The observed magnetic lines below are exactly like a stream of water since we are looking at a divergence-free ***B***. We can see that, where two streams of magnetic lines collide head on, vortices are produced. The magnetic lines meander toward the edge of the tape and then beyond the tape edge.

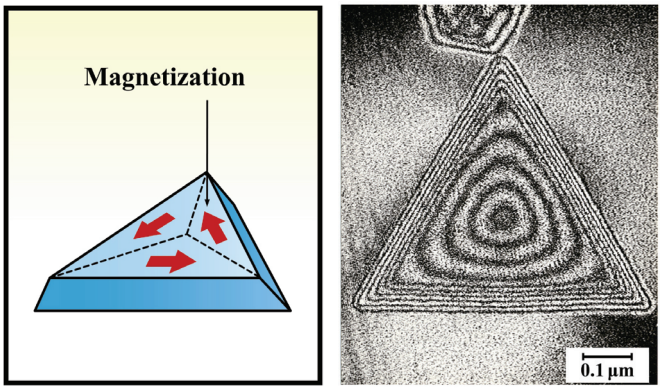


Fig. 8. Triangular cobalt particle. (a) Schematic. (b) Interference micrograph (phase amplification × 2). Phase contours in peripheral part of particle in (b) indicate thickness contours, and those in central part indicate in-plane magnetic lines.

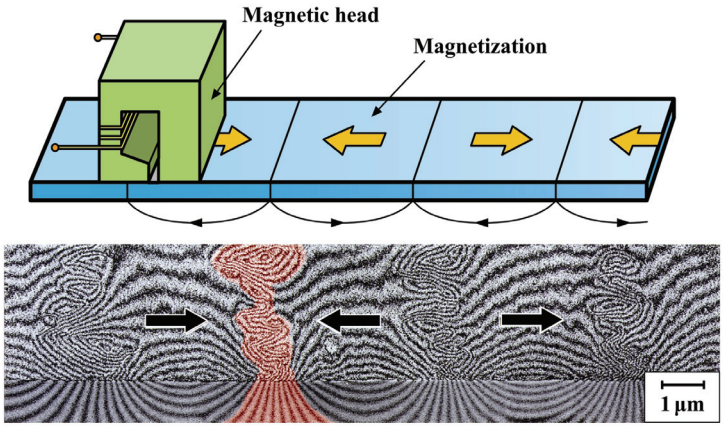
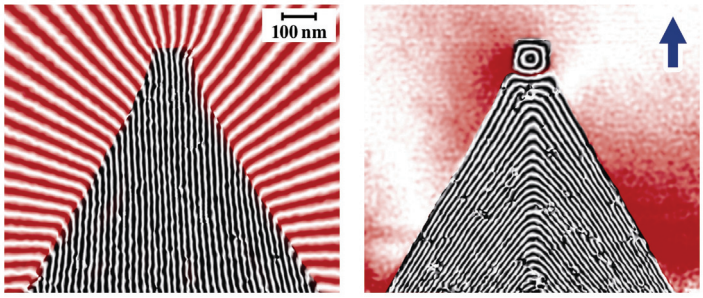


Fig. 9. Magnetic lines inside and outside a recorded magnetic tape. Detailed magnetic lines observed under various conditions, such as tape material and spacing and gap of head, provide information about how higher density recording can be attained.

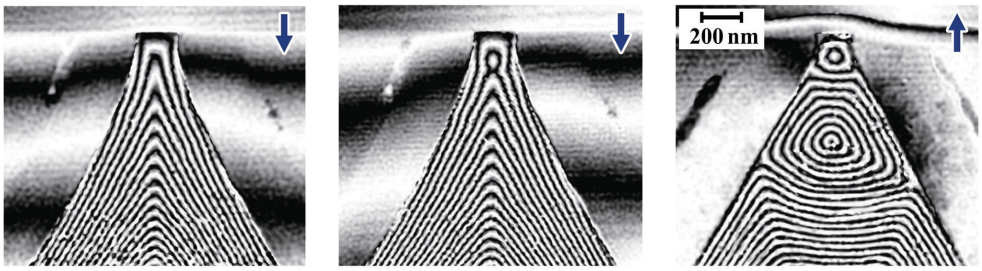
The performance of a magnetic head has been investigated by electron holography (Kim *et al.*, 2008). Magnetic lines from a tiny magnetic head for "perpendicular" high-density

recording are evident in Fig. 10(a). While a high and localized magnetic field is needed, other characteristics are also required. For example, we also have to be able to control the head in any way we want. In some cases, we need to get a zero field outside the head, which is not easy since the density of magnetic lines inside ferromagnets remains constant.
For specific head materials and shapes, magnetic fields leaking from the head become negligible when the applied magnetic field is removed: straight magnetic lines inside the head (Fig. 10(a)) change into "U-turn" lines (Fig. 10(b)). At the pole tip, the magnetic lines form a tiny closed circuit. This remanent state in Fig. 10(b) is reversible, but slightly history-dependent (Hirata et al., 2009), as shown in Figs. 11(a) and (b).



(a) (b)

Fig. 10. Magnetic lines of magnetic head for perpendicular recording. (a) Magnetized state ($\boldsymbol{H}$ = 150 Oe). (b) Remanent state ($\boldsymbol{H}$ = 0 Oe). Magnetic lines from magnetic head provide direct information about recording magnetic lines when magnetic field $\boldsymbol{H}$ is applied and also about leakage magnetic lines.



(a) (b) (c)

Fig. 11. Different remanent states of magnetic heads. (a) Magnetic lines of NiFe head when magnetic field applied in upward direction was removed. (b) A little different magnetic lines though under the same condition as (a). (c) Magnetic lines of NiFeCo when magnetic field applied in upward direction was removed. Remanent states are important since leaking magnetic flux sometimes erases recorded magnetic information.

We also found that the remanent pattern depends on the material. When the head material was changed from "nickel-iron" to "cobalt-nickel-iron", the remanent state became irreversible, and the remanent pattern greatly changed (Hirata *et al.*, 2009). An example is shown in Fig. 11(c): a large closure domain was produced in addition to the smaller one.

A more recent application example is the observation of the nucleation and growth of ferromagnetic domains in colossal magnetoresistance (CMR) materials (Murakami *et al.*, 2009). In CMR, an electric current begins to flow when the magnetic field increases or the temperature decreases. The microscopic mechanism for this is explained as follows: isolated tiny ferromagnets nucleate and coalesce in the 'insulating' phase, thus creating conduction paths for electrons. Using Lorentz microscopy, we investigated how this happens microscopically. When we reduced the sample temperature, a tiny spot suddenly appeared with bright and dark contrast features. The single magnetic domain (Fig. 12(a)) changed after a tremor into a double domain (Fig. 12(b)) and grew, keeping the double-domain form for a while (Fig. 12(c)) since these states are energetically stable. The double domains then coalesced into multiple domains (Fig. 13), until finally expelling insulating domains.

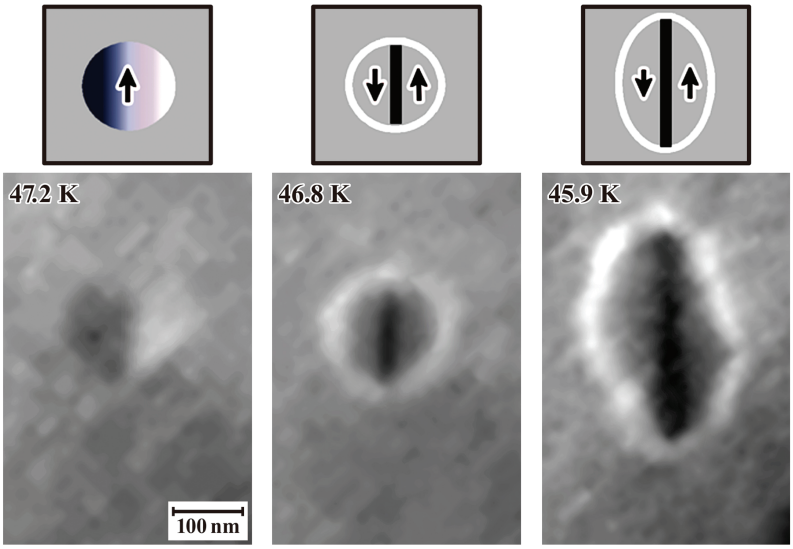


Fig. 12. Lorentz micrographs of $La_{0.25}Pr_{0.375}Ca_{0.375}O_3$ thin film. (a) Single domain at T = 47.2 K. (b) Double domain produced immediately after magnetization reversal at T = 46.8 K. (c) Growing double domain at T = 45.9 K.

We found two kinds of "boundaries", straight and zigzag. The domain walls inside the ferromagnetic phase were smooth, while the boundaries between the ferromagnetic and insulating phases were zigzag. We concluded that, in the insulating domains, there must exist tiny charge-ordered clusters that cannot easily become ferromagnetic. The boundaries between ferromagnetic and insulating domains are affected by these clusters, resulting in zigzag boundaries and their stepwise motion. The boundaries between the two ferromagnetic domains are straight since no clusters exist there.

The existence of the clusters was confirmed not only by electron diffraction but also by Lorentz microscopy. Photographic evidence was obtained when we carefully observed the growth of the ferromagnetic domains while gradually decreasing the sample temperature

(Fig. 14). The insulating domains (Fig. 14(a)) began at 54 K to be invaded by ferromagnetic domains (Fig. 14(b)). The invasion was locally interrupted by tiny clusters, resulting in the formation of zigzag domain boundaries. At 45 K, the ferromagnetic domains occupied almost the whole field of view (Fig. 14(c)), and the insulating domains became smaller. The tiny circular domains indicated by the arrows in Fig. 14(c) are attributed to the clusters. At 35 K, the whole field of view was ferromagnetic and conductive.

We also obtained evidence for the effect of magnetic fields on the growth of ferromagnetic domains. Magnetic fields leaking from the ferromagnetic domains, which cannot be observed by Lorentz micrograph (Fig. 15(a)) but by holographic interference micrograph (Fig. 15(b)), help to locally grow the domains.

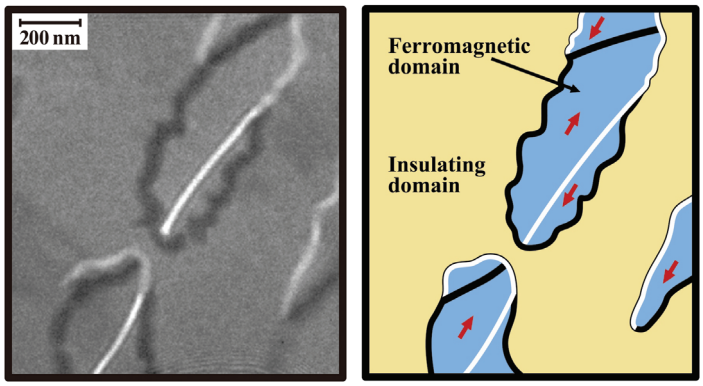


Fig. 13. Coalesced ferromagnetic domains. (a) Lorentz micrograph. (b) Schematic. Boundaries inside ferromagnetic domains are straight, and boundaries between insulating and ferromagnetic domains are zigzag due to existence of charge ordered clusters in insulating domains.

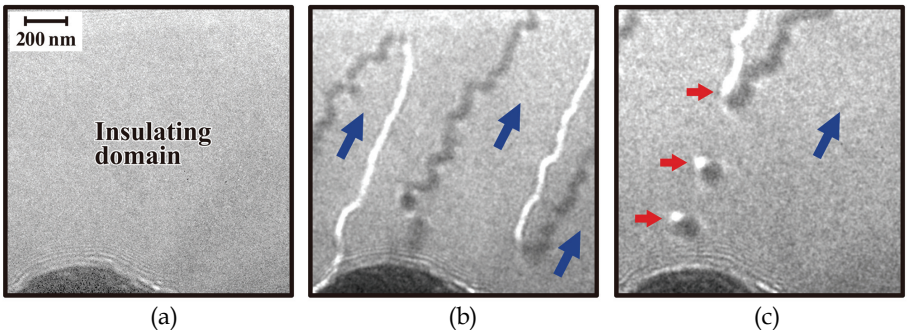


Fig. 14. Evidence of charge-ordered clusters in Lorentz micrographs. (a) T = 60 K, (b) 54 K, (c) 45 K. When sample temperature was decreased, ferromagnetic domains grew and finally occupied the whole region, which was insulating at higher temperatures. Tiny circular insulating domains indicated by arrows in (c) were produced at cluster positions. Whole region then became ferromagnetic.

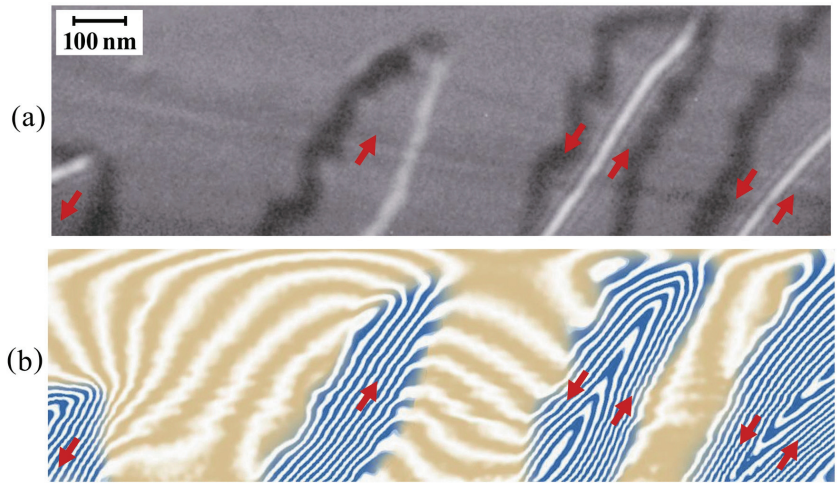


Fig. 15. Growth of ferromagnetic domains in colossal magnetoresistance manganite, $La_{0.25}Pr_{0.375}Ca_{0.375}MnO_3$, thin film at T = 62.4 K. (a) Lorentz micrograph; (b) interference micrograph.

5. Exploring new possibilities

The AB effect celebrated its 50th anniversary in 2009, and symposia on the AB effect were held in Tel Aviv and Bristol. Since the AB effect cannot be understood in terms of classical electrodynamics, a long controversy about it persisted for decades. Even its existence was questioned. Now, however, most physicists regard the AB effect as one of the most fundamental phenomena in quantum mechanics and as related to the deepest concepts of physics. It directly involves the reality of vector potentials, the locality of interactions, the single valuedness of wavefunctions in a multiply-connected space, and so on. Advanced technologies including bright coherent electron beams have made it possible to carry out experiments on the AB effect that were once regarded as thought experiments.

On a personal note, it was the mysterious and yet attractive AB effect in addition to the wave-particle duality of electrons including the theory of observation that motivated me to enter the field of electron wave optics. In 1967, I started research on electron holography, and I had to continually develop brighter, more coherent electron beams. If I had not started this development, the field of electron phase microscopy including electron holography would not be as developed as it is now.

The AB effect was proven to be fundamental to quantum mechanics and also to gauge field theories unifying all fundamental interactions in nature. Therefore, the AB effect is ubiquitous not only in electromagnetism but also in nuclear and gravitational forces. It appears not only in physics but also in chemistry and biology.

In our field of electron microscopy, it has been applied in various ways to the observation of microscopic objects and electromagnetic fields, which are both phase objects for illuminating electrons.

The phase shift of electrons can now be measured with a precision of 1/100 the wavelength. However, if we want to observe the world of atoms and molecules, this precision is not

enough. We need a precision in the phase measurement of at least the order of 1/1,000 the wavelength, which, I believe, is possible given that the precision approaches 1/10,000 the wavelength of light.

6. References

Aharonov, Y. & Bohm, D. (1959). Significance of Electromagnetic Potentials in Quantum Theory. *Phys. Rev.*, Vol. 115, No. 3, 485-491.

Bocchieri, P. & Loinger, A. (1978). Nonexistence of the Aharonov-Bohm Effect. *Nuovo Cimento A*, Vol. 47, No. 4, 475-482.

Bocchieri, P. & Loinger, A. (1981). Comments on the Letter << On the Aharonov-Bohm Effect >> of Boersch *et al*. *Lett. Nuovo Cimento*, Vol. 30, No. 15, 449-450.

Bocchieri, P.; Loinger, A. & Siragusa, G. (1982). Remarks on << Observation of Aharonov-Bohm Effect by Electron Holography >>. *Lett. Nuovo Cimento*, Vol. 35, No. 11, 370-372.

Boersch, H.; Hamisch, H. & Grohmann, K. (1962). Experimenteller Nachweis der Phasenverschiebung von Elektronenwellen durch das magnetische Vektorpotential. *Z. Phys.*, Vol. 169, No. 2, 263-272.

Chambers, R. G. (1960). Shift of an Electron Interference Pattern by Enclosed Magnetic Flux. *Phys. Rev. Lett.*, Vol. 5, No. 1, 3-5.

Fowler, H. A.; Marton, L.; Simpson, J. A. & Suddeth, J. A. (1961). Electron Interferometer Studies of Iron Whiskers. *J. Appl. Phys.*, Vol. 32, No. 6, 1153-1155.

Gabor, D. (1949). Microscopy by Reconstructed Wave-Fronts, *Proc. R. Soc. Lond. A*, Vol. 197, No. 1051, 454-487.

Hirata, K.; Ishida, Y.; Kim, J. J.; Kasai, H.; Shindo, D.; Takahashi, M. & Tonomura, A. (2009). Electron holography observation of in-plane domain structure in writer pole for perpendicular recording heads. *J. Appl. Phys.*, Vol. 105, No. 7, 07D538-1-07D538-3.

Kim, J. J.; Hirata, K.; Ishida, Y.; Shindo, D.; Takahashi, M. & Tonomura, A. (2008). Magnetic domain observation in writer pole tip for perpendicular recording head by electron holography. *Appl. Phys. Lett.*, Vol. 92, No. 16, 62501-1-62501-3.

Miyazawa, H. (1985). Quantum Mechanics in a Multiply Connected Region. *Proc. 10th Hawaii Conf. High Energy Physics 1985*, Hawaii, Univ of Hawaii, pp. 441-458.

Möllenstedt, G. & Bayh, W. (1962). Kontinuierliche Phasenschiebung von Elektronenwellen im Kraftfeldfreien Raum durch das magnetische Vektorpotential eines Solenoids. *Phys., Bl.*, Vol. 18, 299-305.

Murakami, Y.; Kasai, H.; Kim, J. J.; Mamishin, S.; Shindo, D.; Mori, S. & Tonomura, A. (2009). Ferromagnetic domain nucleation and growth in colossal magnetoresistive manganite. *Nat. Nanotech.,* Vol. 5, No. 1, 37-41.

Osakabe, N.; Yoshida, K.; Horiuchi, Y.; Matsuda, T.; Tanabe, H.; Okuwaki, T.; Endo, J.; Fujiwara, H. & Tonomura, A. (1983). Observation of recorded magnetization pattern by electron holography. *Appl. Phys. Lett.*, Vol. 42, No. 8, 746-748.

Peshkin, M. & Tonomura, A. (1989). "The Aharonov-Bohm effect." *Lecture Notes in Physics*, 340, Springer, 978-3540515678.

Roy, S. M. (1980). Condition for Nonexistence of Aharonov-Bohm Effect. *Phys. Rev. Lett.*, Vol. 44, No. 3, 111-114.

Tonomura, A.; Matsuda, T.; Endo, J.; Arii, T. & Mihama, K. (1980). Direct observation of fine structure of magnetic domain walls by electron holography. *Phys. Rev. Lett.*, Vol. 44, No. 21, 1430-1433.

Tonomura, A.; Matsuda, T.; Suzuki, R.; Fukuhara, A.; Osakabe, N.; Umezaki, H.; Endo, J.; Shinagawa, K.; Sugita, Y. & Fujiwara, H. (1982). Observation of Aharonov-Bohm effect by electron holography. *Phys. Rev. Lett.*, Vol. 48, No. 21, 1443-1446.

Tonomura, A. (1983). Observation of magnetic domain structure in thin ferromagnetic films by electron holography. *J. Magn. and Magn. Matls.*, Vol. 31-34, Part 2, 963-969.

Tonomura, A.; Umezaki, H.; Matsuda, T.; Osakabe, N.; Endo, J. & Sugita, Y. (1983). Is magnetic flux quantized in a toroidal ferromagnet?. *Phys. Rev. Lett.* Vol. 51, No. 5, 331-334.

Tonomura, A.; Umezaki, H.; Matsuda, T.; Osakabe, N.; Endo, J. & Sugita, Y. (1984). Electron Holography, Aharonov-Bohm Effect and Flux Quantization, In: *Proc. Int. Symp. Foundations of Quantum Mechanics,* edited by Kamefuchi, S.; Ezawa, H.; Murayama, Y.; Namiki, M.; Nomura, S.; Ohnuki, Y. & Yajima, T., Physical Society of Japan, Tokyo, 1984, 20-28.

Tonomura, A.; Osakabe, N.; Matsuda, T.; Kawasaki, T.; Endo, J.; Yano, S. & Yamada, H. (1986). Evidence for Aharonov-Bohm effect with magnetic field completely shielded from electron wave. *Phys. Rev. Lett.*, Vol. 56, No. 8, 792-795.

Tonomura, A. (2005). Direct observation of hitherto unobservable quantum phenomena by using electrons. *Proc. Natl. Acad. Sci. USA*, Vol. 102, No. 42, 14952-14959.

Wu, T. T. & Yang, C. N. (1975). Concept of nonintegrable phase factors and global formulation of gauge fields. *Phys. Rev. D*, Vol. 12, No. 12, 3845-3857.